
Blue light energy pathways in light harvesting complexes

INAUGURAL-DISSERTATION

TO OBTAIN THE ACADEMIC DEGREE
DOCTOR RERUM NATURALIUM (DR. RER. NAT.)

SUBMITTED TO
THE DEPARTMENT OF BIOLOGY, CHEMISTRY, PHARMACY
OF FREIE UNIVERSITÄT BERLIN

BY
SIMON PETRY
FROM BERLIN, GERMANY
2023

This work was conducted under the supervision of
Dr. Jan Philipp Götze (Freie Universität Berlin)
at the Institute of Chemistry and Biochemistry
from March 2019 until March 2023

1. Reviewer: Dr. Jan Philipp Götze
2. Reviewer: Prof. Dr. Beate Paulus
Date of the defense: 23.05.2023

STATUTORY DECLARATION

Herewith I, Simon Petry, certify that I have prepared and written my thesis with the subject

Blue light energy pathways in light harvesting complexes

independently and that I have not used any sources and aids other than those indicated by me.

Simon Petry
Berlin, March 31, 2023

Danksagung

Diesen Platz meiner Arbeit möchte ich all diesen Menschen widmen, die mich mit ihrer unerschütterlichen Unterstützung während und vor meiner Promotion begleitet haben.

Allen voran danke ich meinem Betreuer Jan Phillip Götze, der mir nicht nur ermöglicht hat in seiner Arbeitsgruppe zu promovieren, sondern auch mein Interesse an biochemischen Systemen geweckt hat. Er hat sich stets die Zeit genommen mir mit wissenschaftlichen Fragen und Problemen zu helfen und die resultierenden Diskussionen waren immer sehr bereichernd. Seine Begeisterung für die Forschung und seine Akzeptanz für neue Ideen, haben mir einen angenehmen Freiraum zur Entfaltung darboten.

Weiterhin möchte ich Beate Paulus meinen Dank dafür aussprechen, dass sie mir bereits im Laufe meines Studiums die theoretische Chemie nähergebracht hat und mir die Möglichkeit gegeben hat Erfahrungen in der Lehre zu sammeln.

Desweiteren gilt mein Dank Jean Christophe Tremblay, der für jede wissenschaftliche Problemstellung immer einen guten Rat parat hatte. Natürlich möchte ihm dafür danken, dass mich auf diese Promotionsstelle aufmerksam gemacht hat.

Ich schätze mich sehr glücklich, dass ich sowohl in der AG Paulus als auch in der AG Keller so warmherzig aufgenommen wurde und so einzigartige Kolleg:innen kennenlernen durfte. Des Weiteren möchte ich mich für das freundliche und entspannte Arbeitsumfeld bei allen aktuellen und ehemaligen Mitgliedern der AGs der theoretischen Chemie bedanken. Ich habe sowohl die wissenschaftlichen Diskussion als auch das feierabendlichen Beisammensein im Pigroom mit euch genossen.

An dieser Stelle möchte ich mich aufrichtig bei allen entschuldigen, die ich aufgrund des Platzes oder meiner persönlichen Vergesslichkeit namentlich nicht erwähne. Ich danke euch, Jan, Denis, Luca, Leon, Joana, Marco, Marius, Ruben, Melli, Max und Franzi für die großartige Zeit und eure Unterstützung. Ich möchte mich auch bei meinen zahlreichen Forschungspraktikant:innen bedanken, an denen ich meine didaktischen Fähigkeiten erproben durfte, und im Besonderen Fabian, Alina, Dilara und Anna.

Ein besonderer Dank geht an Oliver Lemke und Fabian Weber, welche ich als Tutoren während meines Studiums kennengelernt habe und die mich als Freunde durch meine Promotion begleitet haben. Speziell in den letzten Monaten aber auch schon davor haben mich Felix Stein, Stefanie Kieninger, Jennifer Anders und Felix Witte bei sämtlichen Rückschlägen aufgefangen und mich wieder aufgebaut.

Schließlich aber nicht weniger möchte ich mich bei meiner Familie und meinen Freunden bedanken, ohne die ich nicht der Mensch wäre der ich heute bin. Speziell meinem Bruder möchte ich danken, dass er immer für mich da ist. Ich bin unglaublich dankbar euch an meiner Seite zu wissen und euch blind vertrauen zu können. Ich danke euch allen für eure unschätzbaren Beiträge zu meinem Leben und euer Verständnis in den letzten Jahren.

Nicht zuletzt möchte ich mich bei meiner Freundin Lena für die ständige Unterstützung und Ermutigung in den letzten Jahren bedanken. Diese Doktorarbeit wäre ohne ihre unaufhörliche Ermutigung und bedingungslose Unterstützung nicht möglich gewesen. Sie ist meine treibende Kraft, meine größte Vertraute und meine beste Freundin.

Abstract

The analysis of the structural and optical properties of light-harvesting complexes, which are parts of a complicated utilization of light energy, is indispensable to understand their biological “advantage” and their physical-chemical functions. These insights can provide the basis for adaptations of these systems in method and material development for solar fuels, particularly in relation to the transition to renewable energy.

This doctoral thesis presents a hybrid quantum mechanical/molecular mechanical investigation of the light-harvesting complex CP29, which plays a crucial role in the initial steps of photosynthesis in plants and algae. The main objective of this work is to understand the energy transfer mechanism and the electronic properties of CP29 chromophores and selected amino acids using computational methods. The methods used are designed to accurately describe both the structural properties of large biomolecules and the optical properties of individual molecules.

The first part (Paper 1 and Paper 2) of the thesis focuses on the development of a suitable software for the necessary calculations and the comparability of the spectroscopic properties of the main pigments of photosynthetic plants between theory and experiment. The software `gmx2qmmm` is a Python based interface between the quantum chemistry program Gaussian and the software package for simulations of molecular mechanical processes GROMACS. This software was used to perform most of the calculations of this thesis.

The second part (Paper 3) of the thesis investigates the role of the protein environment in the electronic properties of CP29. This is achieved by modeling the protein CP29 using classical molecular dynamics simulations. The electronic properties of CP29 are calculated for different conformations of the protein, and the results are analyzed to understand the electrostatic effect of protein on the electronic properties of the individual chromophores as well as on energy transfer. Therefore, quantum chemical methods based on time-dependent density functional theory were applied and the energy pathways were identified using a Förster resonance energy transfer coupling scheme, with a strong emphasis on the analysis of the blue light energy states of the chlorophylls, which have received little attention in the literature with respect to energy transfer.

The final part of the thesis (Paper 4) investigates the effect of multipole moments on the excitation energy transfer. This is achieved by modeling the interaction between the chlorophyll chromophores of CP29 including the carotenoids as acceptors and comparing the Förster resonance energy transfer coupling scheme with the transition density cube scheme.

Overall, this work provides a comprehensive understanding of the spectroscopic properties and energy transfer mechanism of the light-harvesting complex CP29 using quantum mechanical and molecular mechanical methods. The overarching question of the work is whether blue/higher energy pathways exist between chromophores of CP29, represent a competing process to other decay processes, and whether preferred energy pathways can be identified.

Kurzzusammenfassung

Die Analyse der strukturellen und optischen Eigenschaften von Lichtsammelkomplexen, die Teil einer komplizierten Nutzbarmachung von Lichtenergie sind, ist unerlässlich, um ihren biologischen und physikalisch-chemischen Funktionen zu verstehen. Insbesondere in Bezug auf den Umstieg auf erneuerbare Energien können diese Erkenntnisse die Grundlage für Methoden- und Materialentwicklung neuer solarer Brennstoffe darstellen.

In dieser Dissertation wird eine quantenmechanische/molekularmechanische Untersuchung des Lichtsammelkomplexes CP29 vorgestellt, der eine entscheidende Rolle bei den ersten Schritten der Photosynthese in Pflanzen und Algen spielt. Das Hauptziel dieser Arbeit ist es, den Energietransfermechanismus und die elektronischen Eigenschaften von CP29 mit Hilfe von rechnergestützten Methoden zu verstehen. Die verwendeten Methoden sind darauf ausgelegt, sowohl die strukturellen Eigenschaften großer Biomoleküle als auch die optischen Eigenschaften einzelner Moleküle genau zu beschreiben.

Der erste Teil (Paper 1 und Paper 2) der Arbeit konzentriert sich auf die Entwicklung einer geeigneten Software für die notwendigen Berechnungen und die Vergleichbarkeit der spektroskopischen Eigenschaften der wichtigsten Pigmente photosynthetischer Pflanzen zwischen Theorie und Experiment. Die Software gmx2qmmm ist eine auf Python basierende Schnittstelle zwischen dem Quantenchemieprogramm Gaussian und dem Softwarepaket für Simulationen molekularmechanischer Prozesse GROMACS. Mit dieser Software wurde ein Großteil der Berechnungen durchgeführt.

Im zweiten Teil (Paper 3) der Arbeit wird die Rolle der Proteinumgebung für die elektronischen Eigenschaften von CP29 untersucht. Dazu wird das Protein CP29 mit Hilfe von klassischen Molekulardynamiksimulationen modelliert. Die elektronischen Eigenschaften von CP29 werden für verschiedene Konformationen des Proteins berechnet, und die Ergebnisse werden analysiert, um den elektrostatischen Effekt des Proteins auf die elektronischen Eigenschaften der einzelnen Chromophore sowie auf den Energietransfer zu verstehen. Dazu wurden quantenchemische Methoden auf der Grundlage der zeitabhängigen Dichtefunktionaltheorie angewandt und die Energiepfade mit Hilfe eines Förster-Resonanz-Energietransfer-Kopplungsschemas identifiziert. Dabei lag der Schwerpunkt auf der Analyse der Blaulichtenergiezustände der Chlorophylle, die in der Literatur im Hinblick auf den Energietransfer wenig Beachtung gefunden haben.

Der letzte Teil der Arbeit (Paper 4) untersucht die Auswirkung von Multipolmomenten auf den Anregungsenergietransfer. Dazu wird die Wechselwirkung zwischen den Chromophoren von CP29 einschließlich der Carotinoide als Akzeptoren modelliert und das Förster-Resonanz-Energietransfer-Kopplungsschema mit dem Schema der *transition density cubes* verglichen.

Insgesamt bietet diese Arbeit ein umfassendes Verständnis der spektroskopischen Eigenschaften und des Energietransfermechanismus des Lichtsammelkomplexes CP29 unter Verwendung quantenmechanischer und molekularmechanischer Methoden. Die übergreifende Frage der Arbeit ist, ob Blaulichtenergiepfade zwischen den Chromophoren von CP29 existieren, einen konkurrierenden Prozess zu anderen Zerfallsprozessen darstellen und ob bevorzugte Energiepfade identifiziert werden können.

Contents

List of publications	VIII
List of abbreviations	IX
1 Introduction	1
1.1 Photosynthesis	1
1.2 Light-harvesting	7
1.3 Scope of this work	13
2 Theory	18
2.1 Molecular mechanics	18
2.2 Fundamentals of quantum mechanics	23
2.3 Hartree-Fock method	25
2.4 Density functional theory	29
2.5 Optical properties	41
2.6 QM/MM	47
3 Summary and conclusion	52
4 Publications	58
4.1 Paper 1	58
4.2 Paper 2	79
4.3 Paper 3	98
4.4 Paper 4	114

List of publications

1. "A user-friendly, Python-based quantum mechanics/Gromacs interface: gmx2qmmm"
J. P. Götze, Y.-W. Pi, **S. Petry**, F. Langkabel, F. Witte, O. Lemke
International Journal of Quantum Chemistry **2021**, 121, 3, e26486.
DOI: 10.1002/qua.26486
2. "Spectral characterization of the main pigments in the plant photosynthetic apparatus by theory and experiment"
J. P. Götze, F. Anders, **S. Petry**, F. Witte, H. Lokstein
Chemical Physics **2022**, 559, 111517 ff.
DOI: 10.1016/j.chemphys.2022.111517
3. "Effect of protein matrix on CP29 spectra and energy transfer pathways"
S. Petry, J. P. Götze
Biochimica et Biophysica Acta - Bioenergetics **2022**, 1863, 2, 148521 ff.
DOI: 10.1016/j.bbabi.2021.148521
4. "Impact of structural sampling, coupling scheme and state of interest on the energy transfer in CP29"
S. Petry, J. C. Tremblay, J. P. Götze
BioRxiv preprint server **2023**
DOI: 10.1101/2023.01.25.525376

List of abbreviations

SE	Schrödinger equation
PSII	photosystem II
PSI	photosystem I
OEC	oxygen-evolving complex
LHCII	Light-harvesting complex II
LHC	light-harvesting complex
Chl	chlorophyll
Crt	carotenoid
Lut	lutein
Vio	violaxanthin
Neo	neoxanthin
Zea	zeaxanthin
QM	quantum mechanics
MD	molecular dynamics
MM	molecular mechanics
QM/MM	quantum mechanics/molecular mechanics
VdW	Van-der-Waals
PBC	periodic boundary conditions
BOA	Born-Oppenheimer approximation
HF	Hartree-Fock
SCF	self-consistent-field
LCAO	linear combination of atomic orbitals
MO	molecular orbital
AO	atomic orbital
CI	configuration interaction
MCSCF	multi-configurational self-consistent-field
DFT	density functional theory
DFA	density functional approximation
TD-DFT	time-dependent density functional theory
LR TD-DFT	linear response time-dependent density functional theory
LDA	local density approximation
LSDA	local spin density approximation
GGA	generalized gradient approximation
mGGA	meta-generalized gradient approximation
LC	long-range correction
PES	potential energy surface
FC	Frank-Condon
HT	Herzberg-Teller
FCHT	Frank-Condon-Herzberg-Teller
VR	vibrational relaxation
ISC	intersystem crossing
IC	internal conversion
NAC	non-adiabatic coupling
SOC	spin-orbit coupling
TDM	transition dipole moment
TDC	transition density cube

FRET	Förster resonance energy transfer
EET	excitation energy transfer
LSCF	local self-consistent-field
GHO	generalized hybrid orbital
PCF	point charge field
PC	plastocyanin
Fd	ferredoxin
ATP	adenosine triphosphate
PSU	photosynthetic unit
LUMO	lowest unoccupied molecular orbital
HOMO	highest occupied molecular orbital
IDA	ideal dipole approximation
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
FNR	Fd-NADP ⁺ -oxidoreductase
RuBP	ribulose-1,5-biphosphate
NADPH	nicotinamide adenine dinucleotide phosphate
RuBisCO	RuBP carboxylase-oxygenase
3-PGA	3-phosphoglycerates
G3P	glyceraldehyde 3-phosphate
NPL	non-planarity
BLA	bond length alternation
LCF	link atom correction function
CC	coupled cluster

1 Introduction

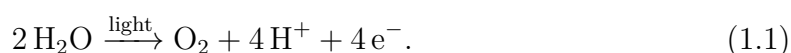
The aim of this chapter is to give the reader a rough overview of the systems studied in this thesis and the used methods. Here, only the application, advantages/disadvantages and implementation of the methods will be described, while the underlying theory of these will be discussed in sec. 2. The reader must be aware that some of the subsequent processes of photosynthesis are still part of current research and therefore many hypotheses and proposed mechanisms are presented. Due to the large number of photosynthetically active organisms, only evolutionary differences will be exemplified in the following, while mainly the systems of higher plants, *i.e.*, see sec. Light-harvesting in PSII, will be described.

1.1 Photosynthesis

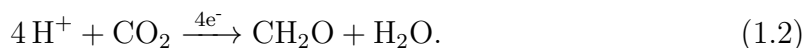
The natural process of photosynthesis dominates the biosphere as the most widespread metabolic process on earth.^[1] Among photosynthetic organisms, oxygenic phototrophs including all known species of cyanobacteria, algae, and higher plants, paved the way for the evolution of animals by enabling the creation of atmospheric oxygen (O₂).^[2] About 3 billion years ago, the ancestors of these organisms triggered the world's greatest natural catastrophe, the transformation of a carbon dioxide atmosphere into an oxygen atmosphere. This change transformed the surface of the earth from a lush aluminosilicate desert to a blossoming green carpet.^{[3],[4]} Life as we know it today would be impossible without this drastic reshaping due to the increase in oxygen level by photosynthetically active organisms.^[5]

The knowledge that plants obtain natural substances from air using light was confirmed in the 18th century. A core feature of photosynthesis was first described by Joseph Priestley, who proved that plants renewed the air “consumed” by candles or mice.^{[6]–[8]} Photosynthesis reduces carbon with the help of light energy and thus represents the opposite process of oxidative carbohydrate consumption. The photosynthetic process fixes the carbon of atmospheric carbon dioxide in the form of carbohydrates. The latter serve as energy sources/storage or building material (cellulose) for the photosynthesizing organisms as well as for all organisms that feed directly or indirectly on the photosynthetically active organisms (various carbohydrate forms). Through this process, between 10¹⁰–10¹¹ tons of carbon are fixed annually, which would correspond to a storage of up to 10¹⁸ kJ of energy per year.^{[9],[10]} In comparison, EU's energy consumption in 2020 was about 37 · 10¹⁵ kJ and the solar energy the Earth receives is approximately 38 · 10²⁰ kJ per year.^[11]

In the early 20th century, it was erroneously assumed that after light absorption, pigments would directly reduce CO₂, which would then combine with water to form carbohydrates.^[12] In fact, the current state of research is that this is a two-step process consisting of a *light-dependent reaction* and a *light-independent reaction*. In the light-dependent step, light energy is used to oxidize water:



The light-independent or “dark” reaction can be summarized in a very simplified fashion as:



This process is however much more involved than **Eq. (1.1)**, which technically only requires a single protein complex (see secs. 1.1.1 and 1.2.4). The location where photosynthesis occurs in eukaryotes is the chloroplast (see Fig. 1). Photosynthetically active regions of a cell may consist of 1 to 100 chloroplasts, which can vary in size and shape, but are typically ellipsoids about 5 micrometers long.^[13] Chloroplasts have a permeable outer membrane and impermeable inner membrane separated by a narrow intermembrane space. The inner membrane envelops the stroma (a concentrated solution of enzymes). Further, the stroma also contains deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and ribosomes, which are involved in the synthesis of chloroplast enzymes and the replication of chloroplasts themselves. A third membrane compartment is formed by the thylakoids, which are enclosed by the stroma — and the thylakoid is the location of the photosynthesis light(-dependent) reactions. A thylakoid compartment is thought to consist of a single highly folded vesicle, although in most organisms the thylakoids resemble stacks of disc-shaped sacs (granathylakoids). The granathylakoids are connected by stromathylakoids. The thylakoid membrane is formed by invagination of the inner membrane of newly forming chloroplasts.^{[14],[15]} Being the main compartment for the photosynthetic light reaction, a thylakoid membrane contains protein complexes that harvest light, transport electrons, and are involved in adenosine triphosphate (ATP) synthesis. In contrast to plants and algae, for prokaryotic organisms such as photosynthetically active bacteria, the apparatus for the light-dependent reaction is localized in the plasma membrane, which often forms invaginations or multilamellar structures.^[16] This gave rise to the theory of endosymbiosis, namely that a predecessor of the chloroplasts was an early cyanobacterium that found its way into an eukaryotic cell, resulting in a symbiotic interaction.^{[17],[18]}

1.1.1 Light-dependent reactions

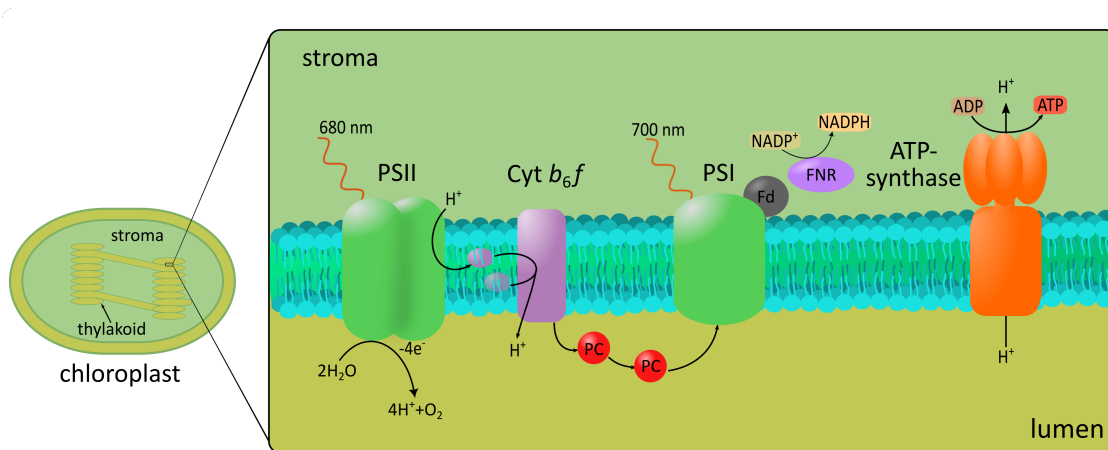


Figure 1: Schematic representation of a chloroplast (left) and of all relevant protein complexes involved in the light-dependent reaction embedded in a thylakoid membrane (right). More detailed schematics of PSII and PSI are shown in Fig. 2 and 3, respectively. This figure is inspired by ref. [13].

The aim of the light-dependent reaction is to convert the radiant energy of sunlight into chemical energy in several reaction steps. The energy is stored in the form of the universal energy carrier ATP and the electron carrier molecule nicotinamide adenine dinucleotide phosphate ($\text{NADP}^+/\text{NADPH}$) (see Fig. 1) They represent the ma-

terial prerequisites for the subsequent light-independent reaction to accomplish the buildup of sugars.^[19]

While photosynthetic bacteria show a cyclic electron transport within one reaction center, plants, algae and cyanobacteria show a non-cyclic electron transport that proceeds through two reaction centers.^{[20],[21]} The reaction centers of green plants are photosystem I (PSI), which reduces NADP^+ , and photosystem II (PSII), which oxidizes H_2O . Both photosystems have a chlorophyll (Chl) dimer in the reaction center, which is called a special pair. In the case of PSI, this dimer absorbs at 700 nm and is therefore named $P700$. In PSII, the dimer absorbs at 680 nm ($P680$) and in the excited form ($P680^+$ radical) represents one of the strongest biological oxidants. Both centers are excited independently by light, but electrons always flow (in terms of reaction order, mediated by various proteins and compounds in between) from PSII to PSI. The spatial separation of the two photosystems prevents energy transfer besides the electron transfer between the systems on the one hand, but on the other hand it allows the chloroplasts to respond to different sunlight exposures. PSII occurs almost exclusively within the stacked membranes (grana) with little contact to the stroma and PSI occurs mainly in the unstacked thylakoids with high contact to the stroma and accordingly to the stroma-solvated NADP^+ .

To facilitate a balanced light influx between PSII and PSI, the possibility exists to regulate the accumulation of light-harvesting complexes (LHCs) in the different domains. In this regard, at high exposure, the accumulation in the unstacked thylakoid is favored by the phosphorylation of specific threonin residues of LHCs by a protein kinase, resulting in more LHCs attaching to PSI. At low exposure, this process is reversed, resulting in more LHCs attaching to PSII.^[13]

The photochemistry of PSII begins when light is absorbed by the pigments (see

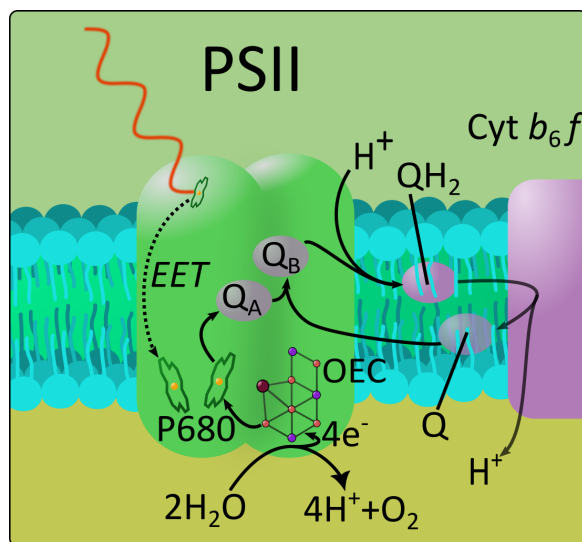


Figure 2: Schematic representation of PSII with the associated reactions and components containing the special Chl pair $P680$, oxygen evolving complex (OEC), quinones ($\text{Q}_{A/B}$, Q) and quinols (QH_2) and the protein complex Cytochrome b_6f . The orange arrow represents absorption of light, while black arrows represent reactions and the black dotted line represents excitation energy transfer (EET).

sec. Light-harvesting in PSII). As indicated in Fig. 2, the excitation energy is transferred via EETs to the special Chl pair $P680$ in the reaction center. Exciting $P680$ to the corresponding excited $P680^*$ results in several redox reactions: An electron transfer from $P680^*$ to pheophytin (20 ps), from pheophytin to quinone A (Q_A ; 300 ps)

resulting in a $P680^+Q_A^-$ and an electron transfer from Q_A^- to a second quinone B (Q_B ; 0.5–1 ms).^[22]

This process represents a cycle that requires two light absorption instances to reduce the quinone to a quinol. The remaining (after each electron transfer) $P680^+$ radical is reduced back to original, neutral $P680$ by electrons coming from the oxygen-evolving complex (OEC), an oxo-bridged tetramanganese-calcium cofactor (Mn_4O_5Ca) cluster via an intermediate redox-active tyrosine residue (100 ns). The reduction of the tyrosine radical occurs through the cluster in 40 μs –2 ms depending on the state of the OEC.^{[22],[23]} After 4 light-induced charge separation events and associated oxidations of the cofactor by the $P680^+$ radical, a triplet state oxygen molecule is released. This cycle was first postulated by Kok and coworkers in 1970. The Kok cycle, or S-state cycle, has the five states S_0 – S_4 , where S_1 represents the resting state of the OEC ($Mn_2(III)Mn_2(IV)$)^[24] and S_0 represents the most reduced form ($Mn_3(III)Mn(IV)$).^[25] The cycle is still not considered to be fully understood today and several experimental and theoretical groups have proposed their own explanations for the observed spectra.^{[26]–[30]}

In total, 2 water molecules are thereby split into 4 protons, 4 electrons and one oxygen molecule by 4 light-induced processes (**Eq. (1.1)**). Both, the protons and electrons are released as membrane bound quinones as described above. Note that the protons bound by the quinones originate from the stroma and not from water splitting (see Fig. 2). Q_A is strongly bound and thus immobile (in both oxidized and reduced forms). Q_B is more weakly bound, but further hydrogen bonding after the first reduction increases the affinity for the binding site.^[31] This ensures that the Q_B^-/Q_BH intermediates are not replaced and are available as substrate for the second reduction. After the second reduction, the final quinol Q_BH_2 forms, which leaves PSII and can further transport electrons to cytochrome *b6f*.^[32] The leaving quinol is replaced by a new nearby quinone. Too much light can cause a decrease in concentration of the quinones, effectively increasing the lifetime of the $P680^+Q_A^-$ and thereby increasing the probability of recombination to a triplet excited state Chl. The triplet excited state Chl may in turn cause the formation of a singlet oxygen and damage the entire complex (see sec. 1.2.5).^[33]

Since the oxidation of plastoquinol is the slowest and thus rate-limiting step, the cytochrome *b6f* complex plays a central role in electron transport. The task of the complex is to transport electrons from the membrane-bound quinone to the peripheral membrane protein plastocyanin (PC), which is localized on the lumen side of the thylakoid membrane, and to release the protons on the lumen side (see Fig. 3).^{[34],[35]} The release of the protons into the stroma results in a proton gradient which is necessary for the ATP synthase to work.

The PC transports the electrons within the lumen to the PSI complex. Excitation of PSI results in charge separation at the special pair $P700$, similar to PSII, but not completely identical. The excitation can occur directly at the special pair as in PSII, which is statistically unlikely or can be transferred there via EET. The electron is transferred to a phylloquinone by two nearby Chls *a*. The charged $P700^+$ is regenerated by oxidation from the reduced PC formed by the cytochrome *b6f* complex.^{[36],[37]} The electron is transferred from the special pair $P700$ through a series of intermediate carriers to the ferredoxin (Fd), a water-soluble electron carrier binding on the stromal side of PSI.^{[38],[39]} Fd reduced by PSI transfers the electron to the Fd-NADP⁺-oxidoreductase (FNR) enzyme, which is responsible for the reduction of

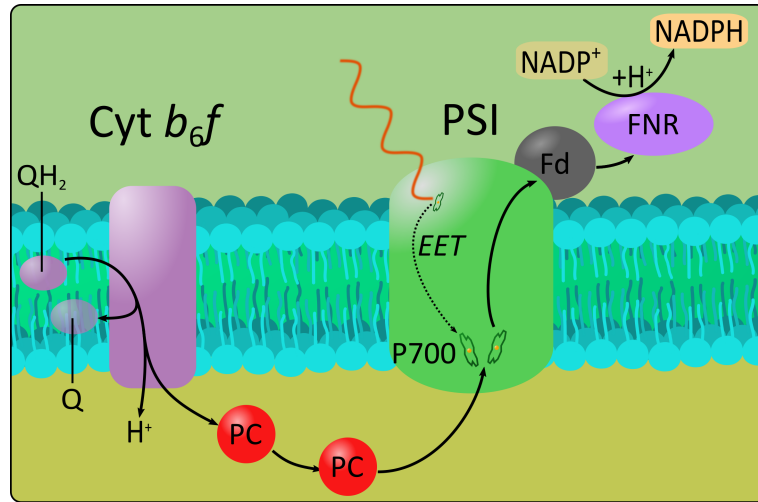
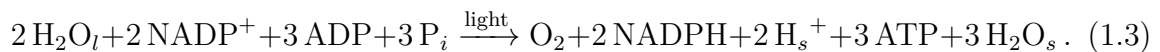


Figure 3: Schematic representation of PSI containing the special Chl pair P700, Cytochrom b_6f complex, ferredoxin (Fd) complex, Fd- NADP^+ -oxidoreductase (FNR) enzyme, plastocyanine (PC), quinone/quinols (Q/ QH_2) and NADP^+ / NADPH . The orange arrow represents absorption of light, while black arrows represent reactions and the black dotted line represents excitation energy transfer (EET).

NADP^+ to NADPH .

The established proton gradient is required to drive the molecular machinery of ATP synthase. The enzyme ATP synthase is composed of 8–20 different subunits, which can be grouped into two complexes: The water-insoluble F_0 complex, which is located in the membrane and regulates proton transport, and the water-soluble F_1 complex, which catalyzes the formation of ATP.^[40] The controlled flux of protons from lumen to stroma fuels the binding of ADP, phosphorylation to ATP and subsequent release. All these highly complex sub-processes can be simplified into a summation formula for the total balance for the formation of one molecule of oxygen with s and l representing stromal and luminal position, respectively:



Note that the stoichiometry for the luminal-stromal proton exchange is not precisely represented.

1.1.2 Light-independent reactions

The light-independent reaction is the second part of photosynthesis and uses the products of the light-dependent reaction (ATP and NADPH) to form sugars. This metabolic pathway, also called Calvin cycle, describes the fixation (incorporation into organic molecules) of carbon atoms from CO_2 and the formation of sugars with a length of 3 to 7 carbon atoms. As illustrated in Fig. 4, the Calvin cycle can be divided into 3 main stages: Carbon fixation, reduction and regeneration of the reactant molecule.^[41] In the carbon fixation, a CO_2 molecule is combined with a five-carbon acceptor molecule, ribulose-1,5-biphosphate (RuBP). This reaction is catalyzed by the enzyme RuBP carboxylase-oxygenase (RuBisCO). The resulting 6-carbon molecule is unstable and decomposes into two 3-carbon molecules, which are 3-phosphoglycerates (3-PGA). Accordingly, two 3-PGA molecules are formed for each CO_2 that is converted. The subsequent reduction step requires ATP and

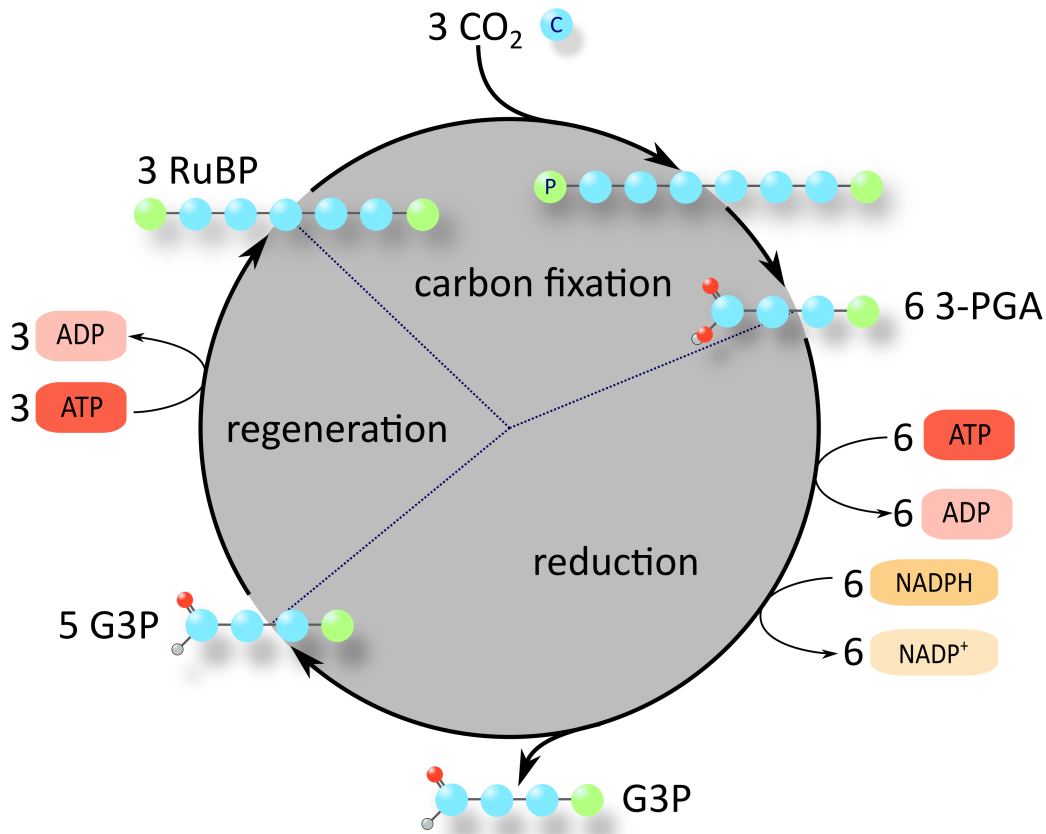


Figure 4: Schematic representation of the Calvin cycle containing a simplified representation of the associated sugar species (3-phosphoglycerates(3-PGA), glyceraldehyde 3-phosphate (G3P), ribulose-1,5-biphosphate (RuBP) and an intermediate 6-carbon molecule) with the color code: Carbon (cyan), oxygen (red), phosphorus (green) and hydrogen (grey).

NADPH to convert 3-PGA into a 3-carbon sugar. This stage takes place in two major steps. First, phosphorylation of the 3-PGA into a doubly phosphorylated molecule, 1,3-biphosphoglycerate, occurs. In this process, ATP is converted back to ADP. In the second step, the 1,3-biphosphoglycerate is reduced by NADPH and loses one of the phosphate groups. The resulting product is a 3-carbon sugar called glyceraldehyde 3-phosphate (G3P).

The fuel for these reactions are the products of the light-dependent reaction, ATP and NADPH. The Calvin cycle uses the energy of light, which is in the form of chemical energy (ATP) and reducing power (NADPH), to build monosaccharide and regenerate ADP and NADP⁺, which are necessary for the light-dependent reaction.

G3P represents a very essential building block of the plant, as it is used for various sugar syntheses, such as glucose or fructose production. Since G3P is used in the Calvin cycle to regenerate RuBP again, at least 3 CO₂ molecules must be fixed to synthesize 6 G3P molecules to regenerate in turn 3 RuBP for one G3P molecule to leave the cycle.^{[42],[43]}

1.2 Light-harvesting

As can be seen from the section before, plants use sunlight to convert the energy into a fuel for their metabolism. As described earlier, the pigment dimers of special pairs in PSI and PSII consist of Chl *a*. But for the understanding that single Chls are not responsible for, e.g., oxygen production, the 1940s years represent a breakthrough. In 1932, Robert Emerson and William Arnold succeeded in showing by experiments on suspension of green algae that about 2500 Chl molecules were involved in the production of one molecule of oxygen.^[20] In 1934 Arnold and Henry Kohn confirmed the existence of such “units” of ~ 2400 Chl molecules per molecule of oxygen produced for various photosynthetic systems and called them “chlorophyll units.”^[44] Two years later, in 1936, Kohn concluded that one Chl unit was more like 500 Chl, which at that time was consistent with the reasoning that four photons were needed to produce one oxygen molecule^[45] and with the current understanding of the Kok cycle.^[46] Between 1936 and 1940, Hans Gaffron and Kurt Wohl determined that a single Chl would take an average of one hour to capture 4–12 photons, which would correspond to the evolution of an oxygen molecule.^{[47],[48]}

From this finding, Gaffron and Wohl concluded that “quantum energy” had to be transferred between molecules. They postulated a model that says wherever the energy is absorbed within the “unit” (nowadays photosynthetic unit (PSU)), it is transferred to the “reaction center” and used for photochemistry. Without the today’s methods to analyze the structure of these complexes available, many studies have been done on the size of these “units”. The ratio of Chls per oxygen evolution was found to differ between organisms.^{[44],[49],[50]} Nowadays, the determination of the molecular structure of such PSUs, which has been identified as reaction centers of PSI and PSII, is feasible and is about 200–400 light-collecting molecules per reaction center of higher plants or green algae.¹ David Mauzarell and Nancy L. Greenbaum highlighted in 1989 that there is a difference between the oxygen forming unit, the total number of Chls per oxygen molecule production (compare definition of Emerson and Arnold, *i.e.*, 2500 Chls for a single O₂), and the “quantum requirement” for O₂ formation.^[50] Furthermore, they pointed out that there is a theoretical limit for PSU size, which corresponds to the limit for antennae containing only one type of chromophores (see sec. 1.2.3). The limit on the size of the light-gathering architecture can be increased by incorporating pigments that absorb at higher energies or smaller wavelengths of light (e.g., Chl *b* or carotenoids (Crts)). This allows the existence of more efficient larger PSUs using heterogeneous antennae, antennae containing different chromophores.^[51] However, Zoe Perrine found in 2012 that, light-gathering antennae are not optimized to achieve maximum quantum efficiency, but rather to capture the maximum light under *both* low- and high-intensity light conditions.^[52] The existence of chlorosomes contradicts the argumentation of a size limit. Chlorosomes are the largest known photosynthetic antenna structures, e.g., the chlorosomes of *Chlorobaculum tepidum* (green sulfur bacteria) are approximately 110–200 nm in length and 40–60 nm in diameter. A single chlorosome of *C. tepidum* contains more than 222500 chromophores, *i.e.*, 200000 bacteriochlorophylls (BChls) *c* molecules, 2500 BChl *a* molecules and 20000 carotenoid molecules.^[53]

Over evolutionary time, this has led to the diversification and optimization of PSUs,

¹The structural findings of PSII are described in more detail in section Light-harvesting in PSII from a historical perspective.

which are composed of numerous different light-gathering complexes. (see sec. 1.2.4) The fundamental mechanism is comprised of (I) highly efficient energy absorption, (II) followed by a rapid EET within and between the light-gathering antennae, (III) culminating in the trapping of the excitation at the special pair in the reaction center — as well as the photochemical processes of the reaction center pigments (see sec. 1.1).

1.2.1 Properties of chlorophylls

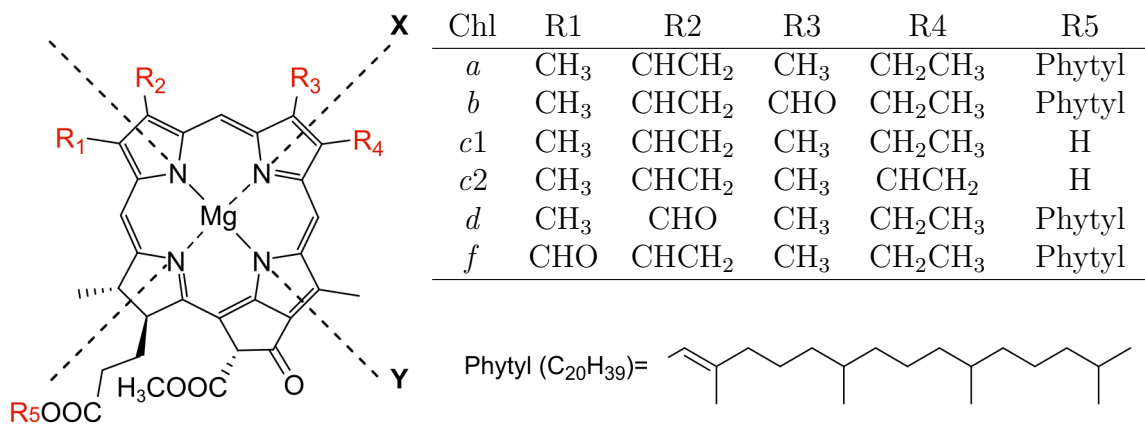


Figure 5: Molecular structure of Chl *a*, *b*, *c1*, *c2*, *d* and *f*, including the X and Y axis of the macrocycle accordingly to the Gouterman’s “four orbital” model^{[54],[55]} and listing of the respective functional groups of the position R1–R5.

The most important photoreceptor pigments in photosynthesis are Chls. This cyclic tetrapyrrole is biosynthetically derived from Protoporphyrin IX (like the heme group of globins). Chl differs from the well-known heme structure by another cyclopentanone ring and by the central metal ion, which is Mg²⁺ in Chl and Fe(II/III) in heme. The basic structure of Chl contains a porphyrin ring with the central Mg ion and a hydrophobic chain (phytol chain). There are the different Chl types *a*, *b*, *c1*, *c2*, *d* and *f* in higher plants (see Fig. 5). Bacterial systems contain similar or identical Chls, but also structurally different bacteriochlorophylls. The different side chains are responsible for tuning the absorption spectrum of the pigment and thus represent evolutionary adaptations to environmental influences or a functional specialization of the respective compound.^[56]

As with porphyrin, the strongest light absorption of Chls is in the visible region of light, due to the highly π -conjugated ring structure. A distinction is made between the Q band, which absorbs between 500–700 nm, and the B or “Soret” band, which absorbs between 400–500 nm (see Fig. 6). The region between the two bands is called the green gap. The basic spectroscopic properties of Chls could be explained by the Gouterman’s “four orbital” model. This states that the electronic transitions with the strongest oscillator strength are described by the π - π^* electronic transitions. These excitations involve the two highest occupied molecular orbitals (HOMOs) and the two lowest unoccupied molecular orbitals (LUMOs) and result in electronic transitions polarized along either the X or Y axis of the macrocycle (see Fig. 5). Accordingly, two transitions result for each of the two absorption bands described, Q_y, Q_x, B_y and B_x.^{[54],[55],[57]} The most abundant Chl *a* is present in all photoactive eukaryotes, microalgae, and cyanobacteria. Chl *b* is present in all land plants, green algae, and euglenozoa. The two Chls have the same basic structure and differ only in one functional group of the

porphyrin ring (Chl *a*: methyl, Chl *b*: aldehyde). Chl *c* can be found in two forms in golden-brown algae.^[58] Both are structurally very different from Chl *a/b* in that they lack a phytyl chain. Among themselves, Chl *c2* and *c1* differ only in R3, which is an alkane in *c1* and an alkene in *c2*.^[59] Chl *d* has been detected in red algae and differs structurally from Chl *b* in the side chain R1 (Chl *a*: CHCH₂, Chl *d*: CHO).^{[60],[61]} Chl *f* can be found in cyanobacteria and are structurally different from Chl *a* in R1 (Chl *a*: methyl, Chl *f*: aldehyde).^{[56],[62],[63]}

1.2.2 Properties of carotenoids

Besides Chls, Crts are the most important chromophores in photosynthesis and are vital components of biological tissue for animals and plants. More than 600 different variants of Crts exist in different organisms.^[65] Crts are divided into carotenes, Crts without oxygen (e.g., α -carotene, β -carotene or lycopene) and xanthophylls, Crts containing oxygen (e.g., lutein (Lut) or zeaxanthin (Zea)). Among other things, they act as building blocks by helping proteins to fold; act as auxiliary pigments in plants; contribute to light-harvesting and energy transport; and contribute to signal transduction in plants and animals. Compared to Chls, Crts absorb characteristically with an onset at 500 nm with various maxima

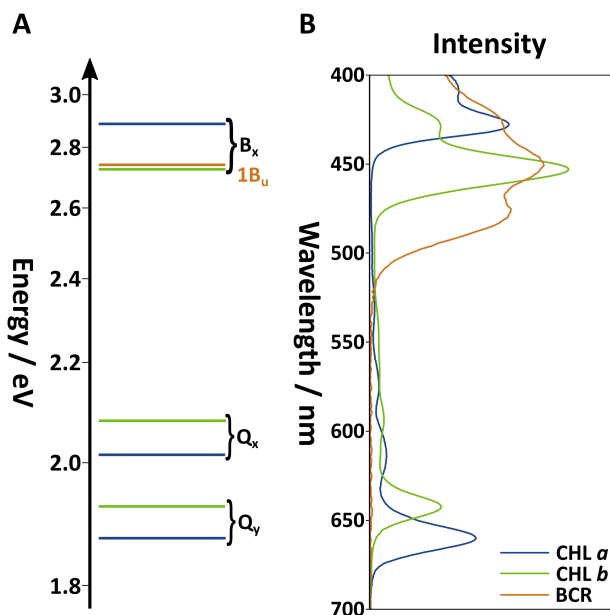


Figure 6: (A) Jablonski diagram of the first three excited states of Chl *a/b* (blue/green) and the $1B_u$ of β -carotene (BCR, orange). (B) Experimental absorption spectra of Chl *a/b* and BCR in acetone using the same color code as in (A).^[64]

centered at about 450 nm and a tail which extends to about 300 nm (see Fig. 6).^{[66],[67]} This very broad and intense band arises from a single strong dipole allowed transition ($1A_g \rightarrow 1B_u$). Crts are proposed to play an essential role for photoprotection (see sec. 1.2.5).^{[68]–[70]} Chl *d* and *f* are supposedly adaptations to biomes which only allow for low available intensities consisting of mostly of far-red light irradiation.^[71]

1.2.3 Requirements for excitation energy transfer

The first experimental evidence of energy transfer between two molecules was obtained in 1922 by Günther Cario and James Franck.^[72] Between 1927–1929, Jean Perrin and Francis Perrin observed energy transfer of fluorophores in solution.^{[73]–[75]} They described that molecules in solution can interact with each other without collision and with a distance exceeding their molecular diameter. They postulated that the reason of electronic energy transfer was an inductive resonant interaction between two transition dipole moments of molecules. Theodor Förster described this interaction, as a semiclassical oscillation of the donor electrons, during deexcitation, inducing oscilla-

tions of the acceptor electrons, which causes electronic excitation (see sec. 2.5.3). For resonant energy transfer to occur between two chromophores, the energy differences of the initial and final states of the donor and acceptor must be equal. As a consequence of this condition, the energy transfer can only take place "downhill". Furthermore, the donor state must be fluorescent, since the transfer rate is proportional to the oscillator strength of the fluorescence spectrum of the donor and the oscillator strength of the absorption spectrum of the acceptor.^{[76],[77]}

Chl *a* is an example of a good EET molecule with *itself* along the Q band. The corresponding small Stokes shift of the Q_y state equals a small difference in the position of the absorption and fluorescence spectra, ensuring energy transfer within the same band. Since only one type of chromophore exists in homogenous PSUs and energy transfer occurs "downhill", absorption can only occur at a finite range and be transported to the reaction center. The chromophores have different biochemical and biophysical properties when associated with different amino acids binding sites^[78] and therefore, this combination represents a way to expand the limit.^[79] As mentioned above, the limit of the size of the PSU can be increased by incorporating pigments that absorb at higher energies or smaller wavelengths (e.g., Chl *b* or Crts). Chl *b* represents an ideal extension for the Chl-based antennae, since the Chl *b* Q band is shifted to smaller wavelengths and the properties in terms of oscillator strengths do not differ greatly from Chl *a*.

For Crts, this reasoning is not so simple, since the absorption takes place in the Soret band region of Chl *b* and accordingly no transfer to the Q band is guaranteed, since it is a resonant energy transfer. However, it has been experimentally demonstrated that energy transfer occurs between xanthophylls and Chls.^{[66],[67],[80],[81]} This can be explained by large Stokes shift or the large difference in the position of the absorption and fluorescence spectra. While the absorption of Crts occurs at 400-500 nm, the excited state fluorescence occurs at 600-700 nm. Accordingly, Crts, among their other functions, could represent a shortcut between the Soret and the Q bands of Chl.^[82]

1.2.4 Light-harvesting in PSII

In this section, the structural design of the PSII and the organization of the antenna complexes are discussed. PSII exists *in vivo* in a thylakoid membrane as a dimer, with each monomer consisting of 19–31 polypeptide subunits depending on the organism.^{[84],[85]} The thylakoid membrane is flexible and thus able to respond to external environmental cues by adjusting the composition or organization of the PSII supercomplex.^{[86],[87]} The core of PSII is composed of the major subunits: PsbA (D1 protein) and PsbD (D2 protein), which comprises much of the reaction center, the two light-harvesting complexes PsbB (CP47) and PsbC (CP43), and the manganese stabilizing protein PsbO.

The photosynthetic reaction center from *Rhodospseudomonas viridis* was the first transmembrane protein whose X-ray structure was solved.^[88] Over the years, better methods for structural resolution have provided greater insight, for example: The 2D crystal structure of the first eukaryotic membrane protein (Light-harvesting complex II (LHCII) from pea) by electron cryo-microscopy (1994; 3Å),^[89] the 2D crystal structure of the PSII core without water oxidizing activity (1998, 8Å),^[90] the active PSII by electron microscopy (2000; 15–30Å),^[91] the X-ray structure of the active PSII (2001; 3.8Å),^{[92],[93]} and the entire PSII-LHCII supercomplex by cryo-electron microscopy (2017, 3.2–2.5 Å).^[83] The totality of these structures and many more allow

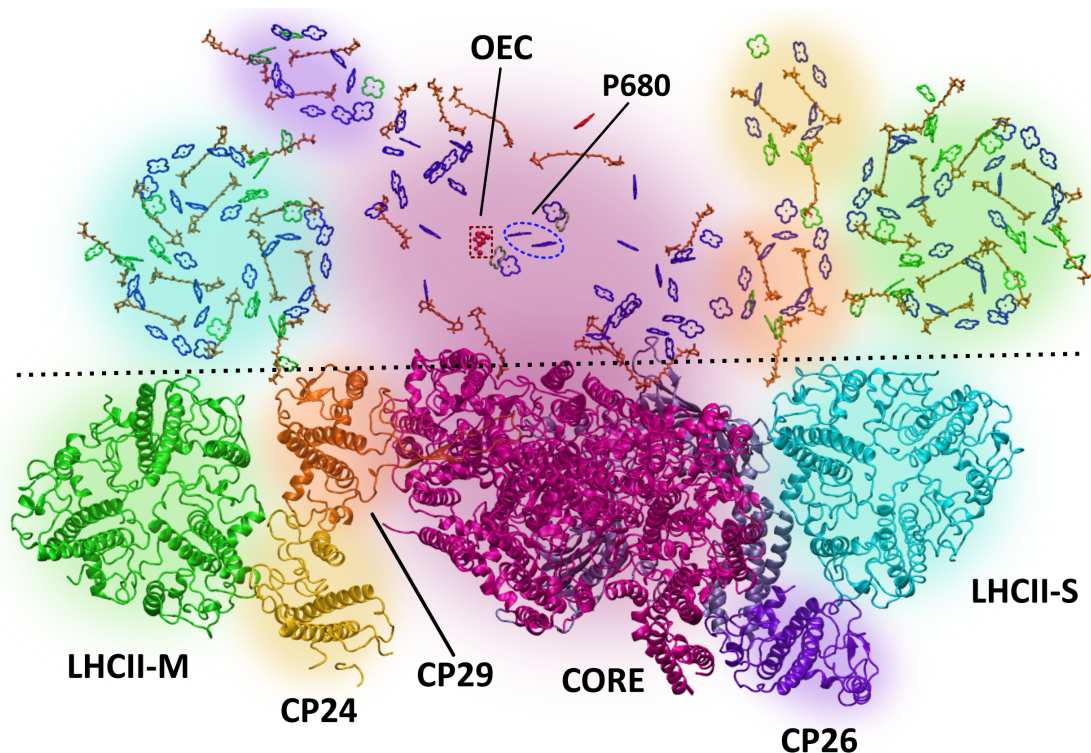


Figure 7: Stacked-C2S2M2 supercomplex containing a dimeric core (C2), two LHCII-S (S2) and two LHCII-M (M2) viewed from the stromal side. The two PSII-LHCII monomers are separated by the black dotted line, with the lower part showing the protein structure in a ribbon diagram and the upper part showing only the chromophores and the acOEC (PDB: 5XNL).^[83] The color code of the complexes is consistent in both parts showing LHCII-M (green), LHCII-S (cyan), CP24 (yellow), CP26 (blue), CP29 (orange) and the core complex (intrinsic subunits: Purple, extrinsic subunits: Grey). The upper part shows Chl *a/b* (blue/green) core ring atoms, Crt (orange), pheophytin (grey), heme (red), P680 (blue-dashed circle) and the OEC (red-dashed box).

precise conclusions to be drawn about the composition of the complexes, the arrangement of the cofactors, and the position of the OEC in the PSII supercomplex.^{[94],[95]} Specifically, the first X-ray structure of PSII represented a major breakthrough, providing the first insight into subunit and cofactor organization.^[92] Because of its high resolution of 1.9 Å, Yasufumi Umena's crystal structure allowed researchers to examine both the water distribution within the subunits and the water molecules that coordinate at the OEC.^[96]

In eukaryotics, the PSII mostly couples to LHCII and minor CP such as CP29, CP26 or CP24, which form an advanced PSII-LHCII supercomplex.^[97] The LHCs all share a very similar protein sequence and folding.^{[98]–[101]} The common motif of the protein subunits is given by 3 transmembrane helices. The LHCII trimers associated with the core can be divided into the 3 classes: Strong(S), moderate(M) and loose(L) depending on their position in the PSII and the "binding strength" to the core(C).^[102] Thus, the C2S2M2 supercomplex is composed of a core dimer (C2), two LHCII S-trimers (S2) bound via a direct contact site to the core and via CP29, CP26 and two LHCII M-trimers (M2) interacting with the core via CP29 and CP24 (see Fig. 7).^{[91],[102]–[105]} The loosely bound LHCII have so far only been found in spinach, which is why their exact position is still unclear.^[102] Compared to the PSII core, which has only Chls *a* and β -carotenoids, the antenna complexes have more different pigments. One LHCII monomer binds 18 pigments: 6 Chls *a*, 8 Chls *b*, 4 xanthophyll: 2 Luts, 1 violaxan-

thin (Vio), 1 neoxanthin (Neo) (see Fig. 7).^[100] The Luts are located in the center of the monomer, while the Vio and Neo are located at the peripheries of the respective monomers. Based on the sequence similarity of all LHCS, most of the Chl-binding amino acids are similar, which can be seen by comparing CP29 and LHCII.^[106] The composition of the chromophores present differs in the individual antenna complexes. For example, CP26 and CP29 have a Vio at the position where a Lut is found in LHCII and CP24 has no Neo. Furthermore, the number and position of Chls *b* in the respective antenna complexes differs. As the isolation of the antenna complexes from the membrane can lead to a loss of pigments, the characterization of these antenna complexes is difficult.^[107] Therefore, the purified CP29 complex was expected to have 8–9 Chls with a Chl *a/b* ratio of ~ 3.0 , while experiments show that embedded CP29 binds 8.5 Chls *a* and 4.5 Chls *b* with a ratio of ~ 1.89 .^{[108]–[111]} Recent studies confirmed the existence of a fourteenth Chl next to the core.^[83] Previous claims that Chl binding sites are promiscuous have so far not found any recent support, and the current consensus is that each Chl site binds usually only one type of Chl.

Although all members of the LHC multigenic family^[112] share common features, each complex shows slightly different biochemical, spectroscopic or functional properties. However, the main functions of LHCS are the light absorption, the EET and the control of the energy flow to avoid photodamage.^[113] Besides LHCII, the CP29 antenna is the most widely studied plant antenna complex. CP29 and CP26 are both directly associated to the core in a bridging position to LHCII.^{[83],[103],[114]} Especially these locations within the PSII structural architecture suggests that CP29 and CP26 play essential roles in ensuring an efficient EET towards the core.^[115] Investigations of isolated antenna complexes showed a rapid EET with a fast thermal equilibration of a few picoseconds.^{[113],[116]–[118]} The overall equilibration of the antenna within the PSII appeared to be over a magnitude longer.^{[119]–[121]} This suggests that the EET among different subunits of the PSII antenna represents the limiting step for photochemical trapping and suggests that a single chromophore or a group of chromophores is responsible for the excited state energy transfer between the complexes.^{[115],[122]} The proteins play an important role in this inter-complex EET, as they do not only act as a scaffold holding the chromophores in place, but also affect the energies of the pigments, which is part of Paper 3.^{[123],[124]}

In addition to the function of light-harvesting, EET and energy conduction, other functions attributed to the minor complex CP29 are PSII macroorganization, photoprotection and its reversible phosphorylation mainly involved in photoinhibition recovery^[125] and state transitions.^{[126],[127]} It was found that among the 3 minor antenna complexes (CP24, CP26 and CP29), CP29 is typically phosphorylated under environmental stress conditions.^{[128]–[131]} Furthermore, the organization of the PSII supercomplex and the orientation in the membrane is carried out by the minor antenna complexes.^[105] Nevertheless, larger PSIIs with 6 LHCIIs per core dimer have been found,^[132] which shows that CP24 are not mandatory for M-trimers to bind to the supercomplex. Based on C2S2M2, the additional trimers bind on the side where the CP24 would be expected. However, the C2S2 found in *algae C. reinhardtii*,^{[91],[133]} which lacks CP24, suggests that CP24 is relevant for the indirect binding of M-trimers.^{[134],[135]} Furthermore, it was shown that plants in which the minor antenna complexes were removed by reverse genetics, the stability was significantly reduced.

1.2.5 The role of antennae in photoprotection

The role of the minor antenna complexes in protecting the photosynthetic apparatus of PSII from excess energy is still not fully understood. However, experiments show that minor antenna complexes play a crucial role for photoprotection.^{[69],[136]–[139]}

Photo-oxidative damage in plants is defined as various forms of light-induced damage to the cell. The light induces the formation of reactive oxygen species, e.g., hydrogen peroxide, superoxide or singlet oxygen.^[140] The main reactive oxygen species in PSII is singlet oxygen^{[141]–[143]} whose photosensitised generation was first demonstrated by Kautsky and de Bruijn.^[144] The electron spin of the excited Chls can rephase and form a energetically lower Chl triplet state. The triplet state can react with oxygen to produce a reactive singlet oxygen, which has been measured to have a lifetime of approximately 3 μ s.^{[145],[146]} Chl triplet states can be populated via intersystem crossing (ISC) from the excited Chl singlet state or in the reaction centers by charge recombination reactions.^{[147],[148]} The charge recombination reactions are favored under physiological CO₂ exposure to high light intensities or drought. These conditions result in a low concentration of quinones and in photoinhibition, *i.e.*, a loss of PSII activity.^[149] Stress can thus result in singlet oxygen which is toxic for the cell as it can react with proteins, pigments, nucleic acids and lipids and is therefore also likely responsible for the degradation of the PSII core.^{[150],[151]}

1.3 Scope of this work

As I have shown above, photosynthesis is one of the most important processes on Earth, as it provides the basis for most of the food we eat and the oxygen we breathe. The initial steps of photosynthesis involve the absorption of light by pigment molecules, which transfer the photonic energy to reaction centers where it is converted into chemical energy (see sec. 1.1). One of the key components in this process is the minor light-harvesting complex CP29, which is responsible for capturing light and transferring the energy. Chls and Crts are arranged in a highly organized manner to efficiently capture light and transfer energy. The mechanisms of energy transfer in LHCs have been the subject of extensive research, because understanding these energy transfer processes is crucial for the development of efficient artificial light-harvesting systems. However, most studies of the energy transfer of Chls are limited to the red light energy pathways (Q band), which are definitely most important for the special pairs P680/P700 of photosystems (see sec. 1.2.4). Due to the shorter lifetime, mainly non-radiative, fast and local decay processes are considered for the blue states (Soret band).

This thesis deals with the question whether resonant energy transfers of blue wavelength energies between chromophores of LHCs exist and represent competing processes in comparison to the non-radiative decays. Since blue light can cause photodamage to the reaction core (see sec. 1.2.5), this work addresses the question if blue light energy transfer pathways actually exist and, as a subsequent question, what mechanisms the PSII supercomplex has to protect the core from photodamage. The idea of this project was thus to show that the organization of the PSII supercomplex transfers more likely red light energy than blue light energy towards the core (as it would make sense biologically). Further, we tested if the minor light-harvesting complex CP29, which is at least the link between LHCI-M and the reaction core (see Fig. 7), has a special role for blocking higher/blue excitations from entering the PSII

core complex (like a protective gatekeeper).

To investigate this hypothesis we used computational methods, which provides a powerful tool for studying the electronic properties of molecules and their interactions with light. In recent years, there has been significant progress in the development of quantum mechanical methods for modeling complex biological systems. These methods allow us to investigate the electronic properties of proteins and their interactions with pigments and other biomolecules.

Computational simulation of biological systems

The study of the properties of biomolecular systems using computational methods has become increasingly relevant in recent decades. The impressive improvements in achievable simulation speed and the underlying physical models allow atomic-level simulations in millisecond range. The dynamics of biomolecules can be modeled with the help of molecular simulations to comprehend the structural and dynamical properties and to substantiate or explain experimental results.

A powerful tool for the analysis of biomolecules and their corresponding properties are molecular mechanics (MM). MM calculations are inexpensive because each atom is treated as a particle. MM force fields have been widely used in molecular dynamics (MD) simulations, propagations of atomic structures using Newton's equation of motion for a system of interacting particles. These force fields are empirical potentials describing bonded and nonbonded interactions such as small-amplitude vibrations, torsions, Van-der-Waals (VdW) interactions, and electrostatic interactions. During the simulation typically no recalculation of the force field parameters takes place. A detailed description of the theory can be found in the Molecular mechanics section. To describe all the dynamical properties of the system, such as vibrations of the bonds, it requires simulation time steps in the femtosecond regime (10^{-15} s) or shorter.^[152] The description of dynamic processes require simulation times from microseconds (10^{-6} s) to seconds or even longer. Table 1 shows examples of typical required simulation times for dynamic processes.^{[153],[154]} The simulation times are strongly dependent on the described system. A big disadvantage of MM simulations is that they are strongly dependent on the choice of the force field. To account for electron effects like polarization or bond rearrangements, special polarizable force fields are needed, since typical MM describes each atom as a charged mass point.^{[155],[156]} In order to describe changes of the bond situation like chemical reactions or hydrogen rearrangements, methods involving a quantum mechanical description are necessary.

Within its limitations, quantum mechanics (QM) provides accurate "many electron functions" and associated observables such as the electric dipole moment by explicitly considering the quantum nature to a certain extend. However, the application of QM is still limited to relatively small systems of up to a few hundred atoms or even smaller when using the highest level of theory due to the high computationally costs.

A solution to these disadvantages is the quantum mechanics/molecular mechanics (QM/MM) method. Following its initial presentation in 1976^[157] the combination of QM and MM was awarded the Nobel Prize in Chemistry in 2013. The hybrid method combines the two methods and their strengths, the accuracy of QM and the low-computational cost of MM. In doing so, the molecular system is divided into two parts: A QM part describing the reactive region, e.g. the active site and its neighbors in an enzyme, at QM level and an MM part describing the influence of the surrounding environment, e.g. the protein environment, at the MM level.

Table 1: Overview of dynamical processes and their respective time regime.^{[152]–[154]}

Dynamic event	time range
vibrational motions	fs to ps
rotational motions	ps to ns
ligand binding/unbinding	ns to s
protein folding/unfolding	ms to s
aggregation	s

The aim of this thesis is to use computational methods to investigate the energy transfer mechanism in CP29. Specifically, we developed a valid workflow to obtain the necessary properties without neglecting the electrostatic environment of the individual chromophores.

Computational workflow

This general workflow was used in all presented publications of sec. 4. The aim of this section is to familiarize the reader with methods that are explained in section Theory and how they were applied in the workflow. Fig. 8 shows a schematic diagram of this workflow with an experimentally acquired structure (e.g., crystal structures or structures obtained by cryogenic electron microscopy) as starting point. These structures have to be repaired if parts of the structures are not resolved by the experiment. Further, if the system under investigation is a membrane protein, it has to be embedded in a membrane. The embedded system goes through a three-step equilibration phase consisting of a NPT, NVT and an energy minimization step (see sec. 2.1). The equilibration is necessary to perform MD simulations under required conditions, so for example the temperature should be above the phase transition temperature of the membrane lipids. Ideally, the correct equilibrium ensemble appropriate for the target thermodynamic conditions is obtained after the equilibration phase.

In the second stage of the workflow procedure in Fig. 8, the system is in a state ready for MD simulation. The MD simulation describes the dynamical behavior of a system and the structures passed are closer to the *in vivo* reality than for example a static crystal structure. The MD run generates a large set of structures, each of which can be extracted. Each extracted structure represents a snapshot of the simulation and therefore a possible conformation of the system. In the following extraction step, snapshots were taken with a constant time step beginning at a sufficient time for the system to be in an equilibrated state. This extraction of snapshots is done every 5 ns without further cluster analysis to obtain random structures.

The snapshots represent structures of the whole system and are used to obtain the initial structures of the QM/MM ensemble. For every chromophore in every snapshot, QM/MM optimizations and time-dependent density functional theory (TD-DFT) calculations were performed. The spectral properties obtained by the TD-DFT calculations were analyzed and used for calculating the resonance energy transfer between the chromophore using a Förster resonance energy transfer (FRET) scheme or a transition density cube (TDC) scheme. To investigate the influence of the electrostatic environment on the respective chromophore, we did the TD-DFT calculations twice, once including the point charge field (PCF) and once excluding the PCF. The PCF represents the electrostatic of surrounding protein scaffold, water molecules and other

chromophores and is technically identical to the MM zone of the QM/MM calculations.

The results of this work will provide a deeper understanding of the energy transfer mechanism in CP29 and the role of electrostatic environment of the chromophores. This knowledge can be used to design new light-harvesting materials and improve the efficiency of artificial photosynthesis devices. It can also contribute to our understanding of the fundamental principles of energy transfer in biological systems. The steady interest in renewable energy and the steady increase in energy consumption have contributed to the rise in interest in studies of photosynthesis in recent decades. Understanding the functioning and mechanisms of the individual components of the light-harvesting machinery could be inspirations or breakthroughs in method or material development.

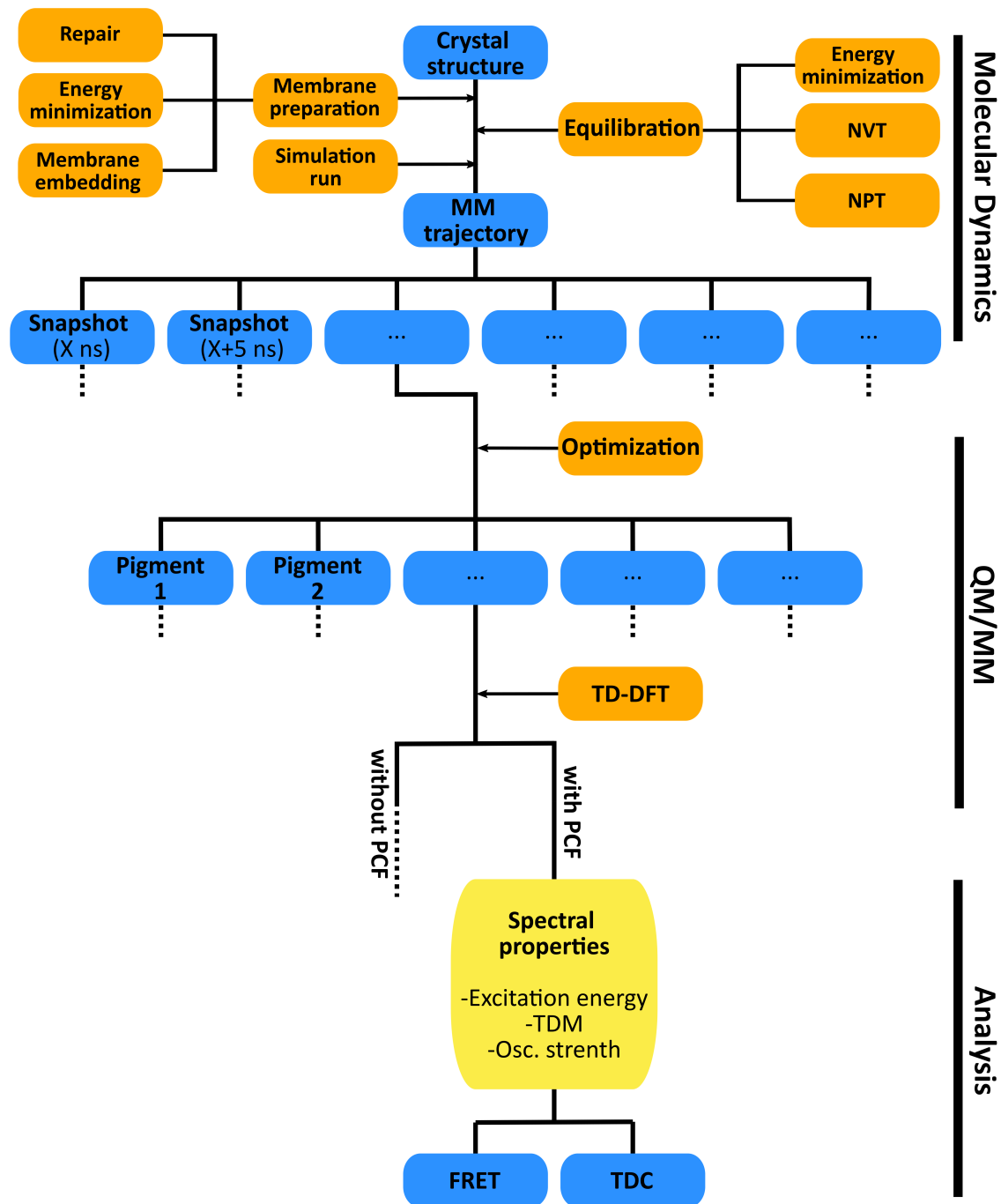


Figure 8: Schematic representation of the general workflow, which was used in the thesis publications.

2 Theory

This section is not intended to be exhaustive, but to provide the reader with the understanding of the theory and the methods used. Since the hybrid method QM/MM is an essential part of the thesis, it is necessary to give an overview of its components QM and MM as well as the method itself.

2.1 Molecular mechanics

“Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.”

“A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed.”

-Sir Isaac Newton

MM is a simple and fast way to evaluate molecular systems.^[158] MM is thus a very efficient way to obtain a large set of subsequent conformational changes, as occurring during a MD. However, MD based on QM^{[159]–[161]} is also possible but prohibitive for larger biological systems such as studied here. In MD simulations a system of atoms is propagated following the laws of classical mechanics. This propagation is based on the integration of Newton’s second law of motion

$$F_i = m_i \frac{d^2 \vec{R}_i}{dt^2} = m_i a_i. \quad (2.1)$$

F_i describes the force acting on the i -th particle with mass m_i and acceleration a_i . Each atom of the simulation is represented as a mass point with the position \vec{R}_i and the charge q_i . In a Cartesian coordinate system, \vec{R}_i is composed of the unit vectors \vec{e}_x , \vec{e}_y and \vec{e}_z (**Eq.** (2.2)). The totality of all positions of all atoms N_{atoms} at the time t describes a configuration of the system and is defined in **Eq.** (2.3) as $\vec{R}(t)$.

$$\vec{R}_i = x\vec{e}_x + y\vec{e}_y + z\vec{e}_z \quad (2.2)$$

$$\vec{R}(t) = \left\{ \vec{R}_i(t) \right\}_{i=1}^{N_{\text{atoms}}} \quad (2.3)$$

To each atom belongs position $\vec{R}_i(t)$, a velocity $\vec{v}_i(t)$ and an acceleration $\vec{a}_i(t)$, which are given by the relations

$$\frac{d\vec{R}_i}{dt} = \vec{v}_i(t) \quad (2.4)$$

and

$$\vec{a}_i(t) = \frac{d\vec{v}_i}{dt} = \frac{d^2 \vec{R}_i}{dt^2}. \quad (2.5)$$

The integration with respect to time t is described by a set of discrete time steps with a time interval Δt . For this purpose, the position and the velocity at the time $t + \Delta t$ are approximated by a Taylor series

$$\vec{R}_i(t \pm \Delta t) = \sum_{n=0}^{\infty} (\pm 1)^n \frac{\vec{R}_i^{(n)}(t)}{n!} \Delta t^n, \quad (2.6)$$

$$\vec{v}_i(t \pm \Delta t) = \sum_{n=0}^{\infty} (\pm 1)^n \frac{\vec{v}_i^{(n)}(t)}{n!} \Delta t^n. \quad (2.7)$$

Here, $\vec{R}_i^{(n)}(t)$ is the n -th derivative of \vec{R}_i at time t . **Eq.** (2.1), (2.4) and (2.5) can be used to update the positions for the atoms by propagating the positions in time with respect to the current velocities and forces. The resulting error scales with Δt^3 neglecting higher order terms.

Numerical integrator

A method to numerically solve Newton's equations of motion is the Verlet algorithm, which was used for the MD simulations of the presented workflow (see sec. 1.3). The Verlet algorithm considers one step forward in time ($t + \Delta t$) and one step backward in time ($t - \Delta t$). The position $\vec{R}_i(t + \Delta t)$ is given by

$$\vec{R}_i(t + \Delta t) = 2\vec{R}_i(t) - \vec{R}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t^2 + O(\Delta t^4). \quad (2.8)$$

Although the local error of the position is $O(\Delta t^4)$, the global (cumulative) error over a constant time interval is $O(\Delta t^2)$. However, the new position depends only on the acceleration or force and no longer on the velocity.

A more robust description is provided by using the leapfrog integrator. Here, the position $\vec{R}_i(t)$ is propagated in increments of $\Delta t/2$ and one step forward and backward in time is considered,

$$\vec{R}_i(t + \Delta t) = \vec{R}_i(t) + \vec{v}_i(t + \frac{\Delta t}{2}) \Delta t + O(\Delta t^3). \quad (2.9)$$

The algorithm describes the velocities at times $t \pm \Delta t/2$,

$$\vec{v}_i(t + \frac{\Delta t}{2}) = \vec{v}_i(t - \frac{\Delta t}{2}) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t + O(\Delta t^3), \quad (2.10)$$

where the position and velocities are calculated at different times and leap over each other (**Eq.** (2.9) and (2.10)). The resulting error scales with Δt^3 . Thus, the velocity is explicitly used to determine the updated position. The kinetic energy E_{kin} can be determined using the average of the velocities $\vec{v}(t + \Delta t/2)$ and $\vec{v}(t - \Delta t/2)$.

$$E_{\text{kin}}(t) = \frac{1}{2} m \vec{v}(t)^2 \quad (2.11)$$

$$\vec{v}(t) = \frac{\vec{v}(t + \Delta t/2) + \vec{v}(t - \Delta t/2)}{\Delta t} \quad (2.12)$$

Regardless of the method, the choice of the time step is essential as the computational cost scales with the number of steps. Too large chosen time steps have a numerical inaccuracy of the integration process and too small chosen time steps result in too short simulation times or too much computational costs. The numerical instabilities caused by too large time steps can lead to atoms coming too close to each other and the resulting energies and gradients are too high. As a direct consequence, the calculated forces between the atoms are too large, which can lead to unphysical behavior. For

biochemical systems, typical time steps are 0.5 fs, since the fastest motion is an X-H bond vibration, which is in the range of 10 fs.^[152] With the help of constraint algorithms, time steps of 2 fs can be achieved. SHAKE and LINCS represent such algorithms which constrain the X-H distance to a fixed value.^{[162],[163]}

Force field

The force F_i acting on the i -th atom depends on the connectivity and the environment and can be represented as

$$F_i(t) = -\nabla_i V(\vec{R}(t)) \quad (2.13)$$

with the negative gradient ∇_i . The potential $V(\vec{R}(t))$ is called a force field and describes a Born-Oppenheimer potential energy surface (PES) on which the movement of atom i takes place. The potential describes all bonded interactions $V_b(\vec{R})$, resulting from the connectivity of the atoms, and non-bonded interactions $V_{nb}(\vec{R})$, which describe the environment.

$$V(\vec{R}(t)) = V_b(\vec{R}(t)) + V_{nb}(\vec{R}(t)) \quad (2.14)$$

The potential describing the bonded forces can be divided into a bond stretching, angle bending and terms describing the rotation of torsional angles (dihedral or improper dihedral)

$$V_b(\vec{R}(t)) = \sum_j^{N_{\text{bonds}}} V(l_j(t)) + \sum_j^{N_{\text{angles}}} V(\theta_j(t)) + \sum_j^{N_{\text{dihed.}}} V(\omega_j(t)) + \sum_j^{N_{\text{imp.dihed.}}} V(\zeta_j(t)). \quad (2.15)$$

A bond stretching is usually described as a harmonic potential, which according to Hooke's law can also be considered as a spring between the two binding particles (see Fig. 9). The bond or spring is represented by a force constant k_{l_j} , which reflects the bond strength, and a reference bond length $l_{0,j}$, which corresponds to the equilibrium distance or the bond length of the bond j if all other terms in **Eq.** (2.15) were zero. The harmonic approximation is sufficient in most cases as the bond lengths rarely deviate strongly from the reference bond length. Similarly, the angle bending potential of three particles can also be described with a harmonic potential (see Fig. 9). The picture of a spring also works here, with the spring connecting the two outer particles of the set. Dihedrals are part of the bonded interactions and are divided into "proper" dihedrals, around bonds connecting groups of atoms, and "improper" dihedrals, which involve the neighbors of the central atom. Improper dihedrals are mostly used to ensure proper planarity or non-planarity of a certain group (e.g. the planarity of an aromatic ring). The potential of the "proper" dihedral is represented as a cosine function, related to the dihedral angle ω_j , and with the help of the force constant k_{ω_j} . The number of minima is defined by the number of bond partners or the hybridization of the bonded atoms. Thus, for two sp^3 hybridized bonded atoms, one uses a triple-well potential. Equivalent to the bond stretching and angle bending term, the improper dihedral term is described by a harmonic potential with a force constant k_{ζ_j} related to a reference dihedral angle $\zeta_{0,j}$. Depending on the force field, further terms can occur, for example to describe coupling of different motions.

The non-bonded interactions usually include the point-charge Coulomb electrostatic

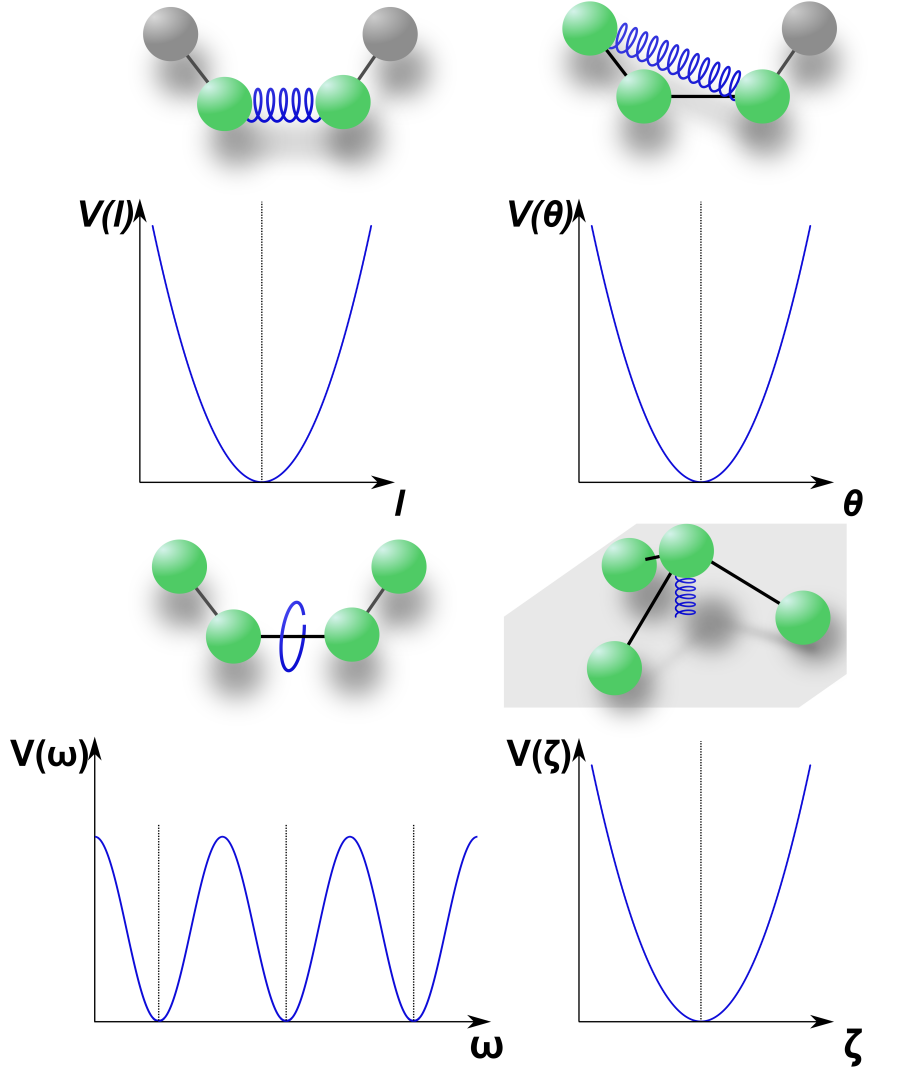


Figure 9: Schematic representation of the bonded interactions: bond stretching (top left), angle bending (top right), rotation (bottom left) and torsional angles (bottom right), and their corresponding force field potentials.

interactions and dispersive or Van-der-Waals (VdW) interactions,

$$V_{\text{nb}}(\vec{R}(t)) = \sum_i^{N_{\text{atoms}}} \sum_{j>i}^{N_{\text{atoms}}} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}(t)} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}(t)} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}(t)} \right). \quad (2.16)$$

Here, the VdW interactions are described by a Lennard-Jones potential (left term in **Eq.** (2.16)), where $R_{ij}(t) = |\vec{R}_i(t) - \vec{R}_j(t)|$. The Lennard-Jones potential is composed of a repulsive term, which decreases with R^{-12} , and an attractive term, which decreases with R^{-6} . The repulsive term is called Pauli repulsion, which includes the Pauli exclusion principle. The attractive term describes dispersive interactions, changes of the electron density causing dipoles and the resulting dipole-dipole interactions. ϵ_{ij} describes the potential depth and σ_{ij} the zero of the Lennard-Jones potential as well as the position of the minimum ($R_{0,ij} = 2^{1/6}\sigma_{ij}$).

The Coloumb potential (right term in **Eq.** (2.16)) describes the electrostatic interaction between atoms, with a point charge q_i attributed to each atom. In contrast to the Lennard-Jones potential, the Coloumb potential represents a long-range interaction which decreases with R^{-1} . The parameter ϵ_0 describes the permittivity of the vacuum.

The most time-consuming part during an MD simulation is the calculation of the non-bonded interactions as they scale with the square of the number of atoms. Therefore, a reduction of the number of interactions is necessary for calculating larger systems efficiently. A common practice to speed up the calculation is to apply a cut-off radius beyond which no interactions are evaluated. Another strategy is Ewald summation, which rewrites the non-bonded interaction potentials as a sum of a short-range term converging quickly in real space and a long-range term converging quickly in Fourier space. Replacing a slowly converging sum by two fast converging sums avoids computational artifacts that would be introduced by cut-off schemes.

Periodic boundary conditions

The most common way to avoid unphysical behaviour of the atoms close to the border of the finite simulation box is to apply periodic boundary conditions (PBC). The simulation box is surrounded by translated copies of itself. Practically this means that if a particle is about to leave the simulation box a replica will automatically enter from the opposite side of the box. The minimum image convention is used to avoid an atom to interact with itself or with several copies of another atom introduced by the PBC. Therefore, the method reduces the non-bonded interactions to the interactions with the closest image of the same atom in different replica.

Ensembles theory

Ensembles represent the set of all states of a system defined by certain thermodynamic parameters such as: constant number of particles (N), temperature (T), volume (V), pressure (P) and energy (E). A statistical ensemble is a set of possible states of a system of particles, which in their entirety describe a single system and is thus a tool of statistical mechanics. There are two classes of ensembles, closed and open systems. Closed systems ensembles include microcanonical (NVE), canonical (NVT), and isothermal-isobaric (NPT) ensembles. In contrast, open systems (grand canonical ensembles) describe systems in which the number of particles can change.

The NVE ensemble is an isolated system in which all states have the same number of particles, the same volume, and the same energy. The microcanonical ensemble is rarely used to describe molecular systems, since it would correspond to a system inside a completely closed box, which does not allow energy and particle exchange with the environment and on which no external fields act.

The NVT ensemble describes an idealized system exchanging energy with a thermal source, where the number of particles (N), the volume (V) and the temperature (T) are constant. An experimental realization of these conditions represents a closed system in a thermal bath. Conversely, NPT is a isothermal–isobaric ensemble corresponding most closely to laboratory conditions.

In MD simulations especially the NVT and the NPT ensemble play a essential role as they are used to equilibrate the system beforehand. Assuming a reasonable starting structure, in terms of structural orientation and solvent orientation, the system may

collapse during an unrestrained dynamic simulation as the solvent and the solute are usually not optimized with respect to each other. To prepare the starting structure for the required temperature and pressure, the equilibration is conducted in two phases (see Fig. 8). In the first phase, a thermostat is used to simulate the system at a constant temperature (NVT). One method is to scale the velocities of the atoms during the simulation such that the average kinetic energy is constant and correspond to the kinetic energy of the given temperature. Note that the thermostat is not a simulation of a system coupled with a thermal reservoir as the kinetic energy is distributed independently of the distance of an atom from the simulation box. In the second phase, a barostat is added to stabilize the pressure and thus the density of the system (NPT). Similar to the thermostat, a barostat scales the volume of the simulation box and consequently the position of the particles. A decrease of the size of the simulation box is a compression of all particles.

2.2 Fundamentals of quantum mechanics

QM represents a fundamental theory for predicting the behavior of atoms, electrons, photons and other subatomic particles. The term *quantum* derives from the Latin quantum, meaning “how much”, and describes the minimum amount of any physical entity involved in a change of state. The hypothesis of quantization states that every physical property can be “quantized” and therefore, the magnitude of the physical property can take only discrete values. The word quantum was also used and characterized by Max Planck. Although he doubted their existence, he needed quanta to reconcile his black-body radiation law with experiments.^[164] In QM, a quantum is not a necessarily particle of materialistic nature, *i.e.*, photons or quasi-particles. The different quanta are defined by discrete values of various physical quantities such as energy or momentum. None of this should be confused with the quanta of energy described by Ludwig Boltzmann, because these were just operands without any physical meaning.

2.2.1 Wave function

In classical mechanics the motion of a particle is described by Newton’s second law. **Eq. (2.1)** allows a precise prediction of the behavior of a particle with given initial conditions. In QM, such a straightforward prediction of a particle’s behavior is not possible. Based on the wave-particle duality, QM postulates for each state of a system with position \vec{r} and time t a function $\Phi(\vec{r}, t)$, which fully describes the system.² The function $\Phi(\vec{r}, t)$ is an equation of state. Furthermore, it was postulated that the wave function of a systems $\Phi(\vec{r}, t)$ is a solution of the time-dependent Schrödinger equation (SE),

$$i\hbar \frac{\partial}{\partial t} \Phi(\vec{r}, t) = \hat{H} \Phi(\vec{r}, t). \quad (2.17)$$

Here the Hamiltonian \hat{H} defines the total non-relativistic energy of the system. For stationary systems, the Schrödinger wave function $\Phi(\vec{r}, t)$ can be separated into a time-dependent $\Theta(t)$ and space-dependent term $\Psi(\vec{r})$.^[165] which leads to a time-dependent

²For simplification the spin dependence is neglected in this chapter and will be discussed in chapter 2.3.

and time-independent SE. The time-independent SE,

$$\hat{H}\Psi(\vec{r}) = E\Psi(\vec{r}) \quad (2.18)$$

contains the time-independent Hamiltonian \hat{H} , which is for molecules composed of the kinetic energies of the nuclei \hat{T}_K and the electrons \hat{T}_E on the one hand and of the Coulomb interactions of the nuclei \hat{V}_{KK} , the electrons \hat{V}_{EE} and the electrons with the nuclei \hat{V}_{EK} on the other hand.

$$\hat{H}(\vec{R}_A, \vec{r}_i) = \hat{T}_K(\vec{R}_A) + \hat{T}_E(\vec{r}_i) + \hat{V}_{KK}(\vec{R}_A, \vec{R}_B) + \hat{V}_{EE}(\vec{r}_i, \vec{r}_j) + \hat{V}_{EK}(\vec{R}_A, \vec{r}_i) \quad (2.19)$$

Solving the time-independent SE represents a $(3(N+n))$ -dimensional problem, because the Hamiltonian depends on all coordinates of the nuclei \vec{R}_A and all coordinates of the electrons \vec{r}_i . N is the number of the atoms and n the number of the electrons.

2.2.2 Born-Oppenheimer approximation

The Born-Oppenheimer approximation (BOA), which was formulated in 1927 by Max Born and Robert Oppenheimer, is an approximation to simplify the time-independent SE for systems containing at least two nuclei and one electron.^[166] The idea was a separation of the coordinates of the nuclei and the electrons to reduce a $(3(N+n))$ -dimensional problem to a $(3N)$ - and a $(3n)$ -dimensional problem. As electrons are much faster than the nuclei because of their lower mass ($\approx 10^4 : 1$), the nuclei can be considered as stationary with respect to the movement of the electrons. Therefore, the wave function can be separated into an electronic and a nuclear part,

$$\Psi_{\text{BOA}}(\vec{r}_i, \vec{R}_A) = \Psi_{\text{el}}(\vec{r}_i; \vec{R}_A) \Psi_{\text{nuc}}(\vec{R}_A). \quad (2.20)$$

While the electronic wave function is dependent on the positions of the electrons \vec{r}_i and parametrically on the positions of the nuclei \vec{R}_A , the nuclear wave function is only dependent on the positions of the nuclei \vec{R}_A . For the electronic Hamiltonian \hat{H}_{el} , the kinetic energy operator of the nuclei (\hat{T}_K) vanishes and the Coulomb interactions operator of the nuclei (\hat{V}_{KK}) is constant.

$$\hat{H}_{\text{el}}(\vec{r}_i, \vec{R}_A) = \hat{T}_E(\vec{r}_i) + \hat{V}_{EK}(\vec{r}_i, \vec{R}_A) + \hat{V}_{EE}(\vec{r}_i, \vec{r}_j) + \hat{V}_{KK}(\vec{R}_A, \vec{R}_B) \quad (2.21)$$

Solving the electronic SE,

$$\hat{H}_{\text{el}}\Psi_{\text{el}}(\vec{r}_i; \vec{R}_A) = E_{\text{el}}(\vec{R}_A) \Psi_{\text{el}}(\vec{r}_i; \vec{R}_A) \quad (2.22)$$

a set of eigenfunctions $\Psi_{\text{el}}(\vec{r}_i; \vec{R}_A)$ and eigenenergies $E_{\text{el}}(\vec{R}_A)$ are obtained. The nuclear SE describes the energy of the nuclei and is composed of the kinetic energy operator of the nuclei (\hat{T}_K) and the eigenenergies ($E_{\text{el}}(\vec{R}_A)$) of the electronic SE.

$$\hat{H}_{\text{nuc}}\Psi_{\text{nuc}}(\vec{R}_A) = E_{\text{nuc}}(\vec{R}_A) \Psi_{\text{nuc}}(\vec{R}_A) = [\hat{T}_K(\vec{R}_A) + E_{\text{el}}(\vec{R}_A)] \Psi_{\text{nuc}}(\vec{R}_A) \quad (2.23)$$

Accordingly, the nuclei are moving along a potential energy surface, which is defined by the solution of the electronic SE. The description of the potential energy for a many-electron-system is complicated, because in such a system every repulsive interaction between electrons and every attractive interaction of electrons with nuclei should be taken into consideration. In general, the BOA introduces extremely small errors. However, the decoupling of the electronic and nuclear motion can introduce larger inaccuracies for some situations, for example at conical intersections.^[167] The BOA is the basis for the concept of PESs, which describe a static solution of **Eq.** (2.22) by varying \vec{R}_A .

2.3 Hartree-Fock method

Despite the presented approximations, it is not possible to calculate chemically relevant systems at this point. The Hartree-Fock (HF) method represents an approximation to solve the electronic SE (**Eq.** (2.22)) for many-electron systems.^[168] Furthermore, the method is a precursor for many common QM methods, and will be used as a basis to explain the approaches actually employed in the thesis publications. For simplification we define $\Psi_{\text{el}} =: \Psi$ in the following. In the HF approach, the electron-electron interaction is replaced by the interaction of one electron with an effective field (mean-field approximation). The mean-field approximation is a powerful tool in quantum chemistry, which reduces a complicated n -particle problem to n simpler one-particle problems. Therefore, the approach requires to develop the many-body wave function $\Psi(\vec{r}_i; \vec{R}_A)$ from orthonormal one-electron functions χ_i . Besides the spatial dependence \vec{r}_i , these one-electron functions depend on the spin σ_i , hence they are also called spin orbitals. The spin orbital $\chi_i(\mathbf{x}_j)$ can then be described as a product of a spatial and a spin function for with the collective variable $\mathbf{x}_j = \{\vec{r}_j, \sigma_j\}$,

$$\chi_i(\mathbf{x}_j) = \psi(\vec{r}_j)\sigma_j. \quad (2.24)$$

The spin function can be one of two orthonormal states α and β ("spin up" and "spin down") with

$$\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1 \quad \text{and} \quad \langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle = 0. \quad (2.25)$$

It is possible to describe Ψ as a determinant of the spin orbitals χ_i ,

$$|\Psi(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \dots, \vec{\mathbf{x}}_n)| = \sqrt{\frac{1}{n!}} \begin{vmatrix} \chi_1(\vec{\mathbf{x}}_1) & \chi_2(\vec{\mathbf{x}}_1) & \dots & \chi_n(\vec{\mathbf{x}}_1) \\ \chi_1(\vec{\mathbf{x}}_2) & & & \vdots \\ \vdots & & & \vdots \\ \chi_1(\vec{\mathbf{x}}_n) & \dots & \dots & \chi_n(\vec{\mathbf{x}}_n) \end{vmatrix} \quad (2.26)$$

This so-called Slater determinant, named after John C. Slater, satisfies the anti-symmetry requirement enforced by the Pauli principle. Each row represents one electron in a different spin orbital and each column accordingly represents one spin orbital, which is occupied by a different electron per row, respectively. The indistinguishability of the electrons can be shown by interchanging two rows or two columns, which leads to a simple sign change rather than a change in value. Slater determinant obeys the Pauli exclusion principle, which can be seen if two columns are equal, because then the value of the determinant is zero and therefore this is no state of the

system, thus prohibiting a scenario where two electrons are sharing both the same location as well as quantum numbers.

In the following, only closed-shell systems are examined. This means that the spin orbitals χ_i have the same spatial function. Therefore, an n -electron system with n different spin orbitals χ_i can be reduced to system with $n/2$ different spatial orbitals. The Fock operator \hat{f} for these systems describes the electronic Hamiltonian consisting of a one-electron term \hat{h} , a Coulomb-interaction operator \hat{J}_a and an exchange operator \hat{K}_a . In shorthand notation for electron 1, the Fock-operator reads

$$\hat{f}(1) = \hat{h}(1) + \sum_a^{n/2} \left(2\hat{J}_a(1) - \hat{K}_a(1) \right) \quad (2.27)$$

with the one-electron term

$$\hat{h}(1) = -\frac{1}{2}\nabla_{r_1}^2 - \sum_A^K \frac{Z_A}{r_{1A}} \quad (2.28)$$

and the two-electron terms describing the interaction of electron 1 with electron 2

$$\hat{J}_a(1) = \int \psi_a^*(2) \frac{1}{r_{12}} \psi_a(2) d\vec{r}_2 \quad (2.29)$$

$$\hat{K}_a(1)\psi_i(1) = \left[\int \psi_a^*(2) \frac{1}{r_{12}} \psi_i(2) d\vec{r}_2 \right] \psi_a(1) \quad (2.30)$$

using $r_{12} = |\vec{r}_1 - \vec{r}_2|$. The interactions define only a part of the situation by means of an effective mean-field potential. The specific interaction of one electron with another electron will be discussed later in sec. 2.3.3. As a result of the variational principle the ground state energy is obtained from the minimization with respect to $d/d\psi$

$$\begin{aligned} \langle \Psi_0 | \hat{H} | \Psi_0 \rangle &= 2 \sum_a \langle \psi_a(1) | \hat{h} | \psi_a(1) \rangle \\ &+ \sum_{a,b} \left(2 \langle \psi_a(1)\psi_a(2) | \frac{1}{r_{12}} | \psi_b(1)\psi_b(2) \rangle - \langle \psi_a(1)\psi_b(2) | \frac{1}{r_{12}} | \psi_b(1)\psi_a(2) \rangle \right). \end{aligned} \quad (2.31)$$

Practically a Lagrange multiplier λ_{ai} is introduced to impose the condition that the ψ are normalised

$$\left[\hat{h}_i + \sum_a \left(2\hat{J}_a + \hat{K}_a \right) \right] \psi_i = \lambda_{ai} \psi_i. \quad (2.32)$$

The electronic SE can be solved for many-electron systems by using the HF method, provided that the form of the spin orbitals is well known. Solving the integro-differential equation is rather difficult and therefore, represents a main disadvantage of the HF approach. As the implementation of these integro-differential equation is quite involved in quantum-chemical software, an expansion of the spatial orbitals in a finite set of L basis functions is used.^[169] Therefore, the molecular orbitals (MOs) ψ_a^{MO} are described by the linear combination of atomic orbitals (LCAO) framework as a combination of atomic orbitals (AOs) ψ_ν^{AO} to generate the required spatial orbitals

for the HF method,

$$\psi_a^{MO} = \sum_{\nu}^L C_{a\nu} \psi_{\nu}^{AO}. \quad (2.33)$$

ψ_{ν}^{AO} represent the basis functions and are atom-centered hydrogen-like orbitals. The coefficients $C_{a\nu}$ define the share of the contribution of the atomic orbitals ψ_{ν}^{AO} of the molecular orbitals ψ_a^{MO} . The shape of the molecular orbitals and their respective energies are obtained exactly for a "complete" basis set of atomic orbitals. Since complete basis sets cannot be used, it is only possible to determine an approximate ground state energy for a finite basis set. The result of the application of the Fock operator to the molecular orbitals ψ_a^{MO} is the energy ϵ_a of the molecular orbitals

$$\hat{f}(1)\psi_a^{MO} = \epsilon_a \sum_{\nu}^L C_{a\nu} \psi_{\nu}^{AO}. \quad (2.34)$$

2.3.1 Roothaan-Hall equation

To determine the spin orbitals, an iterative method (self-consistent-field (SCF) method)^[170] is used in practice. An effective potential is determined by means of a suitable starting set of one-electron wave functions. The solution of the SE of the effective potential yields a new set of one-electron wave functions, which represent a better approximation to the system. Once the calculation reaches a pre-defined threshold, for example a total energy threshold, between two iterations, the result is treated as converged. This method is used to solve the Roothaan-Hall equation, which was established by Clemens C. J. Roothaan and George G. Hall.^{[169],[171]} The Roothaan-Hall equation results from **Eq.** (2.32) and **Eq.** (2.33),

$$\sum_{\nu} F_{\mu\nu} C_{a\nu} = \epsilon_a \sum_{\nu} S_{\mu\nu} C_{a\nu}. \quad (2.35)$$

Here, $F_{\mu\nu}$ is one element of the Fock matrix

$$F_{\mu\nu} = \int \psi_{\mu}^{*AO} \hat{f}(1) \psi_{\nu}^{AO} d\vec{r}_1 \quad (2.36)$$

and $S_{\mu\nu}$ one element of the overlap matrix of the a -th molecular orbital ψ_a^{MO}

$$S_{\mu\nu} = \int \psi_{\mu}^{*AO} \psi_{\nu}^{AO}. \quad (2.37)$$

Taking into account all molecular orbitals the following expression can be obtained

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (2.38)$$

with ϵ as diagonal matrix with the orbital energies ϵ_{aa} as diagonal elements. It is possible to solve the Roothaan-Hall equation iteratively as an eigenvalue equation by means of the mentioned SCF method to obtain the molecular (spin) orbitals and their respective energies.

2.3.2 Limits of HF

The limits of the HF method are related to the lack of consideration of the explicit electron correlation. Even if an infinite basis set is used, the HF equation would not provide the energy exactly, because the equation for each electron is solved individually with an electron being in the mean field of the other electrons. This is the so-called HF limit. One example represents the spatial probability density function of two different electrons $P(\vec{r}_1, \vec{r}_2)$. The function is defined as the product of the probability density of the first and of the second electron

$$P(\vec{r}_1, \vec{r}_2) = P(\vec{r}_1)P(\vec{r}_2). \quad (2.39)$$

This would imply that the probability to find one electron at a position is independent of the position of the other electron. But, according to Coulomb's law two particles with the same charge repel each other with a force that is reciprocal to the squared distance. The probability to find both electrons with the same spin at the same position at the same time is called Fermi heap or Fermi hole. The reason for the HF limit is that the method uses an average potential which cannot describe the explicit electron correlation. To take these electron correlations into account, further methods were developed.

2.3.3 Electron correlation

Although no post-HF methods were used in the thesis publications, a brief introduction methods including electron correlation is given. Usually, a distinction is made between dynamical and non-dynamical (static) correlation.^{[172],[173]} Dynamical correlation describes the instantaneous interaction of two electrons in close spatial proximity, while the static correlation describes the error occurring due to the poor representation of a many-electron system with a single Slater determinant. There are various types of approaches for a correction. A widely used and straightforward approach is the configuration interaction (CI). The CI method is conceptually simple and involves first- and higher-order corrections to the HF wave function that mix in elements of MOs, found in excited states. The CI wave function $|\Psi^{\text{CI}}\rangle$ is a linear combination of several states of the system,

$$|\Psi^{\text{CI}}\rangle = c_0 |\Psi_0\rangle + \sum_{a,a'} c_a^{a'} \Psi_a^{a'} + \sum_{a,b,a',b'} c_{ab}^{a'b'} \Psi_{ab}^{a'b'} + \dots \quad (2.40)$$

Here, Ψ_0 is the reference matrix, which is obtained by the HF method, and $\Psi_a^{a'}$ is a single excited Slater determinant, which describes the excitation of one electron from the spin orbital χ_a into the spin orbital $\chi_{a'}$. The so-called *full*-CI method considers all possible excitations of the system. Since the computational effort can become very high even with a small number of electrons, for larger systems just the excitations to a certain degree are taken into account. This restricted calculation is referred to as *truncated*-CI method. However, the *truncated*-CI method is the missing size consistency.

Another example is the coupled cluster (CC) method, which reduces similar to CI the SE to a matrix eigenvalue problem due to an expansion of the many-body wave function in a basis of Slater determinants. However, CI uses a linear excitation operator,

whereas CC uses a more complicated, exponential one. The Møller–Plesset perturbation theory also represents another post-HF method which is based on the assumption that the Hamiltonian consists of an undisturbed part and a disturbed part arising from electron correlations.^[174] In contrast to the CI method the Møller–Plesset equations does not represent a variational solution of the SE and therefore does not provide a limit for the energy, but it is size consistent. The multi-configurational self-consistent-field (MCSCF) method can be understood as a combination of the CI and the HF methods, because the method varies the set of coefficients of the excited Slater determinants like the CI method and varies the coefficients of the LCAO like the HF method. Therefore, the total electronic wave function with the lowest energy is obtained.

2.4 Density functional theory

Density functional theory (DFT) represents a conceptually different alternative to the HF and the post-HF methods as it is formulated based on the electron density rather than the wave function. In many cases the method provides results comparable to those of the post-HF methods with lower computation costs. However, the results are sensitive to the respective actual implementation of density functional approximation (DFA). The reason for faster computations lies in the fact that any wave function-based method depends on $4n$ variables (one spin coordinate and three spatial coordinates) for an n -electron system. This causes a fast increase of the computational costs for an increase of the number of electrons. In contrast to that, the DFT calculations are just dependent on the location and not on the size of the system. DFT is based on the Hohenberg-Kohn theorems named after Pierre Hohenberg and Walter Kohn,^[175] who showed that a non-degenerate ground state of an n -electron system can be calculated with an electron density function rather than with an electronic wave function.

$$\rho(\vec{r}_1) = n \sum_{\sigma_1=\frac{1}{2},-\frac{1}{2}} \int d_{x_2} d_{x_3} \dots d_{x_n} |\Psi(\vec{r}_1, \sigma_1, \vec{r}_2, \sigma_2, \dots, \vec{r}_n, \sigma_n)|^2 \quad (2.41)$$

Here, the electron density ρ is defined by the integral of the squared wave function $|\Psi|^2$ over all spatial and spin coordinates for all electrons except for one (electron 1 in **Eq.** (2.41)) and finally adding up for the spin coordinate of this electron (σ_1). This provides the function $\rho(\vec{r}_1)$ for the electron 1 in space. However, $\rho(\vec{r})$ is always identical regardless of which electron is taken into account. Since Ψ^2 is symmetric with respect to exchanging two electrons. The volume integral for the density distribution yields the number of electrons of the system

$$\int \rho(\vec{r}) d^3\vec{r} = n. \quad (2.42)$$

2.4.1 Hohenberg-Kohn theorems

P. Hohenberg and W. Kohn proved that **Eq.** (2.41) can be inverted so that the wave function $\Psi_0(\vec{r}_1, \dots, \vec{r}_n)$ can be calculated with the given electron density of the ground state $\rho_0(\vec{r})$. As a result, the ground state energy is a unique functional of the electron

density. Using **Eq.** (2.21), the functional of the energy is

$$E_0 = E[\rho_0] = \langle \Psi[\rho_0] | \hat{H} | \Psi[\rho_0] \rangle = \langle \Psi[\rho_0] | \hat{T} + \hat{V}_{\text{EK}} + \hat{V}_{\text{EE}} | \Psi[\rho_0] \rangle . \quad (2.43)$$

The operator \hat{V}_{EK} can be described as a unique functional of the electron density of a non-degenerate ground state or as a defined potential.

$$\hat{V}_{\text{EK}}[\rho] = \int V(\vec{r})\rho(\vec{r})d\vec{r} \quad (2.44)$$

The kinetic energy operator \hat{T} and the operator of the electron-electron interaction \hat{V}_{EE} are summarized as a universal share of the Hohenberg-Kohn functional $\hat{F}[\rho]$.

$$\hat{F}[\rho(\vec{r})] = \langle \Psi | \hat{T} + \hat{V}_{\text{EE}} | \Psi \rangle \quad (2.45)$$

The second theorem of Hohenberg and Kohn is analogous to the variation principle and was proven in the year 1982 by Mel Levy^[176] for the first time.

$$E_V[\rho] \geq E_V[\rho_0] = E_0 \quad (2.46)$$

According to this, it is possible to calculate the ground state wave function Ψ_0 with a given ground state electron density $\rho_0(\vec{r})$ by using the two Hohenberg-Kohn theorems the other way around. The solution is the ground state wave function. The main problem of the approach is the Hohenberg-Kohn functional. However, there are a lot of different approximations that are specifically tailored for different systems and that that can provide good results with just a small error in comparison to the experiments. This represents the biggest difference between wave function-based methods and DFT. While the HF method and the post-HF methods use exact Hamiltonians and approximate a wave function, DFT uses an exact one-electron density and approximate the Hamiltonians. Accordingly, the wave function-based methods are systematically improvable and provide an exact result at the limit. Although, there is no systematically improve of DFAs the group of John P. Perdew^[177] has evaluated a hierarchy of DFAs (see sec. 2.4.3).

2.4.2 Kohn-Sham equation

The fictitious Kohn-Sham system by Walter Kohn and Lu Jeu Sham^[178] is a system of non-interacting electrons enabling to describe molecules with DFT. To ensure that this non-interacting system is able to recreate the ground state density of an interacting system, an additional so called *external* potential $\nu(\vec{r})$ describing the interaction is introduced, which constitutes $\rho(\vec{r})$ that has an identical shape to that of $\rho'(\vec{r})$ in a system with interactions. Thus, the potential $\nu_0(\vec{r})$ provides the ground state electron density $\rho_0(\vec{r})$. The advantages of this potential is that the electron interactions are not explicitly calculated. However, the calculation of $\nu_0(\vec{r})$ is not possible without basis function

$$\left(-\frac{1}{2}\Delta + \nu_0 \right) \psi_i = \epsilon_i \psi_i . \quad (2.47)$$

ψ_i are the Kohn-Sham spin orbitals, which are not equivalent with the canonical HF orbitals. The total wave function can be represented with a Slater determinant, which

is called the Kohn-Sham determinant. The electron density function is the sum of the square value of each spin orbital

$$\rho(\vec{r}) = \sum_i m_i |\psi_i(\vec{r})|^2. \quad (2.48)$$

Here, $m_i = 0, 1, 2$ characterizes the occupation of the orbitals in the determinant. The electronic ground state of the Kohn-Sham system can be calculated from four terms

$$E_0 = T_0 + \int \nu_0(\vec{r})\rho_0(\vec{r})d\vec{r} + J[\rho_0] + E_{xc}[\rho_0]. \quad (2.49)$$

The kinetic energy of the system is represented by the kinetic energy of the non-interacting electrons

$$T_0 = \sum_{i=1}^N \langle \psi_i | -\frac{1}{2}\nabla^2 | \psi_i \rangle. \quad (2.50)$$

The external potential $\int \nu_0(\vec{r})\rho_0(\vec{r})d\vec{r}$ also contains the electron-nuclei interaction. $J[\rho_0]$ is the classical Coulomb repulsion integral, also known as Hartree potential, defined as

$$J[\rho_0] = \frac{1}{2} \int \int \frac{\rho_0(\vec{r}_1)\rho_0(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2. \quad (2.51)$$

The discrepancy of this non-interacting system and the real one is collected in the exchange-correlation energy functional,

$$E_{xc}[\rho_0] = \Delta T[\rho_0] + \Delta J[\rho_0]. \quad (2.52)$$

The Kohn-Sham equation was determined by examining individual terms while changing the energy.

$$\left\{ -\frac{1}{2}\Delta + \nu + \nu_{\text{coul}} + \nu_{\text{xc}} \right\} \psi_i = \epsilon_i \psi_i \quad (2.53)$$

Here, $\nu_{\text{coul}}(\vec{r})$ is the sum of the Coulomb potentials and $\nu_{\text{xc}}(\vec{r})$ is the first partial derivative of the exchange-correlation energy with respect to the $\rho_0(\vec{r})$

$$\nu_{\text{coul}}(\vec{r}) = \sum_{j=1}^A \hat{J}_j(\vec{r}) \quad (2.54)$$

$$\nu_{\text{xc}}(\vec{r}) = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\vec{r})}. \quad (2.55)$$

The Kohn-Sham equation can be iteratively solved analogously to the Fock equation (**Eq.** (2.32)). Once the iterative process reaches convergence, the external potential ν_0 is obtained as the solution

$$\nu_0 = \nu + \nu_{\text{coul}} + \nu_{\text{xc}}. \quad (2.56)$$

2.4.3 Approximation for the exchange-correlation energy

At this point ν_{xc} or the exchange-correlation energy remains a problem for DFT. The group of John P. Perdew^[177] have evaluated different approximations to the exchange-correlation functionals and created a hierarchy. This hierarchy is metaphorically

known as Jacob's ladder, whereby each single rung represents a better approximation than the previous rung.

First rung: Local density approximation (LDA)

To describe the complex situation of electrons in a molecule, the LDA uses a model of a homogeneous gas in a box. This model is used very often for the treatment of bulk solids, because it avoids surface problems by the imposed periodic boundary conditions.^[179] The approximation attempts to describe the inhomogeneous behavior of the electron density distribution of a molecule as a sum of small local volumes, which can be described approximately with the homogeneous gas theory.

$$E_{xc}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r}) \epsilon_{xc}^{\text{unif}}[\rho(\vec{r})] d\vec{r} \quad (2.57)$$

Here, $\epsilon_{xc}^{\text{unif}}$ is the exact exchange-correlation energy density of a homogeneous gas with the density $\rho(\vec{r})$. A variation of the LDA is the local spin density approximation (LSDA). While the LDA depends on the electron density distribution ρ , the LSDA depends on the two spin density distributions ρ_α and ρ_β . Due to the reliable results LDA is the starting point of DFT, although LDA is a quite simple approximation. The LDA is used mainly for periodic systems like solids, because they show a rather homogeneous electron density throughout the system.

Second rung: Generalized gradient approximation (GGA)

The LDA tends to provide slightly too high binding energies and too low ground state energies of atoms. By taking the inhomogeneity of the system into account in the form of the gradient of the density, this counterbalances the trend. Beginning with the second rung of Jacob's ladder, one speaks of semi-local approximations. This is realized by introducing different functions B_{xc} , which are dependent on the densities ρ_α, ρ_β and their gradients $\nabla\rho_\alpha, \nabla\rho_\beta$.

$$E_{xc}^{\text{GGA}} = \int B_{xc}(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \quad (2.58)$$

The function B_{xc} is described in analogy to a Taylor expansion. Without further parameters, they obtained much better results. One noteworthy functional of this family is the Perdew-Wang functional (PW91).^[180]

Third rung: Meta-generalized gradient approximation (mGGA)

An mGGA functional contains, in comparison to the GGA, more details in the form of the kinetic density function $\tau(\vec{r})$ ^[181]

$$E_{xc}^{\text{mGGA}}[\rho] = \int f[\rho(\vec{r}), \nabla\rho(\vec{r}), \tau(\vec{r})] d\vec{r} \quad (2.59)$$

with

$$\tau(\vec{r}) = \frac{1}{2} \sum_i |\nabla\psi_i(\vec{r})|^2. \quad (2.60)$$

Fourth rung: Hybrid functionals

A hybrid functional is a combination of DFT exchange-correlation functionals and

HF exchange energies.

$$E_{xc}^{\text{hybrid}} = c_x E_x^{\text{HF}} + (1 - c_x) E_x^{\text{GGA}} + E_c^{\text{GGA}} \quad (2.61)$$

The coefficients c_x allow a weighted mixing of both energies and are, for the most hybrid functionals, chosen semi-empirically. One of the most well-known hybrid functionals is the B3LYP functional. It contains several different energy expressions that rely on the above considerations: the exchange energy of the LSDA, the exchange energy of the HF method, the exchange energy of the Becke potential (correction from the year 1988),^[182] the correlation energy of the Lee-Yang-Parr potential^[183] and the correlation energy of the Vosko-Wilk-Nusair potential.^[184]

$$E_{xc}^{\text{B3LYP}} = 0.8 E_x^{\text{LSDA}} + 0.2 (E_x^{\text{HF}}) + 0.72 \Delta E_x^{\text{B88}} + 0.81 E_c^{\text{LYP}} + 0.19 E_c^{\text{VWN}} \quad (2.62)$$

Hybrid functionals combines the advantages of DFT, which approximates the correlation energies and short-range exchange terms very well and HF theory, which gives the exact exchange energy. Although, hybrid functionals are able to model successfully ground- and excited-state properties^{[185],[186]} and represent a systematic performance improvement compared to semilocal DFAs like GGAs or mGGAs, they still suffer in the long-range regime. Due to a density overdelocalization, the asymptote of their exchange potential behaves as $-c_x r^{-1}$, instead of behaving as $-r^{-1}$. For example B3LYP fails for a number of important applications: (I) the polarizability of long chains, (II) excitations using TD-DFT due to the underestimation of the Rydberg excitation energies and the corresponding oscillator strengths and (III) charge transfer (CT) excitations.^{[187]-[189]} The incorrect far-nucleus exchange potential behaves as $-0.2r^{-1}$, instead of $-r^{-1}$ and therefore, it has been suggested that this is responsible for these problems.^[190]

Range-separated functionals

Range-separated functionals circumvent this drawback by additionally including a rangedependence, where specific functionals are applied depending on the inter-particle distance. Tawada and co-workers^[191] have applied the long-range correction (LC) scheme to TD-DFT and thus improved the accuracy of the excitation energies and oscillator strengths of the applications. For this, the electron repulsion operator $\frac{1}{r_{12}}$ is divided into

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\mu r_{12})}{r_{12}} + \frac{\text{erf}(\mu r_{12})}{r_{12}} \quad (2.63)$$

where the range-separation parameter μ is the slope of the two terms. The error function erf smoothly connects the first term describing the short-range interaction and the second term describing the long-range interaction.^[191] From **Eq.** (2.63) it follows that, if $\mu = 0$, LC-DFT calculations correspond to pure (non-LC)-DFT calculations and accordingly $\mu = \infty$ correspond to standard HF calculations.^[190] An alternative formulation of **Eq.** (2.63) was proposed by Yanai *et al.*^[190] by introducing the parameter α and β with

$$\frac{1}{r_{12}} = \frac{\alpha + \beta \text{erf}(\mu r_{12})}{r_{12}} + \frac{1 - [\alpha + \beta \text{erf}(\mu r_{12})]}{r_{12}}. \quad (2.64)$$

The parameter α represents the maximum percentage of HF exchange in the short-range limit, and $\alpha + \beta$ is the corresponding percentage in the long-range limit (see Fig. 10). Therefore, the range-separated exchange energy can be written as:

$$E_x^{\text{RS}} = \alpha E_x^{\text{HF}} + \beta E_x^{\text{HF}} + (1 - \alpha) E_x^{\text{DFT}} - \beta E_x^{\text{DFT}} + E_c^{\text{GGA}}. \quad (2.65)$$

To conserve the hybrid functional short-range behavior, the range-separated exchange correlation energy has to be defined as:

$$E_{xc}^{\text{RS}} = \alpha E_x^{\text{HF}} + [1 - (\alpha + \beta)] E_x^{\text{DFT}} + \beta E_x^{\text{HF}} + \beta E_x^{\text{DFT}} + E_c^{\text{DFT}} \quad (2.66)$$

The most prominent functional of this improvement is the CAM-B3LYP functional with the parameters $\mu = 0.33$, $\alpha = 0.19$ and $\beta = 0.46$. CAM-B3LYP is widely used and has found application in all presented papers of this thesis.

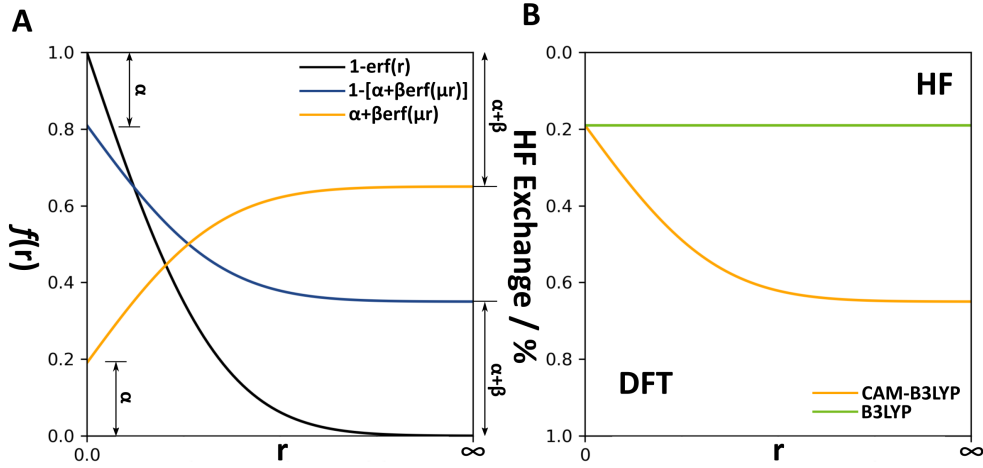


Figure 10: (A) Plots for the functions $f(r) = 1 - \text{erf}(r)$, $f(r) = 1 - [\alpha + \beta \text{erf}(\mu r)]$ and $f(r) = \alpha + \beta \text{erf}(\mu r)$ for $\alpha = 0.19$, $\beta = 0.46$ and $\mu = 0.33$ describing the contribution of $E_x^{\text{DFT}}/E_x^{\text{HF}}$ (blue/orange) of CAM-B3LYP with respect to r . (B) Schematic plots of E_x^{HF} for B3LYP (green) and CAM-B3LYP (orange), apportioned into DFT and HF.^[192]

2.4.4 Time-dependent DFT

DFT as describe above constitutes a mighty tool of quantum chemistry to investigate structural properties. Time-dependent processes and properties related with excited states require an expansion to TD-DFT. The core idea of TD-DFT is that the dynamics of each system can be described as an interaction of fermions and accordingly all details are encrypted in the time-dependent density. The typically used model is a system with n interacting not-relativistic fermions, which are moving through a time-dependent external potential $v_{\text{ext}}(\vec{r}, t)$. The total Hamiltonian of the n -particle system is given as

$$\hat{H}(t) = \hat{T} + \hat{V}(t) + \hat{W}. \quad (2.67)$$

There, the kinetic energy operator \hat{T} is equal to the one used in the time-independent case (Eq. 2.50)

$$\hat{T} = \sum_{i=1}^N -\frac{1}{2} \nabla_i^2. \quad (2.68)$$

In contrast, the potential energy operator $\hat{V}(t)$ is time-dependent

$$\hat{V}(t) = \sum_{j=1}^N v_{\text{ext}}(\vec{r}_j, t). \quad (2.69)$$

The particle-particle interaction operator \hat{W} is given as

$$\hat{W} = \frac{1}{2} \sum_{\substack{j,k \\ j \neq k}}^N w(|\vec{r}_j - \vec{r}_k|). \quad (2.70)$$

For an interaction system the Coulomb interaction is $w(|\vec{r}_j - \vec{r}_k|) = 1/|\vec{r}_j - \vec{r}_k|$ and for a non-interacting system, $w = 0$. The time-dependent many-body SE³

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}_1, \dots, \vec{x}_N, t) = \hat{H} \Psi(\vec{x}_1, \dots, \vec{x}_N, t), \quad (2.71)$$

propagates a given initial state $\Psi(t_0) \equiv \Psi_0$ within the time interval $[t_0, t_1]$ starting at the time t_0 till the final time t_1 . In general, the ground state is chosen to be the initial state and at the time t_0 an external time-dependent potential

$$v_{\text{ext}}(\vec{r}, t) = v_0(\vec{r}) + v_1(\vec{r}, t)\theta(t - t_0) \quad (2.72)$$

initiated by a Heavyside step function $\theta(t - t_0)$ starts to act on the system. This external stimulus $v_1(\vec{r}, t)$ can be a weak, uniform, oscillating electric field in semiclassical dipole approximation

$$v_1(\vec{r}, t) = -\hat{\mu}_{\text{el}}(\vec{r})F_{\text{el}}\cos(\omega t) \quad (2.73)$$

with the electric moment operator $\hat{\mu}_{\text{el}}$, the electric field strength F_{el} , and the field frequency ω . The solution of the time-dependent many-body SE (**Eq.** 2.71) can also be written as an expansion operator with respect to time

$$\Psi(t) = \hat{U}(t, t_0)\Psi_0. \quad (2.74)$$

Two of the most important properties of the operator are the function composition on the one hand and the unitarity on the other hand. The function composition means that the propagation of the system from t_0 to the time t_2 is equal to a propagation from t_0 to the time t_1 followed by a propagation from t_1 to the time t_2

$$\hat{U}(t_2, t_0) = \hat{U}(t_2, t_1)\hat{U}(t_1, t_0), \quad t_2 \geq t_1 \geq t_0. \quad (2.75)$$

Due to the unitarity the normalization is still guaranteed. If the Hamiltonian ($\hat{H}(t) = \hat{H}_0$) is a time-independent operator, a simple expression can be formulated.

$$\hat{U}(t, t_0) = e^{-i\hat{H}(t-t_0)} \quad (2.76)$$

³Note that the simplified notation of the coordinates from section 2.3 is used.

In the case of a time-dependent Hamiltonian, it is significantly more difficult and is given by

$$\hat{U}(t_2, t_0) = \hat{Y} e^{\left\{ -i \int_{t_0}^{t_2} dt' \hat{H}(t') \right\}}. \quad (2.77)$$

Here, \hat{Y} describes the time-ordering operator. Another important part of the expansion operator with respect to time is the time-dependent perturbation. For a time-dependent Hamiltonian written as

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t), \quad (2.78)$$

$\hat{H}_1(t)$ describes the time-dependent perturbation. Hence, the following equation results for the expansion operator with respect to time

$$\hat{U}(t, t_0) = e^{-i\hat{H}_0(t-t_0)} \hat{U}_1(t, t_0). \quad (2.79)$$

In combination with the time-dependent SE (**Eq.** 2.71) the following equation is obtained

$$i\hbar \frac{\partial}{\partial t} \hat{U}_1(t, t_0) = e^{i\hat{H}_0(t-t_0)} \hat{H}_1(t) e^{-i\hat{H}_0(t-t_0)} \hat{U}_1(t, t_0), \quad (2.80)$$

with the starting condition $\hat{U}_1(t_0, t_0) = 1$. The calculation of the solution is carried out iteratively. The approximation of the first order can be obtained by substituting $\hat{U}_1(t, t_0) = 1$ in the right term and integrating over the time.

$$\hat{U}_1(t, t_0) \approx 1 - i \int_{t_0}^t dt' e^{i\hat{H}_0(t'-t_0)} \hat{H}_1(t') e^{-i\hat{H}_0(t'-t_0)} \quad (2.81)$$

The perturbation can be described as an external field $F(t)$ interacting with the observable $\hat{\beta}$ as,

$$\hat{H}_1(t) = F(t) \hat{\beta} \quad (2.82)$$

with $\hat{\beta}$ defined as

$$\hat{\beta}(\tilde{t}) = e^{i\hat{H}_0\tilde{t}} \hat{\beta} e^{-i\hat{H}_0\tilde{t}}. \quad (2.83)$$

Therefore, the approximation of the first order is an expansion operator with respect to time with the form

$$\hat{U}(t, t_0) \approx e^{-i\hat{H}_0(t-t_0)} \left\{ 1 - i \int_{t_0}^t dt' F(t') \hat{\beta}(t' - t_0) \right\}. \quad (2.84)$$

This expression represents the basis of TD-DFT and is essential for linear response theory.

2.4.5 Runge-Gross theorem

The Runge-Gross theorem represents the analogue of the Hohenberg-Kohn theorem for TD-DFT.^[193] The theorem says that the electron density $\rho(\vec{r}, t)$ differs from the electron density $\rho'(\vec{r}, t)$, if both are expansions of the same initial state ψ_0 , but using different external potentials ($v_{\text{ext}}(\vec{r}, t)$ or $v'_{\text{ext}}(\vec{r}, t)$, respectively). The external potentials can be described as Taylor expansions which are evaluated at the initial time

t_0 and differ from each other by more than just one purely time-dependent additive constant.

$$\Delta v_{\text{ext}}(\vec{r}, t) = v_{\text{ext}}(\vec{r}, t) - v'_{\text{ext}}(\vec{r}, t) \neq c(t) \quad (2.85)$$

Therefore, a definite mapping exists of the electron density $\rho(\vec{r}, t)$ at an arbitrary time and the external time-dependent potential $v_{\text{ext}}(\vec{r}, t)$ for a many-body system for each initial state. The Runge-Gross theorem was proven in two parts.

The first part shows that the external potential is a functional of the current density $j(\vec{r}, t)$ using Ehrenfest's theorem.^[194] The expansion in time of the current density can be described by the Ehrenfest theorem

$$i \frac{d}{dt} \langle \Theta(t) | \hat{O}(t) | \Theta(t) \rangle = \langle \Theta(t) | \frac{\delta}{\delta t} \hat{O}(t) + [\hat{O}(t), \hat{H}(t)] | \Theta(t) \rangle, \quad (2.86)$$

for an arbitrary operator $\hat{O}(t)$. The current density is given by

$$j(\vec{r}, t) = \langle \Psi(t) | \hat{j}(\vec{r}) | \Psi(t) \rangle. \quad (2.87)$$

Here, $\hat{j}(\vec{r})$ describes the current density operator

$$\hat{j}(\vec{r}) = -\frac{i}{2} \sum_{s=1}^n \left([\nabla \psi^*(\vec{r})] \psi(\vec{r}) - \psi^*(\vec{r}) [\nabla \psi(\vec{r})] \right). \quad (2.88)$$

As a result, the expansion in time by using the external potential $v_{\text{ext}}(\vec{r}, t)$, which defines $\hat{H}_{v_{\text{ext}}}(t)$, yields

$$i \frac{\partial}{\partial t} = \langle \Psi(t) | [\hat{j}(\vec{r}), \hat{H}_{v_{\text{ext}}}(t)] | \Psi(t) \rangle. \quad (2.89)$$

Because $\Psi(t)$ and $\Psi'(t)$ are expanded from the same initial state $\psi_0(t)$, **Eq.** (2.89) provides

$$\begin{aligned} i \frac{\partial}{\partial t} [j(\vec{r}, t) - j'(\vec{r}, t)]|_{t=t_0} &= \langle \psi_0 | [\hat{j}(\vec{r}), \hat{H}_{v_{\text{ext}}}(t_0) - \hat{H}'_{v_{\text{ext}}}(t_0)] | \psi_0 \rangle \\ &= i \rho(\vec{r}, t_0) \nabla v_{\text{ext}}(\vec{r}, t_0) - v'_{\text{ext}}(\vec{r}, t_0). \end{aligned} \quad (2.90)$$

If both potentials are different at $t = t_0$, the right term of the equation is finite. Therefore $j(\vec{r}, t)$ and $j'(\vec{r}, t)$ have to be different once the time t_0 reached.

The second part of the Runge-Gross theorem shows, by using the continuity equation, that different current densities imply different electron densities.

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla j(\vec{r}, t) \quad (2.91)$$

According to **Eq.** (2.90) and to the continuity **Eq.** (2.91), the following shall apply

$$\frac{\partial^2 \rho(\vec{r}, t)}{\partial t^2} |_{t=t_0} = \nabla [\rho(\vec{r}, t_0) \nabla v_{\text{ext}}(\vec{r}, t_0) - v'_{\text{ext}}(\vec{r}, t_0)]. \quad (2.92)$$

The left term of **Eq.** (2.92) is finite for real systems, if the change of the external potentials $v_\Delta = v_{\text{ext}}(\vec{r}, t_0) - v'_{\text{ext}}(\vec{r}, t_0)$ is not everywhere even.

$$\begin{aligned} & \int d^3\vec{r} v_\Delta \nabla [\rho(\vec{r}, t_0) \nabla v_\Delta] \\ &= \int d^3\vec{r} [\nabla v_\Delta \rho(\vec{r}, t_0) \nabla v_\Delta - \rho(\vec{r}, t_0) |\nabla v_\Delta|^2] \end{aligned} \quad (2.93)$$

$v_{\text{ext}}(\vec{r})$ decays for realistic potentials with at least r^{-1} . According to Green's theorem the first term of right part of **Eq.** (2.93) vanishes. However, the second part stays negative and with an uneven $v_{\text{ext}}(\vec{r}, t_0)$ the integral has to be finite, because the second derivatives of the densities are unequal to zero. Consequently, the external potential is a functional of $\rho(\vec{r}, t)$ and of the initial wave function ψ_0 .

2.4.6 Van-Leeuwen theorem

As shown above, the Runge-Gross theorem allows density functionals to be used meaningfully in the context of time-dependent problems. However, the theorem does not guarantee that the Kohn-Sham ansatz, *i.e.*, the reproducibility of a density of an interacting system by a non-interacting system, is valid.

In the following, only the statements and the implications for TD-DFT of the van-Leeuwen theorem^[195] are described and the proof itself is not given. The van-Leeuwen theorem essentially contributes to the solution of one-to-one correspondence between $v(\vec{r})$ and $\rho(\vec{r})$, namely v -representability, for the time-dependent case. v -representability is a fundamental aspect of ground-state DFT and is part of the Hohenberg-Kohn theorem (see sec. 2.4.1). A distinction is made between the interaction and the non-interacting v -representability depending on whether the density results from a physical (interacting) or a Kohn-Sham (non-interacting) system. If a density exists, which is not v -representable, then an area of the functional $E_{v_0}[\rho]$ would be affected by errors. This would apply, for example, to problems for the description of the derivatives (**Eq.** 2.55)). Ground-state DFT can deal with this problem by using a limited searching algorithm.^{[196],[197]}

Furthermore, the theorem deals with two many-body systems with different particle-particle interactions w and proves, that an unique potential $v'(\vec{r}, t)$ exists, that provides the same time-dependent electron density $\rho(\vec{r}, t)$ for a system with $w'(\vec{r} - \vec{r}')$, like the potential $v(\vec{r}, t)$ would provide for a system with $w(\vec{r} - \vec{r}')$. This allows the use of a Kohn-Sham system in the time-dependent case.

The van-Leeuwen theorem discusses a special case, describing two systems with the same particle-particle interactions, *i.e.*, $w' = w$, which is similar to the Runge-Gross theorem system. However, this case is less general since the densities has to be Taylor expandable around t_0 , while in the Runge-Gross proof only the $v(\vec{r}, t)$ has to be Taylor expandable.^[198]

2.4.7 Linear-response TD-DFT

The most frequently used calculation scheme to derive the time-dependent Kohn-Sham equations is the linear response of a time-dependent density to a small external perturbation. The linear response time-dependent density functional theory (LR TD-DFT) considers the response of the ground state density $\rho_0(\vec{r})$ to an external time-dependent

perturbation $\delta v(\vec{r}, t)$ of the time-independent external field v_0 and is derived from perturbation theory.^[199]

$$v(\vec{r}, t) = v_0(\vec{r}) + \delta v(\vec{r}, t)\delta v(\vec{r}, t \leq 0) = 0 \quad (2.94)$$

The time-dependent electron density can be expressed as

$$\rho(\vec{r}, t) = \rho_0(\vec{r}) + \rho_1(\vec{r}, t) + \rho_2(\vec{r}, t) + \dots \quad (2.95)$$

with the indices of ρ referring to the first, second, ... order of the density and starting from the ground state $\rho_0(\vec{r}, t)$. Higher-order terms, e.g., the quadratic response, are used to describe non-linear effects such as second-harmonic generation.^{[200],[201]} In LR TD-DFT, the time-dependent density is truncated after the first term ($\delta\rho(\vec{r}, t) = \rho(\vec{r}, t) - \rho_0(\vec{r})$). Accordingly, the exchange correlation potential is defined as

$$v_{\text{xc}}[\rho_0 + \delta\rho](\vec{r}, t) = v_{\text{xc}}[\rho_0](\vec{r}) + \int \int dt' dr' f_{\text{xc}}[\rho_0](\vec{r}, t, \vec{r}', t') \delta\rho(\vec{r}', t'), \quad (2.96)$$

where the exchange-correlation kernel f_{xc} is evaluated on the ground state density

$$f_{\text{xc}}[\rho_0](\vec{r}, t, \vec{r}', t') = \left. \frac{\delta v_{\text{xc}}(\vec{r}, t)}{\delta\rho(\vec{r}', t')} \right|_{\rho=\rho_0}. \quad (2.97)$$

Since the exchange-correlation potential is a functional of the initial interacting wavefunction, of the initial Kohn-Sham wavefunction as well as of the entire history of the density, the exchange-correlation kernel is manageable, because it is only a functional of the ground-state density. The response of the ground state to a perturbation of the external field can be described with the point-wise susceptibility $\varsigma[\rho_0](\vec{r}, t, \vec{r}', t')$

$$\delta\rho(\vec{r}, t) = \int \int dt' dr' \varsigma[\rho_0](\vec{r}, t, \vec{r}', t') \delta v(\vec{r}', t'), \quad (2.98)$$

with

$$\varsigma[\rho_0](\vec{r}, t, \vec{r}', t') = \left. \frac{\delta\rho(\vec{r}, t)}{\delta v(\vec{r}', t')} \right|_{v=v_0}. \quad (2.99)$$

ς defines how a small change of the potential at \vec{r}' and t' will affect the density at \vec{r} and t . Accordingly, the function ς_{KS} describes how non-interacting Kohn-Sham electrons of the ground state Kohn-Sham system respond to a change of the Kohn-Sham potential $\delta v_{\text{KS}}(\vec{r}', t')$. Although both functions are different, both have to yield the same response of the density

$$\delta\rho(\vec{r}, t) = \int \int dt' dr' \varsigma_{\text{KS}}[\rho_0](\vec{r}, t, \vec{r}', t') \delta v_{\text{KS}}(\vec{r}', t'). \quad (2.100)$$

with the definition of the potential $v_{\text{KS}}(\vec{r}, t)$

$$v_{\text{KS}}(\vec{r}, t) = v_{\text{ext}}(\vec{r}, t) + v_{\text{H}}(\vec{r}, t) + v_{\text{xc}}(\vec{r}, t), \quad (2.101)$$

and

$$v_{\text{H}}(\vec{r}, t) = \int d\vec{r}' \frac{\rho(\vec{r}', t)}{|\vec{r} - \vec{r}'|}. \quad (2.102)$$

Equating the density change of the non-interacting (**Eq.** (2.98)) with the interacting system (**Eq.** (2.100)) taking the XC kernel into account, a Dyson-like equation for interacting density-density response function in terms of the non-interacting one is yield (in frequency space)⁴.

$$\begin{aligned} \varsigma(\vec{r}, \vec{r}', \omega) &= \varsigma_{\text{KS}}(\vec{r}, \vec{r}', \omega) \\ &+ \int \int d\vec{r}_1 d\vec{r}_2 \varsigma_{\text{KS}}(\vec{r}, \vec{r}_1, \omega) \left[\frac{1}{|\vec{r}_1 - \vec{r}_2|} + f_{\text{xc}}[\rho_0](\vec{r}, \vec{r}_2, \omega) \right] \varsigma(\vec{r}', \vec{r}_2, \omega) \end{aligned} \quad (2.103)$$

This is the central equation of LR TD-DFT as all objects are functions of the ground state density and the equation contains the information of electronic excitation. The function $\varsigma(\vec{r}, \vec{r}', \omega)$ blows up, *i.e.*, has a pole as a function of ω , if ω represents a true transition frequency of the system. The set of poles of the non-interacting density-density response function ς_{KS} refers to the single-particle excitations of the Kohn-Sham system

$$\varsigma_{\text{KS}}(\vec{r}, \vec{r}', \omega) = 2 \lim_{\eta \rightarrow 0^+} \sum_q \left[\frac{\zeta_q(\vec{r}) \zeta_q^*(\vec{r}')}{\omega - \omega_q - i\eta} - \frac{\zeta_q^*(\vec{r}) \zeta_q(\vec{r}')}{\omega - \omega_q - i\eta} \right] \quad (2.104)$$

q is a double index containing all occupied Kohn-Sham orbitals i and all unoccupied (virtual) orbitals a with

$$\omega_q = \epsilon_a - \epsilon_i \quad (2.105)$$

and

$$\zeta_q(\vec{r}) = \psi_i^*(\vec{r}) \psi_a(\vec{r}), \quad (2.106)$$

where ϵ_j is the orbital energy of the Kohn-Sham orbital ψ_j . Therefore, the poles of the first and the second term within the brackets of **Eq.** (2.104) correspond to the excitation and de-excitation energies of the system. Consequently, the non-interacting density-density response function ς_{KS} can be determined purely from ground state Kohn-Sham calculations. Neglecting the Hartree-correlation effects, *i.e.*, $\varsigma = \varsigma_{\text{KS}}$, the allowed transition would be the same as for the ground state Kohn-Sham potential. **Eq.** (2.103) can be solved iteratively, but to circumvent the numerically very demanding procedure and to make it more feasible for implementation in quantum chemical software, the problem can be reformulated into a non-Hermitian eigenvalue problem: The Casida equation is based on a parametrization of the linear density in Fourier space

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \quad (2.107)$$

where

$$A_{ia,jb} = \delta_{ij} \delta_{ab} (\epsilon_a - \epsilon_i) + K_{ia,jb}, \quad (2.108)$$

$$B_{ia,jb} = K_{ia,jb}, \quad (2.109)$$

$$K_{ia,jb} = \int \int d\vec{r} d\vec{r}' \psi_a^* \psi_i \left[\frac{1}{|\vec{r} - \vec{r}'|} + f_{\text{xc}} \right] \psi_b^* \psi_j. \quad (2.110)$$

⁴For practical purposes the following equations are written in frequency space (ω) derived from a Fourier transformation.

The matrices \mathbf{X} and \mathbf{Y} represent the excitations and the de-excitation following

$$\delta\rho(\vec{r}, \omega) = \sum_{i,a} [X_{ia}(\omega)\psi_i(\vec{r})\psi_a^*(\vec{r}) + Y_{ia}(\omega)\psi_a(\vec{r})\psi_i^*(\vec{r})]. \quad (2.111)$$

Neglecting the \mathbf{B} matrix in **Eq.** (2.107) yields the so-called Tamm-Dancoff approximation,^[202] which is equivalent to the CI singles formalism in the HF theory (see sec. 2.3). The Casida equation can be simplified for real-valued Kohn-Sham orbitals and a frequency-independent exchange-correlation kernel by defining matrices \mathbf{C} and \mathbf{Z} with

$$\mathbf{C} = (\mathbf{A} - \mathbf{B})^{1/2}(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})^{1/2} \quad (2.112)$$

and

$$\mathbf{Z} = (\mathbf{A} - \mathbf{B})^{-1/2}(\mathbf{X} + \mathbf{Y}). \quad (2.113)$$

The eigenvalues \mathbf{C} are the square of the excitation energies as noted by Casida^[189]

$$\mathbf{CZ} = \omega^2\mathbf{Z}. \quad (2.114)$$

2.5 Optical properties

2.5.1 Absorption

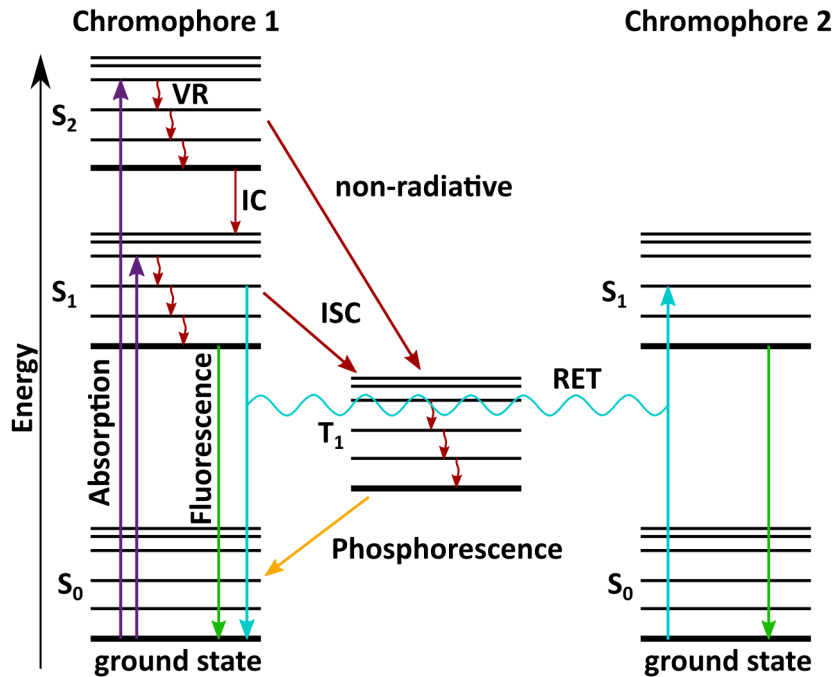


Figure 11: Jablonski diagram: Absorption (purple), fluorescence (green), phosphorescence (orange), non-radiative decays (red) including internal conversion (IC) and intersystem-crossing (ISC), vibrational relaxation (VR), and resonance energy transfer (RET) (cyan).

A photon with the correct wavelength or energy is able to raise a molecular system to an excited state. Absorption spectra of atoms consist of sharp lines, while absorption spectra of molecules show broad bands in the UV/VIS regime. These bands may contain information about the vibrational structure. Each electronic state represents a time-independent wave function and eigenfunction of the electronic Hamiltonian

and has a manifold of vibrational sublevels. While electrons move on the attosecond ($10^{-18}s$) time scale, molecular vibrations are on the femtosecond ($10^{-15}s$) time scale. This observation is essential when considering excited states in quantum chemistry. Due to the difference in time scales, the equilibration of the electronic structure upon photoexcitation occurs before the nuclei can react. As an approximation, this results in a vertical electronic transition and thus, for example, structural changes due to the changed shape of the PES take place only after the excited state has been reached. This quantum mechanical principle is known as the Frank-Condon (FC) principle and is a fundamental tool for understanding photophysical phenomena. The argument is analogous to the one in the BOA. Both approaches are directly linked, since the FC principle is based on the existence of PESs.

In practice, absorption energies are usually overestimated for several reasons. The most obvious reason results from the fact that electronic excitations, e.g. UV/Vis response are not exactly vertical in experiments, but adiabatic, which is especially true for molecules with large structural reorganization. The resulting energy difference between excited states is slightly lowered. More drastic is the fact that excited states, by their nature, have more multireference character than ground states (see sec. 2.3.3). As already discussed, single-reference approaches such as TD-DFT do not take into account static electron correlation. Accordingly, the calculated excited state is usually somewhat higher in energy than the actual excited state. Multi-reference methods such as DFT/MRCI provide an answer to this problem. Furthermore, since excited states are most often associated with weakened bonding interactions, the resulting PESs of an excited state are usually flatter. As a result, the zero-point energy of the ground state is significantly larger than that of the excited state, lowering the excitation energy.

Absorption spectra generated by excitation energies from theoretical calculations are consequently line spectra representing these vertical excitations. To ensure a reasonable comparison between experiment and theory, artificial line broadening is often used, for example with Gaussian functions. Natural homogeneous line broadening is caused by various effects,^[203] of which Heisenberg's uncertainty principle is one

$$\Delta E \Delta t \geq \frac{\hbar}{2}. \quad (2.115)$$

The finite lifetime of the excited state causes an uncertainty in energy ΔE , which causes a broadening of the spectroscopic signal. Accordingly, a molecular line spectrum is experimentally difficult to achieve. Furthermore, absorption spectra can be affected by electron-vibration (vibronic) couplings. Vibronic couplings are caused by the interaction of vibrational modes with electronic modes, resulting in a well-defined fine structure of the absorption band.

To obtain the strength of the vertical interaction that causes the vertical transition between the initial state (i) and the final state (f), the transition dipole moment (TDM) $\mu_{f,i}$ is used

$$\mu_{f,i} = \langle \Psi_f | \hat{\mu} | \Psi_i \rangle. \quad (2.116)$$

Here, $\hat{\mu}$ is the dipole moment operator and Ψ are stationary states of the unperturbed Hamiltonian. The probability for the system to make the transition to the excited

state is expressed by the oscillator strength $f_{i,f}$ with

$$f_{f,i} = \frac{2}{3} \omega_{f,i} |\mu_{f,i}|^2. \quad (2.117)$$

The oscillator strength is associated with the excitation energy $\omega_{f,i}$ and with the eigenvector \mathbf{Z} (Eq. (2.114)). Note that Eq. (2.117) is directly related to Fermi's golden rule No. 2, which will show up later (see Non-radiative energy transfer, Eq. (2.126)).^{[204],[205]}

Frank-Condon-Herzberg-Teller (FCHT) approaches are one way to account for vibronic broadening in theoretical calculations. The lowest order correction involves expanding the transition dipole moment matrix element $\mu_{f,i}$ by linear terms in the coordinates of the normal modes

$$\mu_{f,i}^{\text{FCHT}}(q) = \langle f | \hat{\mu}_{\text{el-vib}} | i \rangle \approx \mu_{f,i}^{(0)} + \sum_k \mu_{k,f,i}^{(1)} (q_k - q_k^i), \quad (2.118)$$

where $q = \{q_k\}$ is the vector of normal mode coordinates at the equilibrium structure q^i of the initial state, $\mu_{f,i}^{(0)}$ is the coordinate-independent component of the dipole element along the field, and $\mu_{k,f,i}^{(1)}$ is the first derivative with respect to q_k . The coordinate-independent term describes the FC transitions, while the vector of dipole derivatives $\mu_{f,i}^{(1)} = \{\mu_{k,f,i}^{(1)}\}$ corresponds to the Herzberg-Teller (HT) couplings.^[206] Accordingly, if a transition is FC-forbidden due to symmetry ($\mu_{f,i}^{(0)} = 0$), the transition may nevertheless be HT-allowed by intensity borrowing.^[207]

2.5.2 Emission

Electronic excited states have short lifetimes. Various processes are responsible for the dissipation of energy of an excited state. Intramolecular photophysical processes can be illustrated in a Jablonski diagram (see Fig. 11). All excited states will eventually release their energy and return to the ground state unless the energy triggers a photochemical process leading, e.g., to homolysis. A distinction is made between radiative and non-radiative transitions. The non-radiative transition within a PES, in which a vibrationally excited state transforms its vibrational energy into heat and transforms into a more energetically favorable vibrational state of the same PES, is called vibrational relaxation (VR). Since VR occurs within an electronic state, VR alone is never able to dissipate the excess energy of an electronic excitation unless aforementioned photochemistry results in the previously excited state to become the new ground state. Two other important non-radiative transitions are internal conversion (IC) and ISC. While the multiplicity is preserved in IC (e.g. S2 \rightarrow S1), it changes in ISC (e.g. S2 \rightarrow T1). The quantum mechanical operators describing IC and ISC are the non-adiabatic coupling (NAC), $d_{fi} = \langle \Psi_f | \nabla | \Psi_i \rangle$ and the spin-orbit coupling (SOC) interaction, $j_{f,i} = a \langle \Psi_f | \hat{I} \cdot \hat{s} | \Psi_i \rangle$, respectively. Here, the SOC matrix elements are described by the angular momentum operator \hat{I} , the spin operator \hat{s} , and the SOC constant a . ISC events have a time scale of 10^{-8} to 10^{-3} s and are significantly slower than IC, which range from 10^{-14} to 10^{-11} s, since spin-flip transitions are spin-forbidden.^[208]

Radiative transitions or luminescence includes all decay processes in which an electronically excited state transitions to a more energetically favorable electronic state,

while dissipating energy through the release of photons. As with IC and ISC, a distinction is made between spin-forbidden (phosphorescence) and spin-allowed (fluorescence) transitions. Therefore, fluorescence transition is much faster (10^{-9} to 10^{-7} s) compared to phosphorescence (10^{-3} to 1s). In general, non-radiative transitions are faster than radiative transitions. Accordingly, IC events lead to a population of the lowest state accessible by NAC before radiative processes occur. Specifically for organic molecules, this so-called "Kasha's rule" has the consequence that fluorescence is associated only with $S_1 \rightarrow S_0$ transitions when the electronic ground state is not accessible by IC.

From a computational perspective, emissions are much more difficult to study than absorptions simply because of the different time scales of the processes involved. Absorption spectra are relatively easily accessible with LR TD-DFT calculations using a single ground state optimized structure of the molecule of interest. Fluorescence transitions, on the other hand, are more difficult to calculate since they require an optimized excited state structure. Since excited PESs have a more shallow profile, they are less accessible. Furthermore, excited-state PESs are quite close to the other states, which can result in crossings or even bands of multiple states. Since BOA breaks down for these situations, the description of such PESs usually requires resource-consuming non-BO treatments.

2.5.3 Non-radiative energy transfer

In addition to monomolecular processes such as emission and non-radiative decays, there are deactivation processes that require more than one molecule and therefore involve transfer of excitation energy between molecules. When an excited donor molecule D^* decays to its ground state with a simultaneous transfer of electronic energy to an acceptor molecule A, it is called an EET. This non-radiative energy transfer is a single-step process (see Fig. 11). It requires the deexcitation of an excited donor molecule $D^* \rightarrow D$ and the excitation of an acceptor molecule $A \rightarrow A^*$ to be isoenergetic and coupled by a suitable interaction. The total electronic coupling V_{DA}^{total} of the transfer is defined by the coupling integral

$$V_{DA}^{\text{total}} = \langle \Psi_f(\vec{r}) | \hat{H}' | \Psi_i(\vec{r}) \rangle . \quad (2.119)$$

\hat{H}' includes the electrostatic interactions of all electrons and nuclei of the donor with those of the acceptor. $\Psi_i = \hat{A}\Psi_{D^*}\Psi_A$ and $\Psi_f = \hat{A}\Psi_D\Psi_{A^*}$ are antisymmetric product wave functions of the initial and final state. At this point it should be emphasized that the initial and final state of a unimolecular process are electronic states of a single molecule, while for a bimolecular process the initial and final state are two states of two different molecules, respectively. The total interaction V_{DA}^{total} can accordingly be regarded as the sum of a Coulomb and an exchange term. This results for a two electron case in

$$V_{DA}^{\text{total}} = V_{DA}^C - V_{DA}^X = \int \chi_{D^*}(1)\chi_A(2) \hat{H}' \chi_D(1)\chi_{A^*}(2) dX_1 dX_2 \quad (2.120) \\ - \chi_{D^*}(1)\chi_A(2) \hat{H}' \chi_D(2)\chi_{A^*}(1) dX_1 dX_2$$

with

$$\Psi_i(\vec{r}) = \frac{1}{\sqrt{2}} (\chi_{D^*}(1)\chi_A(2) - \chi_{D^*}(2)\chi_A(1)) \quad (2.121)$$

and

$$\Psi_f(\vec{r}) = \frac{1}{\sqrt{2}} (\chi_D(1)\chi_{A^*}(2) - \chi_D(2)\chi_{A^*}(1)) . \quad (2.122)$$

1 and 2 referring to the electrons, respectively, and χ_i are spin orbitals (see sec. 2.3). The Coulomb term describes the classical interaction of charge distributions and can be expanded by multipole terms like dipole-dipole, dipole-quadrupole, etc. For sufficient large distances between donor and acceptor r_{DA} the dipole term dominates for allowed transitions. This coupling is related to experimentally measurable quantities^[77] and can be written for the TDMs μ_A and μ_D of the respective molecules as

$$V_{DA}^C(\text{dipole} - \text{dipole}) = V_{DA}^{\text{d-d}} = \frac{\kappa|\mu_D||\mu_A|}{n^2r_{DA}^3} \quad (2.123)$$

where n is the refractive index of the medium, $|\mu_D|$ and $|\mu_A|$ are the magnitudes of the donor and acceptor transition dipoles and κ is the orientation factor of the TDMs of the donor and acceptor. κ in terms of dot product is

$$\kappa = \frac{\vec{\mu}_D \cdot \vec{\mu}_A}{|\mu_D| |\mu_A|} - 3 \left(\frac{\vec{\mu}_D \cdot \vec{r}_{DA}}{|\mu_D| |r_{DA}|} \right) \left(\frac{\vec{\mu}_A \cdot \vec{r}_{DA}}{|\mu_A| |r_{DA}|} \right) . \quad (2.124)$$

Accordingly, from **Eq.** (2.123) for $\langle \Psi_f | V_{DA}^{\text{d-d}} | \Psi_i \rangle^2$ arises

$$\langle \Psi_f | V_{DA}^{\text{d-d}} | \Psi_i \rangle^2 = \frac{\kappa^2 |\mu_D|^2 |\mu_A|^2}{n^4 r_{DA}^6} S_D^2 S_A^2 , \quad (2.125)$$

where $S_D = S_D(E_{D^*}, E_{D^*} - \Delta E)$ and $S_A = S_A(E_A, E_A + \Delta E)$ are the vibrational overlap integrals between the initial vibrational donor state with the energy E_{D^*} and the final state with the energy $E_{D^*} - \Delta E$ and between the initial vibrational acceptor state with the energy E_A and the final state with the energy $E_A + \Delta E$, respectively. According to time-resolved perturbation theory, the rate constant $\Gamma_{i \rightarrow f}$ of a transition between a pair of states is

$$\Gamma_{i \rightarrow f} = 2\pi \left(V_{DA}^{\text{total}} \right)^2 \varrho_E . \quad (2.126)$$

Where ϱ_E is the density of states, given by the number of energy levels per energy unit in the final state at the energy of the initial state. This equation is also referred to Fermi's golden rule No. 2.^{[204],[205]} For the “very weak coupling“ like dipole-dipole interactions one can reformulate **Eq.** (2.126) as

$$\Gamma(E_{D^*}, E_A) = 2\pi \int \langle \Psi_f | V_{DA}^{\text{d-d}} | \Psi_i \rangle^2 dE \quad (2.127)$$

where $\Gamma(E_{D^*}, E_A)$ is the rate of transfer from an excited donor state with initial energy E_{D^*} to an acceptor with initial energy E_A and the integral is over all possible values of the transferred energy E . The total transfer rate of a thermal equilibrium is obtained by introducing Boltzmann factors for the donor ($g(E_{D^*})$) and the acceptor ($g(E_A)$) to

Eq. (2.125) and **Eq. (2.127)**

$$\Gamma_{i \rightarrow f} = \frac{\kappa^2}{n^4 h r_{DA}^6} \int_0^\infty F_D(E) L_A(E) dE. \quad (2.128)$$

Here, $F_D(E)$ is defined as

$$F_D(E) = |\mu_D|^2 \int g(E_{D^*}) S_D^2 dE_{D^*} \quad (2.129)$$

and $L_A(E)$ as

$$L_A(E) = |\mu_A|^2 \int g(E_A) S_A^2 dE_A. \quad (2.130)$$

$F_D(E)$ and $L_A(E)$ are related to experimentally measurable quantities, while $F_D(E)$ is proportional to the normalized fluorescence spectrum of the donor, $L_A(E)$ is proportional to the normalized emission spectrum of the acceptor.

Förster resonance energy transfer (FRET) is based on the assumption that the electronic interaction is described fully by Coulomb interactions and hence that the D – A interactions are fully described by the comparatively long-range dipole-dipole interactions and hence all contributions from higher multipoles can be ignored. As mentioned above, the first assumption is valid if the spatial overlap is negligibly small and the exchange term vanishes. The second assumption is called ideal dipole approximation (IDA) and is primarily valid for comparatively long distances (up to hundreds of Å). In this range of distances r_{DA} , the description of a several oscillating charges with a single gradient vector (dipole) is suitable, because the oscillation distance of the donor $r_D \ll r_{DA}$ as well as for of the acceptor $r_A \ll r_{DA}$. Therefore, the IDA breaks down when r_{DA} is smaller than or the same size as the involved molecules, because the transition density is no longer point-sized and therefore the definition of the center of the transition dipole moment becomes difficult. Furthermore, IDA exaggerates errors for molecules with extended or asymmetric transition moments, e.g. Crts in photosynthetic systems.

Better descriptions of the non-radiative energy transfer are so called beyond-FRET methods that include higher multipoles and take care of the spatial orbital overlap. The transition density cube (TDC) method, developed by Krüger *et al.*,^[209] estimates the Coulomb coupling by using transition densities. The electron transition density is the inverse product of Ψ_i and Ψ_f of the molecular transition. Computationally, transition densities are three-dimensional transition density cubes representing a set of finite-sized volume elements V

$$M_k(x, y, z) = V \int_z^{z+\delta z} \int_y^{y+\delta y} \int_x^{x+\delta x} \int_s \Psi_{ki} \Psi_{kf} ds dx dy dz. \quad (2.131)$$

Therefore, TDC takes the three dimensional shape of the transition moments into account. The Coulomb coupling of the donor and acceptor states is approximately the sum over the Coulomb interactions of their TDCs

$$V_{DA}^{TDC} = \sum_{i,j} \frac{M_D(i) M_A(j)}{r_{i,j}}. \quad (2.132)$$

The accuracy of TDC is limited by the accuracy of the wavefunction and the size of the volume elements. Scholes and coworkers presented a method that takes orbital overlap effects into account and is based on TDCs.^[210] Furthermore, the “transition charge, dipole, and quadrupole from electrostatic potential” (TrEsp-CDQ) method is another way to calculate the Coulombic coupling including higher multipoles. In the original TrEsp (transition charge from electrostatic potential) method, the electronic coupling is evaluated as classical Coulomb coupling interactions between transition charges, but the transition charges are derived from electrostatic potential fitting.^{[211],[212]} TrEsp-CDQ is based on this method, but includes higher order multipoles.

2.6 QM/MM

Although, MD simulations are able to study large, complex, organic and biological systems^{[213]–[216]} as well as inorganic and solid-state systems^{[217]–[219]} using MM force fields are not able to describe electronic structural changes of chemical reactions and electronically excited states. In order to simulate processes involving bond breaking and formation, charge transfer and/or electronic excitations, a QM description is necessary. However, applications of QM are limited to relatively small systems, due to the higher computational cost (see sec. 2.2).

Algorithms that combine QM and MM provide a solution to this problem. Merging of QM and MM can be achieved in several ways. One way is the so-called combined quantum mechanics/molecular mechanics (QM/MM) methodology. QM/MM tries to combine the advantages of both worlds forming a hybrid model, similar to the idea of combining HF and DFT in hybrid density functionals (see sec. 2.4.3).^{[220]–[222]} However, while the hybrid functionals

QM/MM treats a defined region, for example the active site of an enzyme or a chromophore (\mathcal{Q}), with QM methods and the influence of the environment (for example the protein environment, \mathcal{M}) with MM methods. When modeling a system using QM/MM methods, each atom is assigned to at least one of the following groups:

- (i) \mathcal{S} all atoms of the entire System,
- (ii) $\mathcal{Q} := \{x \in \mathcal{S} \mid x \text{ is modeled on the QM level}\}$,
- (iii) $\mathcal{M} := \{x \in \mathcal{S} \mid x \text{ is modeled on the MM level}\}$,
- (iv) $\mathcal{B} := \{x \in \mathcal{S} \mid x \text{ are boundary atoms (see sec. 2.6.2)}\}$,
- (v) $\mathcal{A} := \mathcal{S} + \mathcal{B}$ describing the augmented system.

In most QM/MM schemes $\mathcal{S} = \mathcal{Q} + \mathcal{M}$, but this is not necessarily a disjoint union, since in some QM/MM schemes atoms are present in both groups. \mathcal{B} contains only atoms if covalent bonds are "cut" when the atoms are divided into Q and M and thus a molecule is partially described by QM and MM.

The various QM/MM methods may differ in how exactly the QM and MM regions are coupled. In the following, QM/MM methods are described based on key distinguishing characteristics. The first property class, coupling schemes, describes how the two subsystems are coupled together to correctly represent a physical observable of the whole system. The second property class, boundary atom schemes, manages how atoms that are covalently bonded with atoms of another subsystem are treated. The third property class, electrostatic coupling schemes, determines how the MM and QM systems are electrostatically coupled to each other.

2.6.1 Coupling schemes

Each QM/MM method provides as the main result the potential energy or the forces acting on the atoms. From the potential energy, a correspondent expression for the gradient with respect to nuclear coordinates is usually easily available from which the forces can be determined. When using the forces, however, it is not possible to formulate a global energy function, which is why these schemes are only used in QM/MM methods, whose applications do not have to represent this observable.

Regarding the energy mixing schemes, there are two types, the additive and the subtractive schemes, which, according to the name, combine the energy of the subsystems additively or subtractively to obtain the energy of the entire system.

Additive schemes

For additive schemes, the total potential energy of the entire augmented system \mathcal{A} is composed as the sum of the individual subsystems.

$$V_{\text{QM/MM}}^{\text{aug}}(\mathcal{A}) = V_{\text{Q}}(\mathcal{Q} + \mathcal{B}) + V_{\text{M}}(\mathcal{M}) + V_{\text{Q,M}}(\mathcal{Q}, \mathcal{M}) \quad (2.133)$$

Here $V_{\text{Q}}(\mathcal{Q} + \mathcal{B})$ is the potential energy of the atoms of the QM region including the boundary atoms \mathcal{B} , $V_{\text{M}}(\mathcal{M})$ is the potential energy of the atoms of the MM region and $V_{\text{Q,M}}(\mathcal{Q}, \mathcal{M})$ is the QM/MM coupling energy of both regions (see sec. 2.6.3). Since $V_{\text{QM/MM}}(\mathcal{A})$ contains more atoms than the "original" system \mathcal{S} , if the set of boundary atoms \mathcal{B} is not empty, a corrective term $V_{\text{corr}}(\mathcal{Q}, \mathcal{B})$ is sometimes employed. It can however be neglected, as the error introduced by \mathcal{B} is usually considered identical throughout the PES and thus can be ignored

$$V_{\text{corr}}(\mathcal{Q}, \mathcal{B}) = V_{\text{M}}(\mathcal{B}) + V_{\text{M,M}}(\mathcal{B}, \mathcal{Q}). \quad (2.134)$$

Here $V_{\text{M,M}}(\mathcal{B}, \mathcal{Q})$ describes the coupling energy between the boundary atoms and the QM region on the MM-level and $V_{\text{M}}(\mathcal{B})$ the potential energy of the boundary atoms on the MM-level. Accordingly, for the total potential energy of the whole system \mathcal{S}

$$V_{\text{QM/MM}}(\mathcal{S}) = V_{\text{Q}}(\mathcal{Q} + \mathcal{B}) + V_{\text{M}}(\mathcal{M}) + V_{\text{Q,M}}(\mathcal{Q}, \mathcal{M}) - V_{\text{corr}}(\mathcal{Q}, \mathcal{B}). \quad (2.135)$$

Typically, the correction term is omitted since $V_{\text{M}}(\mathcal{B})$ is negligibly small and $V_{\text{M,M}}(\mathcal{B}, \mathcal{Q})$ is nearly constant. Accordingly, most additive schemes use $V_{\text{QM/MM}}^{\text{aug}}(\mathcal{Q}, \mathcal{M}, \mathcal{B})$.

Subtractive schemes

An alternative to the additive schemes are the subtractive schemes. Instead of summing the individual energies of the subsystems additively, subtractive schemes use the potential energy of the entire system \mathcal{S} and then replace the MM energy of the augmented system \mathcal{A} with the QM energy of the augmented system

$$V_{\text{QM/MM}}(\mathcal{A}) = V_{\text{M}}(\mathcal{S}) - V_{\text{M}}(\mathcal{Q} + \mathcal{B}) + V_{\text{Q}}(\mathcal{Q} + \mathcal{B}). \quad (2.136)$$

Provided that the electrostatic interactions between the QM and the MM region are calculated entirely at the MM level (e.g., using the mechanical embedding, see

sec. 2.6.3), the underlying equations of the additive and subtractive schemes are identical.^[223] Using electrostatic interactions based on the MM level holds

$$V_M(\mathcal{S}) = V_M(\mathcal{Q}) + V_M(\mathcal{M}) + V_{M,M}(\mathcal{M}, \mathcal{Q}) \quad (2.137)$$

and

$$V_M(\mathcal{Q} + \mathcal{B}) = V_M(\mathcal{Q}) + V_M(\mathcal{B}) + V_{M,M}(\mathcal{Q}, \mathcal{B}) \quad (2.138)$$

and (only for mechanical embedding)

$$V_{M,M}(\mathcal{M}, \mathcal{Q}) = V_{Q,M}(\mathcal{Q}, \mathcal{M}). \quad (2.139)$$

Accordingly, **Eq.** (2.136) yields a subtractive energy of

$$\begin{aligned} V_{QM/MM}^{\text{sub}}(\mathcal{A}) &= V_M(\mathcal{Q}) + V_M(\mathcal{M}) + V_{M,M}(\mathcal{M}, \mathcal{Q}) \\ &\quad - V_M(\mathcal{Q}) - V_M(\mathcal{B}) - V_{M,M}(\mathcal{Q}, \mathcal{B}) + V_Q(\mathcal{Q} + \mathcal{B}) \\ &= V_Q(\mathcal{Q} + \mathcal{B}) + V_M(\mathcal{M}) + V_{Q,M}(\mathcal{Q}, \mathcal{M}) - V_{\text{corr}}(\mathcal{Q}, \mathcal{B}) \\ &= V_{QM/MM}^{\text{add}}(\mathcal{S}). \end{aligned} \quad (2.140)$$

This shows that the basic equations of the subtractive schemes are identical to the equations of the additive schemes for mechanical embedding, with the additive schemes providing more flexibility for the individual components of the various electrostatic coupling schemes. Furthermore, additive schemes do not require MM parameters for the QM region, which in many cases includes ligands for which most force fields do not have parameters. Therefore, the description of those ligands is more trivial in QM/MM. Since the subtractive schemes required the MM energy of the entire system \mathcal{S} , parameters for the entire system are necessary. The advantage of the subtractive approaches is simplicity, since it automatically ensures that no interactions are double counted and is independent of the choice of QM and MM software, as long as energies or forces can be calculated. In comparison, the QM/MM software using an additive scheme must ensure that no interactions are counted more than once and, accordingly, MM software is required which allows a QM/MM-tailored selection of the MM terms.

2.6.2 Boundary atom schemes

Ideally, the QM or MM region only encloses complete molecules and not parts of a molecule, so that there is no intersection of a covalent bond. In many cases, this cannot be realized because the interest is in a particular small part of a biological macromolecule such as an active site of an enzyme.

Fig. 12 shows the nomenclature, where the subscript number assigned to each atom is measured by the distance to the boundary and the atom designation Q/M is assigned to the QM/MM region. For such cases, various techniques have been developed to give special consideration to the boundary atoms. Two classes of link-atom schemes and local orbital schemes are distinguished.

In link atom schemes, a so-called link atom is used to saturate the free dangling bond at the Q1 atom. The link atom is typically a hydrogen atom.^{[224]–[227]} However, parameterized atom schemes, e.g., a one-free-valence atom as M1 atom,^[228] "pseudobonds" schemes^[229] or "quantum capping potentials"^[230] schemes can also be used. In the

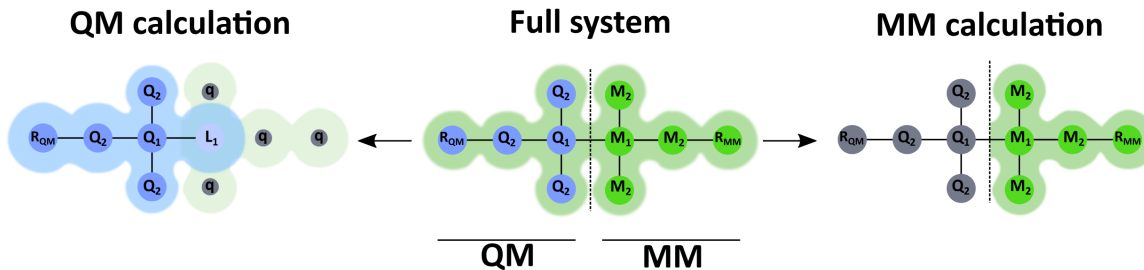


Figure 12: Schematic representation of the separation of the full system in the system of the QM calculation (left) and the system of the MM calculation (right) using the nomenclature of the boundary atoms of the QM and the MM region. The QM calculation includes the link atom L_1 and the point charges q . Note that the charge shift modifications of the QM calculation point charge field are not displayed here.

former, a parameterized semi-empirical Hamiltonian is used and in the latter two, a parameterized effective core potential^{[229],[230]} is used to mimic the properties of the atom to be replaced.

The class of local orbital schemes consists of methods that use localized orbitals of the boundary atoms. An example is the so-called local self-consistent-field (LSCF) algorithm,^{[231]–[233]} which calculates strictly localized bond orbitals of small model components to describe the cut bond. The strictly localized bond orbitals are obtained from the SCF optimization of large molecules to avoid the admixture of other QM basis functions. Another approach, which is modeled after the LSCF method, is the generalized hybrid orbital (GHO) method.^{[234]–[236]} In this method, a set of four sp^3 hybridized orbitals is used, which is assigned to each bound MM frontier atom. The hybridization schemata are thereby determined by the local structure of the three MM atoms bonded to the M1 atom. The fourth orbital represents the bond to the Q1 atom and is called the active orbital. The active orbital and the three auxiliary orbitals are included in the QM calculation, whereas only the active orbital is optimized in the SCF optimization. In the LSCF algorithm as well as in the GHO method it is assumed that the obtained orbitals are transferable.

Each of the presented boundary atom schemes has its advantages and disadvantages. The link atom method is simple and straightforward, but makes the calculation of the QM/MM energy more difficult, because additional atoms are introduced, which are actually not part of the system \mathcal{S} .

2.6.3 Embedding schemes

As already introduced in **Eq.** (2.133), $V_{Q,M}(\mathcal{Q}, \mathcal{M})$ describes the coupling of the two regions (QM and MM). This coupling is composed of the binding interactions, the VdW interactions, and the electrostatic interactions. The binding interactions and the VdW interactions result from the force field of the MM. The electrostatic part can be treated in various ways, called embedding schemes, which fall into three categories. One distinguishes between mechanical embeddings, electrostatic embeddings and polarized embeddings.

In mechanical embedding, one considers the electrostatic interaction between the two regions entirely at the MM level. While this method is computationally efficient and relatively simple to implement (see sec. 2.6.1), it has some drawbacks. For example, the QM density is not polarized by the environment, which can lead to unrealistic

behavior especially for Q atom, MM charge parameters that have been derived for a different environment. Furthermore, MM parameters such as point charges must be available for the atoms of the QM region and since the QM region usually has special molecules or ligands for which no high quality MM parameters are available.

In electrostatic embedding, the atoms of the MM region are explicitly included as point charges in the QM calculation. Thus, neither point charges nor MM parameters are needed for the QM region, nor is the polarization of the QM region neglected. The drawback of these schemes is that MM point charges near the boundary can easily lead to overpolarization of the QM atoms. In recent years, many techniques have been developed to minimize the error of the interactions.

Polarized embedding represents the most realistic of the three presented schemes. In contrast to the electrostatic embedding, not only the polarization of the QM density is taken into account, but also the polarization of the MM region by itself and by the QM as region. To realize this, a polarizable force field is necessary. However, the required self-consistent cycles are computationally very demanding, since one QM evaluation per iteration is necessary.

2.6.4 Employed software

The QM/MM software `gmx2qmmm`, which is part of this thesis and is documented in Paper 1, uses an additive coupling scheme (see Additive schemes). Since the code was mainly developed to study the spectral properties of chromophores, an electrostatic embedding is implemented. This embedding easily enables to turn off the electrostatic influence of the environment acting on a chromophore, because the environment (MM region) is described by a PCF in the QM calculation (see Embedding schemes). In application this means that a TD-DFT calculation using `gmx2qmmm` can be modified simply by deleting the PCF to describe a gas phase calculation. The software uses a sophisticated link atom scheme (for more details see Paper 1) to describe the QM and MM interface (see Boundary atom schemes).

3 Summary and conclusion

A counter-intuitive hypothesis

At the beginning of the final section of this thesis, I would like to point out that our hypothesis, namely that blue light/higher excited energy pathways in light-harvesting complexes exist, is contrary to the common understanding of these complexes. It is generally accepted that the Q band of the Chls is mainly responsible for the energy transfer towards the special pair of the photosystem.^{[106],[237]–[239]} It is further recognised that all variations of chromophores (e.g., Chl *b* and Crts) extend the absorption spectrum, increasing the absorption cross section and thus the light-harvesting spectral range. Further, it is acknowledged that Crts are responsible for non-photochemical quenching to avoid the creation of singlet oxygen.^{[240]–[243]} Since it is well known that tightly assembled chromophore complexes also perform EET for bands other than only their lowest excited band, we aimed to understand if this is also the case for systems like PSII. The idea of this project was thus to show that the organization of the PSII supercomplex is more likely to transfer red light energy than blue light energy towards the core. This would mean that the minor light-harvesting complex CP29 has a special role for blocking higher/blue excitations from entering the PSII core complex. While it may seem so at first, this hypothesis is not contrary to the presented common understanding, but the following results imply that Crts are wavelength converters, connecting the Soret band with the Q band of Chls.

Fig. 13 is a schematic representation of what has been done in this thesis. While the red light energy transfers are already known in the literature, this work presents an investigation of the blue light energy pathways. To analyse biochemical systems such as CP29, it was necessary to have a suitable software such as `gmx2qmmm` that is able to perform QM/MM calculations (see Paper 1). To be sure that the software and the applied methods are able to reproduce the spectroscopic properties of the relevant chromophores, we performed a comparison and benchmark between experiment and theory (see Paper 2). With these tools, we were able to study the electrostatic influence of the protein scaffold on chromophores and have proposed possible energy pathways (see Paper 3). Furthermore, we have created a general workflow which is transferable to other LHCs (see Fig. 8). To confirm that the results are independent of the coupling scheme and to consider higher multipole moments, we applied a different coupling scheme, namely TDC. This scheme additionally allowed us to determine the transfer between Chl and Crt (see Paper 4).

Publication overview

The first project was about the development of a user-friendly, Python-based QM/GROMACS interface: `gmx2qmmm`. The interface links the Gaussian and GROMACS software packages in an additive QM/MM scheme (see sec. 2.6.4). It builds the foundation of this thesis and enables the development and implementation of different tools. With the simplicity of using the same file formats as well as the same input convention as GROMACS, which was used to perform MD simulations, `gmx2qmmm` was the perfect tool to fulfill the requirements of the shown computational workflow (see sec. 1.3). Furthermore, the software was designed to be platform-independent and portable, so that no further recompiling or reconfiguration of the existing software is necessary. Free access to the open-source code enables

users to modify the code according to their own need.

We tested different correction functions to account for the presence of link atoms and showed that the presented link atom correction functions (LCFs) are improvements to QM/MM forces. However, we also showed that LCFs do not improve energy barriers along strong structural changes hence they are designed to improve the situation close to potential minima. For the involved tests, the overall shape of the potential energy surface and therefore the behavior was almost the same between full QM and LCF-corrected QM/MM.

We used the interface to investigate the excitation energies of an optimized structure of the first excited (Q_y) and the ground state of Chl *a*610 and the coordinating glutamine (E198) of CP29 (see sec. 1.2.4) for the following cases: (I) regular QM/MM-optimized structures, (II) same structures neglecting the PCF, and (III) the gas phase-optimized structures. We showed that neglecting the PCF causes a systematic red shift of the absorption energies for the Chl. Our results proved that the interface in fact reproduces the experimentally observed trends with only minor drawbacks.

The second publication compares of theoretical (TD-CAM-B3LYP/6-31G* and DFT/MRCI) and experimental analysis of spectroscopic properties of the most relevant light-harvesting pigments in plants, namely β -carotene, Zea, Vio, Neo, Chl *a* and Chl *b*. To this end, the spectral properties of the optimized ground and excited states of the chromophores were computed and compared with the experimental data. Furthermore, we included calculations of Chls embedded in a protein scaffold to compare with the gas phase and solvated (acetone) case.

Besides the well-known bond length alternation (BLA) analysis of the structural differences of the ground- and excited-states of Crts,^[244] we introduced non-planarity (NPL) as a new parameter to measure the structural differences of electronic states of Chls. The structural differences between the bright $1B_u$ state and the ground state of the Crts is significantly higher than between the excited states and the ground state of Chls. For the small structural changes of the different electronic states of Chls an increasing trend with higher states could be identified qualitatively.

We showed that the applied method, CAM-B3LYP, was less suitable for predicting exact electronic excitation energies while being excellent at reproducing vibrational properties of the investigated electronic states. For Chls, the effect is most evident in the “green gap”, *i.e.*, the energy difference between the Q and Soret band maxima, which is overestimated by CAM-B3LYP and less so by DFT/MRCI. The different Crts only differ in their excitation energies but display analogous subsequent photophysics. This suggests that the Crts fulfill basically identical biological functions, with slightly different state energies. However, the experimentally known exchange of Zea with Vio in LHCII upon light stress may be explained by a diminishing probability of a hypothetical $1B_u(\text{Crt})$ to $Q(\text{Chl})$ transfer due to the energetically higher lying $1B_u$ state in Vio compared to Zea. Furthermore, the DFT/MRCI results show a smaller energetically difference of the $1B_u$ and the $1A_g$ state compared to previous studies.^[245] This implies that $1B_u(\text{Crt})$ to $Q(\text{Chl})$ transfers are more likely than $1A_g(\text{Crt})$ to $Q(\text{Chl})$ transfers.

The structural changes identified by BLA or NPL explain the high reorganization energies of the Crt $1B_u$ states, which is experimentally represented by the Stokes shift, and the study also explains analogously the almost negligible reorganization

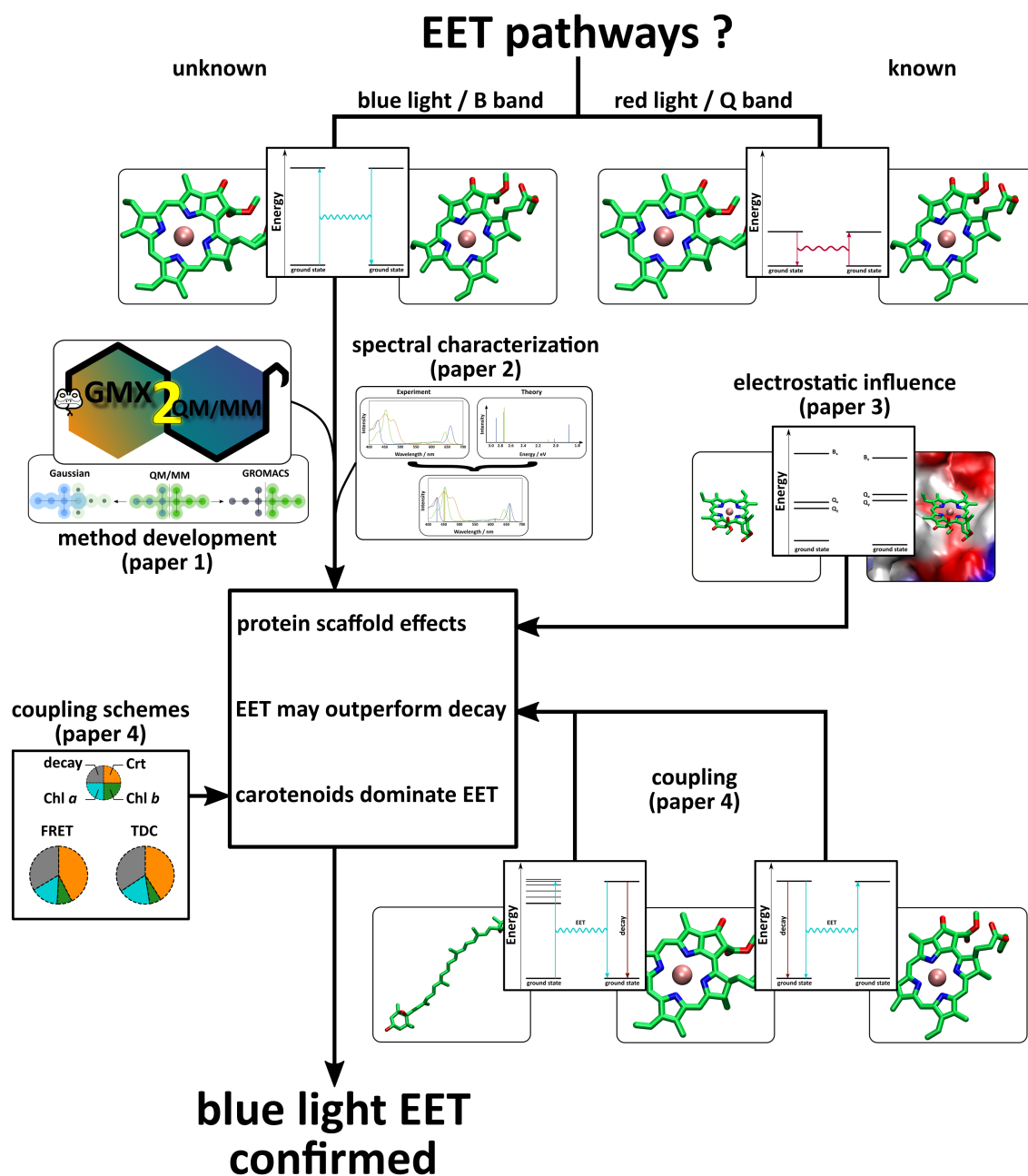


Figure 13: Schematic representation of the summary of the thesis including simplified representations of the known red light energy transfers, of the investigated blue light energy transfers, of the paper 1-4 and three resulting key statements.

energies of the Chl states. The only reason why the $1B_u(\text{Crt})$ to $Q(\text{Chl})$ transfer has so far not been considered is the high energetic difference of the two states, but taking the high reorganization energies of the $1B_u$ state into account, this difference decreases drastically. From a computational perspective, these results suggest a possible EET from the Crt bright state ($1B_u$) to the Chls (Q band) and supports the hypothesis that Crts can act as a shortcut for Chl Soret-to-Q band energy flow.^[82]

The third publication investigates the effects of the electrostatic environment on the spectral properties of Chls of CP29 and the corresponding EET between the

Chls. Here, we performed QM/MM TD-DFT calculations, including and excluding the PCF, of all QM/MM optimized chromophores of several timesteps of an MD trajectory (see sec. 1.3). Although the results of this procedure do not represent actual dynamics, MD snapshots represent a sample of the dynamics of the system and provide more understanding on the flexibility than a single crystal structure. In comparison to previous studies, not only did we analyse the electronic excited states of the Q band but also the energetically higher lying states of the Soret band.

The spectral analysis of the Chls show significant shifts caused by the PCF. However, the net effect of the sum spectrum is negligible, which is astounding, since each Chl can experience strong shifts (up to 154 meV). Furthermore, we saw that Soret and Q band are shifted differently. We found two uniform shifts for all bands corresponding to the Mg-coordinating ligand: (I) a blue shift for Chl-water and (II) a red shift for Chl-glutamine by activating the PCF.

Also in this publication, the EET between Chls was determined based on the relative orientation of the TDMS of the respective donor/acceptor. The PCF mainly influences the coupling via the tuning of the energetic position of the excited states, rather than through varying the TDMS. However, a fine-tuning via changes of the orientation of the TDMS, especially for the states energetically above the Q_y state could be identified. Based on the relative orientation of the TDMS, we observed a trend for the PCF to steer potentially harmful high energy excitation away from the PSII core complex. This effect implies the evolutionary origins and the functional relevance of the CP29 acting as “conduit”. The results are restricted to a FRET model (only considering incoherent EET based only on dipoles) and do not include Crts.

To confirm the results of the third publication (see sec. 4.3), we used the TDC method, which can be interpreted as a beyond-Förster method (see sec. 2.5.3), to include higher multipole moments and to take the spatially challenging Crts as acceptors into account. We compared the rates and coupling calculated with the FRET scheme to those of the TDC scheme. The structural sampling, representing the structures of the MD trajectory, showed a small effect for Chl-Chl transfer of the Q_y states. The corresponding effects for B_x state and the Chls(B_x)-Crts transfer were found to be larger due to the higher flexibility of the energetically higher lying states.

In terms of Chl-Chl Q band couplings and associated rates the TDC and FRET methods vary only slightly, staying mostly within the same order of magnitude, resulting in minor effect on EET acceptor preferences. Comparing Q_y - Q_y rates with the experimental ones, the EET outperforms all decay processes which relates to the experimentally known high quantum yield of photosynthetic light-harvesting.

In comparison to the Q band, excited states of the Crts are possible acceptors for the excited states of Chl B band due to the similar excitation energies. As mentioned above, without taking the high reorganization energies of Crts into account, EET of $1B_u(\text{Crt})$ to $Q(\text{Chl})$ is very unlikely. We restricted the analysis on the ground state optimized Crts and therefore on the Crts acting as acceptors($B_x(\text{Chl})$ - $1B_u(\text{Crt})$) rather than donors ($1B_u(\text{Crt})$ - $Q(\text{Chl})$). The Chl-Crt interaction differs depending on the method, which is not surprising, since the TDC method was specifically set up to amend downsides of the FRET approach, namely the ideal dipole approximation (IDA). However, the FRET method does not systematically over- or

underestimate the rates compared with TDC. Still, all Chls that are spatially close to a Crt, show a transfer efficiency above 50% to Crts. Contrary to the Q band, the decay processes (*i.e.*, the IC between Soret and Q) are competitive processes and are not negligible. Despite the Soret decay being in the fs-ps time range the EET between Chls or Crts mostly outperforms the decay processes. In this publication, we showed that the calculated model indicates that the Soret states are likely highly coupled, as expected from other, artificial systems, but curiously, due to the competition by Chl *b* and Crt, show almost no experimentally observable EET from Chl *a* to Chl *a*, despite Chl *a* being by far the most common chromophore in the system.

Impact and future tasks

Paper 1 represents the tools used and the foundation of the subsequent work, while Paper 2 tests the functionality of the tool by comparing its accuracy to that of common theoretical methods and experimental data. Paper 3 and Paper 4 provide insight into the antenna complex CP29 and offer a scope for possible pathways for red light and blue light energies. With the last two papers, especially Paper 4, I have shown that an analysis restricted only to the first excited state of Chl (Q_y) with respect to energy transfer is an incomplete simplification of the reality. This is in noticeable contrast to the prevailing hypothesis.

Future investigations based on the presented studies may be both of experimental and theoretical nature. Our newly acquired understanding of the electronic influence of chromophores surrounding, paves the way towards experimental investigations, for example targeted mutations of the surrounding amino acids. Combined experimental and theoretical investigations will enable modifications and improvements of energy pathways and may eventually lead to applications in new organic solar cells or bring accessibility to the excess energy. The results of this work are entirely computational and require more experimental studies of the Soret band. An example of this are studies on systems lacking a specific chromophore, such as the studies light-harvesting complexes without β -carotenes.^[246] It should be highlighted that the accessibility of such data is more difficult to obtain experimentally than computationally.

From a theoretical perspective, there are numerous so-called beyond-FRET methods that compensate for its weaknesses or consider even further transfer processes. For example, TDCs (see Paper 4) can be used to calculate excitonic coupling models. However, already the calculated energy levels of the excited states and the calculated transfer rates can be used to build simpler kinetic models.^{[247],[248]} In this case, the reorganization must be taken into account for the emission, either by optimizing the excited states or by shifting the energies with the corresponding Stokes shift. This enables an analysis of the population of the respective states in time, which in turn gives insight into the biological function of the chromophores. Moreover, we have so far only investigated the function of Crts as acceptors and not as donors, since the latter requires an optimization of the excited states, which can only be realized with methods like DFT-MRCI or more involved post-HF methods.

Another aspect is that the presented workflow and methods are easily transferable to all other light-harvesting complexes of the PSII supercomplex. Both FRET and TDC provide direct insight of couplings between complexes, although the complexes are calculated separately. This is because the chromophores are calculated separately

in any case and the relative orientation of the complexes is given. Nevertheless, the consideration of single complexes without the surrounding complexes will result in unnatural behavior as shown for Chl *a*616 (see Paper 3). Our work is limited to the antenna complex CP29, an investigation of the surrounding light-harvesting complexes LHCII and CP24 is necessary to understand their mutual interaction. This would elucidate their function more clearly. To fully understand the functionality and underlying processes, an investigation of the entire PSII supercomplex would be necessary, or at least of all subunits.

In this thesis, the membrane solely consists of DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine) molecules, which is typically used in computational models to represent thylakoid-like membranes.^[249] As the membrane assembly can have a strong influence on the behavior of membrane proteins,^[250] the use of a more accurate description of the membrane in terms of composition based on experimental data could increase the understanding. It may even serve as a new angle for studying the reaction of the complexes to new membrane compositions.

Since the variation of the different chromophores and the variation of the structures of the respective light-harvesting complexes is manageable, approaches such as neural networks or artificial intelligence can be used to draw conclusions about related biological systems without calculating each complex individually. Based on the overall similarity of LHCs (see sec. 1.2.4), it is possible to apply methods that estimate the coupling of two chromophores fully based on an analysis of structural and electronic similarity of predefined data sets. However, selecting criteria for similarity analysis is crucial and must be done carefully. For example in the case of Chl the NPL parameter (see sec. 4.2) could be used as such a criterion.

The results of the presented publications show unambiguously that blue light energy pathways exist and that they compete with other decay processes, *i.e.*, the IC between Soret and Q. This suggests that the current state of the science of light-harvesting needs refinement.

4 Publications

4.1 Paper 1

“A user-friendly, Python-based quantum mechanics/Gromacs interface: gmx2qmmm”

J. P. Götze, Y.-W. Pi, S. Petry, F. Langkabel, J. F. Witte, O. Lemke

International Journal of Quantum Chemistry **2021**, 121, 3, e26486.

DOI: 10.1002/qua.26486

URL: <https://doi.org/10.1002/qua.26486>




Figure 14: Graphical abstract of paper 1.

Contributions

Jan Götze conceived the project and wrote the manuscript with individual parts written by Simon Petry, Fabian Langkabel, Felix Witte and Oliver Lemke. The development of the software and the implementation was carried out by Jan Götze, Yuan-Wei Pi, Simon Petry and Felix Witte. Tests and optimizations of the code performance was done by Simon Petry. The validation of the software with regard to application for spectroscopic properties of Chls embedded in antenna complexes was performed by Simon Petry and of *Lepidium virginicum* water-soluble chlorophyll-binding protein by Oliver Lemke. Furthermore, link atom correction functions were elaborated and implemented by Fabian Langkabel. The conversion of the software from python 2 to python 3 was done by Simon Petry and Yuan-Wei Pi. All authors contributed to the final version of the manuscript.

A user-friendly, Python-based quantum mechanics/Gromacs interface: gmx2qmmm

Jan P. Götze  | Yuan-Wei Pi | Simon Petry | Fabian Langkabel |
Jan Felix Witte | Oliver Lemke

Physikalische und Theoretische Chemie, Freie Universität Berlin, Berlin, Germany

Correspondence

Jan P. Götze, Physikalische und Theoretische Chemie, Freie Universität Berlin, Berlin, Germany.

Email: jgoetze@zedat.fu-berlin.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: 393271229

Abstract

We introduce a Python 2.7 software called gmx2qmmm, which provides an interface between the Gaussian and Gromacs software packages in an additive quantum mechanics/molecular mechanics (QM/MM) scheme. Other QM packages will be added in future releases. The main advantage of gmx2qmmm is its simplicity in terms of input setup and configuration as it maintains the Gromacs file formats, as well as input conventions. It is also designed such that users do not need to reconfigure or recompile any of the interfaced programs. While our main goal was to provide a simplified transition from Gromacs to QM/MM using Gromacs directly as the basis for the MM part, we considered alternative ways to treat the QM/MM boundary. Our software was also developed to test a previously considered way to account for the presence of link atoms, which we term here link atom correction functions (LCFs). We show that LCFs can be good candidates to improve the forces at the QM/MM junction; however, from our data, it is also apparent that LCFs will not necessarily improve energy barriers, likely due to them being tailored to improve the situation close to potential minima. LCFs are, however, trivial to set up and can be used in the future to support the accuracy of QM/MM optimizations and dynamics. We present data on how our interface compares to a full QM description of small polypeptides; furthermore, we investigate the UV/vis spectroscopy of chlorophyll-containing proteins depending on the employed potentials and geometries.

KEYWORDS

chlorophyll-protein complexes, Gromacs, Python, QM/MM

1 | INTRODUCTION

A primary challenge for computational chemistry is to obtain accurate results while maintaining reasonable computational effort. The associated cost can be tailored by the employed computational method and the size of the chosen model system. Typically, costs scale nonlinearly with the system size and desired accuracy. It is, therefore, a common research goal to tune down the scaling exponent as the size of the system is fixed by means of the scientific question at hand.

The last decades have, however, seen a variety of ways to scale down the system size. Hybrid approaches, like the coupling of quantum mechanics/molecular mechanics (QM/MM), are very efficient and were also awarded a recent Nobel prize.^[1–3] Multiple levels of computational theory can be coupled in various ways, not only QM plus MM, but also QM plus QM with possibly multiple layers thereof.^[4] From this perspective, solvent continuum models can be considered part of the hybrid schemes.

In this article, we present a new interface to couple a variety of QM programs with the Gromacs force field package.^[5] Our interface, gmx2qmmm, is a very flexible Python 2.7 script, written to mimic the style of a regular Gromacs executable. Currently, the package only supports the Gaussian package as Merz-Kollman charges^[6] are still computed by the QM program itself; however, support for the Orca^[7] and Turbomole^[8] packages is currently under development. The script is able to perform steepest descent and conjugate gradient and Broyden-Fletcher-Goldfarb-Shanno (BFGS)^[9] optimizations for ground and electronically excited states.

The script uses the parallelization capabilities of its interfaced programs and is, as such, very efficient in using the available computational resources. To enhance efficiency, the code transfers the results from each previous step to the next calculation as the initial guess for the wave function, electronic density, or excited-state coefficients; this way, we retain most of the efficiency of the standalone QM programs (saving about 20%-25% CPU time for each optimization step for our CP29 systems, see below, data not shown).

In the following, we will present a few example calculations, as well as several aspects of the approximations included in the linking of the QM and MM regions. The novel aspects of the presented software are akin, but not identical, to what was recently proposed elsewhere.^[10] The main novelty is to compensate for the artifacts that are being introduced when cutting a covalent bond and introducing hydrogen atoms for bond saturation, that is, link atoms. In contrast to the previous proposal, here, we aim here to include a similar correction for the additive QM/MM scheme.

1.1 | Link atom correction functions

Link atoms, while being an elegant way to circumvent many problems of the QM/MM border regions, are artificial parts of the model system. The problem arising from the introduction of link atoms is how to correct for their presence, that is, how to construct a smooth potential for the QM atoms that mimics a full QM description. Several ways exist to cope with this problem,^[3,10,11] and a full listing is not the purpose of this article.

The main difficulty in introducing a link atom L is its role as a replacement for an MM atom M_1 . L is connected to a QM atom Q_1 and must thus have similar properties to M_1 in its contribution to the potential (see Figure 1). This can be circumvented by freezing the position of L, which will result in problems when considering the existence of multiple link atoms as their relative positions will never change. In our code, we employ a Q_1 -L distance-dependent correction function, which is added to the energy and the forces, such that the resulting potential is closer to the QM properties of the original Q_1 - M_1 bond. We have, for demonstrative purposes, chosen to use the rather standard case of a cut between amino acid backbone and side chain, but it is easy and cheap to extend the list of correction potentials. A similar approach was already taken elsewhere for the cases of amino acids and DNA.^[12,13]

We will, however, see that the use of link atom correction functions (LCFs) is not always beneficial; hence, the main purpose of this paper is to introduce our QM/MM interface for Gromacs. We will, in this context, use and test the LCFs but mainly concern ourselves with the feasibility to directly step from the Gromacs-generated output into a QM/MM calculation without the need to recompile/set up the QM programs. In the following section, we will present the general construction of the interface, as well as shortly present how the LCFs were devised. We will then proceed by presenting some exemplary results of model systems and real use cases, and we will conclude the paper with a discussion and an outlook.

2 | METHODS

2.1 | Interface

The gmx2qmmm script was written using the Python 2.7 language; a Python 3 version is currently under preparation. It currently supports the Gaussian package (tested with Gaussian16^[14] but, to our knowledge, compatible with most earlier versions). Gromacs' compatibility was tested for Gromacs^[5] 201X versions but should be present for all versions that use the Gromacs gmx executable, and even for earlier versions after small changes to the corresponding Gromacs-executing code.

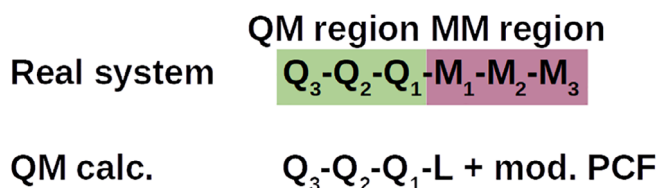


FIGURE 1 $Q_n/M_n/L$ atom definitions, as well as schematic representation of the difference between the full, real system and the actually computed system during the QM calculation. Note that proximity of other cut bonds can reduce the index n at each Q_n and M_n atom. In addition, typically, more than a single Q_2 , M_2 , Q_3 , and M_3 atom exists for each cut bond. A charge-shifted, modified PCF takes care of the charge missing as a result of replacing M_1 by L. PCF, point charge field; QM, quantum mechanics^[11]

Most emphasis was put on making the script extremely user friendly. As such, the script reads and uses only plain-text output files. Gromacs-style help on startup is available as well, and the performed procedures are traced by a logfile. The whole user experience was designed to feel like a smooth transition to just another part of the molecular dynamics (MD) procedure.

Due to this design philosophy, the interface only requires the location of the interfaced executables. This severely simplifies the setup as there is basically no need for one. After unpacking the script, the user provides the path of the Gaussian and Gromacs executables in a path file, which is then provided each time when calling the interface. The path file does not change depending on the calculation, and thus, the writing of this file is the only process that is required to set up the interface for a specific computer system.

2.2 | Running the script

The script is called by a single command line, in the simplest case, by stating

```
python gmx2qmmm
```

in a directory that contains all required files named with the default file names. The actual full syntax for calling the script is

```
python gmx2qmmm -c conf.g96 -p topol.top -qm qm.dat -mm mm.dat -qmmm qmmm.dat -act act.ndx -path path.dat -g logfile
```

This style is very familiar to users of Gromacs, which is intended. In order of appearance, the options specify geometry (.gro or .g96 format), topology, input specific for the QM calculation, input specific for the MM calculation, input specific for the QM/MM calculation (main input file), list of active atoms (ie, nonfrozen atoms, in Gromacs .ndx format), the path of the QM and MM executables, and the name of the output logfile. The command is constructed such that it provides a smooth transition from Gromacs to the interface as it adheres to Gromacs standards.

Input that is nonstandard is separated into specific units (qm/mm/qmmm/path), so those files can, in an ideal case, be set up once and then never be altered anymore, just reused in different combinations. As such, a qm.dat file includes the computational method and basis set and other parameters that are only relevant for setting up the QM input. Note that there are currently no MM-specific alterations possible; the mm.dat file is thus currently only a dummy file. The most important input file is qmmm.dat, which contains all the information on the job itself, starting with the name, the type (single point, optimization), the integrator (steepest descent, conjugate gradient, BFGS), and the corresponding initial parameter values and thresholds. Finally, the path.dat, which is the "setup" file for gmx2qmmm, was already discussed above as the location for the interfaced programs. Details and examples regarding how to run individual jobs can be found in the readme in the base directory of gmx2qmmm, as well as in the /examples/folder delivered with the package. For further details, we would like to direct the users to the manual.

Calling any instance of gmx2qmmm with the -h option will bring up a Gromacs-style help that guides the user toward the individual file types needed.

2.3 | QM/MM potential

The energy and forces of a given geometry are constructed using a regular additive QM/MM approach:

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{Link}} \quad (1)$$

The E_{QM} is obtained from the self-consistent field (SCF) energy of a QM single-point calculation containing all atoms in the QM region, with the MM atoms represented as point charge field (PCF). Depending on the output conventions of the QM program, the SCF energy needs to be corrected for the self-interaction of the PCF or the interaction of the nuclei with the point charges to avoid double counting with the E_{MM} term. In the case of Gaussian16, for example, the self-interaction energy of the PCF is included in the reported SCF energy and thus needs to be removed to form the E_{QM} term that we actually use for our calculations.

E_{MM} is computed by Gromacs for a calculation that contains all atoms of the system but excludes all interactions between QM atoms, as well as all Coulombic interactions between QM and MM atoms. For a system containing no covalent bonding between QM and MM atoms, a simple summation of E_{QM} and E_{MM} yields $E_{\text{QM/MM}}$. To obtain forces, the Cartesian derivatives are computed for each atom as well, using the same setups as for E_{QM} and E_{MM} ; the full-force vector for all atoms is obtained analogously to the energies.

In case of covalent bonds, a boundary correction needs to be introduced to the energies and forces, and we use link (hydrogen) atoms here. This is not the same as the LCF correction potentials mentioned in the introduction; those will be discussed below. The link atom correction in

our QM/MM scheme tries to retain as much of the force field information on the cut bond as possible. Therefore, force field terms corresponding to $Q_1-M_1-M_2$ angles, as well as $Q_1-M_1-M_2-M_3$ and $Q_2-Q_1-M_1-M_2$ dihedrals, are computed during the evaluation of E_{MM} . To account for their now bonded situation, dispersion is not computed for Q_1 and M_1, M_2 , or M_3 ; not for Q_2 and M_1 or M_2 ; and finally, not for Q_3 and M_1 . The scheme is thus very similar to that proposed by Eurenus et al.^[15] There is, however, the distinct difference that our Q–M bonds are treated quantum mechanically and that the LCFs introduce a behavior close to a full QM calculation of the real system (scaling the forces between Q_1 and L to the forces of Q_1 and M_1 , see below). As such, in our QM/MM approach, all atoms, including the boundary atoms, can move freely.

After the above procedure, we still need to adapt the PCF to mimic the original environment dipole field for each QM atom. We decided to follow a charge shift scheme as described elsewhere^[11] but with the simplification of choosing an arbitrary fixed value (+/–0.1944, see Supporting Information for details) for the M_2 -shielding charges, numerically optimizing only their positions to minimize the average deviation from the original dipole moment enacted on each QM atom. We plan for future releases to include the option for analytical charge shift as well.

While the charge shift scheme solves the problem of the overall PCF dipole moment in the QM calculation, we went a step further. We considered emphasizing the force field over the QM approach as the force field does contain information about properties of the M_1 atom. As such, we seek to remove any contributions of the artificially introduced changes to the PCF on the forces and energies, as well as most contributions originating from the L atom. While we do use a standard charge shift scheme to obtain the results of the QM calculation, we modify the results by removing all terms that we consider artificial when using the force field information on the M_1 bonding situation (see Equation 3).

We thus remove Coulomb interactions for forces and energies printed by the QM calculation, which correspond to interactions with L and the PCF, as those are computed by the MM program on the basis of the M_1 atom. The same treatment is enacted on Coulombic interactions resulting from the M_2 -shielding charges and all Q atoms as those shielding charges do not exist in the MM calculation. Finally, if the M_2 charge was modified by the charge shift scheme for the QM calculation, the forces and energies of that difference will be corrected for as well. These corrections obviously require representation of the QM (Q and L) atoms as point charges to create a classical Coulombic energy or force expression; we chose the Merz-Kollman^[6] scheme for this task. The full expression for the QM/MM energy in our scheme thus reads as follows

$$E_{QM/MM} = E_{QM}(Q, L, \text{charge shifted PCF}) + E_{MM}(Q_{vdW}, M, \text{mixed terms } M_{n>1}) - E_{\text{Link}} \quad (2)$$

$$E_{\text{Link}} = E_{\text{Coul.}}(c(L), c(\text{PCF})) + E_{\text{Coul.}}(c(Q_{\text{bonded}}), c(\text{PCF}_{\text{bonded}})) + E_{\text{Coul.}}(c(Q), c(\text{PCF}_{\text{shielding}})) + E_{\text{Coul.}}(c(Q), c(\text{PCF}_{M_1\text{-shifted}})) \quad (3)$$

Forces are computed analogously via the Cartesian negative first derivatives of the respective individual terms.

The resulting scheme in itself would already slightly differ from typical QM/MM approaches, which usually put more trust in the QM calculation than in the MM approach. Our way to remove the additional point charges found in the QM calculation from the final result is not complete (due to the remaining L–QM interactions), but it allows for using the information the force field has on the bond between the QM and the MM zone. In contrast to what was used before elsewhere,^[15] however, we retain the QM forces at the L atom and partially compensate for Q_1 –L bond-related artifacts by introducing the LCFs below.

2.4 | Construction of the LCFs

We scanned the relaxed potential energy surface (PES) along the stretch coordinate between the C_α and C_β carbon atoms of a Ser residue in a Gly-Ser-Gly (GSG) model polypeptide. Initially, we used the CAM-B3LYP^[16] method and a 6-31G* basis^[17] for this task but then rapidly extended to a variety of density functional theory (DFT) and other approaches. This scan was carried out once using a full QM model, and energies and forces were then computed for the QM scan geometries based either on full QM or a QM/MM potential with a force field for the MM part (initially, the AmberGS force field^[18] was used, but others were included as well for the final data base, see below).

The differences between the two obtained PESs were then used as the basis for fitting a correction curve, which we termed our LCF for Ser between C_α and C_β . To demonstrate the dependence on the type of fit, we once used a third-order polynomial potential and another fit that employed a Morse function; yet, only the Morse function proved to be up to the task and was used for the presented purposes. Finally, the fit was carried out for both energies (for the whole system) and forces (along the C_α – C_β bond). The corresponding graphs can be found in the Supporting Information. The effective correction for the total system energy is

$$E^{\text{corr}} = E^{\text{QMMM}} + \sum_{Q_1 M_1 \text{-pairs}} \Delta E^{\text{LCF}}(r_{Q_1 M_1}) \quad (4)$$

$$\Delta E^{\text{LCF}}(r_{Q_1 M_1}) = a \left(e^{-2b(r_{Q_1 M_1} - c)} - 2e^{-b(r_{Q_1 M_1} - c)} \right) + d \quad (5)$$

The parameters a , b , c , and d are specific for each combination of the QM and MM methods and are found in the database.csv file of our package, so they can be freely edited and added if additional methods are required. To obtain corrected forces, the Cartesian forces are projected onto the internal coordinate connecting Q_1 and M_1 and then corrected analogously to the energy term correction via

$$F^{\text{corr}}(r_{Q_1M_1}) = F^{\text{QMMM}}(r_{Q_1M_1}) + \Delta F^{\text{LCF}}(r_{Q_1M_1}) \quad (6)$$

$$\Delta F^{\text{LCF}}(r_{Q_1M_1}) = a' \left(e^{-2b'(r_{Q_1M_1} - c')} - 2e^{-b'(r_{Q_1M_1} - c')} \right) + d' \quad (7)$$

The parameters a' , b' , c' , and d' are again specific for each combination of the QM and MM methods and are available for the user to edit. The resulting force correction ΔF across the QM/MM boundary is not directly added to the internal forces; instead, it is projected back to Cartesian coordinates and equally distributed to Q_1 and M_1 .

The results corresponding to the use of different methods/basis sets or force fields will not be discussed in detail. Note that the database does have the capability to be extended to the method combination of choice so that the user can obtain a consistent LCF for his combination of methods, even if such a combination is not yet existent. The user can also choose to deactivate the LCF treatment altogether. In short, the user is allowed to use any method or force field that is available in the QM and MM programs within their respective limitations.

2.5 | Small polypeptides

Model polypeptides (GXG, with X being any standard amino acid apart from glycine and proline) were created from scratch in a linear arrangement and fully optimized using our QM setup. While GSG was used as the training system, all other systems were tested based on their performance when using the GSG parameters for the displacements along the cut bond between the side chain and backbone.

Additional test runs with GAG and GVG comprised of a simple rotational scan about the cut bond; see Figure 2. As (currently) no coordinate scan feature is available in gmx2qmmm, the scans were performed in the QM software. The resulting coordinates were transferred to the QM/MM interface to compare the QM-optimized geometries on the basis of QM and corrected/unmodified QM/MM single-point energies.

2.6 | LvWSCP

The *Lepidium virginicum* water-soluble chlorophyll-binding protein (LvWSCP) represents a primary use case for our interface. In a previous study,^[19] we generated a variety of MD trajectories with Gromacs, and gmx2qmmm is used here in this article to perform a spectroscopic analysis.

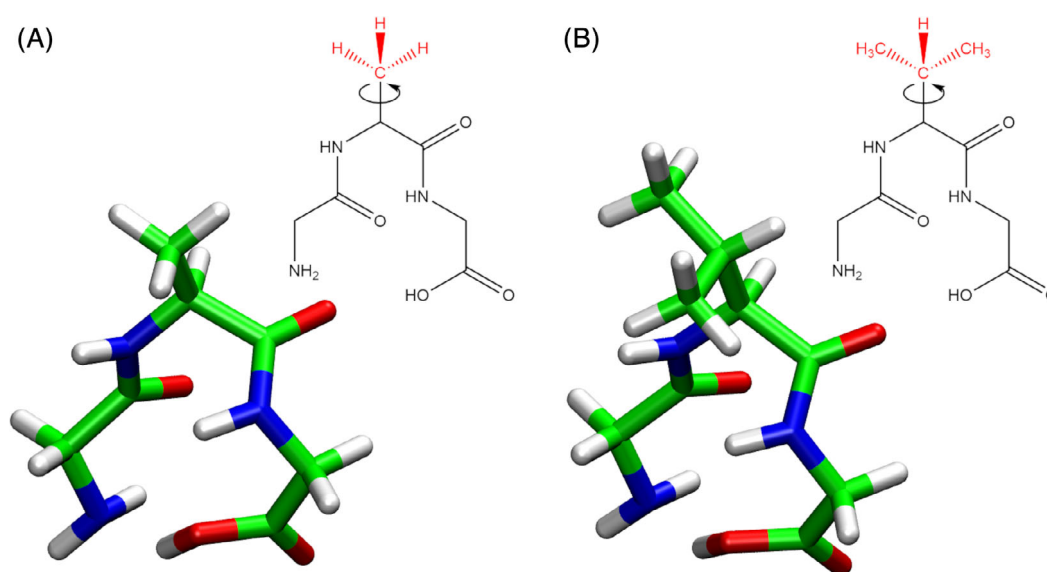


FIGURE 2 Graphical representation of the model systems for the rotational scans and 3D structures showing the potential minima. A, Gly-Ala-Gly; B, Gly-Val-Gly

The full analysis of the corresponding data will be the content of an upcoming paper. Here, we present the differences between the optical spectra computed on the basis of different geometries. Thus, the interface is tested in terms of optimizing ground-state geometries and the setup of time-dependent DFT (TD-DFT)^[20,21] single points.

The Gromacs MD setup is described in detail elsewhere.^[19] From snapshots taken every 1 ns from the first 25 ns of our simulation trajectory, we computed TD-CAM-B3LYP/6-31G* vertical absorption spectra of the LvWSCP chlorophyll (Chl) a compounds. The structures were treated in three different ways before spectral analysis: (a) no further structural modification, (b) optimization using the force field potential, or (c) optimization using QM/MM. Details on the exact setup can be found in the Supporting Information. The QM zone consisted of a Chl a molecule without the phytol chain and the coordinating P38 backbone segment.

2.7 | CP29 in a thylakoid membrane

Finally, we computed Chl a and b optical spectra depending on the presence or absence of a CP29 protein environment. The corresponding setup is described in the Supporting Information. The effect of the protein environment was tested based on the levels of the geometries, the spectra, and the combination thereof. We also performed excited-state optimizations using both the gas phase and the QM/MM potential. This model system thus extends the test toward excited-state geometries and fluorescence calculations compared to the LvWSCP case.

3 | RESULTS

3.1 | Energies, forces, and PES of small polypeptides

The resulting errors in energies and forces along the Q_1 – M_1 side chain-backbone bonds are illustrated in Figure 3 by comparing the corrected potentials to the exact QM potentials along all points of the scan. It can be seen that the LCFs greatly improve the general behavior of the potential, for both energies and forces for all shown cases. To make sure that no artifacts from the closed-shell approach arise, we also tested unrestricted and restricted open-shell approaches, the results of which are found to be identical in all three approaches (data not shown).

To see how our potential holds up against a full QM description of related coordinates, as well as how much our correction influences the resulting PES, we compared the rotational potentials about the cut side chain-backbone bond for GAG and GVG. The resulting PES can be seen in Figure 4.

While GAG is very symmetric in the QM regime, both corrected and uncorrected QM/MM fail to accurately reproduce the slope toward the rotational barrier. Interestingly, however, they fail in different rotational directions. This is because the scan was actually performed unidirectionally (coming from the left side of the graph), which means that some hysteresis of the curve can be expected due to the asymmetry arising from the backbone direction. Once the scan passes the minimum, the corrected QM/MM performs better until it again accumulates some error toward the peak maximum. Overshooting like this can be expected as our LCFs were trained for the situation close to the PES minimum. Such a scan is indeed an extreme case as placing observables of interest directly inside the QM/MM boundary is a constructed case, which should not be applied in reality.

The findings from GAG become even more illustrated for the GVG case. In contrast to GAG, the correction is slightly better for the slope left of the GVG minimum structure; however, QM/MM generally overshoots the predicted energies along the rotational scan. This is likely due to the clash of QM and MM regions, which can have dramatic effects on the energy values when, for example, a negative point charge of the PCF gets close to the QM region. As GVG has an extensive QM region/side chain, such clashes are to be expected along the rotation of the side chain in contrast to the GAG case. Still, the overall shape of the curve is identical.

It should be noted, however, that these results may highlight a general shortcoming of QM/MM when cutting close to the backbone. Our LCFs do not change this picture in one way or the other as their effect on the rotational barrier is minute (see Supporting Information). It is also found that the curve based on pure force field energies (red line in Figure 4) will generally follow the QM curve, although it already exhibits a small trend toward the QM/MM results. As we disentangle the energy contributions of our QM/MM setup, we see that the bias toward higher energies arises from the link energy term, which exhibits a maximum at about 240°.

In our QM/MM boundary setup, akin (although not equal) to Eurenus et al.,^[15] we do remove all Coulomb interactions between the GVG side chain and the 1 to 4 bonded atoms, such as the backbone carbonyl carbon and the backbone nitrogen atom. Yet, we reintroduce the interaction of these atoms via their dihedral energy terms. Previous studies found that this can be avoided by removing any cross-border MM terms in the boundary treatment,^[22] although this would encourage freezing of the boundary atoms, which we actively seek to avoid. However, our study further fortifies the QM/MM caveat that using one method (MM) to obtain geometries and using another (QM/MM) to perform more detailed analyses requires a careful selection of the boundary region; this problematic conformation is an extreme case for which the Merz-Kollman scheme likely does not provide proper charges for all participating 1 to 4 terms (see Supporting Information for further discussion). Further testing is, however, needed to provide a definite answer.

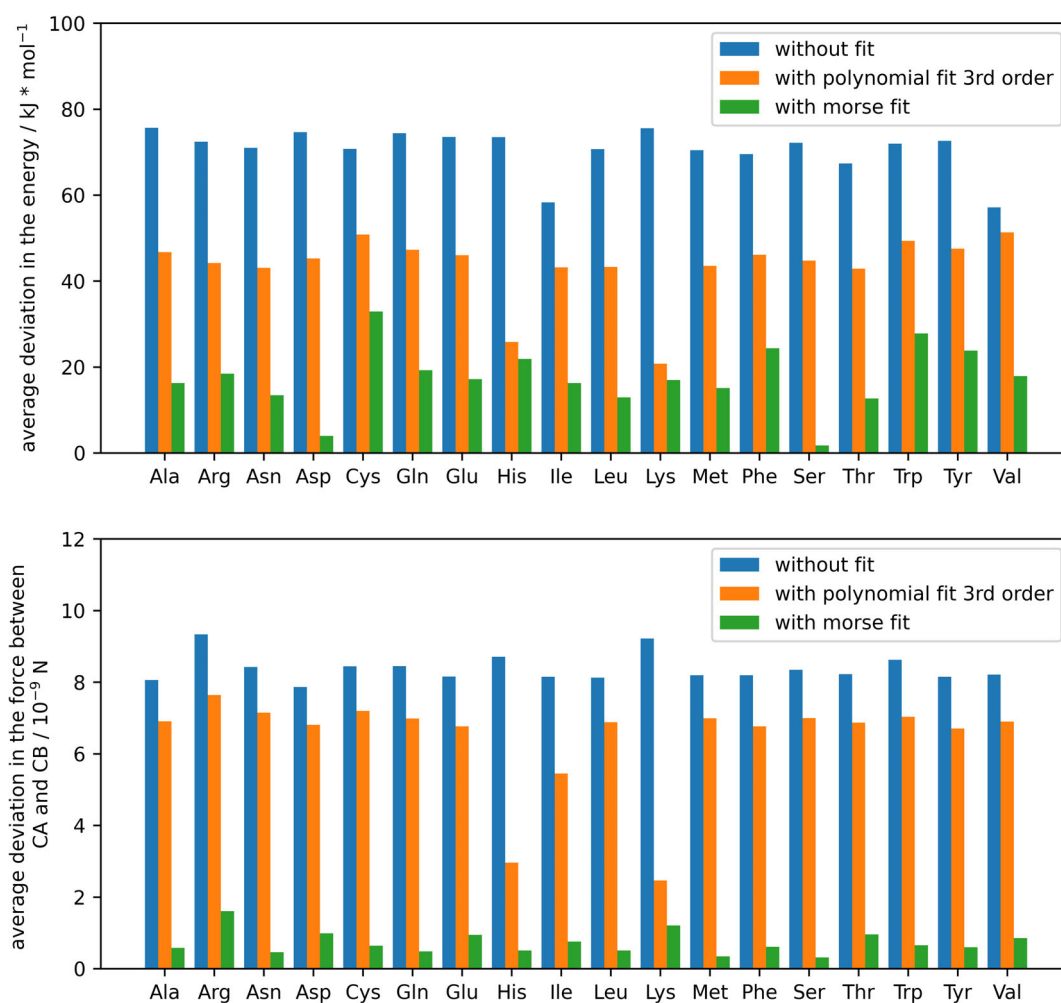


FIGURE 3 Comparison between uncorrected and corrected QM/MM (CAM-B3LYP/6-31G*:AmberGS) stretching potentials along the QM/MM cut C—C bond between backbone and side chain for various amino acids, and average absolute errors from exact QM potentials shown. Top: energies, bottom: forces (norm). Note that LCF parameters were obtained for Ser and applied unchanged to all other shown cases. Gly and Pro not included due to the respective molecular structure. LCF, link atom correction function; QM/MM, quantum mechanics/molecular mechanics

In conclusion, we see that the LCFs indeed improve the linear potentials, even for those cases for which they were not parametrized. However, some related coordinates may suffer in quality, although the error when comparing to the exact (QM) PES is only negligibly changed upon introduction of the LCFs.

3.2 | Geometries and spectra of LvWSCP

The spectra computed for LvWSCP Chl a, optimized under different conditions, can be found in Figure 5. Binned splines, integrating the state oscillator strengths of a specific state over all computed snapshots, result in the apparent broadening of the peaks; see Supporting Information for details. The non-Gaussian character of the broadening can be seen best for the case without optimization—lowest energy state (broad band at 630–700 nm). We chose this type of broadening as an improved way to keep the conformation-dependent character of the energy and inhomogeneous intensity spread for each individual state.

In comparison to the experimental spectra, we can see that none of the computational spectra agree well, especially for the higher-energy Soret band (<500 nm). The good agreement with the experiment on the lowest state (~650 nm) is lifted when transitioning to the supposedly improved QM/MM description. The overall “bad” agreement is, however, due to basing the analysis on vertical excitation spectra, and the good agreement with the lowest state is likely a fortuitous cancellation of errors. A proper approach for obtaining the experimental spectra would include the analysis of the normal modes in each state, leading to a vibrational broadening. The effect of vibrational broadening can only be roughly mimicked on the basis of a ground-state MD simulation as, at best, the properties of the ground-state nuclear wave function can be approximated by the molecular mechanics

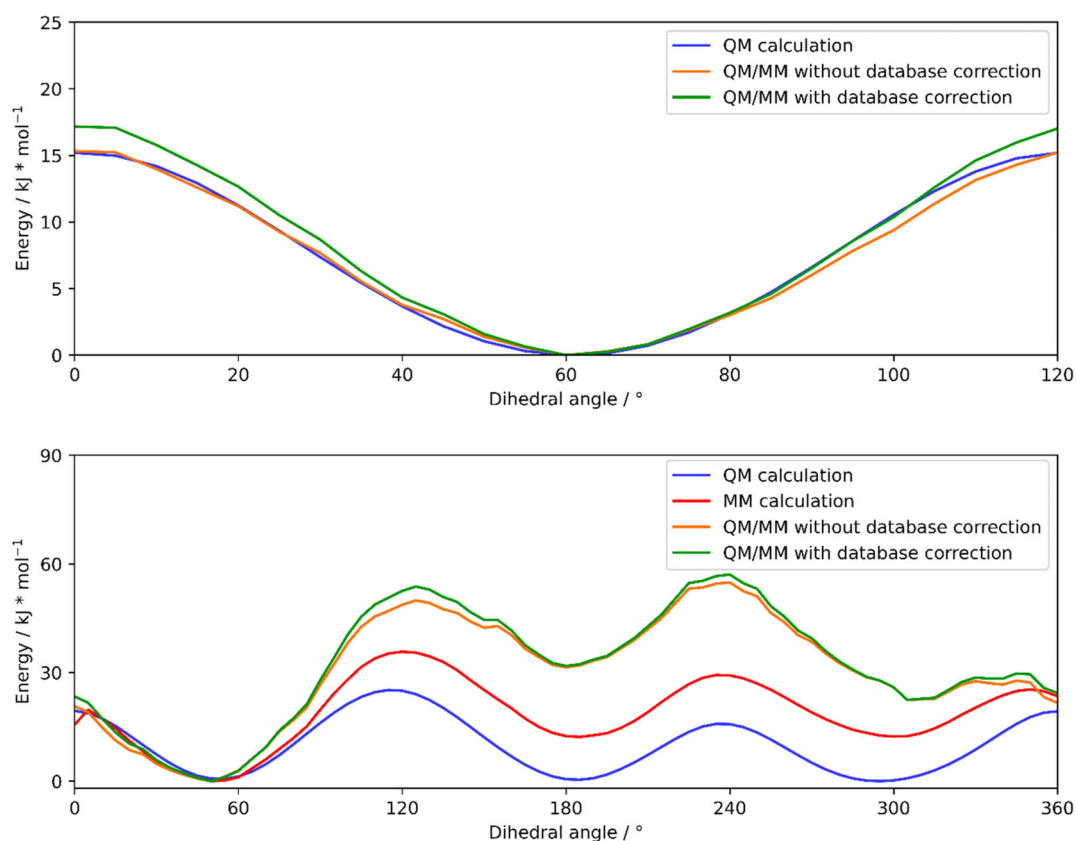


FIGURE 4 Comparison between full QM, MM or uncorrected and corrected QM/MM (CAM-B3LYP/6-31G*:AmberGS) rotational potentials about the QM/MM cut C–C bond between backbone and side chain for GAG (top) and GVG (bottom); all energies based on the full QM scan relaxed geometries. QM/MM, quantum mechanics/molecular mechanics

force field.^[23–25] The computational spectra, however, clearly demonstrate the importance of the choice of structural ensemble for computing the energies and intensities of the transitions. A detailed discussion can be found in the Supporting Information.

To understand the spectral shifts and their structural origin, one has to recall that we optimize the ground state, and thus, the energy of that state is strongly stabilized, while the other states can change unpredictably. This must, on average, result in a larger gap in energy, leading to the blue shift in the spectra. That QM/MM and MM optimization do not agree on the shifts indicates that there are structural features that can only be well described using QM. The changes in the geometries are exemplarily visualized in Figure 6. It can be seen that the Mg ion, which our chosen MM parameters basically force to be coplanar with the porphyrin ring, is especially coming out of plane and thus assuming a more realistic penta-coordinated geometry. As the properties of the Mg ion are known to be crucial for the spectrum,^[26] this suggests that QM/MM treatment of the system is needed to obtain proper spectral features. It should, however, be noted that MM parameters for bound porphyrins are state-of-the-art for other cases such as heme groups.^[27]

The change in relative intensity can be expressed as a difference in the length of the transition dipole vectors. Indeed, more in-depth analysis indicates that not only the lengths but also the orientation of the vectors has changed between QM/MM and MM structures (data not shown). This has consequences for the coupling between chromophores in LvWSCP and is the subject of an upcoming paper as the corresponding results and discussion would be far beyond the scope of the present article. However, this further supports our previous assessment that optimization of Chl a beyond the MM potential is required for a proper biophysical analysis.

3.3 | Spectra of CP29

The CP29 case was used to demonstrate the impact of QM/MM optimization in the ground and excited states of Chl on TD-DFT excitation energies. Two Chl units were arbitrarily chosen to account for a variance in the immediate protein environment as well, and we only present the results for one in the main article and the other in the Supporting Information. The optimized was the lowest Q band state (Q_y , the first excited state). Ground- and excited-state spectra were computed for the following cases: regular QM/MM-optimized geometries, same geometries but without PCF, and gas phase-optimized geometries starting from the respective QM/MM geometries (Figure 7A–C, respectively).

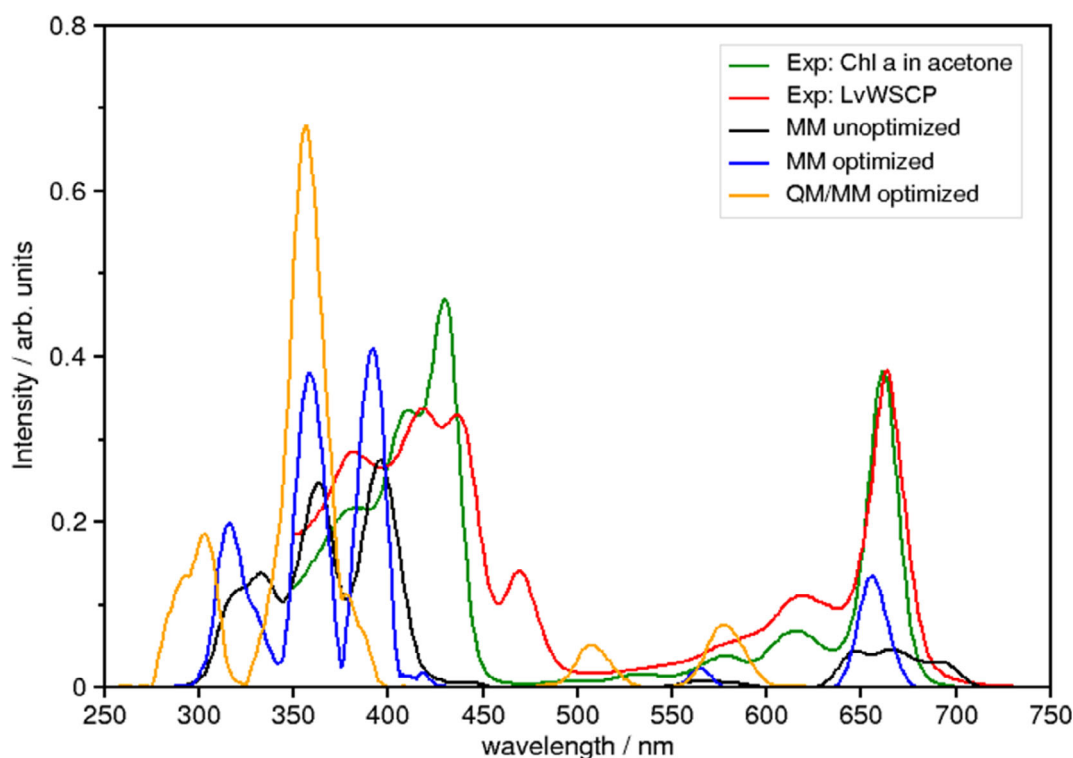
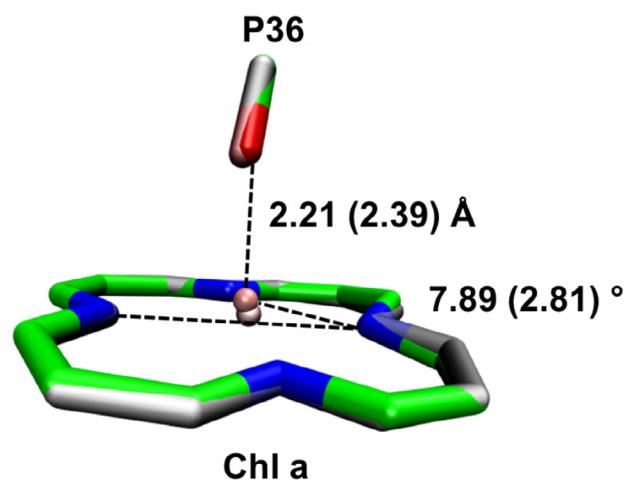


FIGURE 5 TD-CAM-B3LYP/6-31G* vertical excitation spectra of WSCP Chl a based on different structural ensembles, broadened using binned splines. Experimental spectra given for comparison, arbitrarily scaled in intensity, with kind permission of Dr Heiko Lokstein, Charles University, Prague. WSCP, water-soluble chlorophyll-binding protein

FIGURE 6 Comparison of the structural changes in an exemplary snapshot, focusing on the Chl a structure (only coordinating carbonyl and Mg-porphyrin ring shown) between nonoptimized (white) and QM/MM-optimized (colorized) geometries. QM/MM, quantum mechanics/molecular mechanics



Comparing the general effect of optimizing the Q_y state vs the ground state throughout all depicted spectra, we can see the expected general red shift of the Q_y -state excitation energy (intrinsic Stokes shift). The Q_x state (second excited state), as well as several Soret band states (<450 nm), decreases in energy as well, although some Soret states are also observed to increase in energy. Given that there is no straightforward way to predict the corresponding state geometries/energies without explicit optimization of each state, further analysis is omitted here. It can, however, be seen that our interface is capable of providing small, systematic trends in the Chl a spectra; those are discussed in more detail in the Supporting Information.

4 | SUMMARY

We present an interface called `gmx2qmmm` that generates an additive QM/MM scheme between Gromacs and Gaussian. It employs state-of-the-art link atom corrections, including an optional potential to correct for the bond at the QM/MM boundary, and is directly coupled to Gromacs

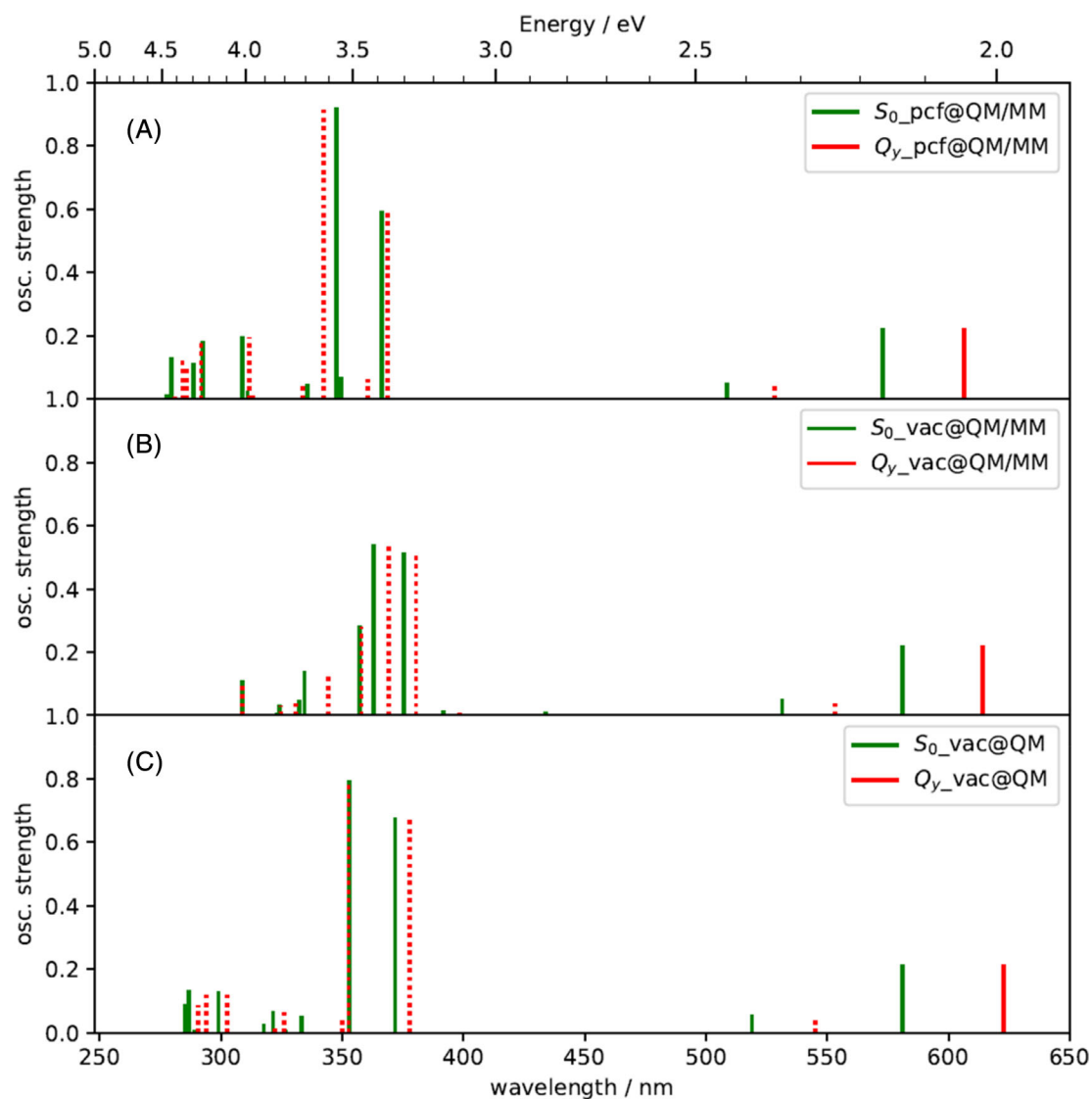


FIGURE 7 TD-CAM-B3LYP/6-31G* line spectra of a CP29 Chl a molecule (labeled 610 in the CP29 chain “r” of the crystal structure with pdb identifier 5XNL) molecule, E198 coordinated. Colors indicate different optimized geometries: green: ground-state minimum; red: Q_y -state minimum. Lowest 12 excited states computed. QM region consisted of the Chl a (phytyl chain omitted) and the coordinating Glu198 residue; see Supporting Information for details. Dashed lines indicate that oscillator strength is computed for a ground-state excitation despite the geometry being optimized for a different state; such intensities are given for comparative purposes only and are experimentally not available. A, Spectra at QM/MM geometries with point charge field present; B, spectra at same geometries as A but without point charges; C, spectra after optimization in the gas phase. QM/MM, quantum mechanics/molecular mechanics

input/output standards. As such, it is capable of directly using the same files and programs that the preceding MD simulation used, thus eliminating any possible methodological inconsistencies due to switching the underlying software.

The main advantages of our package are the extremely trivial input conventions. We focused on giving the impression that the user would be simply using another gmx tool. Furthermore, the tool directly produces only text files, which are automatically archived but are human readable. The important results are concatenated in a log file, leaving the actual QM and MM outputs to be only useful in case of errors or special applications. The whole purpose of the package is to allow less-experienced users to access the full power of QM/MM simulations with only the knowledge of how to use Gromacs.

Our test calculations show that we can perform all tasks that are, at this stage, expected from such an interface, like optimizations and spectral calculations. Optimizations are not limited to the ground-state PES but include the option to optimize any state that the QM method can represent, such as excited states. While we show that our spectroscopic results are promising, we also see that QM/MM rotational potentials do not mimic the full QM description, at least not on the basis of the QM geometries. The minima are preserved, but barriers are vastly overestimated, hinting at a deeper problem with amino acid flexibility when described with QM/MM.

Our next version updates will include a standalone Merz-Kollman^[6] computation of the point charges to be consistent between QM packages, allowing the introduction of other QM software. Furthermore, we seek to include better optimization schemes, normal mode analysis, and coordinate scans.

ACKNOWLEDGMENTS

Jan P. Götze and Simon Petry are currently funded by project no. 393271229 of the Deutsche Forschungsgemeinschaft (DFG). The authors acknowledge the North-German Supercomputing Alliance (HLRN) for providing computational resources.

AUTHOR CONTRIBUTIONS

Jan Philipp Götze: Conceptualization; funding acquisition; methodology; project administration; software; supervision; visualization; writing-original draft. **Yuan-Wei Pi:** Software. **Simon Petry:** Methodology; software; visualization. **Fabian Langkabel:** Conceptualization; data curation; formal analysis; methodology; software; validation; visualization. **Jan Witte:** Methodology; software. **Oliver Lemke:** Data curation; investigation; validation; visualization.

RESOURCES

The gmx2qmmm software is available at github.com/gmx2qmmm/.

ORCID

Jan P. Götze  <https://orcid.org/0000-0003-2211-2057>

REFERENCES

- [1] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.
- [2] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700.
- [3] H. M. Senn, W. Thiel, *Angew. Chemie Int. Ed.* **2009**, *48*, 1198.
- [4] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, M. J. Frisch, *J. Mol. Struct. THEOCHEM* **1999**, *461–462*, 1.
- [5] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1–2*, 19.
- [6] B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431.
- [7] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.
- [8] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *Rev. Comput. Mol. Sci.* **2014**, *4*, 91.
- [9] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley & Sons, New York **1987**.
- [10] L. Cao, U. Ryde, *Front. Chem.* **2018**, *6*, 89.
- [11] P. Sherwood, A. H. De Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sj??voll, A. Fahmi, A. Sch??fer, C. Lennartz, *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1.
- [12] M. Hitznerberger, T. S. Hofer, *J. Comput. Chem.* **2015**, *36*, 1929.
- [13] M. Hitznerberger, M. Ratanasak, V. Parasuk, T. S. Hofer, *Theor. Chem. Acc.* **2016**, *135*, 47.
- [14] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian16*, Gaussian, Inc., Wallingford, CT **2016**.
- [15] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, M. Hodoscek, *Int. J. Quantum Chem.* **1996**, *60*, 1189.
- [16] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.* **2004**, *393*, 51.
- [17] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.
- [18] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198.
- [19] O. Lemke, J. P. Götze, *J. Phys. Chem. B* **2019**, *123*, 10594.
- [20] C. Jamorski, M. E. Casida and D. R. Salahub, *J. Chem. Phys.*, **1996**, *104*, 5134
- [21] M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, **1998**, *108*, 4439.
- [22] P. Amara, M. J. Field, *Theor. Chem. Acc. Theor., Comput. Model. (Theoretica Theor. Chim. Acta)* **2003**, *109*, 43.
- [23] J. P. Götze, B. Karasulu, W. Thiel, *J. Chem. Phys.* **2013**, *139*, 234108.
- [24] B. Karasulu, J. P. Götze, W. Thiel, *J. Chem. Theory Comput.* **2014**, *10*, 5549.
- [25] F. Santoro, R. Improta, A. Lami, J. Bloino, V. Barone, *J. Chem. Phys.* **2007**, *126*, 084509.
- [26] A. B. J. Parusel, S. Grimme, *J. Phys. Chem. B* **2000**, *104*, 5395.
- [27] L. Castro, L. Crawford, A. Mutengwa, J. P. Götze, M. Bühl, *Org. Biomol. Chem.* **2016**, *14*, 2385.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Götze JP, Pi Y-W, Petry S, Langkabel F, Witte JF, Lemke O. A user-friendly, Python-based quantum mechanics/Gromacs interface: gmx2qmmm. *Int J Quantum Chem.* 2021;121:e26486. <https://doi.org/10.1002/qua.26486>

A user-friendly, python-based QM/Gromacs interface: gmx2qmmm

Jan P. Götze, Yuan-Wei Pi, Simon Petry, Fabian Langkabel, Jan Felix Witte and
Oliver Lemke

Supporting information

LCF construction

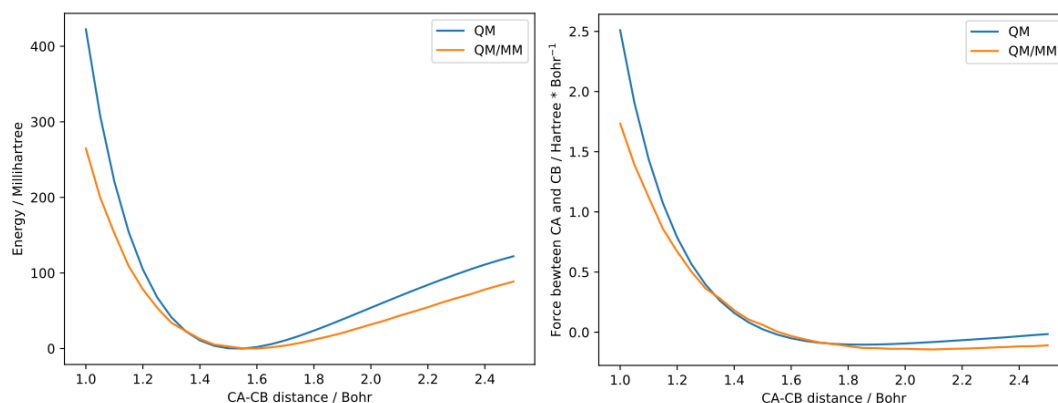


Figure S1: Comparison between full QM (QM: CAM-B3LYP/6-31G*) and QM/MM (MM: AmberGS) stretching potentials along the C-C bond between serine backbone and side chain. Left: Energies, right: forces (norm).

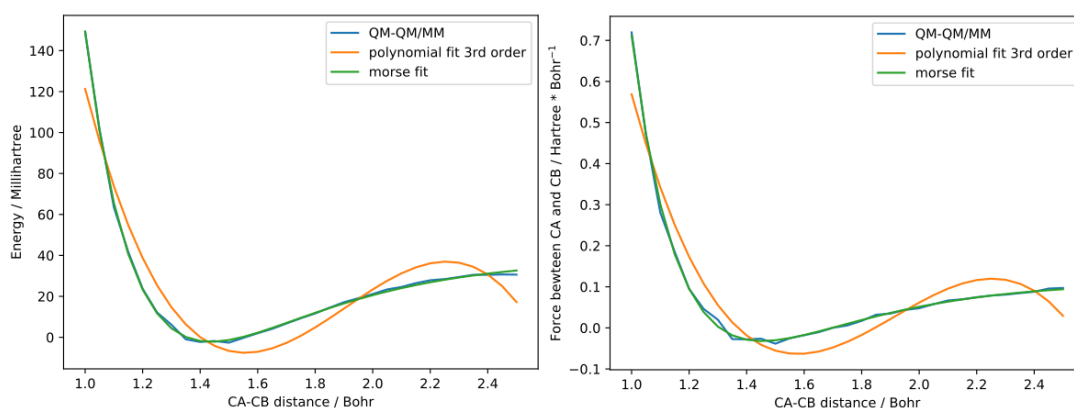


Figure S2: Fitting of the QM-QM/MM differences for energies (left) and forces (norm, right) for the case shown in Figure S1. A Morse potential obviously outperforms a third-order polynomial in both cases.

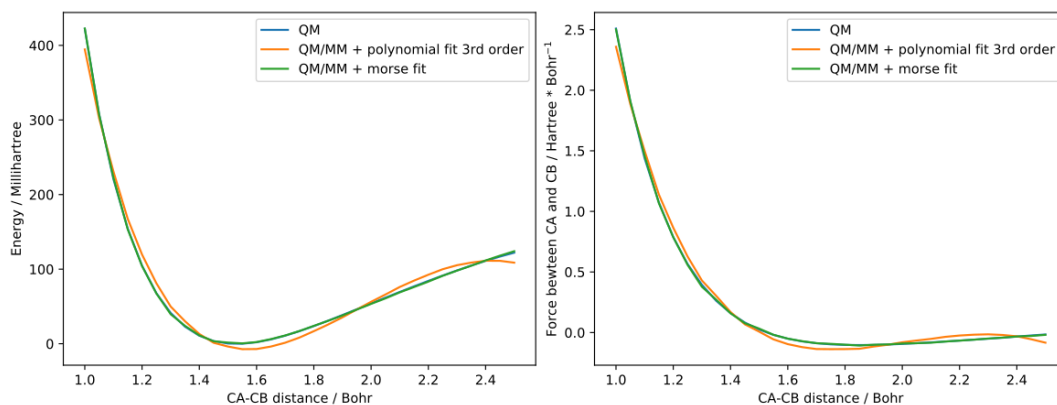


Figure S3: The resulting corrected potentials, combining the data from Figure S1 and Figure S2, compared to the exact QM potential. Left: Energies, right: forces (norm). The Morse corrected potential is virtually identical to the QM potential.

Value of charge shift charges

The charges shielding the M_2 atom charge in the PCF were set to a fixed value of ± 0.1944 , based on the results of test calculations using a preliminary approach. That preliminary approach optimized both charge and distance during the computation of the PCF modifications. The reasoning that led us to arrive at the fixed value is given below.

The main purpose of the PCF modifications is to warrant the integrity of the dipole moment at each Q atom, despite the removal of M_1 atoms and their replacement by L atoms (see Figure S4). At first, we aimed towards an approach that analytically solves the corresponding problem (as described elsewhere¹). However, it turned out that a numerical approach would actually yield better results, i.e. the error between the expected dipole field and the obtained dipole field would be lower when numerically optimizing the charge positions and values. An exemplary test case was the BLUF domain studied for an earlier publication,² here we obtained an error in the modified dipole field with the numerical approach that was 4% smaller than that of the field provided by the analytical approach. This is likely due to our approach taking into account all Q atoms during the optimization process, not only those close to the L atom connected to the M_2 atom.

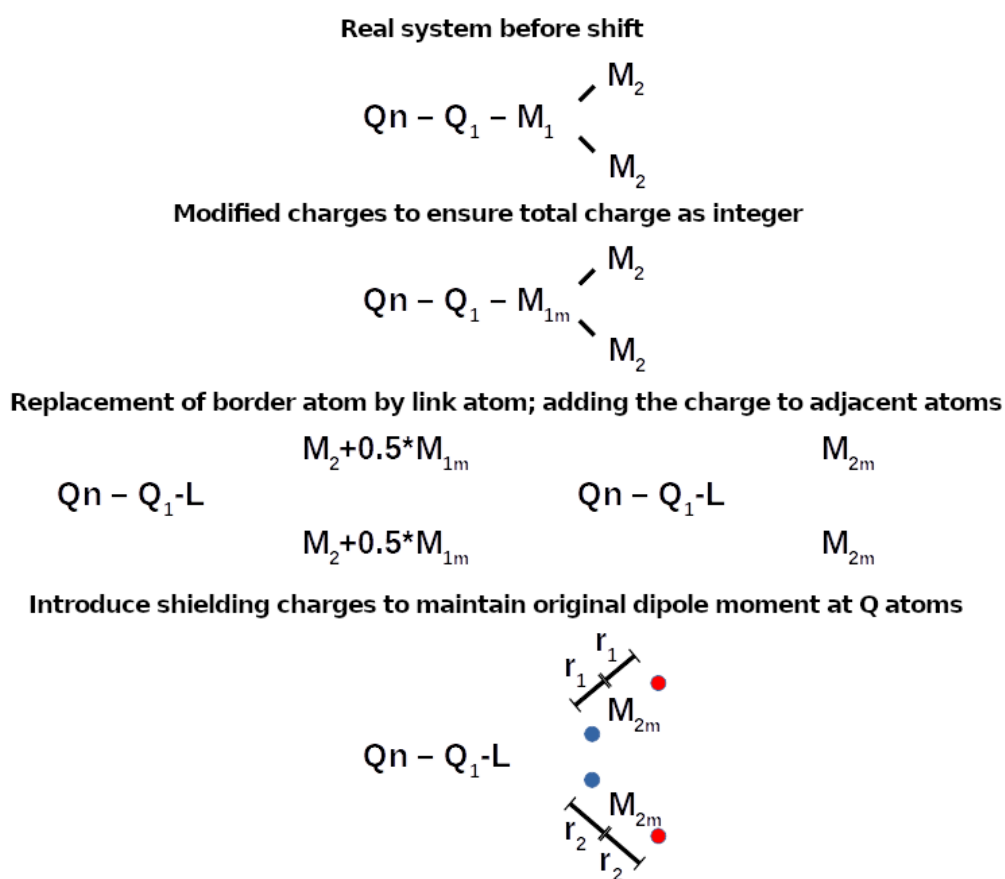


Figure S4: Overview on the charge shift scheme by Sherwood et al.¹ In the present work, the value of the shielding charges is fixed at ± 0.1944 and the set of r values is numerically optimized to ensure a proper representation of the original, real system dipole moment of the PCF at each Q atom.

Unfortunately, the numerical approach would sometimes run into problems when the shielding charges are small. Then, the optimizer would in some cases artificially inflate the charge value and the distances at the same time. We thus selected a fixed charge value, found for a real case (above BLUF domain system) for realistic distances; i.e., closer to M_2 than to L. Fixing the value changed the error by $< 0.1\%$ and thus turned out irrelevant for the actual resulting dipole field. So far, all observed systems

provide improved dipole fields using above approach, despite the arbitrary choice of the shielding charge strength. For the rare case that distances become too large, a warning flag can fire and interrupt the program; apart from unrealistic test cases (due to massive relocated charges to M_2), this has never happened for any studied case.

GVG energies far from the minimum structure

To further investigate the deviations for the GVG case from the full QM case, we have performed a variety of additional energy analyses based on the geometries used for the GVG energy graph in the main paper. The results are shown in Figure S5.

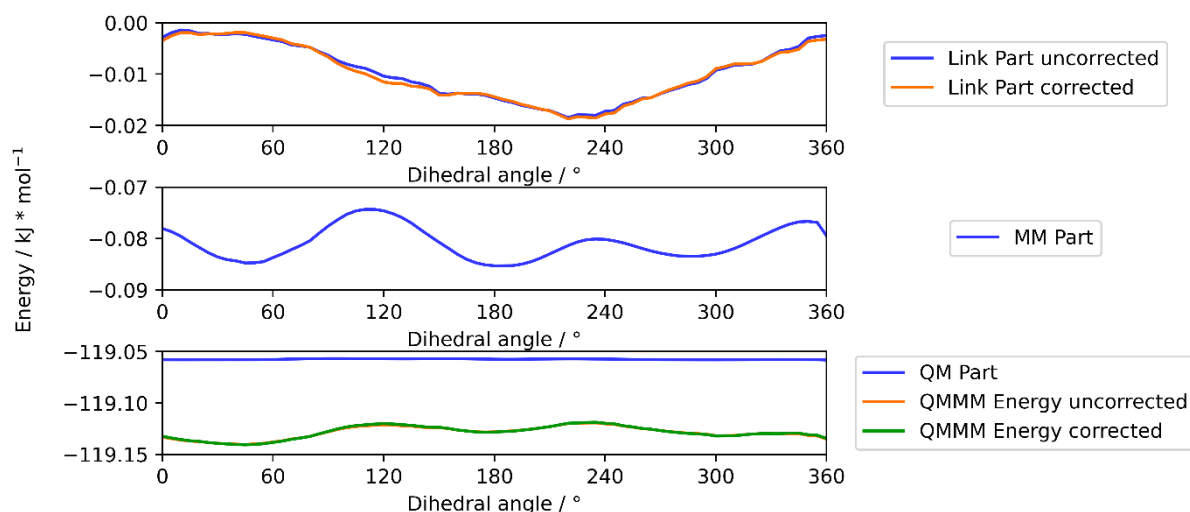


Figure S5: Energy contributions to the QM/MM energy, with (corrected) and without (uncorrected) LCFs active. Note that the link part is shown with the sign according to Eq. (2) of the main paper.

It can be seen that the effect of the LCFs is minimal. The MM energy dominates the overall energy of the scan maxima and minima; the QM energy is comparatively constant. The link energy exhibits a maximum at 240 degrees. This would indicate that either (i) the corresponding geometry is either physically unfavored, which is contradicted by the results of the full MM single points shown in the main paper, or (ii) the corresponding geometry is specifically badly described by the link potential employed in our approach. The geometry in question is shown in Figure S6.

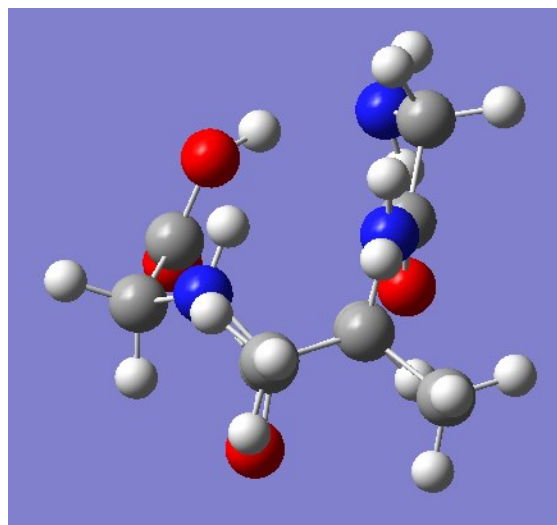


Figure S6: Extreme point along the link energy graph for GVG at 240° rotation.

In the depicted geometry, we find that various atoms are close to each other that are interacting based on their Coulomb potentials; furthermore, this geometry is the only eclipsed position that allows the methyl groups of the valine side chain to also remain close to other 1-4 bonded atoms (the 120° shifted geometries would clash with the backbone hydrogen and thus have to distort the methyl group orientation). We thus identify the reason for the high link atom contribution at 240° the spatial proximity of various 1-4 bonded groups, of which each are “corrected” via a mixed MM charge-QM charge Coulomb term. For such extreme cases, however, a more sophisticated approach needs to be taken, as the MM charges would probably be affected as well, hence such a strongly packed situation would require repolarization of the MM charges, which is not possible using a static force field approach.

LvWSCP setup

QM/MM

For the QM/MM-optimization of a chlorophyll a bound to WSCP, the system was split into three layers. The inner layer, which was optimized on a QM-level (CAM-B3LYP/6-31G*)^{3,4}, consisted of the porphyrine ring of a chlorophyll (phytyl chain was cut at the first carbon-carbon-bond after the ester) and the magnesium coordinating P36 residue (backbone was cut at the C_α of the P36 and the C_α of the neighboring A37 neglecting side chains). The central layer, which was optimized on the MM-level (see below) using a steepest descent algorithm, contained all residues, solvent molecules and ions within a range of 1.2 nm with respect to the chlorophyll's magnesium ion. For the outer layer, which was kept frozen during the optimization, the remaining protein residues as well as all solvent molecules and ions located within a range larger than 1.2 nm and smaller than 4 nm with respect to the chlorophyll's magnesium ion were chosen. The remaining molecules with a distance larger than 4 nm were neglected. Molecules that were shared between two layers were assigned to the respective inner one. The QM/MM-optimization was carried on for 20 optimization steps using Gaussian 16⁵ and the GROMACS simulation package 2016.1.⁶

MM

An optimization solely on MM-level was carried out using the steepest descent algorithm implemented in GROMACS 2016.1 (force tolerance: 100.0 kJ/mol/nm, step size = 0.01 nm, radius of non-bonded partner list = 1.2 nm). For the protein the CHARMM36 force field,^{7,8} for chlorophyll parameters were used that are described elsewhere.⁹⁻¹³

Snapshots/MD and absorption spectra

As starting structures for optimizations, 25 snapshots were taken at equal distance of 1 ns from a simulation of WSCP containing four chlorophylls; detailed information and simulation setup are found elsewhere.¹⁴ Thus, 100 absorption spectra were calculated (one spectrum for each chlorophyll per snapshot). The absorption spectra were calculated using TD-DFT^{15,16} and the QM approach stated above calculating the first ten excited states. The system was defined analogously to the inner layer used in the QM/MM-optimization. The central and the outer layer were treated as a point charge field during the calculation. Following each TD-DFT calculation, the excited states were classified according to the orbital contribution as described elsewhere,¹⁷ to allow a state-wise assignment of the peaks throughout the different snapshots. Same excitations between different calculations were grouped.

For the plotting of the absorption spectra, all excitation within a group were binned using a bin size of 10 nm and the cumulative sum of the oscillator strengths was calculated. The spectra were normalized by the total number of calculations. To obtain a smooth spectrum a spline function with a resolution of 1 nm was used.

Spectroscopic analysis

Foremost, it can be seen that optimized spectra, regardless of the used potential, are exhibiting a much smaller broadening of the peaks. This indicates that the structures after optimization should exhibit a smaller structural deviation; however, such a statement is difficult to support by our QM/MM data as the size and participating atoms of the active region varies between snapshots, while the QM region remains constant in size. For purely MM optimized geometries, however, RMSD can be quantified easily and is found to go down (2.212 Å optimized vs. 2.243 Å non-optimized).

The second insight from the optimization is that the MM potential will not drastically change the average peak energies compared to the non-optimized case. For the most important peaks, namely Q_y , Q_x (lowest two excitations) and the two Soret states (higher excitations) with highest oscillator strength, we find an average/maximum difference between no optimization and MM potential optimization of 5.5/10 nm or 31/57 meV. Combined with the insights from the previous paragraph, the thermalized MD simulation appears indeed to simulate an ongoing, stable dynamics in close vicinity to a MM potential minimum; a structure that can be found through the optimization process.

The comparison between the MM and the QM/MM optimization (see main paper) shows that the QM/MM potential results in a general optical blue shift of all bands compared to the MM potential, as well as a redistribution of relative intensities. In both the Q and Soret bands of the spectrum, the second state in each band gains oscillator strength compared to the energetically preceding states upon QM/MM optimization. Due to the fact that intensities are affected by vibrational broadening as well (see above), we however refrain from further analysis at this point.

CP29 setup

MD Setup

The structure of light-harvesting CP29 was isolated from the high-resolution electron microscopy structure of a stacked C2S2M2-type PSII-LHCII supercomplex from *Pisum sativum* (pdb entry: 5XNL, res. 2.7 Å) resolved by Su *et al.*¹⁸ The repaired structure was embedded in a membrane of DOPC to reproduce the *in vivo* architecture. The membrane is created using the website "<http://memgen.uni-goettingen.de/>" and passes the following steps. During the orientation phase the complex is orientated along the z-axis, while the bilayer lies in the x-y plane. Afterwards the lipid is packed around the complex by cutting a sphere out of the middle of the bilayer and placing the complex in. The lipids are expanded and then slowly contracted back until area per lipid of 0.769 nm² is reached, which is comparable to the experimental results of a DOPC membrane with 0.725 nm².¹⁹ During the shrinkage, only the lipids are flexible and the complex is frozen. After adding 12 neutralizing potassium ions, a general MD procedure takes place with an energy minimization, a short NVT (100 ps with a time step of 2 fs) equilibration phase followed by a longer NPT phase (1000 ps / 1 ns with a time step of 2 fs). A MD simulation with 50 million steps with a time step for integration of 0.002 ps which corresponds 100 ns was performed on the HLRN system. The AMBER99SB*-ILDNP force field²⁰ has been used for the proteins, lipid11 for the lipids and *ad hoc* for the cofactors. For the chlorophylls and the carotenoids, the set of parameters is reported elsewhere.^{21,22} Water molecules were treated using the TIP3P model.²³⁻²⁵

QM/MM setup

For the QM/MM optimization, a snapshot of the MD Simulation at 100 ns was taken as an initial structure. Two (Resid: 608 and 610) of the fourteen chlorophylls of the CP29 were exemplary isolated.

Both chlorophyll calculations were split into four layers. The QM layer, which was optimized on a QM level (CAM-B3LYP/6-31G*),^{3,4} consisted the porphyrine ring of a chlorophyll (phytyl chain was cut at the first carbon-carbon-bond after the ester) and the magnesium coordinating residue. The coordinating residues are water for the chlorophyll b (resid 608) and glutamate for the chlorophyll a (resid 610). The inner layer, which is subdivided in a frozen and a non-frozen part (active layer), included all atoms within a range of 4 nm with respect to the chlorophyll's magnesium ion. The active layer, which was optimized on the MM level (see below), using a steepest descent algorithm, contained all residues, solvent molecules and ions within a range of 0.3 nm with respect to all atoms in the QM layer. All atoms above the 4 nm radius were kept frozen and are described by the outer layer. Furthermore, all solvent molecules with a distance larger than 2 nm were neglected. The QM/MM-optimization was carried on for 99 optimization steps using Gaussian16⁵ and the Gromacs simulation package 2016.1.²⁶

The optimized structures were used to calculate the influence of the PCF and the geometry of the ground and the first excited state. Therefore, a geometry optimization followed by a time-dependent DFT (TD-DFT)^{15,16} calculation using the gmx2qmmm interface was performed ($S_0_pcf@QM/MM$, $Q_y_pcf@QM/MM$). Based on this previous optimized structure a TD-DFT calculation was performed by removing the surrounding PCF for the two lowest states ($S_0_vac@QM/MM$, $Q_y_vac@QM/MM$). Furthermore, a geometry optimization followed by a TD-DFT calculation in vacuum using GAUSSIAN was performed, based on the pre-optimized gmx2qmmm structures ($S_0_vac@QM$, $Q_y_vac@QM$).

MM setup

An optimization solely on MM-level was carried out using the steepest descent algorithm implemented in Gromacs 2016.1 (force tolerance: 100.0 kJ/mol/nm, step size = 0.01 nm, radius of non-bonded partner list = 3 nm). The used parameters are described in the MD-Setup (see above).

Spectral analysis

Comparing the spectra A and B of CP29 Chl a (main paper), we see that the presence of the PCF induces a blue shift in the excitation and fluorescence spectra. This implies that the PCF is more beneficial for the relative stability of the ground state than for the other states; this is what we would readily expect from a properly optimized ground state geometry (for the green bars) or a Q_y PCF that is also similar to the ground state PCF configuration (for the red bars, both dashed and solid). The latter is supported by the fact that the energetic shift of the Q_y excitation upon Q_y optimization is almost not affected by the presence of the PCF. Note that the spectrum B (main paper) contains a few low-intensity charge transfer states occurring between 380 and 475 nm, which are due to the absence of stabilizing countercharges from the PCF.

Looking at the spectra C (main paper), we can see that further relaxing the geometry in comparison to the spectrum shown for CP29 Chl a spectra B (main paper) will lead to a red shift of the excitations corresponding to the optimized states. All other states, however, will respond with a blue shift compared to the QM/MM geometry. This would indicate that the preferred geometry of the ground and first excited state in the gas phase is further away from any other excited state minimum compared to when the geometry is in its native CP29 geometry.

As a consequence, activation of the PCF appears to undo most effects that the change in QM versus QM/MM geometry has (cf. earlier work on LHCI²⁷; we have added the case of Chl b in Figure S7 for comparison with the earlier publication for the interested reader). Indeed, the spectra A and C of CP29 Chl a (main paper) look much more similar to each other as to spectra B; this is astonishing as they are methodically/structurally much closer to B. Especially similar are the positions of the Q band excitations (lowest two excited states) and the Soret band oscillator strengths. The corresponding

analysis will be expanded in a future publication, including more Chl residues for comparison. For the time being, we consider this to be a very convincing test of our QM/MM interface, being able to unravel the different effects of the protein structure on the spectra of Chl a.

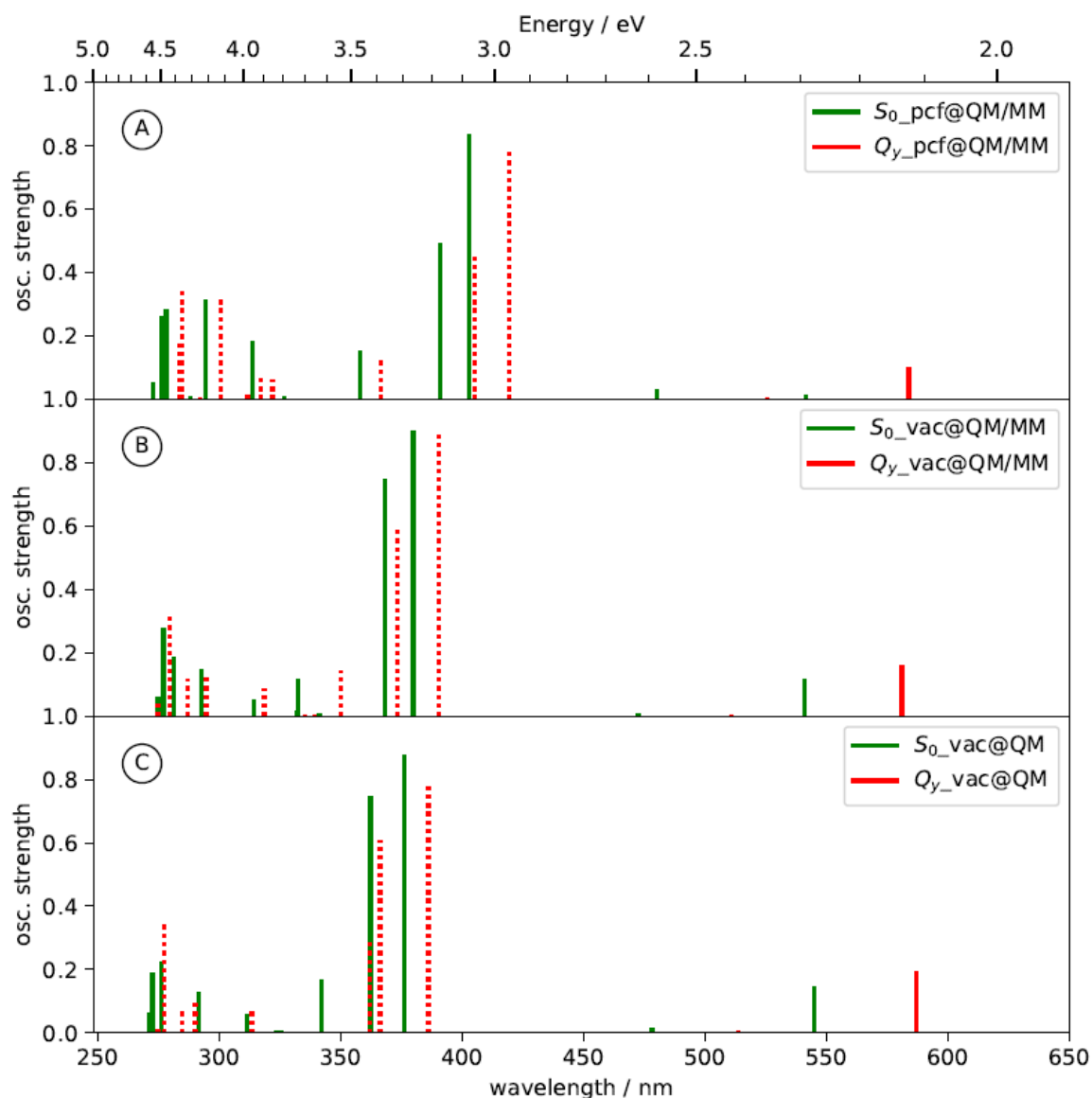


Figure S7: TD-CAM-B3LYP/6-31G* line spectra of a CP29 Chl b molecule (labeled 608 in the CP29 chain “r” of the crystal structure with pdb identifier 5XNL) molecule, water coordinated. Colors indicate different optimized geometries: green: ground state minimum; red: Q_y state minimum. Lowest 12 excited states computed. QM region consisted of the Chl b (phytyl chain omitted) and the coordinating water, see text for details. Dashed lines indicate that oscillator strength is computed for a ground state excitation, despite the geometry being optimized for a different state; such intensities are given for comparative purposes only and are experimentally not available. A: Spectra at QM/MM geometries with point charge field present; B: Spectra at same geometries as A, but without point charges; C: Spectra after optimization in the gas phase.

References

1. P. Sherwood, A. H. De Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjöqvist, A. Fahmi, A. Schfer and C. Lennartz, *J. Mol. Struct. THEOCHEM*, **2003**, 632, 1–28.

2. T. Mathes and J. P. Götze, *Front. Mol. Biosci.*, **2015**, 2.
3. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, **2004**, 393, 51–57.
4. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **1973**, 28, 213–222.
5. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, **2016**.
6. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, **2005**, 26, 1701–1718.
7. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. Mackerell, *J. Chem. Theory Comput.*, **2012**, 8, 3257–3273.
8. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. Mackerell, *Nat. Methods*, **2017**, 14, 71–73.
9. F. Guerra, S. Adam and A.-N. Bondar, *J. Mol. Graph. Model.*, **2015**, 58, 30–39.
10. K. Kuczera, J. Kuriyan and M. Karplus, *J. Mol. Biol.*, **1990**, 213, 351–373.
11. N. Foloppe, M. Ferrand, J. Breton and J. C. Smith, *Proteins Struct. Funct. Genet.*, **1995**, 22, 226–244.
12. A. Damjanović, I. Kosztin, U. Kleinekathöfer and K. Schulten, *Phys. Rev. E*, **2002**, 65, 031919.
13. N. FOLOPPE, J. BRETON and J. C. SMITH, in *PHOTOSYNTHETIC BACTERIAL REACTION CENTER II: STRUCTURE, SPECTROSCOPY, AND DYNAMICS*, ed. Breton, J and Vermeglio, A, PLENUM PRESS DIV PLENUM PUBLISHING CORP, NEW YORK, **1992**, vol. 237, pp. 43–48.
14. O. Lemke and J. P. Götze, *J. Phys. Chem. B*, **2019**, 123, 10594–10604.
15. C. Jamorski, M. E. Casida and D. R. Salahub, *J. Chem. Phys.*, **1996**, 104, 5134–5147.
16. M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, **1998**, 108, 4439–4449.
17. O. Lemke, Freie Universität Berlin, **2020**.
18. X. Su, J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang and M. Li, *Science (80-.)*, **2017**, 357, 815–820.
19. H. I. Petrache, S. Tristram-Nagle, K. Gawrisch, D. Harries, V. A. Parsegian and J. F. Nagle, *Biophys. J.*, **2004**, 86, 1574–1586.
20. A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard and R. M. Lanigan, *Proteins Struct. Funct. Bioinforma.*, **2014**, 82, 195–215.
21. M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, **2003**, 24, 129–142.
22. L. Zhang, D.-A. Silva, Y. Yan and X. Huang, *J. Comput. Chem.*, **2012**, 33, 1969–1980.

23. W. L. Jorgensen and J. D. Madura, *Mol. Phys.*, **1985**, 56, 1381–1392.
24. M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, **2000**, 112, 8910–8922.
25. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, **1983**, 79, 926–935.
26. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, **2015**, 1–2, 19–25.
27. D. Kröner and J. P. Götze, *J. Photochem. Photobiol. B Biol.*, **2012**, 109, 12–19.

4.2 Paper 2

“Spectral characterization of the main pigments in the plant photosynthetic apparatus by theory and experiment”

J. P. Götze, F. Anders, S. Petry, J. F. Witte, H. Lokstein

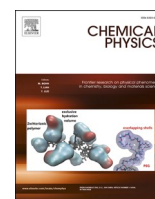
Chemical Physics **2022**, 559, 111517 ff.

DOI: 10.1016/j.chemphys.2022.111517

URL: <https://doi.org/10.1016/j.chemphys.2022.111517>

Contributions

Jan Götze conceived the project and wrote the manuscript. The calculations of the antenna embedded pigments were conducted by Simon Petry. Furthermore, the elaboration and implementation of the non-planarity parameter as well as the implementation of the excited state optimizations for QM/MM calculation in gmx2qmmm was done by Simon Petry. The QM calculations of the isolated structures were performed by Florian Anders and the DFT/MRCI calculations were performed by Felix Witte. The comparison of the embedded and isolated calculation were performed by Simon Petry. All experimental work was carried out by Heiko Lokstein. All authors contributed to the final version of the manuscript.



Spectral characterization of the main pigments in the plant photosynthetic apparatus by theory and experiment[☆]

Jan P. Götze^{a,*}, Florian Anders^a, Simon Petry^a, Jan Felix Witte^a, Heiko Lokstein^b

^a Institut für Chemie und Biochemie, Freie Universität Berlin, Arnimallee 22, D-14195 Berlin, Germany

^b Department of Chemical Physics and Optics, Charles University, Ke Karlovu 3, 121 16 Prague, Czech Republic

ARTICLE INFO

Keywords:

Photosynthesis
Light harvesting pigments
Chlorophylls
Carotenoids
UV/vis spectroscopy
Quantum chemistry

ABSTRACT

Comprehensive experimental spectroscopic and TD-CAM-B3LYP/6-31G*, DFT/MRCI libraries of the most relevant plant light harnessing pigments are presented. Included are β -carotene, zeaxanthin, violaxanthin, lutein, neoxanthin, chlorophyll *a* and chlorophyll *b*. We employ vibrational broadening with various approximations to enable direct comparison of computational to experimental data. Spectroscopic shifts based on continuum solvents and protein matrices were computed as well, employing CPCM or (for chlorophylls) QM/MM models. We also provide insight into the transient behavior after photoexcitation through optimization of the corresponding excited states.

It is found, that the investigated carotenoids only differ in their excitation energies but display analogous subsequent photophysics. Vibrational broadening of the carotenoid spectra compares well to experimental data regardless of the actual approach chosen. This appears to be a result of a strong gradient at the Franck-Condon point. Thus, it is not trivial to assess which computational method is actually suited for these purposes. Further, Q and Soret band reorganization energies are very similar in both, chlorophylls *a* and *b*. Both, TD-DFT, as well as DFT/MRCI predict the existence of an intramolecular charge transfer state in the Soret region of chlorophylls. The existence/absence of this state has previously not been experimentally corroborated. However, measurements of chlorophyll *a* absorption spectra in different solvents support this assumption. The results suggest, that the relaxation of the chlorophyll Soret bands might be affected by the presence of such an intramolecular charge transfer state.

1. Introduction

Plants, and all other photosynthetic organisms obtain their primary energy from sunlight via electronic excitation of (bacterio)chlorophylls, (B)Chls and/or carotenoids (Crts)[1,2]. High quality spectroscopic datasets for many Chls and Crts are available, either for isolated molecules or for ones located in their native environments. Their interpretation may be in some cases straightforward through analogies arising from simple, highly symmetric model systems.

While isolated Chl or Crt molecules may be addressed by employing porphyrin[3] or linear polyene analogies,[4] their overall structure is often intrinsically not symmetric. Protein-embedded chromophores are subject to further interactions with the environment which act as

additional symmetry-breaking elements. The most relevant plant pigments, Chls *a* and *b*, and the bacterial analogue BChl are already structurally symmetry-broken via chemical modifications. Crts, on the other hand, are often found to be structurally symmetric, yet many organisms have evolved Crts like peridinin[5,6] that cannot be described by a simple C_{2h} point group. This line of thought was discussed recently for Crts by Fiedor and coworkers[7]. Finally, due to the large variety and intrinsic chirality of protein environments, Chls and Crts rarely find themselves in a situation that would support symmetry[8,9]. There are only few notable exceptions of highly-symmetric arrangements like in the basal plate of chlorosomes [10].

Abbreviations: Chl, Chlorophyll; Crt, Carotenoid; DFT, Density functional theory; MRCI, Multireference Configuration Interaction; QM, Quantum mechanics; MM, Molecular mechanics; IC, Internal conversion; EET, Excitation energy transfer; FC, Franck-Condon; AH, Adiabatic Hessian; VG, Vertical Gradient; VH, Vertical Hessian; BLA, Bond length alternation.

[☆] This article is a contribution to the special issue celebrating the 60th birthday of Professor Dr. Peter Saalfrank. To Peter, who is always eager to learn new things.

* Corresponding author.

<https://doi.org/10.1016/j.chemphys.2022.111517>

Available online 28 March 2022
0301-0104/© 2022 Elsevier B.V. All rights reserved.

1.1. Current theoretical approaches

Symmetry has been a way to model systems that cannot be explicitly simulated by a more affordable, and, at the same time, reliable approach. Fortunately, recent years have provided a drastic increase in the quality of computational methods for optical spectra in photosynthesis [11–13]. Key components are the improved spectroscopic accuracies of the employed quantum mechanical (QM) approaches, but also introducing the environment via coupled QM/molecular mechanics approaches [14–16]. This article also covers another crucial problem when representing the experimentally measured spectra in a theoretical fashion: vibrational broadening [17–19].

Photosynthetic pigment spectra exhibit complex patterns of peaks and shoulders that arise also from the vibrational modes of electronic states. To be comparable to experiments, we require not only the energy differences between ground and excited states (the vertical excitation energy, E_v); we also need to include the transitions between actual molecular (combined nuclear-electronic) states. While this can computationally spiral out of control quickly and become costly even for medium-sized molecules like Chls and Crts, several approaches have been devised to overcome the scaling for hundreds of degrees of freedom and combinations and overtones of those, interacting with each other. We have provided an overview on the quality of these approaches for flavin systems several years ago [18].

Any interpretation that relies on symmetry arguments will thus always neglect the tendency of (B)Chls and Crts to be symmetry broken, possibly for specific functions. The effect can be sorted into two different classes: (i) Color change, i.e., changes in absorption energy and/or intensity and (ii) pathway modification by changing how internal conversion (IC) or excitation energy transfer (EET) may take place.

1.2. Explicit modelling of carotenoids and chlorophylls

With both concepts in mind, for controlling the spectra and/or the excitation energy pathways, we can analyze different photosynthetic chromophores and ask: Why is a certain chromophore at its specific location in a protein scaffold? Do they only differ spectrally, or is also their molecular structure capable of more functions? Experimentally, there are even indications alluding to that removing all native carotenoid species from an genetically engineered plant system and replacing them by a foreign Crt (astaxanthin) still may lead to a possibly functional phenotype [20] *In vitro* reconstitution of light harvesting complexes (LHCs) results in constructs which can be somewhat stable [21].

However, most LHCs reconstituted with non-native pigment complements, including Crts, or LHCs isolated from Chl or Crt deficient mutants, lose stability and their functionality (at least partially) [22–24].

Combining these insights can lead to new computational approaches, namely replacement studies, which should provide the functions in question regardless of the Crt in the model. Previous attempts to do so in the context of photosynthetic pigments [25] might have simply been unsuccessful due to presuming possibly incorrect mechanism(s) [26].

As a first step in the direction outlined above, in this study we provide complete libraries for the most relevant photosynthetic pigments of plants. We compare the results of state-of-the-art theoretical approaches to assess the different properties of each pigment within the same model methodology. Furthermore, we address the direct comparability to experimental data and the basic effect (i.e., energy shift) of a protein scaffold. Introducing the protein scaffold allows us to investigate the effect of environmental symmetry breaking. Note, that in this article, we restrict this to Chls since the effect of symmetry breaking environmental effects has been covered extensively in our previous articles [26–28].

2. Methods

2.1. Isolation and measurement of the absorption spectra of the pigments

Pure pigments were isolated as described before [23] with slight modifications: A Waters chromatography system as in a previous study [22] with a 1 ml injection loop was used. Individual pigments were collected, dried in dim light under a stream of nitrogen, solved in the respective solvents, and used immediately for absorption measurements.

2.2. Gas phase structures

The spectra of Chl *a* and Chl *b*, as well as those of the Crts β -carotene (Bcr), zeaxanthin (Zea), violaxanthin (Vio), lutein (Lut) and neoxanthin (Neo), were investigated employing the same methodology for all cases. To obtain geometries of the isolated molecules, initial Crt coordinates were taken from the CP29 complex (Lut, Vio, Neo; PDB code 3PL9 [8]; also Zea by deletion of the Vio epoxy groups) or from the PSII core complex (Bcr; PDB code 6KAC [29]). Chl coordinates were obtained for Chl *a* from the ChEBI [30] and Chl *b* via modification of Chl *a* using the molden [31] software; the latter was done to obtain minimal initial structural difference between the two compounds. Phytol chains of Chls were omitted. Missing hydrogen atoms were added using the program Avogadro [32]. The structures were then optimized using the CAM-B3LYP method [33] with a 6-31G* basis set [34]. We further performed analogous calculations (including following steps) for selected systems employing CAM-B3LYP with various dispersion correction approaches, [35,36] and also the ω B97XD [37] functional; those calculations did not provide qualitatively different results (see supporting information, SI). The optimized structures were checked to be real minima via normal mode analysis, which is also required to perform the vibrational broadening. All corresponding calculations were performed using Gaussian16 [38].

2.3. Linear response spectra and excited state optimizations

Successfully optimized structures were the basis for a time-dependent density functional theory (TD-DFT) single point calculation, [39,40] with 8 states of interest for the Chls and 5 for the Crts. For the Q and Soret (or “B”) states of Chl, the excited state minima were computed; in some cases, this was found to be not reliable due to crossings with energetically close charge transfer (CT) states, which might be methodological artifacts. However, it is known that B band IC is much faster than Q band relaxation (i.e., fluorescence), [41,42] in accordance with Kasha’s rule [43]. For Crts, TD-DFT does not provide the correct first excited state energy but is performing well for the bright S_2 state. Hence, we optimized the corresponding excited state geometries and can, thus, only make indirect statements about the first excited state.

To (partially) amend this, we employed multireference DFT (DFT/MRCI) to the stationary points of our study for all systems and all states, which yields proper energies for the first and second excited states. Note, that DFT/MRCI apparently improves the Chls’ state energies as well, as we will see below. DFT/MRCI excited states were computed at the B3LYP/MRCI/def2-SVP level [44–46] using the stand-alone DFT/MRCI program [47] interfaced with ORCA (4.2.0) [48] employing the R2016 Hamiltonian. A reference space of 16 electrons and 14 orbitals with a default cut-off of 1.0 Hartree was chosen. All single and double excitations were incorporated. The reference space was refined once, i.e., two DFT/MRCI calculations were performed in succession. Initial testing found that a second refinement of the reference space did not yield any significant change in excitation energies, oscillator strengths, or transition dipole moments.

2.4. Vibrational broadening

As we have discussed for the case of flavins before, all approaches towards vibrational broadening have their advantages and shortcomings. For some Crts, this was shown as well. However, the current implementation in Gaussian is versatile with respect to the approach taken, allowing us to compare the approaches of vertical gradient (VG), vertical Hessian (VH) and adiabatic Hessian (AH) using the corresponding default values [17,19,49]. VH often requires remapping of the ground and excited state Hessians since the excited state Hessian may contain negative force constants, which is un-physical. A strategy available in Gaussian16 is to remove a set of related, low-frequency modes. The exact number of removed modes can be found for each case in the SI. Furthermore, AH requires a successful excited state optimization for each excited state in the spectrum. In addition to the VG/VH/AH variants, we employ time-independent (TI) and time-dependent (TD) Franck-Condon approaches; test calculations show negligible Herzberg-Teller influence (data not shown) as was already observed elsewhere [50]. TD-broadened spectra were computed at a temperature of 0 K.

It turned out, that only the TD-VG and TD-VH combinations yield reliable spectra throughout all presented cases. Thus, we concentrate in the main article on those. A large part of our discussion will focus on why, in the case of Crts, VH and VG may actually produce incorrect results. Results obtained with the not included AH calculations will be used for comparative purposes where appropriate.

2.5. Solvent effects

For improving the comparison to the experimental data, a continuum model was employed (CPCM[51]). We have chosen acetone as solvent for this study since the corresponding experimental spectra were available from our own work.

Of the computational steps listed in the previous subsections, we only repeated the ground state optimization and the corresponding normal mode test, as well as the TD-DFT single point at the ground state minimum. Though technically possible, it is physically meaningless to generate a corresponding TD-DFT gradient for a VG/VH approach. The same issues arise for AH when trying to obtain the optimized excited state structure in the presence of a solvent: The corresponding spectra are red shifted due to the overestimation of the excited state gradient. Requesting gradients in the solvent framework also includes an overestimated adaptation of the solvent, resulting in too low excitation energies. *In vitro*, the solvent is not fast enough to fully adapt during the time scale of excitation, but this is, to our best knowledge, not covered in any QM package so far.

Thus, we opted to rely only on the shift predicted by a vertical TD-DFT calculation, which in the Gaussian software includes the changes in the solvent fast component field. As a result, the spectra displayed below are, for the solvent effect, simply shifted by a state-specific energy value which arises from the difference between the vertical excitation energy with or without the solvent treatment.

2.6. Protein-induced shift

Like in the solvent case, not all computations are physically meaningful to repeat in a QM/MM protein environment. Consequently, we provide data from ground state and excited state QM/MM structures using the QM/MM interface gmx2qmmm with the same setup which has been used in another study [16,52]. For each of the relaxed structures (true minima cannot be verified due to lack of normal mode analysis in the QM/MM framework), we provide TD-DFT spectra as well.

2.7. Structural analysis criteria

For Crts, the spectral properties are dominated by the relative car-

bon-carbon distances along the chain, or bond length alternation (BLA). BLA for each non-terminal carbon atom n in the Crt conjugated chain is defined as

$$BLA_n = (-1)^{|n|} \left(\vec{r}_{n,+} - \vec{r}_{n,-} \right)$$

with n starting at 0 at the center of the conjugated chain and running twice, once for each side of the Crt chain (denoted by different signs in the figures throughout the article). $\vec{r}_{n,+}$ and $\vec{r}_{n,-}$ indicate the following and preceding conjugated bond lengths for each center n . BLA provides insight into the spectroscopic properties of Crt state structures which are virtually identical yet have distinctively different excitation energies. Note, that we do not agree with recent approaches to label BLA as a systemwide property; BLA is changing along the chain[13,26] and is a property that is drastically reduced in usefulness when summed up to a single value.

During the course of this study, we found that Chls display a behavior similar to (some) Crts: the Chls' ground and excited state minima do not differ to an extent which can be perceived by the human eye. Thus, we seek to introduce new parameters, one of which is the non-planarity (NPL) of the system. In contrast to BLA, NPL is in its current form a holistic criterion; as an aromatic system such as Chl does not have "edges" or termini like Crts, we consider this to be justified for the time being. NPL is defined as

$$NPL = \sum_{i=1}^n |\Delta z_i|$$

with n being all conjugated atoms of the aromatic system (nitrogen, carbon, oxygen; also, the Mg ion), and Δz being the minimum distance to the plane defined by the normal vector resulting from the closest distance between the Chl Na/Nc and Nb/Nd vectors; with the plane intersection located at the center of the closest distance between those vectors. An explanatory illustration is provided in the SI. Another parameter which was tested in this regard is the change in the edge size of the Na-Nb-Nc-Nd squares, $r(4N)$.

3. Results and discussion

3.1. Geometries

The ground state minimum structures of the investigated compounds can be found in Fig. 1. A comparison between the ground state structures and the excited state structures can be found in the SI; it has been omitted here since the structures are for all presented cases virtually identical, at least on the presented scale, regardless of electronic state. For the Crts, this was discussed before.[28] Neo is an exception in this regard; the more intricate structure of Neo results in changes similar to what was found for peridinin earlier.[5] Since Neo and peridinin have similar features, such as the presence of the sp-hybridized carbon atom, this is not entirely surprising.

A BLA analysis of the individual Crt/state structures can be found in Fig. 2. BLA is found to be equivalent for all Crt cases, with some slight asymmetry between the "+" and "-" sides arising from our initial structure for Bcr, Lut and Neo. Overall, however, it is found that formation of $1B_u$ results in a much more diverse BLA as compared to the $1A_g$ state. Interestingly, BLA changes are much more pronounced at the center of the chain. This supports our statement above that BLA is not a system-wide criterion and should be handled on a bond-by-bond basis. Ground state BLA does not depend on the presence of a solvent. For the Neo case, it is especially interesting to see that the BLA is apparently simply "cut off" at the "+" side (towards the sp-hybridized carbon). This might indicate that BLA can be retained along the whole chain even for shorter conjugated chains, i.e., higher excitation energies.

The structures of Chl *a* and *b* presented in Fig. 1 are very similar, regardless of being in the gas phase or in a solvent, and for the

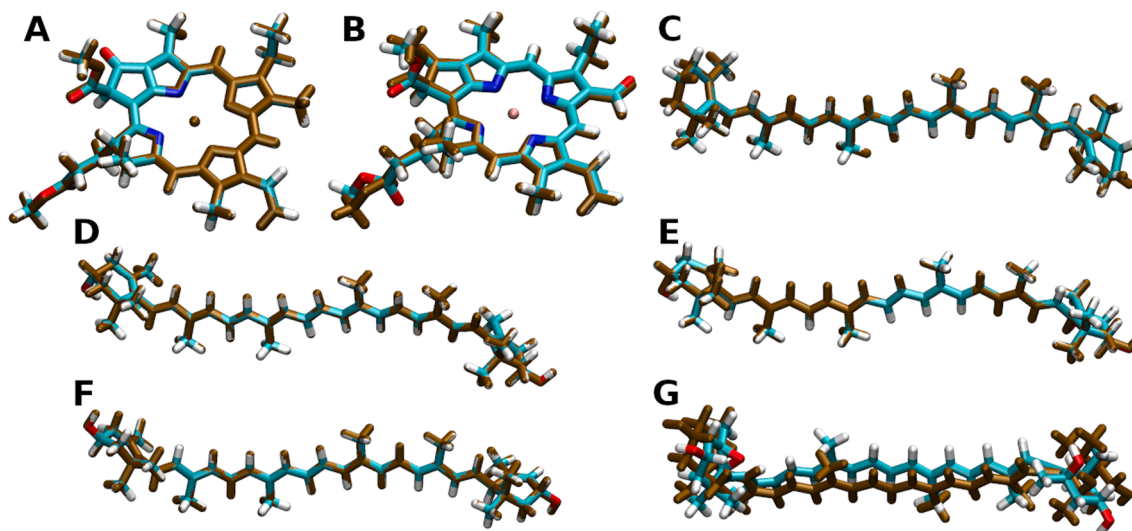


Fig. 1. CAM-B3LYP/6-31G* ground state structures of (A) Chl a, (B) Chl b, (C) Bcr, (D) Zea, (E) Vio, (F) Lut and (G) Neo. Gas phase (regular colors) and acetone (brown color) structures shown.

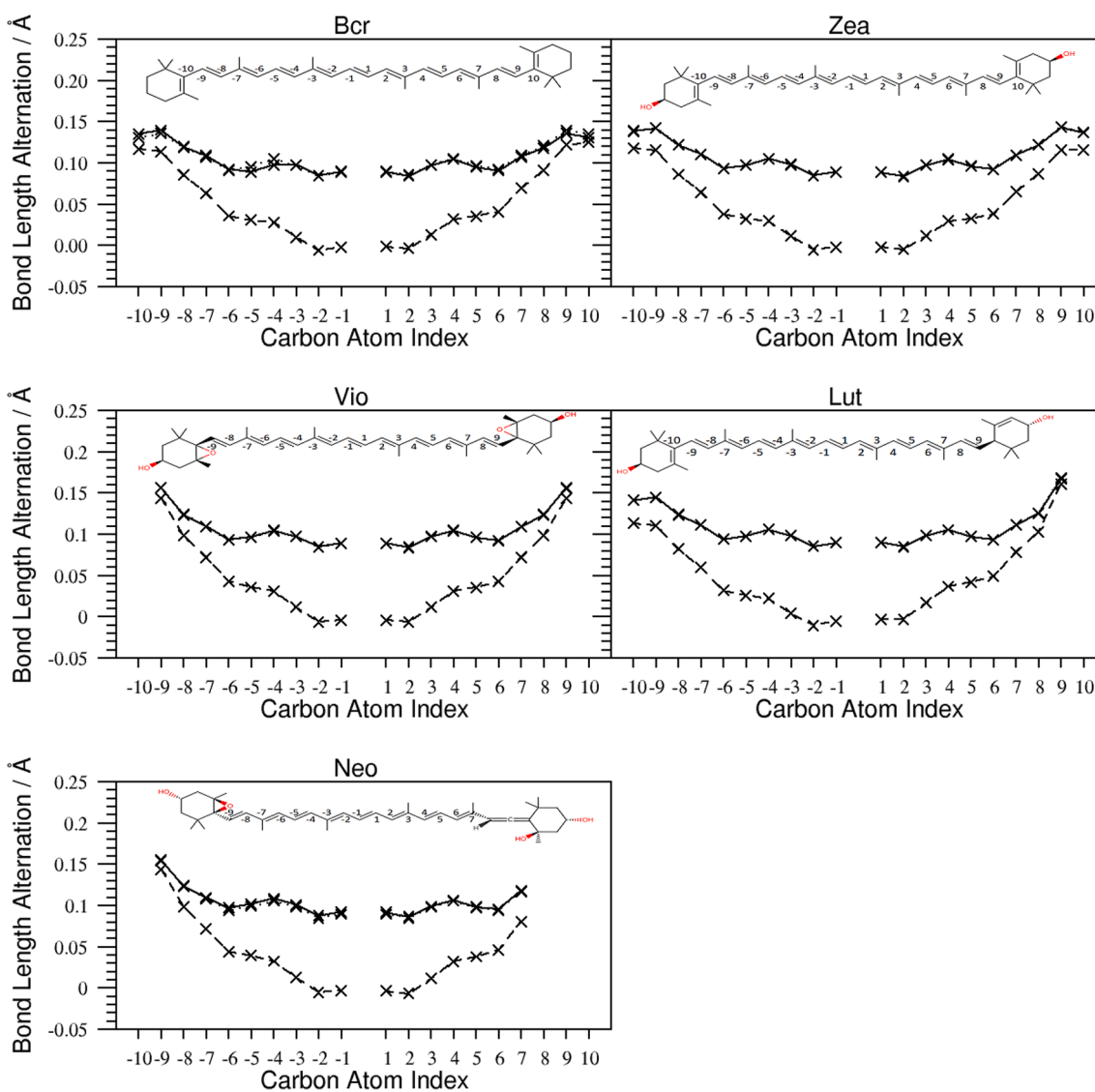


Fig. 2. Bond length alternation (BLA) for the Crt (TD-)DFT minima; ground state in gas phase (solid line) and solvent (dotted line), the $1B_u$ state in the gas phase (dashed line).

comparison between ground and excited state structures (see SI). Chls however contain more electronic states of interest in the UV/Vis region than Crts. To characterize the individual Chl *a/b* states in a similar manner as BLA does for Crts, we provide the NPL criterion and the distance between the Chl core nitrogen atoms as defined earlier (Table 1). It can be seen that NPL indeed shows an increasing trend with higher states, although the change is not continuous. Interestingly, there is also a trend for some states to become even more planar than the ground state, most prominently Chl *b* in the gas phase. This finding somewhat reduces the usefulness of the NPL criterion, though it apparently works well when Chl *b* is embedded in a protein (see below). In the same manner, taking the size of $r(4N)$, higher states mostly show a larger overall inter-nitrogen distance than the ground state, but, again, not in a continuous fashion between excited states. For the protein embedded Chls (see also below), the distance increases with higher states. The B_y states of Chl *b* shows a slightly smaller $r(4N)$ compared to the B_x state, while the B_x states of Chl *a* shows slightly smaller distance compared to the Q_x states. Overall, ΔNPL appears much more consistent than the change in $r(4N)$.

For the Chls, we have spectra of QM/MM excited state optimized structures available as well, as shown in the SI. Just as for the gas phase and CPCM computed structures, optimized excited state Chl QM/MM structures are virtually indistinguishable with respect to the actual electronic state and are thus omitted here. We find however as discussed above that NPL provides structural parameters which increase between Q and B states for all tested Chl *a/b* QM/MM structures. It was also found, that the NPL might span several orders of magnitude, as the case of 606*b* can range from no change (Q_y) to changes in the pm range (all other states). The observed differences between gas phase vs. QM/MM changes may arise from the more elaborate Chl coordination, since it is known that simple charge-nullifying environments will not lead to strong changes in excited state properties, as was shown for the Q and B bands previously [53,54].

3.2. Chlorophyll spectra

The computed Chl spectra can be found in Fig. 3. VG and VH generally exhibit similar results, indicating that the Hessians of the ground and excited states do not vary too much. This assessment agrees with earlier VG computations, especially for larger molecules [55]. Physically, the larger the number of normal modes, the less impactful should the changes due to a single-electron excitation be, so this effect is easily understood.

CAM-B3LYP apparently overestimates the extent of the “green gap”, i.e., the energy difference between the Q and Soret band maxima. DFT/MRCI (see Fig. 4) performs much better in this respect, shifting the gas phase Chl *a* (Chl *b*) DFT vertical energy gap between Q_y and B_x of 1.28 (1.06) eV to only 1.06 (0.82) eV. The experimental green gap is 1.01 (0.80) eV in acetone, almost a perfect match with the DFT/MRCI data; although 1.14 (0.91) eV have been reported for the gas phase [56]. We, however, note that all methods perform within their expected error ranges, DFT/MRCI outperforming DFT with less than 0.1 eV difference to the experiments.

Table 1

Changes (in pm) for Chl geometry parameters of excited state optimized structures compared to the respective ground state minimum structures. Chl *a*, *b* in the gas phase; other systems (604*a*, 606*b* and 607*b*) are arbitrarily chosen Chl *a*, *b*, *b* sites from a QM/MM optimization within the CP29 protein scaffold. $r(4N)$ corresponds to the total distance between the core Chl nitrogen atoms, along the edges of a Na, Nb, Nc, Nd square, with $\Delta r(4N) = r(4N, \text{excited state}) - r(4N, \text{GS})$.

	Q_y		Q_x		B_x		B_y		CT	
	ΔNPL	$\Delta r(4N)$	ΔNPL	$\Delta r(4N)$	ΔNPL	$\Delta r(4N)$	ΔNPL	$\Delta r(4N)$	ΔNPL	$\Delta r(4N)$
Chl <i>a</i>	-0.04	-0.01	0.09	0.04	0.19	0.02	n/a	n/a	n/a	n/a
Chl <i>b</i>	-0.03	-0.01	-0.23	0.04	-0.07	0.03	n/a	n/a	n/a	n/a
604 <i>a</i>	-0.14	-0.00	0.09	0.06	-0.07	0.00	0.21	0.08	0.54	0.08
606 <i>b</i>	0.00	0.00	-1.19	0.03	2.37	0.07	2.11	0.05	2.02	0.07
607 <i>b</i>	0.29	0.00	0.47	0.05	0.61	0.06	0.50	0.05	0.47	0.05

The green gap with VH vibrational broadening is even larger than vertical TD-DFT, 1.34 (1.15) eV, indicating that post-DFT treatment is necessary for obtaining realistic, relative Chl state energies. The Q-to-B state energy difference is also overestimated by other work employing a DFT approach: Earlier work reports the TD-CAM-B3LYP spectra for a different basis set [57]. The gas phase vertical Q_y/B_x gap for Chl *a* (Chl *b*) was calculated to be 1.46 (1.22) eV, about 0.2 eV larger than our values. Apparently, a larger basis set is beneficial for a better prediction of the green gap, however, we have shown a decade ago that there are limits to this improvement [27].

The solvents were found to affect the states differently, but always induced a shift to lower energies, which can be seen when comparing the regular to the mirrored spectra in Fig. 3. Note, that CAM-B3LYP as well as DFT/MRCI predict the presence of an ICT state in the Soret region. It is unclear if this ICT state is an artifact or not, [11] yet the Soret states are much more sensitive to the solvent, which is however not the focus of this study. Currently, we assume that this might be due to the mixing with the nearby ICT state. Hence, Q band shifts are smaller as compared to the shifts of the Soret states.

The vibrational progression of the Chl spectra is of the same high quality as was already shown elsewhere [11]. CAM-B3LYP thus again demonstrates its ability to predict the vibrational patterns of chromophores well, both in ground and excited states.

In the protein environment, we find that the QM/MM geometry and the point charge field representing the protein environment only can drastically affect the Chl *a* and *b* spectra (see SI). Most prominently, we find a general red shift going from the protein environment to the gas phase for all states. An exception are the Q bands of the Chls *b*, which are blue shifted. This effect is most significant for the optimized ground state structures of the Chls *b*. The corresponding results are investigated in more detail elsewhere [52].

3.3. Spectral properties of Chl excited states

We have indicated above that the structures of Chls are not severely affected by the actual electronic states. The Jablonski diagrams in Fig. 4 show that the Chl excited state properties are similar throughout all investigated geometries as well. Note, that there is also no immediate indication for a state crossing in all presented cases, with the exception of the states in the tightly populated Soret region. In the Soret region, the B_y state was often crossing with an ICT state, preventing an analogous analysis for B_y as for the other states. Hence, B_y optimization always led to a complex B_y/ICT state crossing region with no clear exit to a local B_y state minimum.

Comparing TD-DFT and DFT/MRCI, it can be seen that the Q state energies agree much better than the Soret state energies. However, qualitatively, all states that are present in the results obtained with one method are also found with the other. TD-DFT states are generally higher in energy, a phenomenon which we assign to DFT/MRCI providing a non-negligible amount of multi-electron character to each state. This character is higher for the excited states than for the ground state (ground state: about 10%, excited states: > 15%; detailed data omitted here), hence the energy difference drops in comparison to a

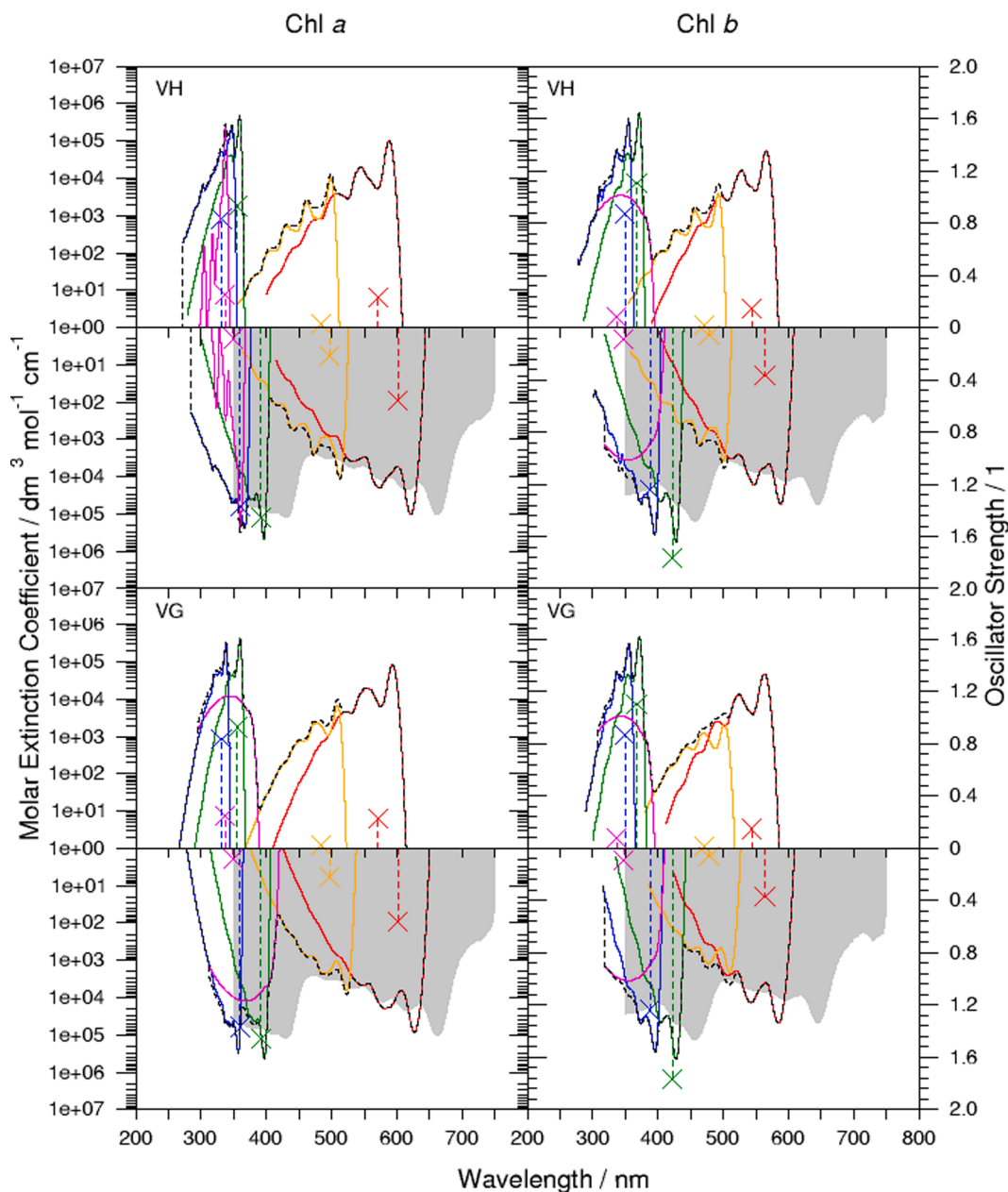


Fig. 3. Computed Chl a (left) and Chl b (right) VH (top) or VG (bottom) spectra. Gas phase (regular) and acetone (mirrored) spectra are shown. TD-DFT oscillator strengths and vertical excitation energies are given for each state as dashed bars. Experiments are shown as grey backgrounds, scaled to arbitrary units to match the computed Q_y intensity. Sum spectra are shown as black, dashed lines. Individual states are colored, red: Q_y , orange: Q_x , green: B_x , blue: B_y , violet: CT state.

single-reference method like TD-DFT. Oscillator strengths, however, are well preserved between TD-DFT and DFT/MRCI.

Before entering the discussion of the Crt results, it must be highlighted that Chl states apparently display almost negligible reorganization energies (in the range of 0.2–0.3 eV, see Fig. 4). This agrees well with the fact that experimental Chl spectra exhibit a dominant peak at the lowest energy onset of the spectrum, with only relatively small vibrational shoulders/side peaks. It also means that the main contribution to the Chl spectra for all states considered here is the 0–0 transition, and that, hence, the vertical excitation energy should not be too different from the 0–0 peak maximum.

3.4. Carotenoid spectra

The computed Crt spectra are shown in Fig. 5. The results for Bcr, Zea and Vio agree with previously published data.[13,58] All presented Crt

spectra exhibit similar vibrational progressions. This is especially interesting in the case of Neo, since it is significantly different in molecular structure. Yet, these differences apparently only result in a blue shift as compared to the other Crts studied here. Moreover, it is important to note, that Bcr and Zea have identical conjugated systems, yet differ slightly in excitation energies, with Zea displaying the higher values. This is experimentally known, and we can assign this to the smaller Zea “box”, when employing a particle-in-a-box model, due to the electron-rich terminal OH-group oxygen atoms.

Similar to Chls, we find for Crts that TD-DFT excitation energies are higher than those of DFT/MRCI. When comparing these observations to earlier work, which used an older version of the DFT/MRCI Hamiltonian,[13] we find that the $2A_g$ energy predicted here is closer to the $1B_u$ transition. For instance, for Vio, the newer DFT/MRCI Hamiltonian used here shifts the $2A_g$ excitation energy up by 0.28 eV. We will see later in this article, if this has any consequence for the Crt photophysics,

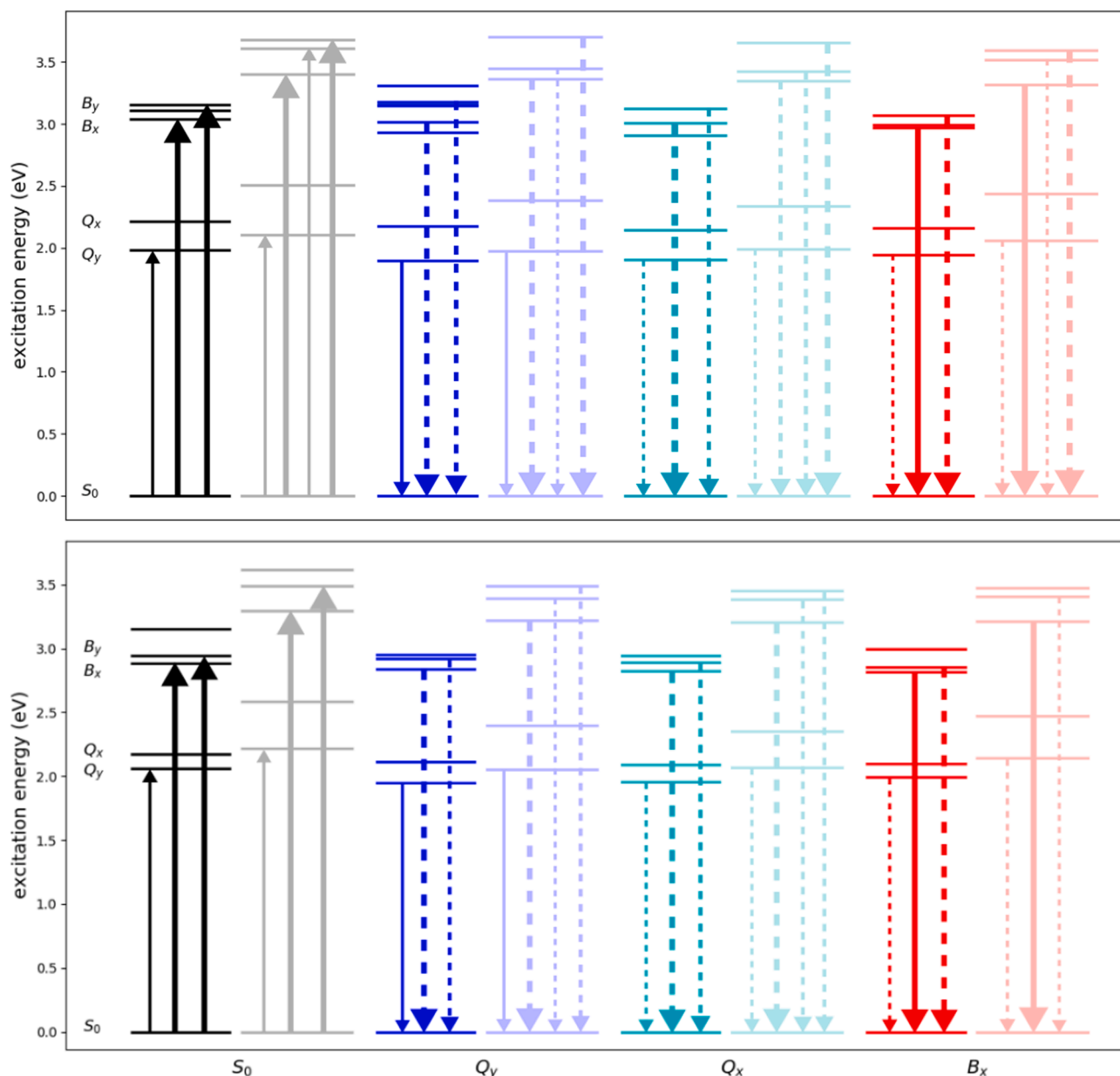


Fig. 4. Jablonski diagrams of Chl a (top) and Chl b (bottom) after nuclear (TD-)DFT relaxation of the indicated electronic state. DFT/MRCI single points (opaque, left) and TD-DFT single points (transparent, right) shown. Oscillator strengths for transitions from the optimized state are indicated as solid, from other states as dashed arrows. Thickness of arrows qualitatively corresponds to relative transition intensities.

since this would put the $1B_u$ and $2A_g$ states much closer together. Energy differences between the states in ground state geometry range from 0.08 eV for Vio to only 0.01 eV for Neo.

In the next section we will discuss in more detail, that the 0–0 energies of each spectrum presented here can vary strongly. For the spectra, this has an interesting consequence with regard to the computational robustness. Taking Lut as an example, VH 0–0 is 2.29 eV, VG 0–0 is 2.67 eV. Optimization of the state shows that the “true”, AH 0–0 energy is 2.49 eV. Still, despite both, VH and VG providing strongly differing 0–0 energies and both being off by about 0.2 eV, the spectra of Lut look reasonable. They do not fall out of line with the other Crt spectra, even though most VG/VH Crt spectra presented here have 0–0 energies close to the true AH values (full details of AH calculations omitted). For Lut, the 0–0 excitation does not even occur in the actual vibrational envelope, as 2.29 eV (541 nm) is far outside the spectra shown. We will provide an explanation for the robustness of the Lut spectra despite their wrong 0–0 energies further down in the article.

3.5. Properties of carotenoid excited states

Crt Jablonski diagrams are provided in Fig. 6. Many of the following arguments regarding their analysis have been presented already elsewhere, [13,19,28,58] so we will only list the most important factors.

In contrast to Chls, Crts exhibit a strong drop in excitation energy along the excited state gradient (0.64 to 0.72 eV for TD-DFT; 0.44 to 0.54 eV for DFT/MRCI), which is a combined effect of mostly reorganization energy and ground state destabilization. [13,26] In the present work, we find that the gradient at the FC point depends on the Crt in question; thus, the harmonic approximation at the FC point (i.e., the basis of VH/VG) is of different quality. E.g., for Vio, the true excited state minimum structure yields only a 0–0 energy difference to VH of 0.03 eV, meaning that the harmonic approximation holds very well. Zea, in contrast, yields a difference of 0.23 eV between 0 and 0 of VH and AH; the harmonic approximation at the FC point for Zea is not as valid as for Vio. This is also reflected in the inability to obtain an AH spectrum for

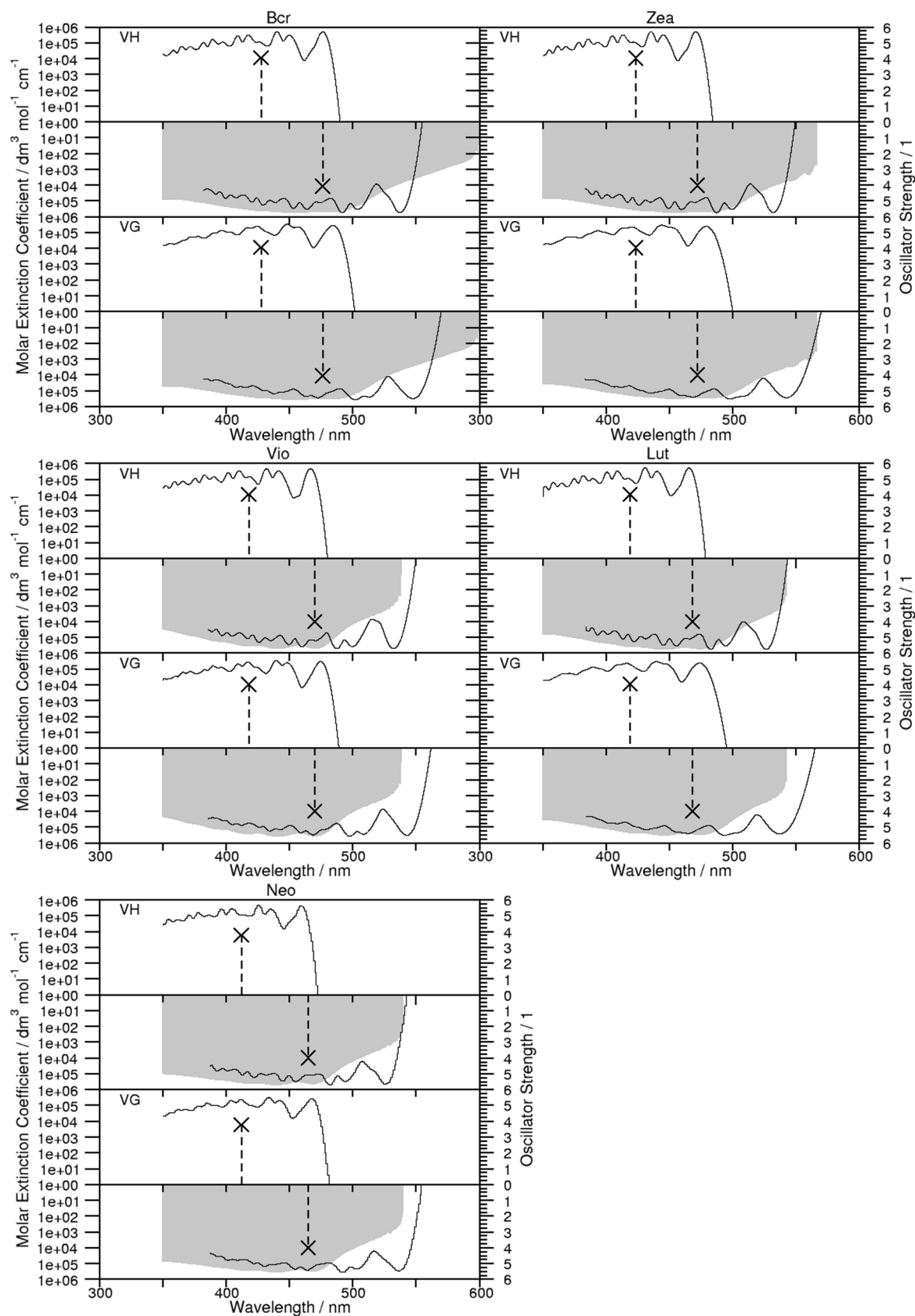


Fig. 5. Computed Crt VH (top) or VG (bottom) spectra ($1A_g \rightarrow 1B_u$ transition). Gas phase (regular) and acetone (mirrored) spectra are shown. TD-DFT oscillator strengths and vertical excitation energies are given as dashed bars. Experiments are shown as grey backgrounds, scaled to arbitrary units to match the computed intensity of the computed maximum (third envelope peak).

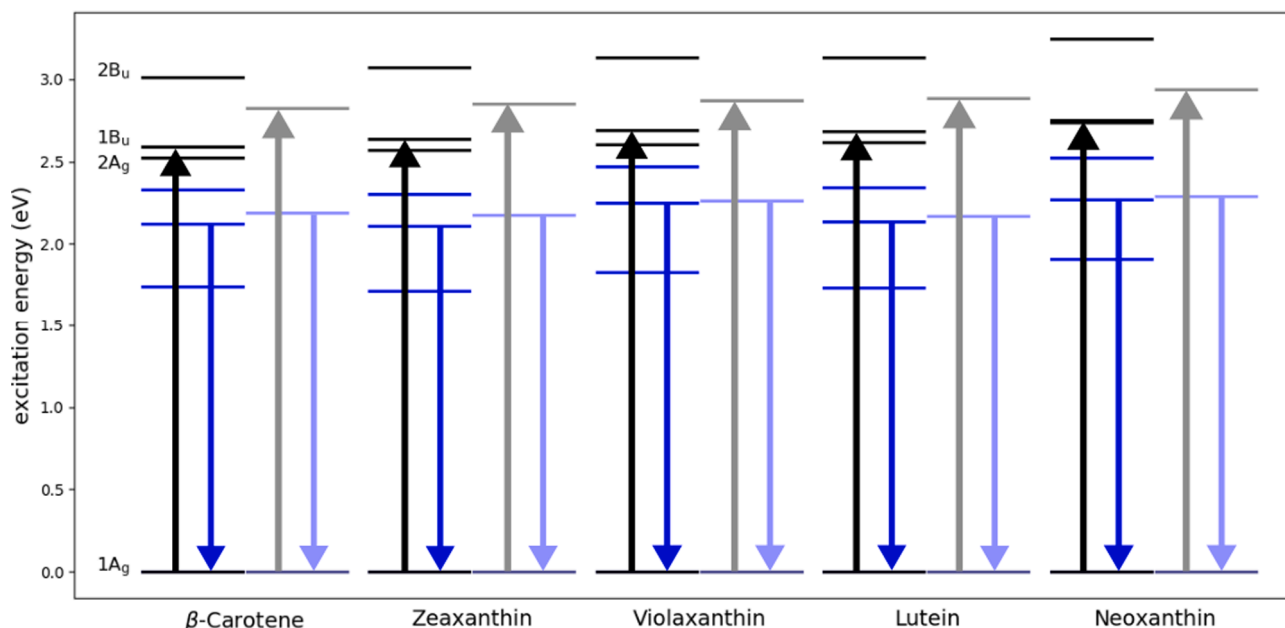


Fig. 6. Jablonski diagrams of all studied Crts after nuclear (TD-)DFT relaxation of the ground state (black lines) or the brightly absorbing $1B_u$ state (blue lines). DFT/MRCI single points (opaque, left) and TD-DFT single points (transparent, right) are shown. TD-DFT only for the $1B_u$ state due to method limitations. Oscillator strengths for transitions from the optimized state are given as solid arrows.

Zea with significant progression. The molecular origin of this failure is the rotation of the terminal rings along the excited state gradient. [13,58].

As stated in the last section, the VG/VH Crt spectra are apparently computed well, despite cases such as Lut, with clearly wrong VG/VH 0–0 energies. The reason for the robustness of the VG/VH approaches for Crts is that the 0–0 transition is of no concern for the spectral envelope in the harmonic approximation. A seemingly good spectrum will be obtained as long as there are enough other states in the FC region that provide a vibrational progression with similar quality. The spectrum is of course physically incorrect, as it does not contain the proper n - m transitions (see elsewhere [59] for experimental 0–0 transitions of Crts). The computed spectrum only appears to be fine since it is in the correct region and has a correct shape.

For Crts, the spectral envelope is mostly defined by the overlap between vibrations dominated by conjugated C–C bonds. [13,17,58] In harmonic approximation, these vibronic states have the same energy difference, regardless of the actual quantum number, and thus overtones are simply replaced by other overtones of the same character when shifting the 0–0 energy. As long as the gradient at the FC point is correct, and the Hessian matrix of the excited state is a reasonable description of the dominant modes, the computed spectrum will match the experiment in appearance. This insight basically applies to all cases for which the reorganization energy is so large that the adiabatic excited state minimum is pushed to the edge/out of the FC region.

Coming back to the Crt state properties, we find that the $2A_g$ state does not exhibit a crossing along the $1B_u$ gradient (in contrast to other Crts, like peridinin [5,60]). This indicates that the molecular coordinate stabilizing the $1B_u$ state is also energetically beneficial for $2A_g$, actually even more so, as the energy gap between $1B_u$ and $2A_g$ increases. As such, the newer DFT/MRCI Hamiltonian gives more theoretical support for the fast $1B_u \rightarrow 2A_g$ IC process in Crts than the older approach, as $1B_u$ and $2A_g$ are found to be much closer in energy, up to degenerate in the case of Neo. State characters however are nearly unaffected by the similarity of the state energies. In the most extreme case, Neo, the bright state has 86% single excitation character regardless of the Neo geometry. $2A_g$ does also not change much in composition, only increasing in multi-electron excitation character from 57% at the ground state to 63% at

the excited state minimum structure. Apparently no state-mixing is present as was found for peridinin. [5,60].

4. Summary and conclusion

We have computed, measured and discussed the spectroscopic properties of all Chls and Crts relevant to plant photosynthesis. The employed computational method, CAM-B3LYP, turns out to be less suitable for the actual electronic energies than anticipated. While being excellent at reproducing the vibrational properties of the investigated electronic states, the actual representation of Chl electronic state energy differences is unrealistic. For Crts, we find a better agreement with the experiments (as has already been shown earlier [13]). It appears that the approach to employ a combination of TD-DFT and a higher-level method, such as DFT/MRCI, remains the better way to go. This was also shown for Chl lately, [11] though not in combination with DFT/MRCI.

Another technical note: it was found that the VG/VH approaches may provide spectra that are only superficially correct. As long as the reorganization energy of a compound is large and not drastically underestimated due to a small gradient, the spectrum will likely appear to be fine. Great care thus must be taken by comparing the results of the predicted VG/VH 0–0 energies and the 0–0 energies of either experiment or a computed adiabatic minimum. The need to employ an AH check when doing VG/VH spectra drastically limits the usefulness of these approaches, as a main selling point (no need to optimize an excited state minimum) is not retained anymore.

In terms of Crt photophysics, there is no computational indication that these compounds should exhibit any difference in their relaxation/IC patterns. We observe slightly changed state energies between all Crts, and possibly interesting ways to retain BLA parameters with shorter chain length in the case of Neo. Thus, it remains the most likely theoretical assessment that Crts fulfill basically identical roles regardless of Crt type, with only slightly adjusted state energies. Our earlier proposal that the Zea to Vio exchange in LHCII upon light stress aims to reduce a potential $1B_u \rightarrow Q$ transfer is supported by our results above [26]. Yet, the role of $2A_g$ remains not tractable by our computational approach. For consolidating experiment and theory, it would be crucial to assess

whether the $2A_g$ state is actually formed *in vivo* before excitation energy transfer to Chl(s). From a computational perspective, it appears much more likely that the $1B_u$ state, being a bright state, interacts readily with Q_x (or even Q_y). The only reason why this has so far not been considered is that the energy of $1B_u$ was located far above the Q band region, but we hope that this contribution finally shows that (a relaxed) $1B_u$ state might be much more flexible than that. A potential $1B_u/Q_x$ interaction reminiscent of the pathways found for PCP[5,6,60] might be much more universal than currently assumed.

For Chls, we find that all our approaches support the existence of an ICT state near the Soret states. The existence of this state must be corroborated and analyzed in more detail in further studies. Such a state might drastically improve B to Q (likely Q_x) IC.

CRedit authorship contribution statement

Jan P. Götzte: Conceptualization, Funding acquisition, Writing – original draft, Supervision, Validation. **Florian Anders:** Methodology, Visualization, Investigation. **Simon Petry:** Methodology, Visualization, Investigation. **Jan Felix Witte:** Methodology, Software, Visualization, Investigation. **Heiko Lokstein:** Conceptualization, Writing – review & editing, Validation, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was funded by the Deutsche Forschungsgemeinschaft, project number 393271229. H.L. acknowledges financial support by the Charles University Center of Nano- and Bio-Photonics (UNCE/SCI/010) and the Czech Science Foundation, GAČR (22-17333S). The computing facilities of Freie Universität Berlin (ZEDAT) are acknowledged for providing computational resources [61]. J.F.W. thanks the research grant PA 1360/16-1 of the Deutsche Forschungsgemeinschaft for financial support.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemphys.2022.111517>.

References

- H. Lokstein, G. Renger, J. Götzte, Photosynthetic Light-Harvesting (Antenna) Complexes—Structures and Functions, *Molecules* 26 (11) (2021) 3378.
- R. Croce, H. van Amerongen, Light Harvesting in Oxygenic Photosynthesis: Structural Biology Meets Spectroscopy, *Science* (80-) 369 (6506) (2020) eaay2058.
- M. Gouterman, Spectra of Porphyrins, *J. Mol. Spectrosc.* 6 (1961) 138–163.
- J.H. Starcke, M. Wormit, A. Dreuw, Nature of the Lowest Excited States of Neutral Polyenyl Radicals and Polyene Radical Cations, *J. Chem. Phys.* 131 (14) (2009) 144311, <https://doi.org/10.1063/1.3246350>.
- J.P. Götzte, B. Karasulu, M. Patil, W. Thiel, Vibrational Relaxation as the Driving Force for Wavelength Conversion in the Peridinin-Chlorophyll a-Protein, *Biochim. Biophys. Acta - Bioenerg.* 1847 (12) (2015) 1509–1517.
- M. Krikunova, H. Lokstein, D. Leupold, R.G. Hiller, B. Voigt, Pigment-Pigment Interactions in PCP of *Amphidinium carterae* Investigated by Nonlinear Polarization Spectroscopy in the Frequency Domain, *Biophys. J.* 90 (1) (2006) 261–271.
- L. Fiedor, A. Dudkowiak, M. Pilch, The Origin of the Dark S 1 State in Carotenoids: A Comprehensive Model, *J. R. Soc. Interface* 16 (158) (2019) 20190191, <https://doi.org/10.1098/rsif.2019.0191>.
- X. Pan, M. Li, T. Wan, L. Wang, C. Jia, Z. Hou, X. Zhao, J. Zhang, W. Chang, Structural Insights into Energy Regulation of Light-Harvesting Complex CP29 from Spinach, *Nat. Struct. Mol. Biol.* 18 (3) (2011) 309–315.
- J. Standfuss, A.C. Terwisscha van Scheltinga, M. Lamborghini, W. Kühnbrandt, Mechanisms of Photoprotection and Nonphotochemical Quenching in Pea Light-Harvesting Complex at 2.5 Å Resolution, *EMBO J.* 24 (5) (2005) 919–928.
- J.T. Nielsen, N.V. Kulminkaya, M. Bjerring, J.M. Linnanto, M. Rätsep, M. Ø. Pedersen, P.H. Lambrev, M. Dorogi, G. Garab, K. Thomsen, et al., In Situ High-Resolution Structure of the Baseplate Antenna Complex in *Chlorobaculum tepidum*, *Nat. Commun.* 7 (1) (2016) 12454.
- A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsák, D.A. Pantazis, Accurate Computation of the Absorption Spectrum of Chlorophyll a with Pair Natural Orbital Coupled Cluster Methods, *J. Phys. Chem. B* 124 (40) (2020) 8761–8771.
- S. Knecht, C.M. Marian, J. Kongsted, B. Mennucci, On the Photophysics of Carotenoids: A Multireference DFT Study of Peridinin, *J. Phys. Chem. B* 117 (44) (2013) 13808–13815.
- J.P. Götzte, W. Thiel, TD-DFT and DFT/MRCI Study of Electronic Excitations in Violaxanthin and Zeaxanthin, *Chem. Phys.* 415 (2013) 247–255.
- P. Sherwood, A.H. De Vries, M.F. Guest, G. Schreckenbach, C.R.A. Catlow, S. A. French, A.A. Sokol, S.T. Bromley, W. Thiel, A.J. Turner, et al., QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis, *J. Mol. Struct. THEOCHEM* 632 (2003) 1–28.
- S. Dapprich, I. Komáromi, K.S. Byun, K. Morokuma, M.J. Frisch, A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives, *J. Mol. Struct. THEOCHEM* 461–462 (1999) 1–21.
- J.P. Götzte, Y.-W. Pi, S. Petry, F. Langkabel, J.F. Witte, O. Lemke, A User-friendly, Python-based Quantum Mechanics/Gromacs Interface: Gmx2qmmm, *Int J Quantum Chem* 121 (3) (2021), <https://doi.org/10.1002/qua.v121.310.1002/qua.26486>.
- S. Banerjee, D. Kröner, P. Saalfrank, Resonance Raman and Vibronic Absorption Spectra with Duschinsky Rotation from a Time-Dependent Perspective: Application to β -Carotene, *J. Chem. Phys.* 137 (22) (2012) 22A534.
- B. Karasulu, J.P. Götzte, W. Thiel, Assessment of Franck-Condon Methods for Computing Vibrationally Broadened UV-Vis Absorption Spectra of Flavon Derivatives: Riboflavin, Roseoflavin, and 5-Thioflavin, *J. Chem. Theory Comput.* 10 (12) (2014) 5549–5566.
- J.P. Götzte, B. Karasulu, W. Thiel, Computing UV/Vis Spectra from the Adiabatic and Vertical Franck-Condon Schemes with the Use of Cartesian and Internal Coordinates, *J. Chem. Phys.* 139 (23) (2013), 234108.
- P. Xu, V.U. Chukhutsina, W.J. Nawrocki, G. Schansker, L.W. Bielczynski, Y. Lu, D. Karcher, R. Bock, R. Croce, Photosynthesis without β -Carotene. *Elife* (2020) 9.
- R. Croce, M.G. Müller, R. Bassi, A.R. Holzwarth, Chlorophyll b to Chlorophyll a Energy Transfer Kinetics in the CP29 Antenna Complex: A Comparative Femtosecond Absorption Study between Native and Reconstituted Proteins, *Biophys. J.* 84 (4) (2003) 2508–2516.
- H. Lokstein, H. Härtel, P. Hoffmann, G. Renger, Comparison of Chlorophyll Fluorescence Quenching in Leaves of Wild-Type with a Chlorophyll-b-Less Mutant of Barley (*Hordeum vulgare* L.), *J. Photochem. Photobiol. B Biol.* 19 (3) (1993) 217–225.
- H. Lokstein, L. Tian, J.E.W. Polle, D. DellaPenna, Xanthophyll Biosynthetic Mutants of *Arabidopsis thaliana*: Altered Nonphotochemical Quenching of Chlorophyll Fluorescence Is Due to Changes in Photosystem II Antenna Size and Stability, *Biochim. Biophys. Acta - Bioenerg.* 1553 (3) (2002) 309–319.
- H. Rogl, R. Schödel, H. Lokstein, W. Kühnbrandt, A. Schubert, Assignment of Spectral Substructures to Pigment-Binding Sites in Higher Plant Light-Harvesting Complex LHC-II †, *Biochemistry* 41 (7) (2002) 2281–2287.
- A. Dreuw, M. Wormit, Simple Replacement of Violaxanthin by Zeaxanthin in LHC-II Does Not Cause Chlorophyll Fluorescence Quenching, *J. Inorg. Biochem.* 102 (3) (2008) 458–465.
- J.P. Götzte, D. Kröner, S. Banerjee, B. Karasulu, W. Thiel, Carotenoids as a Shortcut for Chlorophyll Soret-to-Q Band Energy Flow, *ChemPhysChem* 15 (15) (2014) 3392–3401.
- D. Kröner, J.P. Götzte, Modeling of a Violaxanthin-Chlorophyll b Chromophore Pair in Its LHCII Environment Using CAM-B3LYP, *J. Photochem. Photobiol. B Biol.* 109 (2012) 12–19.
- J.P. Götzte, Vibrational Relaxation in Carotenoids as an Explanation for Their Rapid Optical Properties, *J. Phys. Chem. B* 123 (10) (2019) 2203–2209.
- X. Sheng, A. Watanabe, A. Li, E. Kim, C. Song, K. Murata, D. Song, J. Minagawa, Z. Liu, Structural Insight into Light Harvesting for Photosystem II in Green Algae, *Nat. Plants* 5 (12) (2019) 1320–1330.
- J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, S. Turner, N. Swainston, P. Mendes, C. Steinbeck, ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites, *Nucleic Acids Res.* 44 (D1) (2016) D1214–D1219.
- G. Schaftenaar, J.H. Noordik, Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures, *J. Comput. Aided. Mol. Des.* 14 (2000) 123–134.
- M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform, *J. Cheminform.* 4 (1) (2012) 17.
- T. Yanai, D.P. Tew, N.C. Handy, A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP), *Chem. Phys. Lett.* 393 (1–3) (2004) 51–57.
- P.C. Hariharan, J.A. Pople, The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies, *Theor. Chim. Acta* 28 (3) (1973) 213–222.
- S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, *J. Chem. Phys.* 132 (15) (2010), 154104.
- S. Grimme, S. Ehrlich, L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, *J. Comput. Chem.* 32 (7) (2011) 1456–1465.

- [37] J.-D. Chai, M. Head-Gordon, Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections, *Phys. Chem. Chem. Phys.* 10 (44) (2008) 6615.
- [38] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, et al., Gaussian16 Revision A.03. (2016).
- [39] M.E. Casida, C. Jamorski, K.C. Casida, D.R. Salahub, Molecular Excitation Energies to High-Lying Bound States from Time-Dependent Density-Functional Response Theory: Characterization and Correction of the Time-Dependent Local Density Approximation Ionization Threshold, *J. Chem. Phys.* 108 (11) (1998) 4439–4449.
- [40] M.E. Casida, D.R. Salahub, Asymptotic Correction Approach to Improving Approximate Exchange-Correlation Potentials: Time-Dependent Density-Functional Theory Calculations of Molecular Excitation Spectra, *J. Chem. Phys.* 113 (20) (2000) 8918–8935.
- [41] D. Leupold, K. Teuchner, J. Ehlert, K.-D. Irrgang, G. Renger, H. Lokstein, Two-Photon Excited Fluorescence from Higher Electronic States of Chlorophylls in Photosynthetic Antenna Complexes: A New Approach to Detect Strong Excitonic Chlorophyll a/b Coupling, *Biophys. J.* 82 (3) (2002) 1580–1585.
- [42] D. Leupold, K. Teuchner, J. Ehlert, K.D. Irrgang, G. Renger, H. Lokstein, Stepwise Two-Photon Excited Fluorescence from Higher Excited States of Chlorophylls in Photosynthetic Antenna Complexes, *J. Biol. Chem.* 281 (35) (2006) 25381–25387.
- [43] M. Kasha, Characterization of Electronic Transitions in Complex Molecules, *Discuss. Faraday Soc.* 9 (1950) 14.
- [44] A.D. Becke, A New Mixing of Hartree-Fock and Local Density-functional Theories, *J. Chem. Phys.* 98 (2) (1993) 1372–1377.
- [45] I. Lyskov, M. Kleinschmidt, C.M. Marian, Redesign of the DFT/MRCI Hamiltonian, *J. Chem. Phys.* 144 (3) (2016), 034104.
- [46] F. Weigend, R. Ahlrichs, Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy, *Phys. Chem. Chem. Phys.* 7 (18) (2005) 3297.
- [47] M. Kleinschmidt, C.M. Marian, M. Waletzke, S. Grimme, Parallel Multireference Configuration Interaction Calculations on Mini- β -Carotenes and β -Carotene, *J. Chem. Phys.* 130 (4) (2009), 044708.
- [48] F. Neese, Software Update: The ORCA Program System, Version 4.0. *WIREs Comput. Mol. Sci.* 8 (1) (2018), e1327.
- [49] F. Santoro, R. Improta, A. Lami, J. Bloino, V. Barone, Effective Method to Compute Franck-Condon Integrals for Optical Spectra of Large Molecules in Solution, *J. Chem. Phys.* 126 (8) (2007), 084509.
- [50] N. Zamzam, J. van Thor, Excited State Frequencies of Chlorophyll f and Chlorophyll a and Evaluation of Displacement through Franck-Condon Progression Calculations, *Molecules* 24 (7) (2019) 1326.
- [51] G. Scalmani, M.J. Frisch, Continuous Surface Charge Polarizable Continuum Models of Solvation, I. General Formalism, *J. Chem. Phys.* 132 (11) (2010), 114110.
- [52] S. Petry, J.P. Götze, Effect of Protein Matrix on CP29 Spectra and Energy Transfer Pathways, *Biochim. Biophys. Acta - Bioenerg.* 1863 (2) (2021), 148521.
- [53] B.F. Milne, Y. Toker, A. Rubio, S.B. Nielsen, Unraveling the Intrinsic Color of Chlorophyll, *Angew. Chemie Int. Ed.* 54 (7) (2015) 2170–2173.
- [54] M.H. Stockett, L. Musbat, C. Kjær, J. Houmøller, Y. Toker, A. Rubio, B.F. Milne, S. Brøndsted Nielsen, The Soret Absorption Band of Isolated Chlorophyll a and b Tagged with Quaternary Ammonium Ions, *Phys. Chem. Chem. Phys.* 17 (39) (2015) 25793–25798.
- [55] T. Petrenko, F. Neese, Efficient and Automatic Calculation of Optical Band Shapes and Resonance Raman Spectra for Larger Molecules within the Independent Mode Displaced Harmonic Oscillator Model, *J. Chem. Phys.* 137 (23) (2012), 234107.
- [56] E. Gruber, C. Kjær, S.B. Nielsen, L.H. Andersen, Intrinsic Photophysics of Light-harvesting Charge-tagged Chlorophyll a and b Pigments, *Chem. – A Eur. J.* 25 (39) (2019) 9153–9158.
- [57] M.R. Preciado-Rivas, D.J. Mowbray, K. Lyon, A.H. Larsen, B.F. Milne, Optical Excitations of Chlorophyll a and b Monomers and Dimers, *J. Chem. Phys.* 151 (17) (2019), 174102.
- [58] J. Cerezo, J. Zúñiga, A. Requena, F.J. Ávila Ferrer, F. Santoro, Harmonic Models in Cartesian and Internal Coordinates to Simulate the Absorption Spectra of Carotenoids at Finite Temperatures, *J. Chem. Theory Comput.* 9 (11) (2013) 4947–4958.
- [59] H.A. Frank, J.A. Bautista, J.S. Josue, A.J. Young, Mechanism of Nonphotochemical Quenching in Green Plants: Energies of the Lowest Excited Singlet States of Violaxanthin and Zeaxanthin, *Biochemistry* 39 (11) (2000) 2831–2837.
- [60] N.L. Wagner, J.A. Greco, M.M. Enriquez, H.A. Frank, R.R. Birge, The Nature of the Intramolecular Charge Transfer State in Peridinin, *Biophys. J.* 104 (6) (2013) 1314–1325.
- [61] FU ZEDAT HPC Service, 2020.

Spectral Characterization of the Main Pigments in the Plant Photosynthetic Apparatus by Theory and Experiment

Jan P. Götze, Florian Anders, Simon Petry, Jan Felix Witte and Heiko Lokstein

Table of contents:

- Spectra obtained with other QM methods
- QM/MM vertical spectra for selected CP29 Chl residues
- Modes removed from total sets in the VH approach
- Geometries
- NPL illustration

Spectra obtained with other QM methods

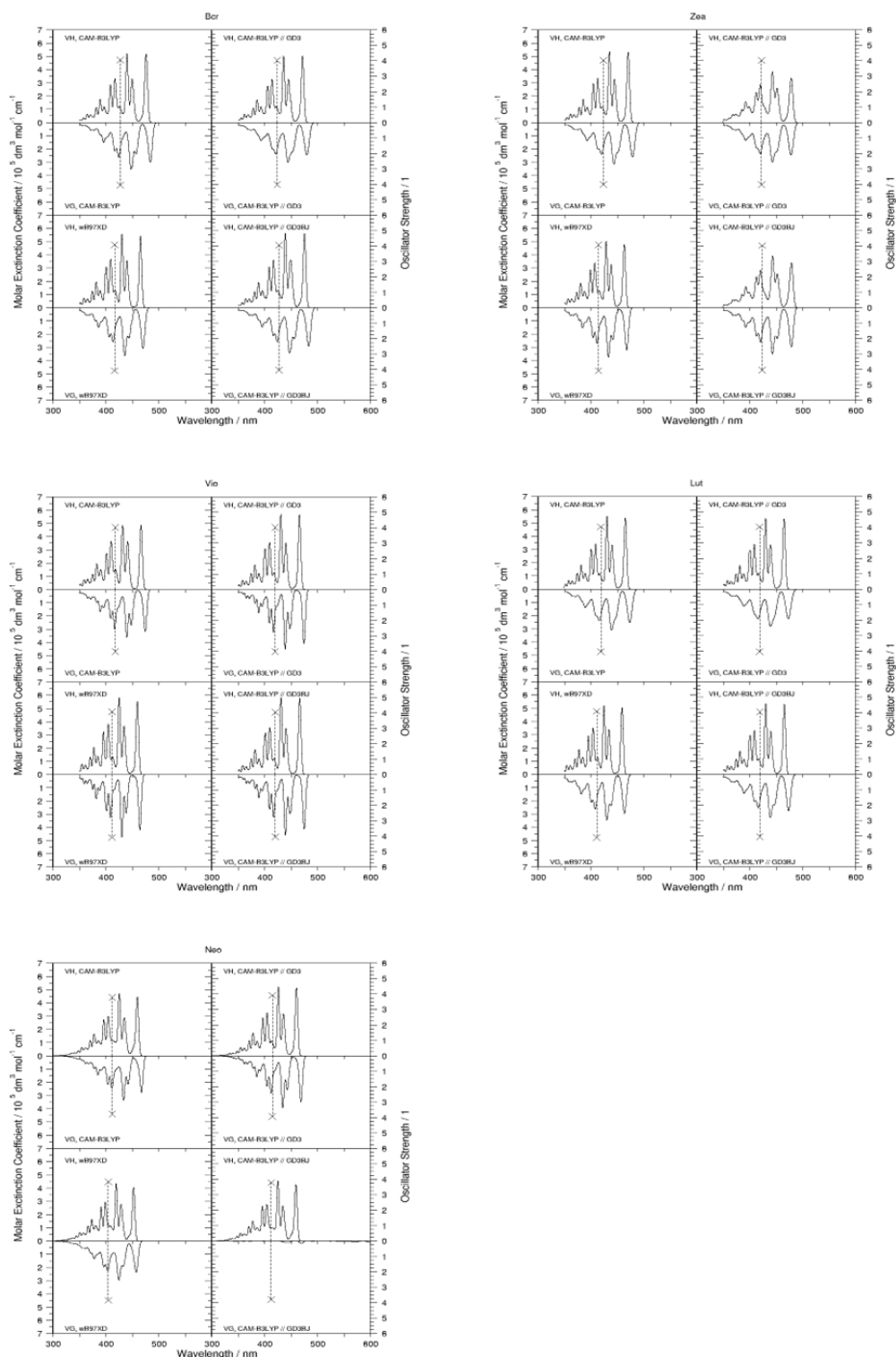


Figure S1: Crt spectra (TD-FC broadened; vertical excitation as dashed line) computed with CAM-B3LYP and various dispersion corrections, as well as with ω B97XD. Spectra are virtually identical and depend more on the choice of TD-FC approach (VG/VH) than on the density functional.

QM/MM vertical spectra for selected CP29 Chl residues

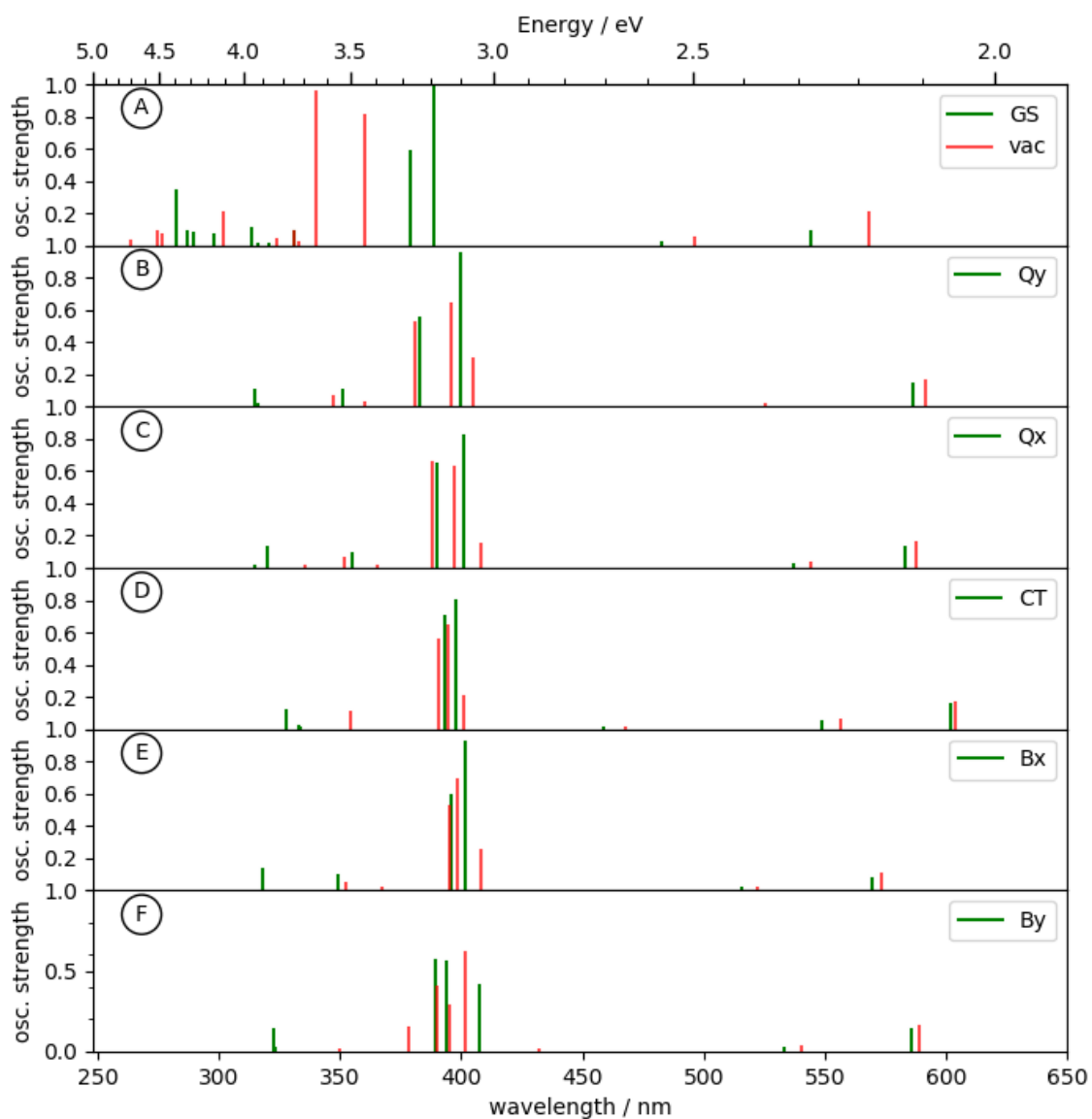


Figure S2: Chl 606b vertical excitation spectra computed with CAM-B3LYP, as embedded in the protein matrix (green) and with the same geometry in gas phase (red) for the ground state (GS), the excited states Q_y , Q_x , B_x , B_y and a charge transfer state (CT).

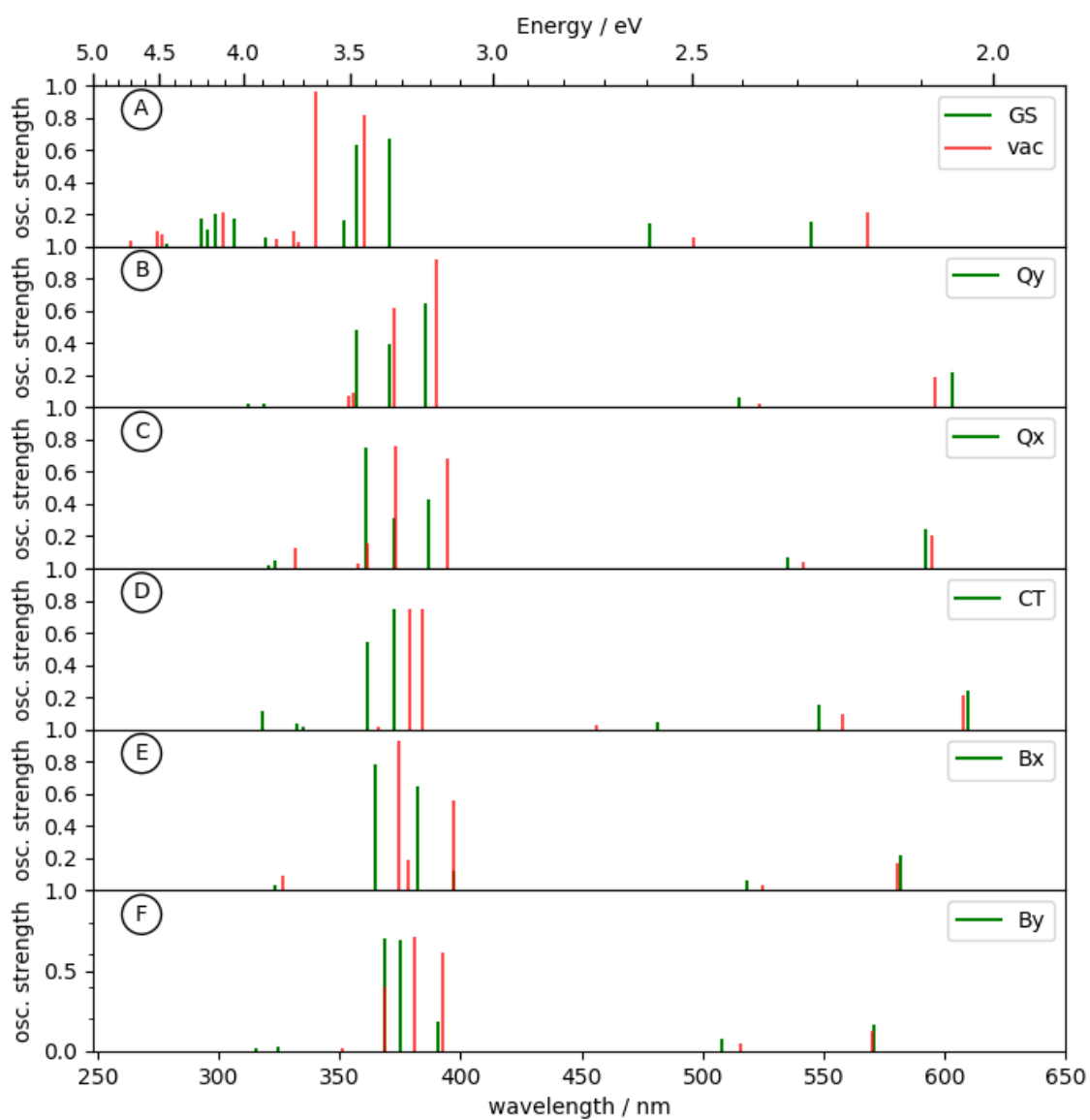


Figure S3: Chl 607b vertical excitation spectra computed with CAM-B3LYP, as embedded in the protein matrix (green) and with the same geometry in gas phase (red) for the ground state (GS), the excited states Q_y, Q_x, B_x, B_y and a charge transfer state (CT).

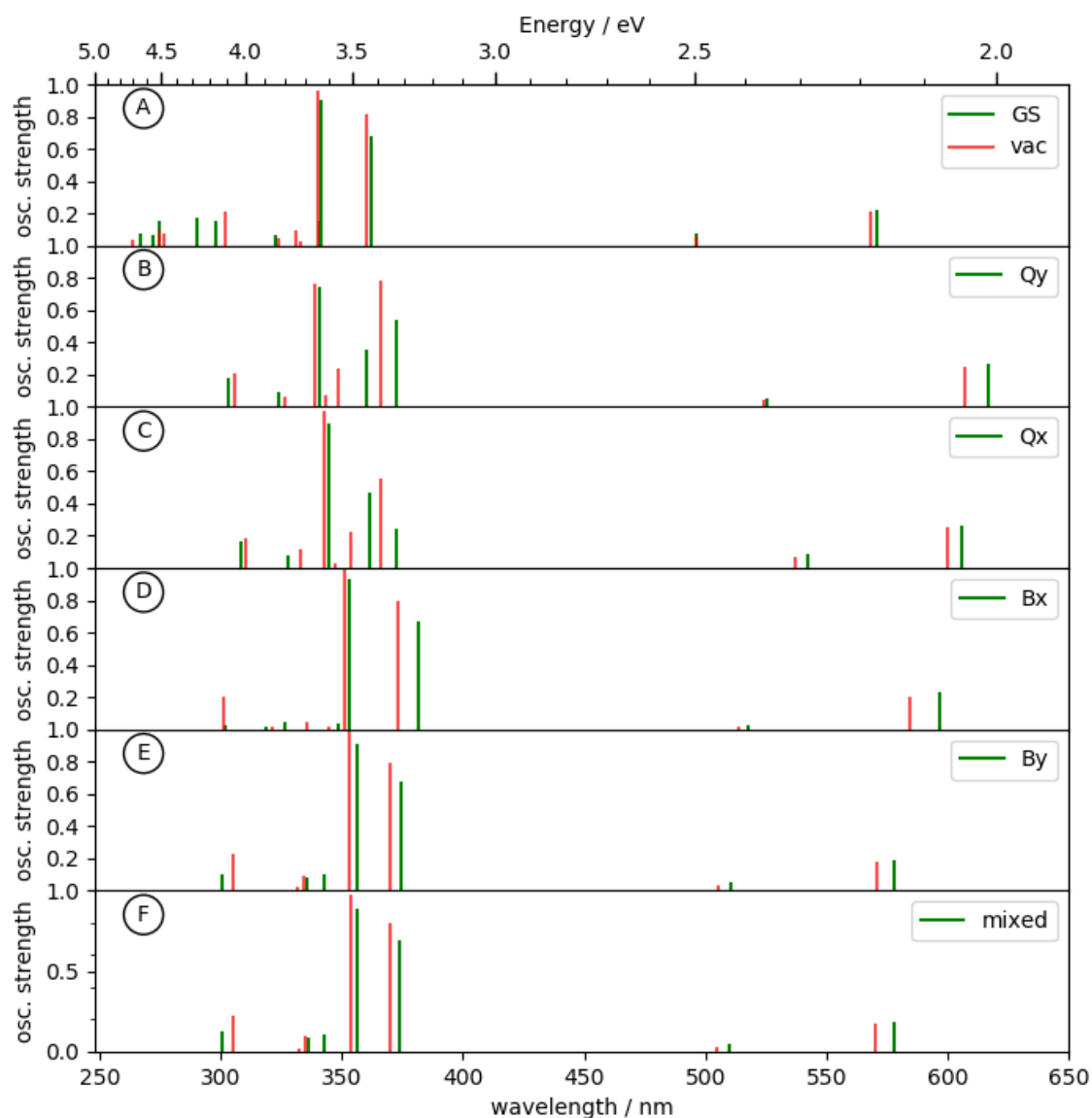


Figure S4: Chl 604a vertical excitation spectra computed with CAM-B3LYP, as embedded in the protein matrix (green) and with the same geometry in gas phase (red) for the ground state (GS), the excited states Q_y, Q_x, B_x, B_y and a mixture of a charge transfer state (CT) with the B_y (mixed). Note that the optimization of both, B_y and the mixed state apparently end up in the same structure.

Modes removed from total sets in the VH approach

Method	Bcr	Zea	Vio	Lut	Neo
CAM-B3LYP	58	33	55	31	32
CAM-B3LYP w. GD3	34	10	31	35	34
CAM-B3LYP w. GD3BJ	29	10	36	35	35
ω B97XD	55	35	34	36	36

Geometries

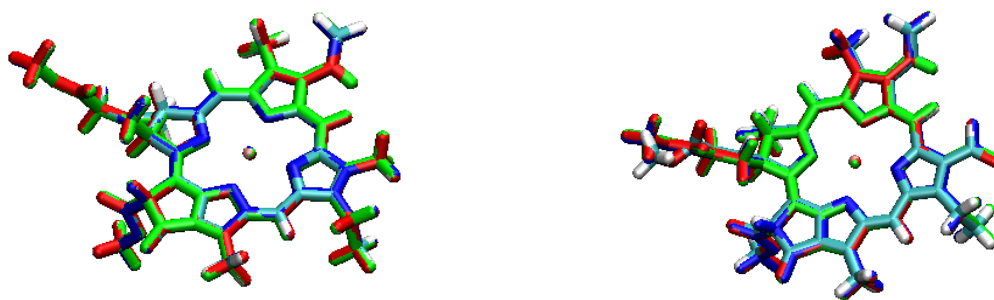


Figure S5: Chl a (left) and b (right) CAM-B3LYP/6-31G* optimized structures in the gas phase for the ground state (colored), Q_y (blue), Q_x (red) and B_x (green). States are virtually identical in structure.

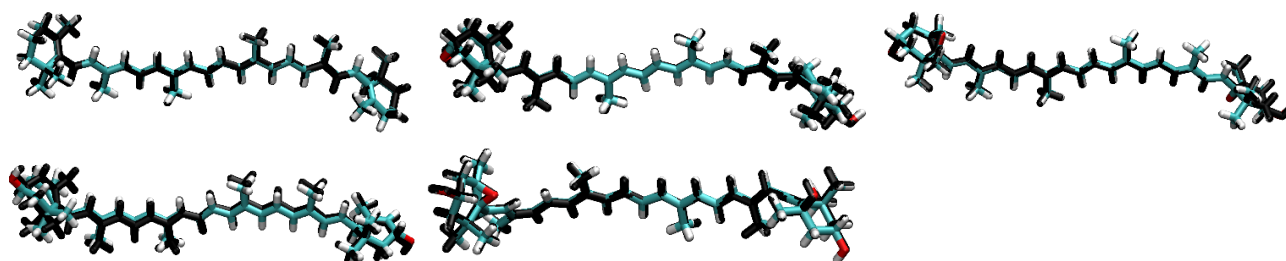


Figure S6: Crt structures (top, left to right: Bcr, Zea, Vio, bottom: Lut, Neo) for the ground state (colored) and the $1B_u$ excited state (blue). Structures are virtually identical regardless of state.

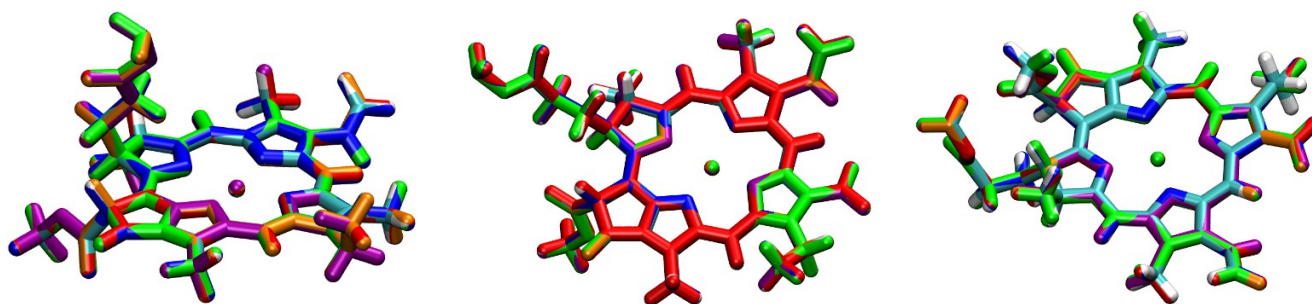


Figure S7: Chl a 604 (left), Chl b 606 (middle) and Chl b 607 (right) CAM-B3LYP/6-31G* optimized structures embedded in the protein matrix for the ground state (colored), Q_y (blue), Q_x (red) and B_x (green), B_y (orange) and CT (purple). States are virtually identical in structure.

NPL illustration

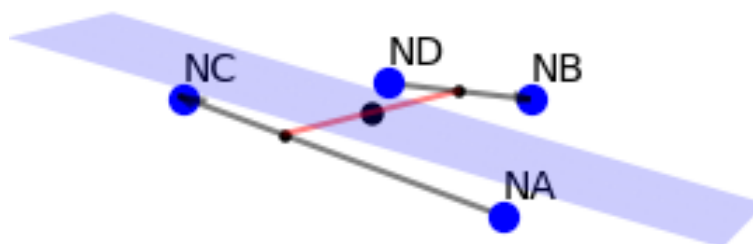


Figure S8: Definition of the plane to obtain deviation from the planarity parameter. The plane is located at the center of the shortest connection between the NA-NC and NB-ND lines; with the plane perpendicular to the connection.

4.3 Paper 3

“Effect of protein matrix on CP29 spectra and energy transfer pathways”

S. Petry, J. P. Götze

Biochimica et Biophysica Acta - Bioenergetics **2022**, 1863, 2, 148521 ff.

DOI: 10.1016/j.bbabi.2021.148521

URL: <https://doi.org/10.1016/j.bbabi.2021.148521>

Contributions

The project was conceived by Simon Petry with the help of Jan Götze. Simon Petry designed and performed all QM/MM and QM calculations and wrote the manuscript. The evaluation and interpretation of the results were carried out by Simon Petry. All authors contributed to the final version of the manuscript.



Contents lists available at ScienceDirect

BBA - Bioenergetics

journal homepage: www.elsevier.com/locate/bbabio

Effect of protein matrix on CP29 spectra and energy transfer pathways

S. Petry, J.P. Götze*

Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany

ARTICLE INFO

Keywords:

Light harvesting complex
Excitation energy transfer
Transition dipole moment
CP29
Blue light pathways
Chlorophyll

ABSTRACT

We investigate energy transfer pathways between strongly coupled chlorophylls (Chls) in the CP29 (LHCII B4.1) antenna complex of *Pisum sativum*, including the possibility of higher energy states. We test for the environmental effects caused by the protein, membrane and solvent using a hybrid QM/MM approach. Classical molecular dynamics simulations of the full CP29 complex embedded in a DOPC membrane have been performed, followed by calculations of the time dependent DFT spectra of all Chls at several timesteps. The relative orientations of transition dipole moments (TDMs) were specifically analyzed, including and excluding the point charge field (PCF) of the surrounding environment.

The PCF is found to drastically shift the spectra of specific Chls, while the majority of Chls is mostly unaffected. The net effect on the sum spectrum is however found to be negligible: The few strong changes in Chl spectra cancel each other due to being opposite in sign. We further find that the spectra of the Chls coordinating to water show a blue shift upon introduction of the environment. Conversely, the spectra of the Chls coordinating to glutamine show a red shift upon activation of the PCF.

As the main influence of the PCF for tuning the couplings, we identify the energetic position of the individual chromophores. The fine-tuning, especially for states energetically above the Q_y state, is however controlled by the changes in the TDM orientations. We also find an indication for the PCF to steer potentially harmful high energy excitations away from the PSII core complex.

1. Introduction

Photosynthesis represents a highly complex network capturing sunlight and converting it into various different carriers of chemical energy, while producing oxygen as a by-product. An essential part of this photosynthetic machinery is represented by the light-harvesting complexes (LHCs), pigment binding proteins acting in different roles as antennae [1,2] and “conductors” [3–5]. It can be assumed that the closer these complexes are to the PSII core, the stronger is the focus on the transport function vs. the harvesting function [1]. Although photosynthesis has been known for about two centuries and has been comprehensively studied from both experimental and theoretical views [6,7], the flow of high-energy excitations in biological LHCs is still not fully understood. While chlorophylls (Chls) strongly absorb in the blue spectral region, most models represent this simply via a fast internal conversion (IC) to lower excited states. This is likely an oversimplification, as the Chls are strongly coupled. The resulting pathways may have implications for experiments that involve other blue light absorbers such as carotenoids.

Improved measurement techniques like X-ray diffraction and electron microscopy have recently provided high-resolution structures of PSII and the surrounding antenna complexes. This information is essential to understand the system, because the protein matrix determines the organization of the pigments by stabilizing the binding pigments [8]. Furthermore, the flexibility of the protein allows different stable conformations, which tune the optical properties of the pigments.

The antenna complex CP29 [9,10], which is part of the PSII supercomplex and present in plants and green algae, contains 14 Chls (ten Chls a and four Chls b) as well as 3 carotenoids (neoxanthin, lutein and violaxanthin) [11,12]. The secondary and the tertiary structure of CP29 is very similar to the major LHC of PSII (LHCII), it however does not form a trimer as LHCII does. As these systems are highly complex and adjusted to their environmental conditions, protective mechanisms have evolved to adapt to different incoming light conditions. Dynamic quenching mechanisms may deactivate the underlying Förster resonance energy transfer (FRET), which represents transition dipole moment (TDM) interactions between the different Chls. The singlet-excited Chls are thus quenched before they can convert into triplet states by this protective

* Corresponding author.

E-mail addresses: s.petry@fu-berlin.de (S. Petry), jgoetze@zedat.fu-berlin.de (J.P. Götze).

<https://doi.org/10.1016/j.bbabio.2021.148521>

Received 5 July 2021; Received in revised form 25 October 2021; Accepted 2 December 2021

Available online 8 December 2021

0005-2728/© 2021 Elsevier B.V. All rights reserved.

mechanism.

Computational modelling of these complexes is challenging: Molecular mechanics (MM) force fields are not accurate enough to describe their mostly electron-based mechanisms and, on the other hand, the protein-chromophore complexes are too large to be described by quantum mechanics (QM). Fortunately, hybrid methods are available to describe a specific portion on a QM level and the large remainder using a less expensive MM approach (QM/MM) [13,14]. QM/MM combines the accuracy and bonding flexibility of the QM approach with the possibility of calculating the complete system using the MM methodology.

For Chl, the description using QM/MM is necessary, as the spectroscopic properties can only be obtained via a proper QM description. Graczyk et al. have shown that various Chl spectra can be well reproduced using the combination of the density functional CAM-B3LYP [15,16] with a 6-31G(d) basis [17]. Chl in LHCII systems was also treated using CAM-B3LYP by our previous work [5,18–20]. The Chl properties regarding their lowest energy states (the Q band, or the two lowest excited singlet states) have been in the context of light harvesters and CP29 described before [12]. However, in the context of the high energy states (Soret band, or the blue states higher than the Q band), we need to develop new approaches to analyze potential pathways due to the much higher number of involved donor (site/state)-acceptor (site/state) possibilities. It is thus the purpose of this article to provide a first venture into the potential energy exchanges between all optical states of Chls.

2. Experimental methods

2.1. MD setup

The high-resolution electron microscopy structure of a stacked C2S2M2-type PSII-LHCII supercomplex from *Pisum sativum* (pdb entry: 5XNL, res. 2.7 Å) resolved by Su et al. [21] was used to obtain the structure of CP29. The complex was embedded in a double layer membrane consisting of 445 1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine (DOPC) lipid molecules. DOPC membranes have been used to represent the thylakoid membrane in a simplified way [22]. The area per lipid was 0.769 nm², which is comparable to the experimental results of a DOPC membrane with 0.725 nm² [23]. The system was neutralized using twelve potassium ions.

MD simulations were performed using the GROMACS simulation package 2018.6 [24]. After an energy minimization, a short NVT equilibration phase of 100 ps followed by a longer NPT phase of 1 ns, we performed a MD simulation with 50 million steps. For all dynamics, a time step for integration of 2 fs was employed. The AMBER99SB*-ILDNP forcefield [25] has been used for the proteins and lipid11 [26] for the lipids. For the Chls and the carotenoids the set of parameters is reported in ref. [27–29]. The production run was performed in a NPT ensemble at T = 300 K using leap-frog integration [30] and the TIP3P water model [31]. The procedure is oriented on the publication of Jurinovich et al. [12] which studies the low energy bands of the system.

2.2. QM/MM setup

The QM/MM calculations were performed with the gmX2qmmm interface [14] using Gaussian 16 [32] and the GROMACS simulation package 2018.6 [24]. From the MD simulation, 11 snapshots were taken at every 5 ns beginning at 50 ns to obtain the initial structures of the QM/MM ensemble. From each snapshot, each of the 14 Chls was used as center of a separate, 4-layered QM/MM calculation. Layers were, starting from the inside: The QM layer, which is optimized the CAM-B3LYP/6-31G* method/basis set [15–17]; the inner layer, which is optimized on the same force field level as in the MD simulation, but is subdivided in a non-frozen (active layer) and a frozen part; and an outer layer, which is a fully not interacting and frozen area. The last layer is kept allowing the original MD run files to be used for consistency

between the MD and QM/MM calculations.

The QM-layer consisted of a Chl with the phytyl chain cut at the first carbon-carbon bond after the ester group, and the magnesium coordinating residue, if present within a cutoff radius of 0.3 nm. To ensure a consistent approach, we kept this cutoff radius even for those Chl geometries which result in carotenoids as Mg-coordination partners or no coordination at all. However, these cases are not used for our analysis because the interaction of Chl with carotenoids will not be part of this study. A list of the coordinating groups can be found in the Supporting information (SI).

The inner (interacting) layer included all atoms within a range of 4 nm with respect to the Chl's magnesium ion. The active region, namely all residues, solvent molecules and ions within a range of 0.3 nm of QM layer atoms, was optimized using a steepest descent algorithm (force tolerance: 100.0 kJ/mol/nm, step size 0.01 nm, 3 nm radius of non-bonded partner list). To further reduce computational costs all solvent molecules with a distance larger than 2 nm to the magnesium of the QM Chl were removed from the system beforehand.

After QM/MM optimization, the structures were used for TD-DFT [33] calculations of the first twelve lowest states to simulate vertical absorption spectra. The TD-DFT calculations were performed twice, once including and once excluding the PCF obtained from the QM/MM procedure, to study the impact of the protein environment. Note that due to our model only including CP29, the QM/MM optimization and TD-DFT calculation of Chl 616, which is at the interface between CP29 and the PSII CP47 antenna, were based on the crystal structure to retain a realistic orientation.

2.3. Coupling analysis

There are various advanced coupling schemes available which have been applied to light harvesting systems [34,35]. However, we seek to employ a simple scheme that is easily expanded for multiple states as well as can be separated easily into the components of relative state energies, chromophore distance and orientation. Therefore, to get a first impression of the low and high energy state chromophore coupling, the square value of the orientation factor κ was calculated (see below). This restriction of our analysis arises from the fact that we show in the beginning of our results section (see below) that one cannot trust the calculated spectra to the extent of being exact in terms of the full spectral envelope. Therefore, it is not possible to compute the spectral overlap of the donor emission spectrum and the acceptor absorption spectrum, which would complete a full Förster treatment of the coupling. We thus only include κ^2 and assume unity for absorption/emission overlap. We also aim to use the results of this study for comparison with inter-complex coupling studies for which Förster transfer is the only feasible mechanism.

The orientation factor κ , a measure for the possibility of FRET interaction k_{ET} , is defined as:

$$k_{ET} \sim \kappa^2 \quad (1)$$

$$\kappa = \underbrace{\hat{\mu}_D \cdot \hat{\mu}_A}_{\kappa_1} - 3 \underbrace{(\hat{\mu}_D \cdot \hat{R})(\hat{\mu}_A \cdot \hat{R})}_{\kappa_2} \quad (2)$$

With the normalized TDMs $\hat{\mu}_i$ and the inter-chromophore displacement \hat{R} , κ^2 can range between 0 and 4. The approach also has the advantage that the uncertainty about the TDM sign in TD-DFT is of no relevance due to the squaring of κ . While the TDM orientation for the Q band states is known, it becomes much more uncertain for the case of the Soret states or for large distortions by the PCF.

Our QM/MM scheme combined with above approach yields a series of interaction matrices, in total 1848, for 12 excited states \times 14 Chls \times 11 frames, minus 288 frames/Chl combinations that were removed due to structural criteria mentioned above. The complete set of 1428 interaction matrices was obtained once for the QM/MM situation (with PCF) and once for the gas phase situation (no PCF). Each matrix represents the

interaction of a donor (site/state) of a specific frame with all states of all other Chls.

While the matrices described above deliver a statement on the general possibility to enter a target state for a specific MD frame, we required an overview on the full MD trajectory. We thus devised a Boolean matrix that is true for all cases for which κ^2 is higher than the average value κ^2 at the dynamic limit, where the rotation of the chromophores is faster than the lifetime of the excited state ($\kappa^2 > 2/3$). Also, the donor energy of that specific frame must be higher than the energy of the acceptor for the entry to be “true”. These Boolean matrices thus represent the possibility of downhill energy transfer with high probability for a given frame.

To obtain the picture on all frames from the MD trajectory, the Boolean matrices were summarized to form count matrices which represent the number of “true” counts for each donor/acceptor combination, effectively reducing the matrices from 1560 to 168 and at the same time representing all frames in one matrix. Due to the simplicity of our coupling model and the statistically low number of QM/MM frames, we decided to restrict our analysis to the differences between the PCF and gas phase by subtracting the count matrices for both cases. These difference matrices are the main product of our Förster-type analysis and represent how much the electrostatic part of the protein environment favors (positive values) or disfavors (negative values) a specific donor/acceptor interaction. A short “recipe” for the interaction matrix is (I) calculate κ^2 for each chromophore interacting with all other chromophores, (II) calculate a Boolean matrix that is true, if the energy criterium is fulfilled ($E_D > E_A$) and the threshold criterium is fulfilled ($\kappa^2 > 2/3$), (III) summarize all Boolean matrices of each frame, (IV) subtract the gas phase sum matrix from the PCF sum matrix.

To facilitate analysis of the Chl couplings, we created arbitrary clusters of Chls. As indicated in Fig. 1 we chose to divide the CP29 Chls in a stromal and a luminal cluster on the PSII-oriented side (PSII-L and PSII-S clusters) and in a stromal and a luminal cluster oriented towards LHCII (LHCII-L and LHCII-S clusters), while the Chl 602, 615, 608 remain ungrouped. The clustering is (only) based on the relative proximity of the Chls towards each other and the neighboring protein complexes. The PSII-S/PSII-L/LHCII-S/LHCII-L clusters are indicated by the blue/green/red/yellow boxes in the difference interaction matrices

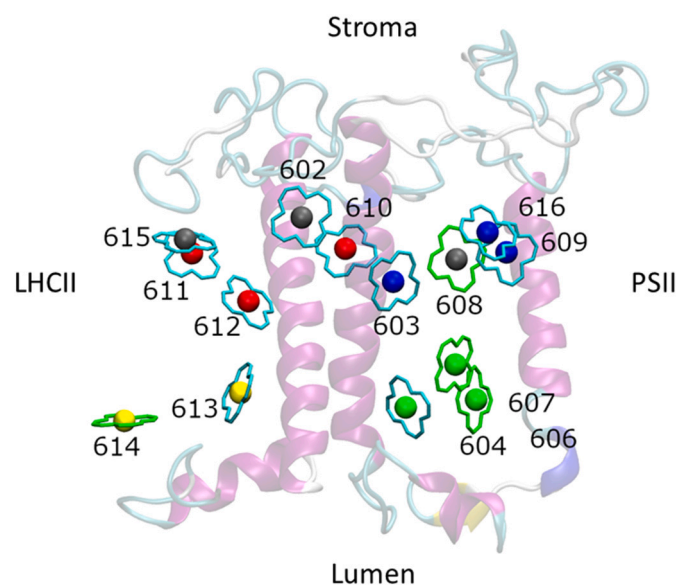


Fig. 1. Last frame of our CP29 MD trajectory, only showing the protein backbone and Chl a/b (cyan/green) core ring atoms. Chl labelling as in the original structure. Mg ion associated colours (near LHCII/PSII, luminal: yellow/green; stromal: red/blue; grey: unclustered) indicate Chl groups as used for coupling analysis.

shown throughout the paper.

3. Results

3.1. Optical spectra

The resulting sum spectra adding up all QM/MM frames for the individual Chls are shown in Fig. 2 beginning with the ten Chl a, followed by the four Chl b and ending with the sum of all of them as well as the experimental spectrum in grey [36]. To account for the method-intrinsic error of CAM-B3LYP, the theoretical spectra are shifted 299 meV to lower energies, which corresponds to the difference of the Q band maxima of the sum and the experimental spectrum.

Comparing the gas phase to the PCF situation, each spectrum is influenced differently, which is reasonable due to the environment being different for each case. Shifts are found however to be similar in quality, though not quantity, for Soret and Q band for the individual Chls, showing that the PCF acts in a state-specific manner in terms of the extent of the shifts. Though some Chl spectra are shifted significantly upon introduction of the PCF (e.g., Chl a 610), gas phase and PCF sum spectra are very similar. This is due to only few Chl spectra shifting strongly, and those that do, shift with opposite signs (e.g., Chls b 606 & 607 red vs. Chls b 608 & 614 blue), virtually canceling each other. This might indicate that despite individual shifts being beneficial for coupling (see next section), retaining the “unbound” sum spectrum could be beneficial.

In Fig. 2, two systematic shifts can be identified (see SI for overlap spectra and shift calculation). Those can be attributed to the Chl coordination, namely water or glutamine, despite these residues being present in both PCF and gas phase calculations. A Chl a (604) and two Chls b (608, 614) are coordinated to a water molecule and show a red shift going from vacuum to PCF. The shift is -6 , -22 , -6 meV (604a, 608b, 614b) for the Q band and -36 , -159 , -50 meV (604a, 608b, 614b) for the Soret band. The overall shift is larger than for the Q band and thus in accordance with our statement above that the PCF has a stronger effect on the Soret band.

Conversely, Chls a (602, 609, 610) and Chl b (606) are coordinated to glutamine, and their spectra are all blue shifted when including the PCF. The shift is $+15$, $+13$, $+32$, $+44$ meV (602a, 609a, 610a, 606b) for the Q band and $+62$, $+112$, $+154$, $+43$ meV (602a, 609a, 610a, 606b) for the Soret band. Again, we find a stronger effect on the Soret band than on the Q band, this will be shown in even more detail at the end of the next section. For glutamine, it can be expected that due to the strong charge of the coordinating residue, the omission of the environment in the gas phase has an adverse effect on the ground state energy, leading to the observed blue shift when re-introducing the PCF.

3.2. Energies

Table 1 shows the comparison of the average Q_y state energies of the fourteen Chls with previous work. Our calculated average Q_y state energies are blue shifted on average by $+0.196$ eV, $+0.284$ eV, $+0.121$ eV compared with the site energies of Jurinovich et al. [12], Mascoli et al. [3], Maity et al. [37], respectively.

The comparison with the Chl site energies of Mascoli et al. [3] shows two major differences in the energetic order. Our excitations arising from Chl 603 and 609 are higher in energy, possible due to the spatially close Chl 616, which is not covered in the work of Mascoli. Furthermore, we predict the Chl 615 excitation to be lower in energy. Apart from these differences, the energetic order within the presented clusters is identical to the findings of Mascoli et al.; i.e., $E_{Q_y}(603) < E_{Q_y}(609)$ in PSII-S, $E_{Q_y}(613) < E_{Q_y}(614)$ in PSII-L, $E_{Q_y}(611) < E_{Q_y}(610) < E_{Q_y}(612)$ in LHCII-S and $E_{Q_y}(604) < E_{Q_y}(607) < E_{Q_y}(606)$ in LHCII-L.

Compared to the computational results of Jurinovich et al. [12], our results do not match the experimental excitation energies as well in absolute terms. However, as we have seen in the last paragraph that the

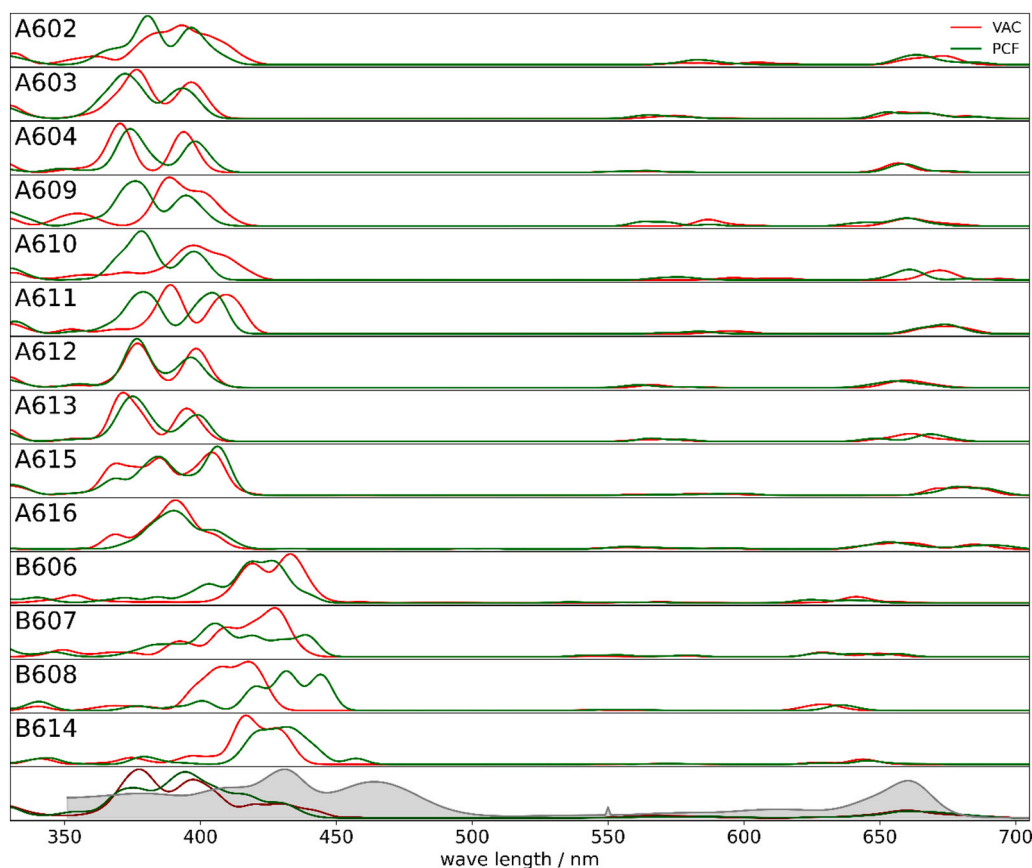


Fig. 2. Comparison of TD-CAM-B3LYP/6-31G⁺ sum spectra of all QM/MM frames for each individual CP29 Chl a/b molecule (labelled 608–616, located in chain “r” of the crystal structure with PDB identifier 5XNL [21]), including their coordinating residue. Bottom spectrum is sum over all Chls (full CP29 spectrum). Green line: spectra with PCF; red line: spectra in gas phase. Lowest 12 excited states shown, Gaussian broadened (7.5 nm FWHM). The bottom spectrum also shows the experimental spectrum of CP29 [36] as grey area. The theoretical spectra are shifted by 299 meV to higher wave lengths to match the experimental Q_y excitation.

Table 1

Comparison of the average site energies (eV) along the MD trajectory with the works of Jurinovich et al. [12], Mascoli et al. [3] and Maity et al. [37].

Chl	Jurinovich	Maity	Present work	Mascoli (exp.)
602a	1971	2057	2.158	1862
603a	1973	2060	2.170	1862
604a	1969	2034	2.176	1901
609a	1958	2060	2.190	1868
610a	1975	2028	2.172	1869
611a	1976	2041	2.145	1862
612a	1974	2060	2.187	1873
613a	1971	2034	2.168	1872
615a	1986	2050	2.118	1.887
616a	–	–	2.209	–
606b	2042	2082	2.262	1999
607b	2035	2107	2.237	1947
608b	2037	2163	2.253	1987
614b	2.046	2111	2.228	1.980

energetic order is reproduced well for the Chl clusters by our approach. It thus appears that a QM/MM optimization of the Chl and its environment is necessary to obtain the correct energetic hierarchy. Non-QM/MM optimized structures from earlier work seem to deliver a more homogenous picture of the Q_y states. As for the match in absolute values, it is well known that vertical energies need to be slightly blue shifted [38]; also, in our case here, the blue shift is appearing after stabilization of the ground state due to the QM/MM optimization.

The site energies of Maity et al. [37] are closer to the results of our study but display a different state order. The energetic order of Maity et al. within our clustering scheme shows slight differences: E_{Q_y}(603) = E_{Q_y}(609) in PSII-S, E_{Q_y}(610) < E_{Q_y}(611) < E_{Q_y}(612) in LHCI-S and E_{Q_y}(604) < E_{Q_y}(606) < E_{Q_y}(607) in LHCI-L.

As described above, the PCF has a crucial influence on the energetic

position of the Chl excitations. Fig. 3 shows this influence by comparing the Q_y and Q_x state energies with PCF and in the gas phase for the individual clusters. For Q_y, we observe only minute changes, mostly affecting Chl 606 and possibly 610; for Q_x, however, changes are most more pronounced and affect all presented Chl clusters.

We further show in Figs. 4 and 5 the effect of the PCF on the Soret state energies, namely the lowest two excitations. This region is obviously much more responsive to a change in environment, which might be related to an energetically charge transfer (CT) state mixing with the bands. The existence of this CT state is so far dependent on the computational method employed and experimentally not covered. However, we find that even states with low CT character in the Soret band will respond much stronger than Q band states to a change in PCF (state character analysis not shown); only few states remain unaffected when changing the environment. This was also found already for another system, for which the environment had a much stronger impact on the Soret band than on the Q band [14,20]. We thus decided to analyze the effects of the PCF on the coupling between Chl states including the whole Chl UV/Vis spectrum (Q and Soret bands).

3.3. Transition dipole moment

For all difference interaction matrices, we find that the interactions within each cluster are for most only weakly influenced by the PCF. Small changes of ±2 counts are observed, apart from a few notable exceptions. The cluster-cluster interaction is more affected by the PCF, which enables us to identify trends for clusters to act as donor or acceptors.

3.3.1. Q_y state

For the Q_y state, shown in Fig. 6, the donor role of the PSII-S cluster is enhanced by 7.0 counts on average per Chl. This Chl triad shows

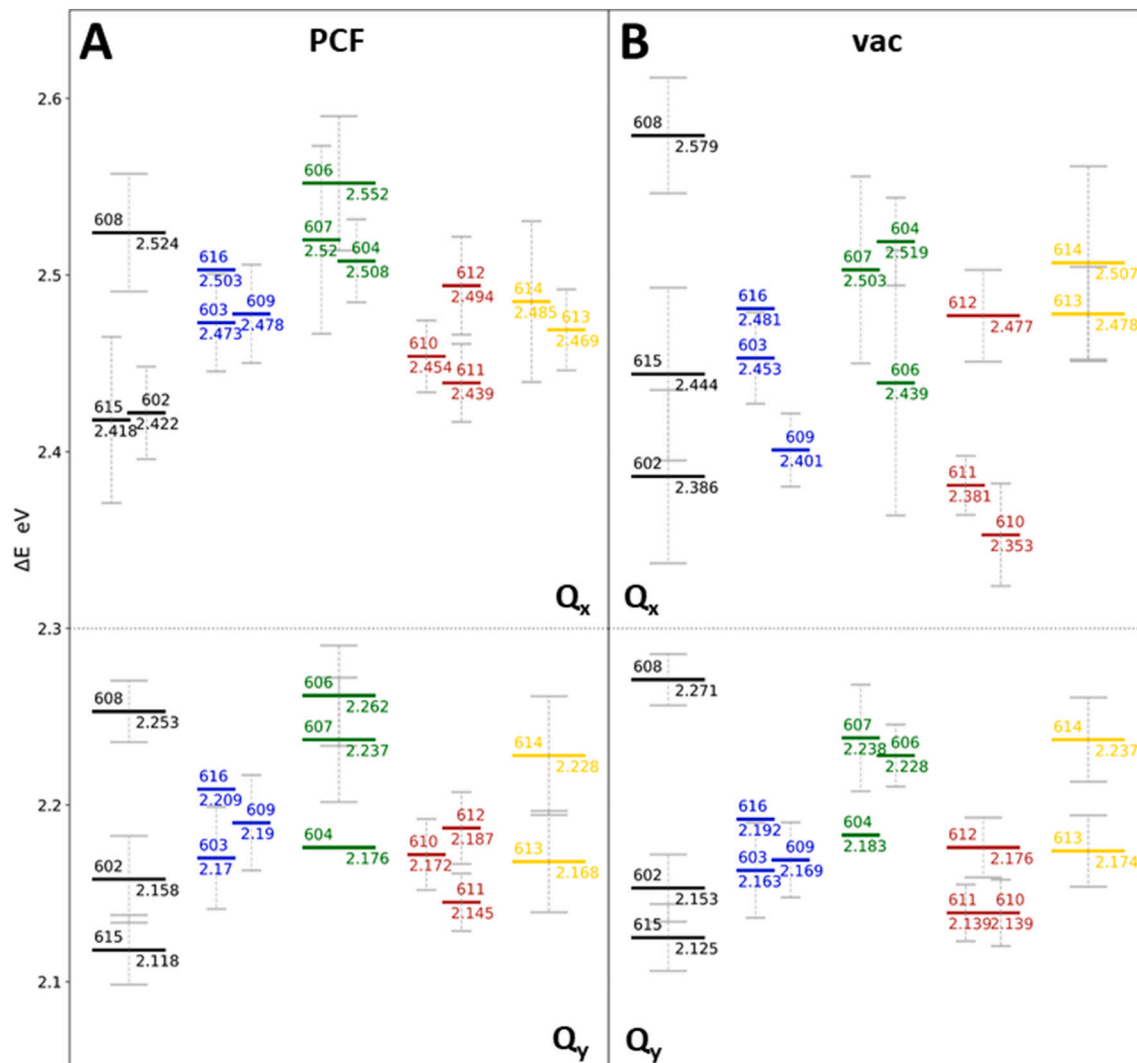


Fig. 3. Chl vertical excitation energies in the Q_y (bottom) and the Q_x (top) region including PCF (A) and in gas phase (B) using the introduced colour code for the clusters (PSII-S: blue, PSII-L: yellow, LHCII-S: red, LHCII-L: green). The thick line of for each Chl/state corresponds to the average energy along the 100 ns MD simulation with their corresponding standard deviation as grey vertical range bar.

individual Chls with heterogenous changes in donor activity. Chl 609 is enhanced acting as a donor by 13 counts, noting that 6 counts are arising from changes towards the PSII-L cluster, especially towards Chl 604 with 4 counts. As Chl 616 is based on the X-ray coordinates, this needs to be taken with caution.

The donor role of the PSII-L cluster is suppressed by 6.7 counts on average per Chl. Chl 604 is exclusively suppressed acting as a donor by -10 counts and at the same time, its acceptor role is suppressed by -9 counts.

The LHCII-oriented clusters are less affected, except Chl 610 of the LHCII-S and Chl 614 of the LHCII-L cluster. While the acceptor role Chl 610 is enhanced by 8.0 counts and the donor role is equally enhanced/suppressed by 5.0/ -8.0 counts, the acceptor role of Chl 614 is suppressed by -13.0 counts and its donor role is exclusively enhanced by 12.0 counts.

Apart from the general assessment above, we refrain from a further case-by-case analysis as the number of interactions is still large. However, comparing Fig. 6A and B shows that the Chls are mainly unaffected by the orientation of the TDMs, as the energetic order remains constant between PCF and gas phase. We find that Chl 610 is mostly affected by the relative energetic position, as its enhanced donor role only appears when taking the energy shift into account. Conversely, Chl 614 is also

improving as acceptor due to the energetic position. However, most changes we observe in the full scheme Fig. 6A do not appear in the TDM-only scheme and are thus originating from the energetic changes between PCF and gas phase.

3.3.2. Q_x state and $Q_{x/y}$ interactions

In comparison to the Q_y state, the Q_x state can energetically downhill donate to the Q_x and Q_y state. The theoretically possible $Q_y \rightarrow Q_x$ arrangement was not found in our data set. The internal interactions within each cluster are more strongly affected by the PCF compared to the Q_y case but remain weak: The interactions within the clusters themselves only show small enhancements/suppressions of 1.6/ -1.0 counts per cluster.

For the Q_x - Q_x state interactions, the donor role of the LHCII-L cluster is suppressed by -9.0 counts on average per Chl, while its acceptor role is enhanced by 14 counts on average per Chl in the cluster. This also applies for the unclustered, stromal Chls with an average suppression of their donor role by -10.6 counts and an average enhancement of the acceptor role by 12.6 counts per Chl. The PCF is shifting the energies of the Q band states of LHCII-L Chls 613a, 614a and of the unclustered Chls 615a, 608b to lower energies (-8 , -16 , -47 , -37 meV) making these states more accessible as acceptors. The Q states of Chl 602 are averagely

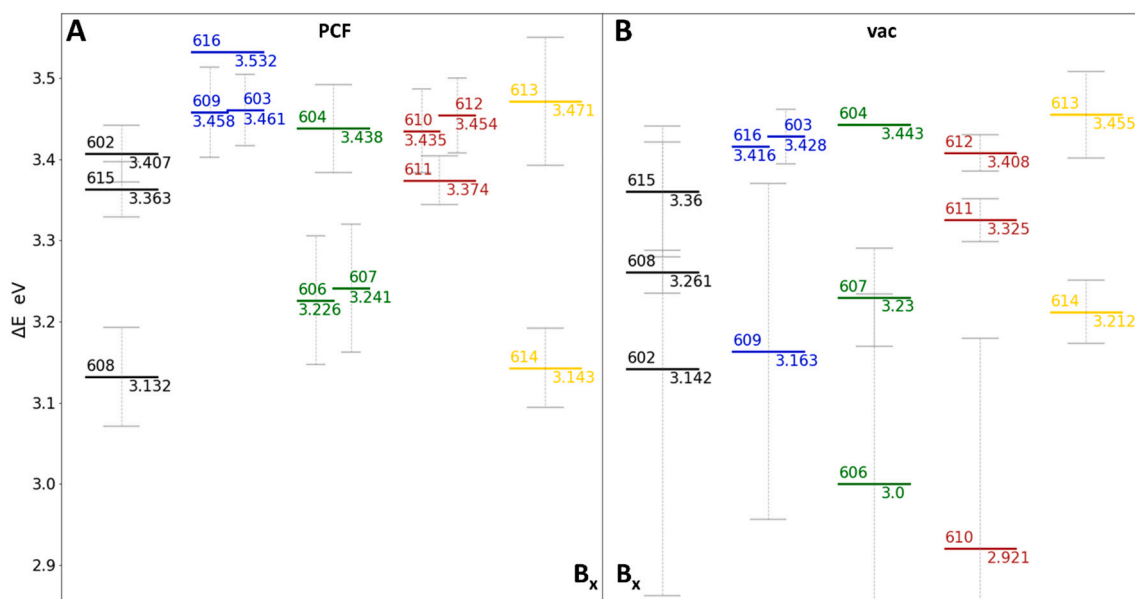


Fig. 4. Chl vertical excitation energies in the B_x region (third TD-DFT excited state) including PCF (A) and in gas phase (B) using the introduced colour code for the clusters (PSII-S: blue, PSII-L: yellow, LHCII-S: red, LHCII-L: green). The thick line of for each Chl/state corresponds to the average energy along the 100 ns MD simulation with their corresponding standard deviation as grey vertical range bar.

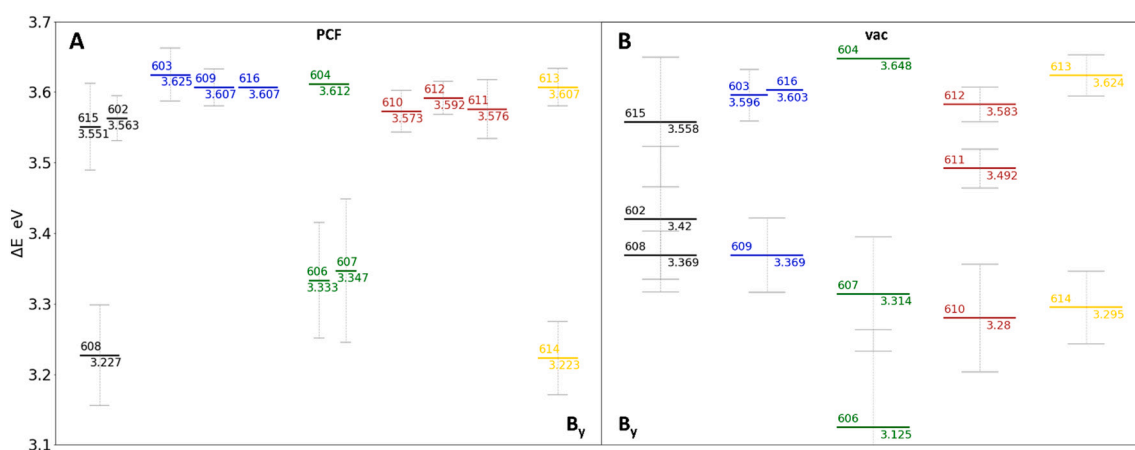


Fig. 5. Chl vertical excitation energies in the B_y region (fourth TD-DFT excited state) including PCF (A) and in gas phase (B) using the introduced colour code for the clusters (PSII-S: blue, PSII-L: yellow, LHCII-S: red, LHCII-L: green). The thick line of for each Chl/state corresponds to the average energy along the 100 ns MD simulation with their corresponding standard deviation as grey vertical range bar.

shifted by 21 meV to higher energies. Although the shifts on the unclustered Chls are inverse to each other the effect on their acceptor and donor role is similar: The donor roles of 602 and 608 are suppressed by -10 and -9 counts. The emission of 615 is suppressed by -13 counts, especially towards the Chl 609 of the PSII-S cluster by -6 counts.

The donor role of the PSII-S cluster is enhanced by 14 counts on average per Chl, while its acceptor role is suppressed by -12.3 counts on average per Chl. However, the picture of the individual Chls is different and just the donor role of Chl 603 is similarly enhanced/suppressed by $4/-7$ counts. The donor roles of Chls 609/616 are enhanced by $21/17$. At the same time, the acceptor role of the Chls 609/616 are suppressed by $-22/-12$.

In the PSII-L cluster, Chl 606 is most affected by the PCF. The donor role is enhanced by 16 counts and the acceptor role is suppressed by -23 counts. In the LHCII-S cluster, a similar situation applies for the Chl 610 with an enhanced donor role by 11 counts and a suppressed acceptor role by -12 counts.

Adding the Q_x - Q_y interaction to the Q_x matrix (see Fig. 7), only small

changes to the picture arising from Q_x alone can be identified. We find that some effects that were observed for the Q_y state interactions translate into the Q_x - Q_y state picture: Especially the increases and decreases in the acceptor role Chl 614 with various partners was already displayed for the Q_y state and is analogously found for Q_x - Q_y transitions. A new strongly controlled interaction set is found for Chl 613 as acceptor, which appears only when looking at Q_x - Q_y transitions, indicating that Chl 613 may be a primary Q_x acceptor for specific Chl residues.

3.3.3. Soret band (B_x state)

We analyze the Soret band analogously to the Q band. This assumes that the Soret band consists only of two relevant states (B_x and B_y); also, we do not consider B-Q energy transfer to not venture into the discussion which photophysical process is faster (internal conversion or energy transport). Considering only the B states, we assume that the corresponding states are close in energy and character to at least provide competition between these processes.

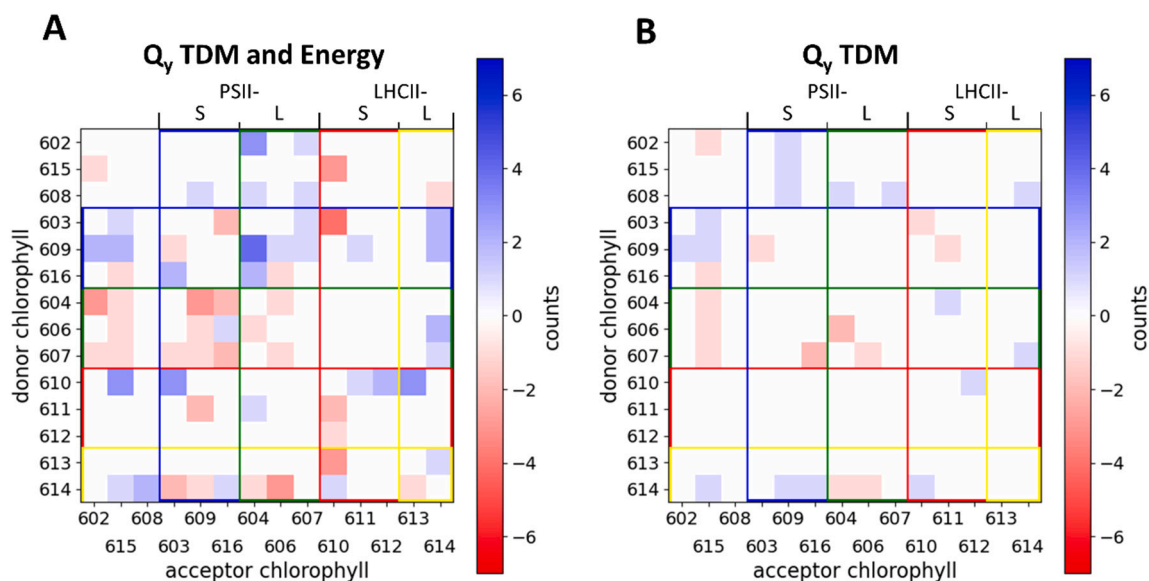


Fig. 6. PCF-vac. interaction matrix for the Q_y state. A: including both the TDM orientations and energy shifts; B: only accounting for changes in TDM orientation.

For the B_x state, the LHCII-L cluster's acceptor role is enhanced by 10.5 counts per Chl, while the donor role is suppressed with -11.5 counts per Chl. The donor role of the PSII-S cluster is enhanced by 7 counts on average per Chl. This includes 9 counts of the PSII-S-LHCII-L cluster interaction. However, this is mainly related to the Chl 609, whose donor role is enhanced by 12 counts including 6 counts of the Chl 609a-615a interaction. The LHCII-S cluster is similar and shows an increased donor role with 19 counts per Chl, especially towards the non-clustered Chls. This is related to a red shift of the B_x states of the LHCII-L cluster and a blue shift of the B_x states of Chl 602, 609, 610, 611, as Fig. 8 shows that energy shifts are required for these interactions to appear.

Within this framework, Chl 610 is the most affected by the PCF; its donor role is enhanced with 39 counts and its acceptor role is suppressed by -25 counts. Again, this is mainly due to the energetic position, as the energetic position is too low without PCF for the state to act as a donor. The donor role of the PSII-L cluster is suppressed by -8.3 counts on average per Chl. While the acceptor role of Chl 606 is suppressed with -21 counts and the acceptor role Chl 607 is equally enhanced/suppressed with $7/-15$ counts on average per Chl, the acceptor role of Chl 604 is solely enhanced by 13 counts including 6 counts of interaction with Chl 610. The non-clustered Chl 615a and 608b show an enhancement of their acceptor roles and a suppression of their donor roles. However, it is the opposite for Chl 602, which is related to a blue shift of the B_x state of Chl 602 and a red shift of Chl 615 and Chl 608.

Comparing Fig. 8A and B shows that the B_x is much more affected by the orientation of the TDMs than the Q band. E.g. like the Q_y state, Chl 610 is in parts controlled by the relative energetic position; its enhanced donor role only appears when taking the energy shift into account. But the orientation of the TDMs plays a significant role for Chl 610: While not accounting for the energies (see Fig. 8B), the effect of supporting Chl 610 as a donor turns into that of supporting Chl 610 as an acceptor; but the corresponding changes in TDM orientations remain the same regardless, as the pattern stays the same (see Fig. 8A). Furthermore, the enhancements/suppressions of the interactions of the PSII-L cluster Chls 606b and 607b are mainly defined by the orientation of the TDMs and not by the changes in relative energetic position.

Like in the Q_y state, the PSII-L cluster is suppressed to act as donor. In contrast to Q_y , however, the acceptor role of PSII-L for B_x is suppressed as well, making it less likely to accept B_x energy from any cluster. We have seen above that for the Q_y state, PSII-L was at least enhanced to accept energy from the non-clustered Chls and the PSII-S cluster. Furthermore, the suppression of the acceptor role of the PSII-L cluster is

not only related to an energetic shift rather to a change of the TDM orientation.

Generally, B_x energy is apparently steered to remain in the stromal side of CP29. There might be a slight preference of leaving the stromal side to the LHCII-L cluster, which already acted as a weakly supported acceptor in the Q band regime for all other clusters.

3.3.4. Soret band (B_y state and $B_{y/x}$ interactions)

The B_y state also shows some enhanced effect of the PCF within clusters, in contrast to the other discussed states. Furthermore, for the B_y state the clusters do not show uniform enhancement/suppression of a specific role. The donor roles of the Chls 602a, 609a, 610a and 606b are enhanced by the PCF by 23, 65, 34 and 11 counts, respectively. It must be mentioned that especially the donor role of Chl 609 is enhanced by 65 counts, which is different for the B_x state.

The acceptor role of the PSII-L cluster is more enhanced by 20.3 counts on average per Chl compared to the B_x state. The acceptor role of Chl 610 is suppressed by -34 counts. This is similar for the B_x and Q_x states, but contrary to the Q_y state. This can, again, be assigned to the massive shift in energy. Still, even though the actual role (donor/acceptor) of the Chl 610 B band is very much dependent on the PCF and the corresponding state energies, we find in Fig. 9D interactions that are suppressed via TDM orientation; most prominently Chl 604 B_y is completely TDM-suppressed when trying to donate to the Chl 610 B band, regardless of B_x or B_y .

Other small changes can be identified when also taking the B_y - B_x interaction into account. For example, Chl 615 shows an enhancement/suppression of the donor role with $7/-13$ counts and of the acceptor role with $36/-7$ counts, while the exclusion of B_x shows an enhancement/suppression of the donor role with $10/-9$ counts and of the acceptor role with $26/-3$ counts. Both states are red shifted by 67 meV. This makes Chl 615 more similar to 608 including the B_x interactions and more similar to 602 excluding these interactions. The PSII-S, PSII-L and LHCII-L cluster do not change qualitatively, except that the acceptor role of 606 (PSII-L cluster) is less enhanced and more suppressed, while the acceptor role of 607 is more enhanced. The behavior of 609 is reinforced with a suppression of its acceptor role with -38 counts and an enhancement of the donor role with 108 counts. Generally, the Chl 609 may be essential for B_y steering due to its donor role being strongly supported by the PCF.

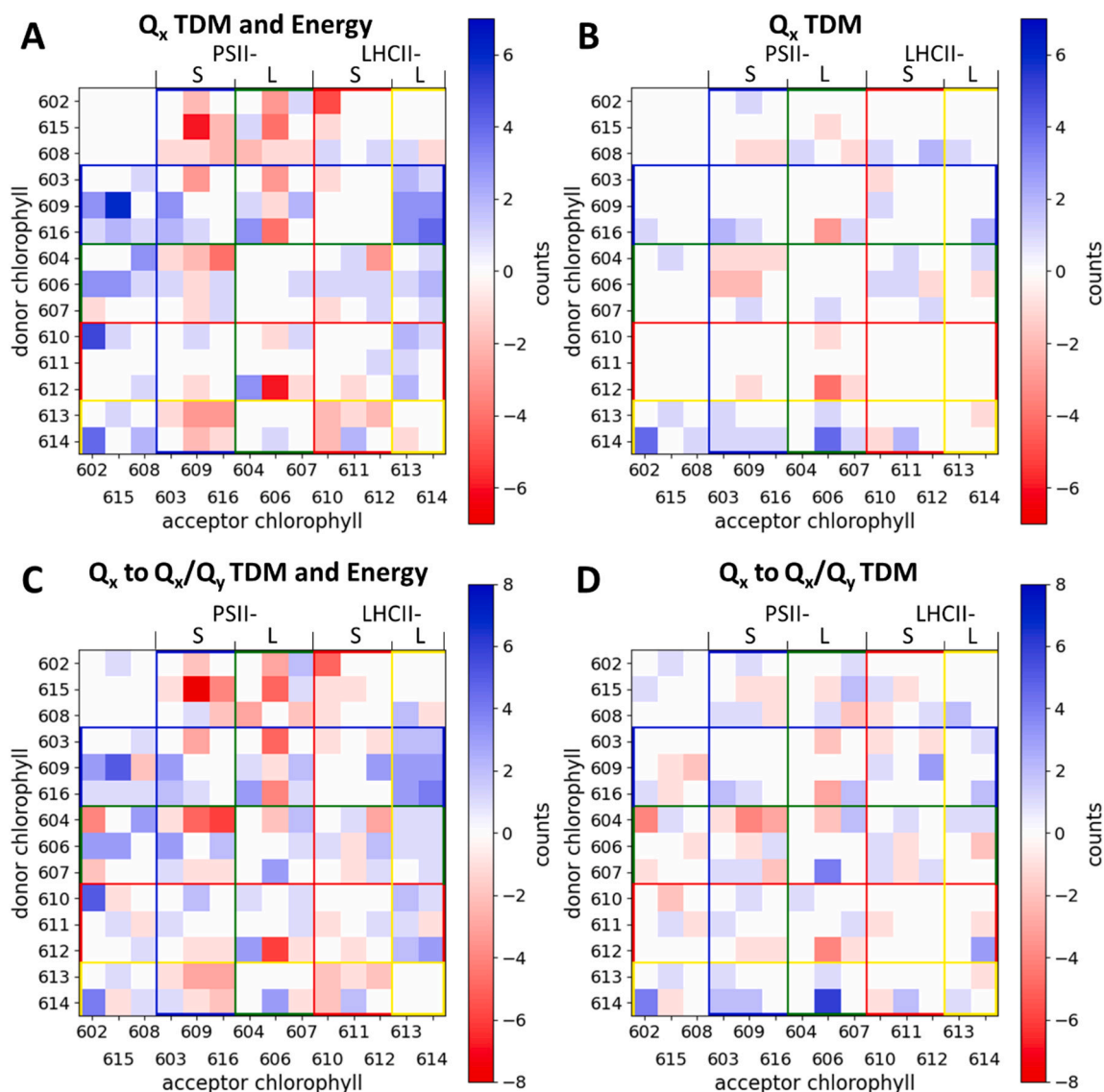


Fig. 7. PCF-vac. Interaction matrices of (A) Q_x - Q_x interactions only and (C) Q_x - Q_x/Q_y interactions. Schemes neglecting the energy shift (TDM only) also presented (B, D).

4. Conclusion

In this study, we have presented the effect of QM/MM optimization and the electrostatic environment on CP29 Chl spectra. We have employed a simplified Förster scheme to tackle the vast number of possible energy transitions when including also energetically higher states. It might be possible to use this scheme also for obtaining explicit coupling strengths, but our statistical basis in this work does not suffice for this application. By increasing the number of frames, the energy pathways should be clearly defined.

With the presented difference interaction matrices comparing the interaction in a protein matrix and in vacuum, we can identify whether interactions are suppressed or enhanced by the PCF and whether the role of a Chl or a cluster of Chls acting as donor or acceptor changes. For example, we have shown that the luminal LHCII-oriented Chls (LHCII-L cluster above) mainly act as acceptors for any kind of state, and most Chls in the system; most prominently however for the Soret band energy. Primary donor Chls (or rather, Chls for which the PCF shifts the role more to being donors) were found to be, for Q_y/Q_x , the stromal PSII-oriented Chls. While the neighboring, stromal, ungrouped Chls have shown a donor character for the Q_y , they act as acceptors for Q_x . At first

glance, this seems to be counterintuitive at first w.r.t. energy transfer to the PSII core complex from the stromal Chls of CP29. It becomes however much less problematic when realizing that their preferred acceptor partners are simply other stromal Chls, even more so than the LHCII-L cluster. We find that the presence of the PCF preferentially supports the luminal PSII-oriented Chls as acceptors for Q_y (and B_x) energy. Yet for the Q band in general, the stromal Chls seem to be the preferred transfer target.

This picture is somewhat different for the Soret band. While our data may be affected by the presence of potentially artificial intramolecular CT states, we can still see that the donation between stromal Chls is still given, although to a weaker extent. For Soret, however, the mentioned LHCII-L cluster becomes the primary acceptor, spatially distant from both stromal Chls and PSII. Since it is known that Soret excitations reduce the life time of the PSII core complex [39], removing these interactions from the stromal Chl ring appears beneficial.

This may reveal the evolutionary origins and the functional relevance of the CP29 acting as “conduit”. To confidently state this, we currently investigate the interactions with the present carotenoids as well as the inter complex interaction. Furthermore, the analysis of the protein matrix represented by the PCF showed that the absorption

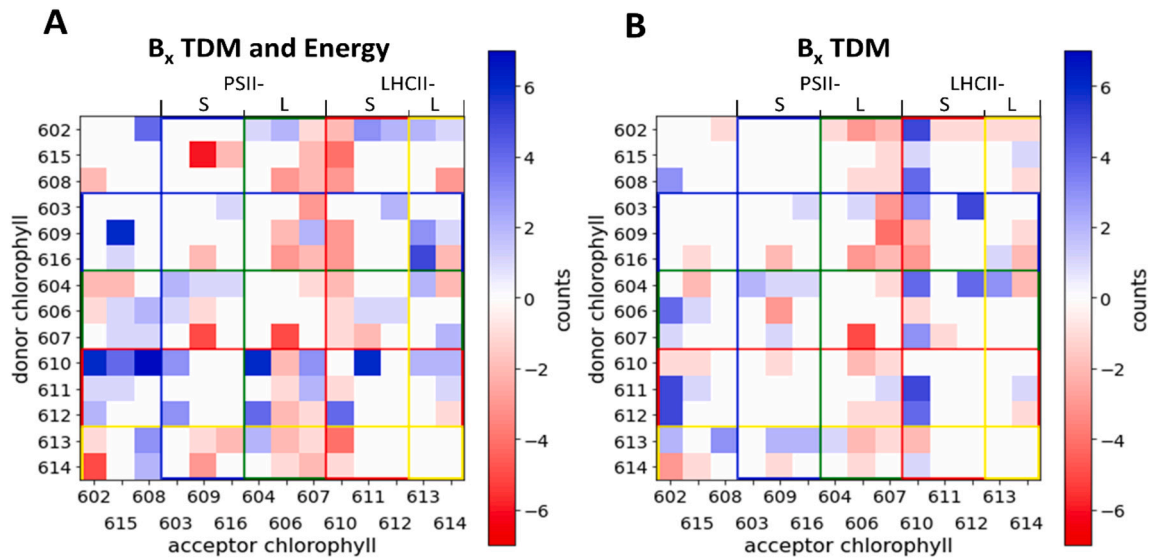


Fig. 8. PCF-vac. interaction matrix for the B_x state. A: including both the TDM orientations and energy shifts; B: only accounting for changes in TDM orientation.

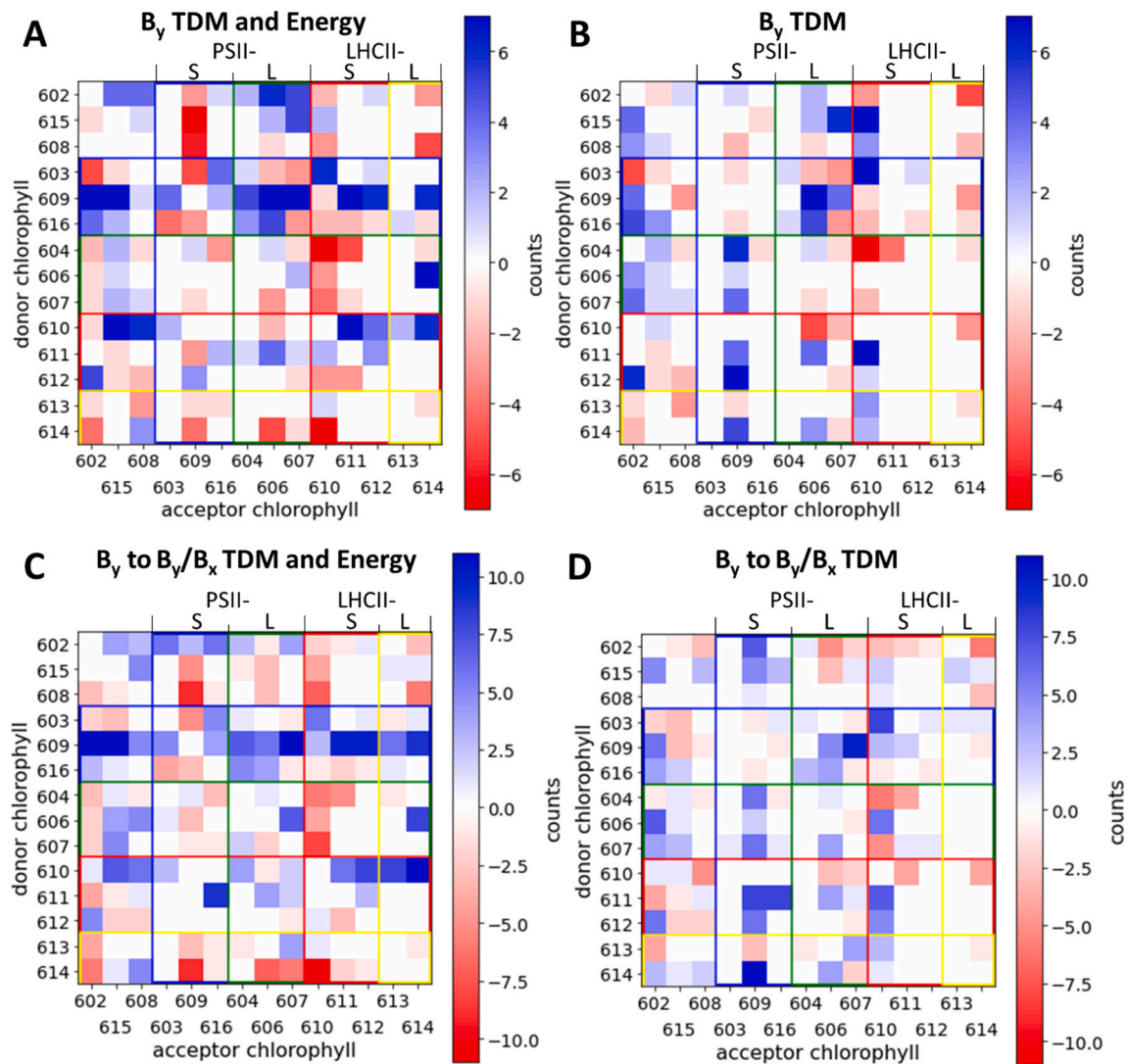


Fig. 9. PCF-vac. Interaction matrices of (A) B_y-B_y interactions only and (C) B_y-B_{y/x} interactions. Schemes neglecting the energy shift (TDM only) also presented (B, D).

spectrum of all Chls is unaffected during the MD simulation, while the individual Chl spectra are shifted depending on the coordinating molecule. We identify the main influence for the couplings is indeed to be found in the energetic position of the individual chromophores depending on the presence of the PCF, but the fine-tuning especially for the Soret states is controlled by the changes in the TDMs.

Declaration of competing interest

Jan Philipp Götz reports financial support was provided by German Research Foundation.

Acknowledgements

The authors gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft, project “Blue light pathways in light harvesting complexes”, project no. 393271229.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbabo.2021.148521>.

References

- R. Croce, H. Van Amerongen, Light-harvesting in photosystem I, *Photosynth. Res.* 116 (2013) 153–166, <https://doi.org/10.1007/s11120-013-9838-x>.
- V. Chukhutsina, X. Liu, P. Xu, R. Croce, Light-harvesting complex II is an antenna of photosystem I in dark-adapted plants, *Nat. Plants* 6 (2020) 1–9, <https://doi.org/10.1038/s41477-020-0693-4>.
- V. Mascoli, V. Novoderezhkin, N. Liguori, P. Xu, R. Croce, Design principles of solar light harvesting in plants: functional architecture of the monomeric antenna CP29, *Biochim. Biophys. Acta - Bioenerg.* 2020 (1861) 1–13, <https://doi.org/10.1016/j.bbabo.2020.148156>.
- J. Nugent, M. Evans, Structure of biological solar energy converters - further revelations, *Trends Plant Sci.* 9 (2004) 368–370, <https://doi.org/10.1016/j.tplants.2004.06.005>.
- E.E. Ostroumov, J.P. Götz, M. Reus, P.H. Lambrev, A.R. Holzwarth, Characterization of fluorescent chlorophyll charge-transfer states as intermediates in the excited state quenching of light-harvesting complex II, *Photosynth. Res.* 144 (2020) 171–193, <https://doi.org/10.1007/s11120-020-00745-8>.
- A.N. Glazer, Comparative biochemistry of photosynthetic light-harvesting systems, *Annu. Rev. Biochem.* 52 (1983) 125–157, <https://doi.org/10.1146/annurev.bi.52.070183.001013>.
- J.-D. Rochaix, Regulation and dynamics of the light-harvesting system, *Annu. Rev. Plant Biol.* 65 (2014) 287–309, <https://doi.org/10.1146/annurev-arplant-050213-040226>.
- M. Lapillo, E. Cignoni, L. Cupellini, B. Mennucci, The energy transfer model of nonphotochemical quenching: lessons from the minor CP29 antenna complex of plants, *Biochim. Biophys. Acta - Bioenerg.* 1861 (2020), 148282, <https://doi.org/10.1016/j.bbabo.2020.148282>.
- E.L. Camm, B.R. Green, Fractionation of thylakoid membranes with the nonionic detergent octyl- β -d-glucopyranoside: resolution of chlorophyll-protein complex II into two chlorophyll-protein complexes, *Plant Physiol.* 66 (1980) 428–432, <https://doi.org/10.1104/pp.66.3.428>.
- S. Caffarri, R. Croce, L. Cattivelli, R. Bassi, A look within LHCII: differential analysis of the Lhcb1–3 complexes building the major trimeric antenna complex of higher-plant photosynthesis, *Biochemistry* 43 (2004) 9467–9476, <https://doi.org/10.1021/bi036265i>.
- R.E. Blankenship, Early evolution of photosynthesis, *Plant Physiol.* 154 (2010) 434–438, <https://doi.org/10.1104/pp.110.161687>.
- S. Jurinovich, L. Viani, I.G. Prandi, T. Renger, B. Mennucci, Towards an ab initio description of the optical spectra of light-harvesting antennae: application to the CP29 complex of photosystem II, *Phys. Chem. Chem. Phys.* 17 (2015) 14405–14416, <https://doi.org/10.1039/C4CP05647G>.
- P. Sherwood, A.H. de Vries, M.F. Guest, G. Schreckenbach, C.R.A. Catlow, S. A. French, A.A. Sokol, S.T. Bromley, W. Thiel, A.J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S.C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis, *J. Mol. Struct. THEOCHEM* 632 (2003) 1–28, [https://doi.org/10.1016/S0166-1280\(03\)00285-9](https://doi.org/10.1016/S0166-1280(03)00285-9).
- J.P. Götz, Y.-W. Pi, S. Petry, F. Langkabel, J.F. Witte, O. Lemke, A user-friendly, Python-based quantum mechanics/Gromacs interface: gmx2qmmm, *Int. J. Quantum Chem.* 121 (2021), e26486, <https://doi.org/10.1002/qua.26486>.
- P.C. Hariharan, J.A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, *Theor. Chim. Acta* 28 (1973) 213–222, <https://doi.org/10.1007/BF00533485>.
- T. Yanai, D.P. Tew, N.C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP), *Chem. Phys. Lett.* 393 (2004) 51–57, <https://doi.org/10.1016/j.cplett.2004.06.011>.
- A. Graczyk, J.M. Żurek, M.J. Paterson, On the linear and non-linear electronic spectroscopy of chlorophylls: a computational study, *Photochem. Photobiol. Sci.* 13 (2014) 103–111, <https://doi.org/10.1039/C3PP50262G>.
- D. Kröner, J.P. Götz, Modeling of a violaxanthin–chlorophyll b chromophore pair in its LHCII environment using CAM-B3LYP, *J. Photochem. Photobiol. B Biol.* 109 (2012) 12–19, <https://doi.org/10.1016/j.jphotochem.2011.12.007>.
- J.P. Götz, D. Kröner, S. Banerjee, B. Karasulu, W. Thiel, Carotenoids as a shortcut for chlorophyll Soret-to-Q band energy flow, *ChemPhysChem* 15 (2014) 3392–3401, <https://doi.org/10.1002/cphc.201402233>.
- O. Lemke, J.P. Götz, On the stability of the water-soluble chlorophyll-binding protein (WSCP) studied by molecular dynamics simulations, *J. Phys. Chem. B* 123 (2019) 10594–10604, <https://doi.org/10.1021/acs.jpcc.9b07915>.
- X. Su, J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang, M. Li, Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex, *Science* 357 (2017) 815–820, <https://doi.org/10.1126/science.aan0327>.
- F.J. Van Eerden, M.N. Melo, P.W.J.M. Frederix, S.J. Marrink, Prediction of thylakoid lipid binding sites on photosystem II, *Biophys. J.* 113 (2017) 2669–2681, <https://doi.org/10.1016/j.bpj.2017.09.039>.
- H.I. Petrache, S. Tristram-Nagle, K. Gawrisch, D. Harries, V.A. Parsegian, J. F. Nagle, Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt, *Biophys. J.* 86 (2004) 1574–1586, [https://doi.org/10.1016/S0006-3495\(04\)74225-3](https://doi.org/10.1016/S0006-3495(04)74225-3).
- M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.
- A.E. Aliev, M. Kulke, H.S. Khaneja, V. Chudasama, T.D. Sheppard, R.M. Lanigan, Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics, *Proteins Struct. Funct. Bioinforma.* 82 (2014) 195–215, <https://doi.org/10.1002/prot.24350>.
- Å.A. Skjerve, B.D. Madej, R.C. Walker, K. Teigen, LIPID11: a modular framework for lipid simulations using Amber, *J. Phys. Chem. B* 116 (2012) 11124–11136, <https://doi.org/10.1021/jp3059992>.
- S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.* 29 (2008) 1859–1865, <https://doi.org/10.1002/jcc.20945>.
- L. Zhang, D.-A. Silva, Y. Yan, X. Huang, Force field development for cofactors in the photosystem II, *J. Comput. Chem.* 33 (2012) 1969–1980, <https://doi.org/10.1002/jcc.23016>.
- M. Ceccarelli, P. Procacci, M. Marchi, An ab initio force field for the cofactors of bacterial photosynthesis, *J. Comput. Chem.* 24 (2003) 129–142, <https://doi.org/10.1002/jcc.10198>.
- W.F. Van Gunsteren, H.J.C. Berendsen, A leap-frog algorithm for stochastic dynamics, *Mol. Simul.* 1 (1988) 173–185, <https://doi.org/10.1080/08927028808080941>.
- W.L. Jorgensen, Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water, *J. Am. Chem. Soc.* 103 (1981) 335–340, <https://doi.org/10.1021/ja00392a016>.
- M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J. V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian16 {R} edition {C}.01, 2016.
- E. Runge, E.K.U. Gross, Density-functional theory for time-dependent systems, *Phys. Rev. Lett.* 52 (1984) 997–1000, <https://doi.org/10.1103/PhysRevLett.52.997>.
- O. Kühn, S. Lochbrunner, *Quantum Dynamics and Spectroscopy of Excitons in Molecular Aggregates*, 2012.
- V. Sláma, L. Cupellini, B. Mennucci, Exciton properties and optical spectra of light harvesting complex II from a fully atomistic description, *Phys. Chem. Chem. Phys.* 22 (2020) 16783–16795, <https://doi.org/10.1039/D0CP02492A>.
- B. Voigt, K.-D. Irrgang, J. Ehlert, W. Beenken, G. Renger, D. Leupold, H. Lokstein, Spectral substructure and excitonic interactions in the minor photosystem II antenna complex CP29 as revealed by nonlinear polarization spectroscopy in the frequency domain, *Biochemistry* 41 (2002) 3049–3056, <https://doi.org/10.1021/bi0155854>.
- S. Maity, P. Sarngadharan, V. Daskalakis, U. Kleinekathöfer, Time-dependent Atomic Simulations of the CP29 Light-harvesting Complex, 2021.
- J.P. Götz, B. Karasulu, W. Thiel, Computing UV/vis spectra from the adiabatic and vertical Franck-Condon schemes with the use of Cartesian and internal coordinates, *J. Chem. Phys.* 139 (2013), 234108, <https://doi.org/10.1063/1.4844055>.
- M.A.K. Jansen, A.K. Mattoo, M. Edelman, D1–D2 protein degradation in the chloroplast, *Eur. J. Biochem.* 260 (1999) 527–532, <https://doi.org/10.1046/j.1432-1327.1999.00196.x>.

Supporting Information (SI)

Effect of protein matrix on CP29 spectra and energy transfer pathways

S. Petry, J. P. Götze

Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin,
Germany

This supporting information is divided in 3 sections. Section 1 contains the magnesium coordinating residue that is part of the QM layer along the 100 ns MD simulation, section 2 describes the used method to calculate the shifts of the optical spectra and the resulting shifts of sum spectra of the individual Chls and section 3 is dedicated to tables containing the vertical excitation energies for the first five excited states and their standard deviations.

Coordination:

Table 1: Coordinating residues

Optical Spectra:

Figure 1: BFGS shifts

Table 2: Calculated shifts

Energies:

Table 3 Table 6: Avg. energy levels

Coordination

Table 1: Coordinating residues within the QM-region for the investigated snapshots along the MD simulation, whereas the abbreviation BB stands for backbone and MEM for membrane. The order of the Chls is similar to the introduced clustering.

	a602	a603	a609	a610	a611	a612	a613	a615	a604	b606	b607	b608	b614
50 /ns	GLU	HIS	GLU	GLU	LHG	HIS	none	BB	SOL	GLU	-	SOL	SOL
55 /ns	GLU	HIS	GLU	GLU	LHG	HIS	none	BB	SOL	GLU	-	SOL	-
60 /ns	GLU	HIS	GLU	GLU	LHG	HIS	none	BB	SOL	GLU	MEM	SOL	SOL
65 /ns	GLU	HIS	GLU	GLU	LHG	HIS	none	BB	SOL	GLU	SOL MEM	SOL	SOL
70 /ns	GLU (ARG)	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	MEM	SOL	SOL
75 /ns	GLU (ARG)	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	MEM	SOL	SOL
80 /ns	GLU (ARG)	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	MEM	SOL	SOL
85 /ns	GLU (ARG)	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	SOL MEM	SOL	SOL
90 /ns	GLU	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	SOL MEM	SOL	SOL
95 /ns	GLU	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	SOL MEM	SOL	SOL
100 /ns	GLU (ARG)	HIS	GLU	GLU	LHG	HIS	GLN	BB	SOL	GLU	-	SOL	-

Optical spectra

Like indicated in the main article, vertical absorption spectra were simulated using TD-DFT calculations, once including and once excluding the PCF. Since the individual bands consist of more than one state, it is not trivial to assess the actual shift between gas phase and PCF for the Q and Soret regions as a whole. To quantify the shifts of the Q and Soret band, we employed the quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS).[1–4] The method was used to shift the computed bands for the two cases (gas/PCF) onto each other resulting in a shift value for each band. These values are selectively presented in the main article, e.g. Chl 604a shows a red shift going from vacuum to PCF of -6/-36 meV for the Q/Soret band, in the context of shifts resulting from different Chl ligands. As the optimization criterion, the root mean square error of the difference of the intensities was minimized. Find below exemplary plots of the (histidine coordinated) Chl 603 calculated spectrum embedded in the protein matrix (PCF), in the gas phase (vac) and the shifted vac spectrum using the BFGS method (see Figure 1).

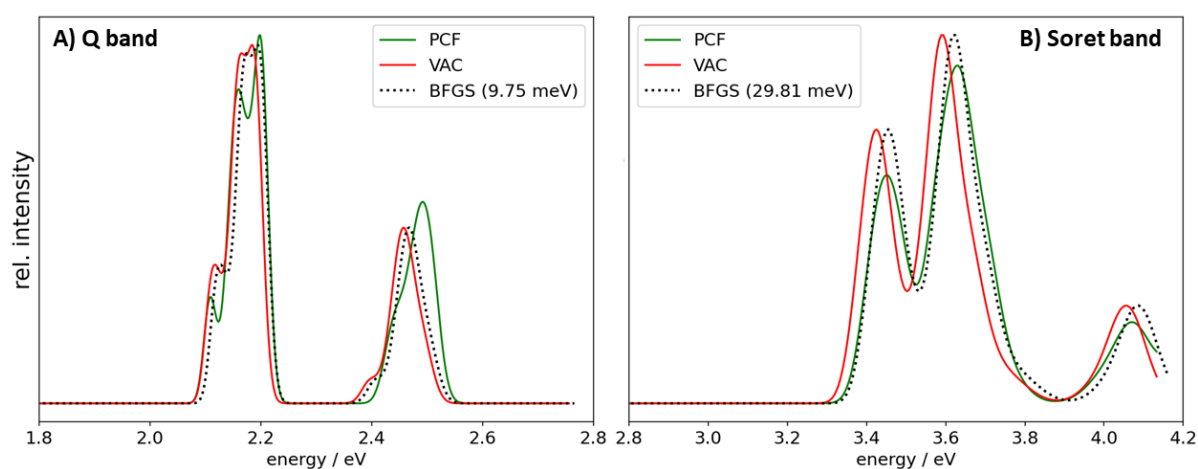


Figure 1: Sum spectra of Chl a 603 calculated in a PCF (green), in the gas phase (red) and the shifted vac spectra using the BFGS method (black dotted) for the Q band (A) and for the Soret band (B).

Table 2: Calculated shifts for the Q and the Soret band using the BFGS method.

Chl	Q band shift / meV	Soret band shift / meV
a602	14.95	61.53
a603	9.75	29.81
a604	-6.03	-36.38
a609	12.8	111.74
a610	31.79	153.63
a611	6.38	62.97
a612	11.76	6.81
a613	-19.22	-23.18
a615	-10.11	-15.78
b606	43.87	42.88
b607	-5.17	52.02
b608	-22.34	-158.88
b614	-6.55	-49.83

Energies

Table 3: Average energies levels (eV) along the MD trajectory computed in the PCF for the first five excited states.

a602	a603	a604	a609	a610	a611	a612	a613	a615	a616	b606	b607	b608	b614
2.158	2.170	2.176	2.190	2.172	2.145	2.187	2.168	2.074	2.209	2.262	2.237	2.253	2.228
2.422	2.473	2.508	2.478	2.454	2.439	2.494	2.469	2.379	2.503	2.552	2.520	2.524	2.485
3.407	3.453	3.405	3.420	3.413	3.374	3.409	3.405	3.257	3.420	3.215	3.241	3.132	3.143
3.549	3.584	3.587	3.542	3.554	3.576	3.556	3.577	3.437	3.532	3.326	3.347	3.227	3.223
3.645	3.644	3.664	3.622	3.603	3.675	3.629	3.649	3.591	3.607	3.632	3.548	3.529	3.574

Table 4: Standard deviation of the energies levels (eV) along the MD trajectory computed in the PCF for the first five excited states. ¹ Only one frame.

a602	a603	a604	a609	a610	a611	a612	a613	a615	a616 ¹	b606	b607	b608	b614
0.025	0.029	0.009	0.027	0.020	0.016	0.020	0.029	0.140	n/a	0.028	0.035	0.017	0.034
0.026	0.028	0.023	0.028	0.020	0.022	0.028	0.023	0.134	n/a	0.038	0.053	0.033	0.046
0.035	0.031	0.018	0.039	0.019	0.030	0.028	0.031	0.335	n/a	0.052	0.079	0.061	0.049
0.019	0.050	0.034	0.048	0.031	0.042	0.044	0.037	0.345	n/a	0.067	0.101	0.071	0.052
0.048	0.035	0.059	0.029	0.035	0.045	0.060	0.035	0.191	n/a	0.062	0.063	0.072	0.082

Table 5: Average energies levels (eV) along the MD trajectory computed in the gas phase for the first five excited states.

a602	a603	a604	a609	a610	a611	a612	a613	a615	a616	b606	b607	b608	b614
2.153	2.163	2.183	2.169	2.139	2.139	2.176	2.174	2.113	2.192	2.228	2.238	2.271	2.237
2.386	2.453	2.519	2.401	2.353	2.381	2.477	2.478	2.434	2.481	2.439	2.503	2.579	2.507
3.063	3.428	3.443	2.931	2.724	3.3325	3.408	3.434	3.325	3.416	2.813	3.193	3.261	3.212
3.342	3.593	3.640	3.290	3.159	3.492	3.583	3.610	3.502	3.603	3.090	3.306	3.369	3.295
3.433	3.672	3.703	3.369	3.280	3.657	3.684	3.662	3.577	3.623	3.158	3.544	3.625	3.590

Table 6: Standard deviation of the energies levels (eV) along the MD trajectory computed in the gas phase for the first five excited states. ¹ Only one frame.

a602	a603	a604	a609	a610	a611	a612	a613	a615	a616 ¹	b606	b607	b608	b614
0.019	0.027	0.008	0.021	0.019	0.016	0.017	0.020	0.042	n/a	0.018	0.030	0.015	0.024
0.049	0.026	0.025	0.021	0.029	0.017	0.026	0.027	0.057	n/a	0.075	0.053	0.033	0.055
0.268	0.033	0.017	0.061	0.102	0.027	0.022	0.020	0.161	n/a	0.246	0.120	0.026	0.039
0.159	0.032	0.021	0.050	0.068	0.028	0.025	0.031	0.143	n/a	0.139	0.080	0.034	0.052
0.119	0.047	0.045	0.053	0.077	0.038	0.069	0.022	0.114	n/a	0.128	0.103	0.047	0.074

References

- [1] C.G. BROYDEN, The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations, *IMA J. Appl. Math.* 6 (1970) 76–90.
<https://doi.org/10.1093/imamat/6.1.76>.
- [2] R. Fletcher, A new approach to variable metric algorithms, *Comput. J.* 13 (1970) 317–322.
<https://doi.org/10.1093/comjnl/13.3.317>.
- [3] D. Goldfarb, A family of variable-metric methods derived by variational means, *Math. Comp.* 24 (1970) 23–26.
- [4] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comp.* 24 (1970) 647–656.

4.4 Paper 4

“Impact of structural sampling, coupling scheme and state of interest on the energy transfer in CP29”

S. Petry, J. C. Tremblay, J. P. Götze

BioRxiv preprint server **2023**

DOI: 10.1101/2023.01.25.525376

URL: <https://doi.org/10.1101/2023.01.25.525376>

Contributions

The project was conceived by Simon Petry with the help of Jan Götze. Simon Petry designed and performed all QM/MM and QM calculations and wrote the manuscript. Furthermore, Simon Petry implemented the TDC method with the help of Jean Christophe Tremblay. The evaluation and interpretation of the results were carried out by Simon Petry. All authors contributed to the final version of the manuscript.

References

- [1] P. G. Falkowski, J. A. Raven, *Aquatic Photosynthesis*, Princeton University Press, **2007**.
- [2] P. Sánchez-Baracaldo, T. Cardona, *New Phytol.* **2020**, *225*(4), 1440–1446.
- [3] G. C. Dismukes, R. E. Blankenship, *The Origin and Evolution of Photosynthetic Oxygen Production*, p. 683–695, Springer Netherlands, Dordrecht, **2005**.
- [4] D. J. Vinyard, G. M. Ananyev, G. C. Dismukes, *Annu. Rev. Biochem.* **2013**, *82*(1), 577–606.
- [5] R. E. Blankenship, *Photosynthetic Organisms and Organelles*, Blackwell Science, Oxford, **2002**, chapter 2, p. 11–25.
- [6] J. Priestley, *Experiments and Observations on Different Kinds of Air: Volume 1: The Second Edition, Corrected*, vol. 1 of *Cambridge Library Collection - Physical Sciences*, Cambridge University Press, **2014**.
- [7] J. Priestley, *Experiments and Observations on Different Kinds of Air: Volume 2: The Second Edition*, vol. 2 of *Cambridge Library Collection - Physical Sciences*, Cambridge University Press, **2014**.
- [8] J. Priestley, *Experiments and Observations on Different Kinds of Air: Volume 3*, vol. 3 of *Cambridge Library Collection - Physical Sciences*, Cambridge University Press, **2014**.
- [9] R. E. Blankenship, D. M. Tiede, J. Barber, G. W. Brudvig, G. Fleming, M. Ghirardi, M. R. Gunner, W. Junge, D. M. Kramer, A. Melis, T. A. Moore, C. C. Moser, D. G. Nocera, A. J. Nozik, D. R. Ort, W. W. Parson, R. C. Prince, R. T. Sayre, *Science* **2011**, *332*(6031), 805–809.
- [10] M. I. Hoffert, K. Caldeira, G. Benford, D. R. Criswell, C. Green, H. Herzog, A. K. Jain, H. S. Kheshgi, K. S. Lackner, J. S. Lewis, H. D. Lightfoot, W. Manheimer, J. C. Mankins, M. E. Mauel, L. J. Perkins, M. E. Schlesinger, T. Volk, T. M. L. Wigley, *Science* **2002**, *298*(5595), 981–987.
- [11] e. European Union, „Gross available energy, EU, 1990-2020“, online data code: nrg_bal_s, **2023** [last access **24.03.2023 21:25**].
- [12] S. Ruben, M. Randall, M. Kamen, J. L. Hyde, *J. Am. Chem. Soc.* **1941**, *63*(3), 877–879.
- [13] D. Voet, J. G. Voet, C. W. Pratt, *Lehrbuch der Biochemie*, Wiley-VCH, **2002**.
- [14] B. E. S. Gunning, O. M. Schwartz, *Funct. Plant Biol.* **1999**, *26*(7), 695–708.
- [15] C. W. Mullineaux, *Trends Plant Sci.* **2005**, *10*(11), 521–525.
- [16] „Chapter 19 - Mitochondria, Chloroplasts, Peroxisomes“ in *Cell Biology (Third Edition)* (Ed.: T. D. Pollard, W. C. Earnshaw, J. Lippincott-Schwartz, G. T. Johnson), 3. ed., Elsevier, **2017**, p. 317–329.

- [17] D. Mauzerall, A. Chivvis, *J. Theor. Biol.* **1973**, *42*(2), 387–395.
- [18] L. O. Björn, G. C. Papageorgiou, R. E. Blankenship, Govindjee, *Photosynth. Res.* **2009**, *99*(2), 85–98.
- [19] A. Stirbet, D. Lazár, Y. Guo, G. Govindjee, *Ann. Bot.* **2019**, *126*(4), 511–537.
- [20] R. Emerson, W. Arnold, *J. Gen. Physiol.* **1932**, *16*(2), 191–205.
- [21] N. Nelson, W. Junge, *Annu. Rev. Biochem.* **2015**, *84*, 659–683.
- [22] A. Klauss, M. Haumann, H. Dau, *Proc. Natl. Acad. Sci.* **2012**, *109*(40), 16035–16040.
- [23] M. Haumann, P. Liebisch, C. Müller, M. Barra, M. Grabolle, H. Dau, *Science* **2005**, *310*(5750), 1019–1021.
- [24] B. Forbush, B. Kok, M. P. McGloin, *Photochem. Photobiol.* **1971**, *14*(3), 307–321.
- [25] L. V. Kulik, B. Epel, W. Lubitz, J. Messinger, *J. Am. Chem. Soc.* **2007**, *129*(44), 13421–13435.
- [26] V. Krewald, M. Retegan, D. A. Pantazis, *Top. Curr. Chem.* **2016**, *371*, 23–48.
- [27] K. Åhrling, S. Peterson, S. Styring, *Biochemistry* **1997**, *36*(43), 13148–13152.
- [28] J. Messinger, J. H. A. Nugent, M. C. W. Evans, *Biochemistry* **1997**, *36*(37), 11055–11060.
- [29] J. Messinger, J. H. Robblee, W. O. Yu, K. Sauer, V. K. Yachandra, M. P. Klein, *Biochemistry* **1997**, *119*(46), 11349–11350.
- [30] V. Krewald, M. Retegan, F. Neese, W. Lubitz, D. Pantazis, N. Cox, *Inorg. Chem.* **2016**, *55*(2), 488–501.
- [31] C. A. Wraight, M. R. Gunner, *The Acceptor Quinones of Purple Photosynthetic Bacteria — Structure and Spectroscopy*, p. 379–405, Springer Netherlands, Dordrecht, **2009**.
- [32] F. Müh, C. Glöckner, J. Hellmich, A. Zouni, *Biochim. Biophys. Acta Bioenerg.* **2012**, *1817*(1), 44–65, photosystem II.
- [33] K. Asada, *Plant Physiol.* **2006**, *141*(2), 391–396.
- [34] J. P. Whitelegge, H. Zhang, R. Aguilera, R. M. Taylor, W. A. Cramer, *Mol. Cell. Proteomics* **2002**, *1*(10), 816–827.
- [35] L. A. Malone, P. Qian, G. E. Mayneord, A. Hitchcock, D. A. Farmer, R. F. Thompson, D. J. K. Swainsbury, N. A. Ranson, C. N. Hunter, M. P. Johnson, *Nature* **2019**, *575*(7783), 535–539.
- [36] I. Caspy, M. Fadeeva, S. Kuhlert, A. Borovikova-Sheinker, D. Klaiman, G. Masrati, F. Drepper, N. Ben-Tal, M. Hippler, N. Nelson, *Biochem. J.* **2021**, *478*(12), 2371–2384.

- [37] G. Hanke, P. Mulo, *Plant Cell Environ.* **2013**, *36*(6), 1071–1084.
- [38] P. Marco, M. Kozuleva, H. Eilenberg, Y. Mazor, P. Gimeson, A. Kanygin, K. Redding, I. Weiner, I. Yacoby, *Biochim. Biophys. Acta Bioenerg.* **2018**, *1859*(4), 234–243.
- [39] Y. Lima-Melo, M. Kilic, E.-M. Aro, P. J. Gollan, *Front. Plant Sci.* **2021**, *12*.
- [40] P. Neupane, S. Bhujju, N. Thapa, H. K. Bhattarai, *Biomol. Concepts* **2019**, *10*(1), 1–10.
- [41] M. Calvin, *J. Chem. Educ.* **1949**, *26*(12), 639–657.
- [42] T. D. Sharkey, *Photosynth. Res.* **2019**, *140*(2), 235–252.
- [43] M. E. A. Carvalho, C. A. Labate, C. Barboza da Silva, P. R. de Camargo e Castro, R. A. Azevedo, *Plant Growth Regul.* **2022**, *97*(3), 477–484.
- [44] W. Arnold, H. I. Kohn, *J. Gen. Physiol.* **1934**, *18*(1), 109–112.
- [45] O. Warburg, E. Negelein, *Über den Einfluß der Wellenlänge auf den Energieumsatz bei der Kohlensäureassimilation*, p. 444–469, Springer Berlin Heidelberg, Berlin, Heidelberg, **1928**.
- [46] H. I. Kohn, *Nature* **1936**, *137*, 706.
- [47] A. M. Gilmore, T. L. Hazlett, P. G. Debrunner, *Photosynth. Res.* **1996**, *48*, 171–187.
- [48] K. Wohl, *New Phytol.* **1940**, *39*(1), 33–64.
- [49] G. H. Schmid, H. Gaffron, *J. Gen. Physiol.* **1968**, *52*(2), 212–239.
- [50] D. Mauzerall, N. L. Greenbaum, *Biochim. Biophys. Acta Bioenerg.* **1989**, *974*(2), 119–140.
- [51] R. MacColl, D. S. Berns, *Photochem. Photobiol.* **1978**, *27*(3), 343–349.
- [52] P. Zoeë, N. Sangeeta, S. Richard T., *Algal Res.* **2012**, *1*(2), 134–142.
- [53] J. Shively, G. Cannon, S. Heinhorst, J. Fuerst, D. Bryant, E. Gantt, J. Maupin-Furlow, D. Schüler, F. Pfeifer, R. Docampo, C. Dahl, J. Preiss, A. Steinbüchel, B. Federici, *Intracellular Structures of Prokaryotes: Inclusions, Compartments and Assemblages*, third edition ed., Academic Press, Oxford, **2009**, p. 404–424.
- [54] M. Gouterman, *J. Mol. Spectrosc.* **1961**, *6*(C), 138–163.
- [55] M. GOUTERMAN, „1 - Optical Spectra and Electronic Structure of Porphyrins and Related Rings“ in *The Porphyrins* (Ed.: D. DOLPHIN), Academic Press, **1978**, p. 1–165.
- [56] M. Chen, *Annu. Rev. Biochem.* **2014**, *83*, 317–340.
- [57] M. Gouterman, *J. Chem. Phys.* **1959**, *30*(5), 1139.

- [58] C. Büchel, *Biochim. Biophys. Acta Bioenerg.* **2020**, 1861(4).
- [59] S. W. Jeffrey, *J. Phycol.* **1976**, 12(3), 349–354.
- [60] B. Grabowski, F. X. Cunningham, E. Gantt, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98(5), 2911–2916.
- [61] W. M. Manning, H. H. Strain, *J. Biol. Chem.* **1943**, 151(1), 1–19.
- [62] F. Gan, S. Zhang, N. C. Rockwell, S. S. Martin, J. C. Lagarias, D. A. Bryant, *Science* **2014**, 345(6202), 1312–1317.
- [63] M. Chen, M. Schliep, R. D. Willows, Z. L. Cai, B. A. Neilan, H. Scheer, *Science* **2010**, 329(5997), 1318–1319.
- [64] J. P. Götze, F. Anders, S. Petry, J. F. Witte, H. Lokstein, *Chem. Phys.* **2022**, 559, 111517.
- [65] G. Britton, S. Liaaen-Jensen, H. Pfander, *Carotenoids: Handbook*, Birkhäuser Basel, **2012**.
- [66] R. Croce, M. G. Müller, R. Bassi, A. R. Holzwarth, *Biophys. J.* **2001**, 80(2), 901–915.
- [67] E. J. G. Peterman, C. C. Gradinaru, F. Calkoen, J. C. Borst, R. van Grondelle, H. van Amerongen, *Biochemistry* **1997**, 36(40), 12208–12215.
- [68] A. V. Ruban, R. Berera, C. Iliaia, I. H. van Stokkum, J. T. Kennis, P. A. A., H. van Amerongen, B. Robert, P. Horton, R. van Grondelle, *Nature* **2007**, 450(7169), 575–578.
- [69] C. Tian, K. R. Clauser, D. Öhlund, S. Rickelt, Y. Huang, M. Gupta, D. R. Mani, S. A. Carr, D. A. Tuveson, R. O. Hynes, *Proc. Natl. Acad. Sci.* **2019**, 116(39), 19609–19618.
- [70] A. V. Ruban, *FEBS Lett.* **2018**, 592(18), 3030–3039.
- [71] D. J. Nürnberg, J. Morton, S. Santabarbara, A. Telfer, P. Joliot, L. A. Antonaru, A. V. Ruban, T. Cardona, E. Krausz, A. Boussac, A. Fantuzzi, A. W. Rutherford, *Science* **2018**, 360(6394), 1210–1213.
- [72] G. Cario, J. Franck, *Z. fur Phys. Chem.* **1922**, 11(1), 161–166.
- [73] J. Perrin, *C. R. Acad. Sci.* **1927**, 184, 1097–1100.
- [74] M. N. Berberan-Santos, *Pioneering Contributions of Jean and Francis Perrin to Molecular Luminescence*, p. 7–33, Springer Berlin Heidelberg, Berlin, Heidelberg, **2001**.
- [75] F. Perrin, *Ann. Phys.* **1932**, 10(17), 283–314.
- [76] T. Förster, *Sci. Nat.* **1946**, 33(6), 166–175.
- [77] T. Förster, *Ann. Phys.* **1948**, 437(1-2), 55–75.

- [78] S. Petry, J. P. Götze, *Biochim. Biophys. Acta Bioenerg.* **2022**, *1863*(2), 148521.
- [79] E. Wientjes, G. Roest, R. Croce, *Biochim. Biophys. Acta Bioenerg.* **2012**, *1817*(5), 711–717.
- [80] D. Siefermann-Harms, *Biochim. Biophys. Acta Bioenerg.* **1985**, *811*(4), 325–355.
- [81] Y. Koyama, T. Miki, X. Wang, H. Nagae, *Int. J. Mol. Sci.* **2009**, *10*(11), 4575–4622.
- [82] J. P. Götze, D. Kröner, S. Banerjee, B. Karasulu, W. Thiel, *ChemPhysChem.* **2014**, *15*(15), 3392–3401.
- [83] X. Su, J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang, M. Li, *Science* **2017**, *357*(6353), 815–820.
- [84] L.-X. Shi, W. P. Schröder, *Biochim. Biophys. Acta Bioenerg.* **2004**, *1608*(2), 75–96.
- [85] L.-X. Shi, M. Hall, C. Funk, W. P. Schröder, *Biochim. Biophys. Acta Bioenerg.* **2012**, *1817*(1), 13–25, photosystem II.
- [86] J. M. Anderson, W. S. Chow, J. De Las Rivas, *Photosynth. Res.* **2008**, *98*(1-3), 575–587.
- [87] T. K. Goral, M. P. Johnson, A. P. Brain, H. Kirchhoff, A. V. Ruban, C. W. Mullineaux, *Plant J.* **2010**, *62*(6).
- [88] J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, *Nature* **1985**, *318*, 618–624.
- [89] W. Kühlbrandt, *Curr. Opin. Struct. Biol.* **1994**, *4*(4), 519–528.
- [90] K.-H. Rhee, E. P. Morris, J. Barber, W. Kühlbrandt, *Nature* **1998**, *396*(6708), 283–286.
- [91] J. Nield, O. Kruse, J. Ruprecht, P. Da Fonseca, C. Büchel, J. Barber, *J. Biol. Chem.* **2000**, *275*(36), 27940–27946.
- [92] A. Zouni, H. Witt, J. Kern, P. Fromme, N. Krauss, W. Saenger, P. Orth, *Nature* **2001**, *409*(6821), 1476–4687.
- [93] A. Zouni, R. Jordan, E. Schlodder, P. Fromme, H. Witt, *Biochim. Biophys. Acta Bioenerg.* **2000**, *1457*(3), 103–105.
- [94] B. Loll, J. Kern, W. Saenger, A. Zouni, J. Biesiadka, *Nature* **2005**, *438*(7070), 1476–4687.
- [95] K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber, S. Iwata, *Science* **2004**, *303*(5665), 1831–1838.
- [96] Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya, *Nature* **2011**, *473*(7345), 55–60.

- [97] X. Wei, X. Su, P. Cao, X. Liu, W. Chang, M. Li, X. Zhang, Z. Liu, *Nature* **2016**, *534*(7605), 69–74.
- [98] Z. Liu, H. Yan, K. Wang, T. Kuang, J. Zhang, L. Gui, X. An, W. Chang, *Nature* **2004**, *428*(6980), 287–292.
- [99] L. Valkunas, J. Chmeliov, T. P. Krüger, C. Iliaia, R. Van Grondelle, *J. Phys. Chem. Lett.* **2012**, *3*(19), 2779–2784.
- [100] T. P. Krüger, C. Iliaia, M. P. Johnson, A. V. Ruban, R. Van Grondelle, *Biochim. Biophys. Acta Bioenerg.* **2014**, *1837*(7), 1027–1038.
- [101] H. van Amerongen, R. van Grondelle, L. Valkunas, *Photosynthetic Excitons*, World Scientific, **2000**.
- [102] E. J. Boekema, H. Van Roon, J. F. Van Breemen, J. P. Dekker, *Europ. J. Biochem.* **1999**, *266*(2), 444–452.
- [103] S. Caffarri, R. Kouřil, S. Kereiche, E. J. Boekema, R. Croce, *EMBO J.* **2009**, *28*(19), 3052–3063.
- [104] J. P. Dekker, E. J. Boekema, *Biochim. Biophys. Acta* **2005**, *1706*(1-2), 12–39.
- [105] A. E. Yakushevskaya, W. Keegstra, E. J. Boekema, J. P. Dekker, J. Andersson, S. Jansson, A. V. Ruban, P. Horton, *Biochemistry* **2003**, *42*(3), 608–613.
- [106] H. van Amerongen, R. Croce, *Photosynth. Res.* **2013**, *116*, 251–263.
- [107] S. Caffarri, R. Croce, J. Breton, R. Bassi, *J. Biol. Chem.* **2001**, *276*(38), 35924–35933.
- [108] T. Henrysson, W. P. Schröder, M. Spangfort, H.-E. Kerlund, *Biochim. Biophys. Acta Bioenerg.* **1989**, *977*(3), 301–308.
- [109] P. Dainese, R. Bassi, *J. Biol. Chem.* **1991**, *266*(13), 8136–8142.
- [110] E. Giuffra, D. Cugini, R. Croce, R. Bassi, *Europ. J. Biochem.* **1996**, *238*(1), 112–120.
- [111] A. Pascal, C. Gradinaru, U. Wacker, E. Peterman, F. Calkoen, K.-D. Irrgang, P. Horton, G. Renger, R. van Grondelle, B. Robert, H. van Amerongen, *Europ. J. Biochem.* **1999**, *262*(3), 817–823.
- [112] S. Jansson, *Trends Plant Sci.* **1999**, *4*(6), 236–240.
- [113] R. Croce, M. G. Müller, R. Bassi, A. R. Holzwarth, *Biophys. J.* **2003**, *84*(4), 2508–2516.
- [114] B. Hankamer, J. Barber, E. J. Boekema, *Annu. rev. plant physiol. plant mol. biol.* **1997**, *48*(1), 641–671.
- [115] T. Reinot, M. Jassas, A. Kell, A. P. Casazza, S. Santabarbara, R. Jankowiak, *J. Chem. Phys.* **2021**, *154*(8), 085101.

- [116] C. C. Gradinaru, A. A. Pascal, F. van Mourik, B. Robert, P. Horton, R. van Grondelle, H. van Amerongen, *Biochemistry* **1998**, *37*(4), 1143–1149.
- [117] V. I. Novoderezhkin, M. A. Palacios, H. van Amerongen, R. van Grondelle, *J. Phys. Chem. B* **2005**, *109*(20), 10493–10504.
- [118] R. Croce, M. G. Müller, S. Caffarri, R. Bassi, A. R. Holzwarth, *Biophys. J.* **2003**, *84*(4), 2517–2532.
- [119] E. C. Engelmann, G. Zucchelli, F. M. Garlaschi, A. P. Casazza, R. C. Jennings, *Biochim. Biophys. Acta Bioenerg.* **2005**, *1706*(3), 276–286.
- [120] K. Broess, G. Trinkunas, A. van Hoek, R. Croce, H. van Amerongen, *Biochim. Biophys. Acta* **2008**, *1777*(5), 404–409.
- [121] S. Caffarri, K. Broess, R. Croce, H. van Amerongen, *Biophys. J.* **2011**, *100*(9), 2094–2103.
- [122] S. Caffarri, T. Tibiletti, R. C. Jennings, S. Santabarbara, *Curr. Protein Pept. Sci.* **2014**, *15*, 296–331.
- [123] G. Zucchelli, D. Brogioli, A. P. Casazza, F. M. Garlaschi, R. C. Jennings, *Biophys. J.* **2007**, *93*, 2240–2254.
- [124] T. Renger, F. Müh, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3348–3371.
- [125] E. Bergantino, P. Dainese, Z. Cerovic, S. Sechi, R. Bassi, *J. Biol. Chem.* **1995**, *270*(15), 8474–8481.
- [126] T. Mikko, P. Mirva, S. Marjaana, S. Sari, M. Paula, V. Julia, V. Alexander, A. Yagut, A. Eva-Mari, *Plant Mol. Biol.* **2006**, *62*(4), 779–793.
- [127] R. Tokutsu, M. Iwai, J. Minagawa, *J. Biol. Chem.* **2009**, *284*(12), 7777–7782.
- [128] M. Hansson, A. V. Vener, *Mol. Cell. Proteomics* **2003**, *2*(8), 550–559.
- [129] W.-J. Liu, Y.-E. Chen, W.-J. Tian, J.-B. Du, Z.-W. Zhang, F. Xu, F. Zhang, S. Yuan, H.-H. Lin, *Biochim. Biophys. Acta Bioenerg.* **2009**, *1787*(10), 1238–1245.
- [130] Y.-E. Chen, S. Yuan, J.-B. Du, M.-Y. Xu, Z.-W. Zhang, H.-H. Lin, *Biochemistry* **2009**, *48*(41), 9757–9763.
- [131] Y.-E. Chen, Z.-Y. Zhao, H.-Y. Zhang, X.-Y. Zeng, S. Yuan, *J. Exp. Bot.* **2013**, *64*(5), 1167–1178.
- [132] B. Drop, M. Webber-Birungi, S. K. M. Yadav, A. Filipowicz-Szymanska, F. Fusetti, E. J. Boekema, R. Croce, *Biochim. Biophys. Acta Bioenerg.* **2014**, *1837*(1), 63–72.
- [133] M. Iwai, Y. Takahashi, J. Minagawa, *Plant Cell* **2008**, *20*(8), 2177.
- [134] L. Kovács, J. Damkjær, S. Kereiche, C. Iliaia, A. V. Ruban, E. J. Boekema, S. Jansson, P. Horton, *Plant Cell* **2006**, *18*(11), 3106–3120.

- [135] S. De Bianchi, L. Dall'Osto, G. Tognon, T. Morosinotto, R. Bassi, *Plant Cell* **2008**, *20*(4), 1012–1028.
- [136] J. Chmeliov, A. Gelzinis, E. Songaila, R. Augulis, C. D. P. Duffy, A. V. Ruban, L. Valkunas, *Nat. Plants* **2016**, *2*(5), 16045–16052.
- [137] L. Nicol, W. J. Nawrocki, R. Croce, *Nat. Plants* **2019**, *5*(11), 1177–1183.
- [138] Z. Guardini, M. Bressan, R. Caferri, R. Bassi, L. Dall'Osto, *Nat. Plants* **2020**, *6*(3), 303–313.
- [139] L. Dall'Osto, S. Cazzaniga, M. Bressan, D. Paleček, K. Židek, K. K. Niyogi, G. R. Fleming, D. Zigmantas, R. Bassi, *Nat. Plants* **2017**, *3*(5), 17033–17042.
- [140] B. Halliwell, J. M. C. Gutteridge, *Free Radicals in Biology and Medicine*, Oxford University Press, **2015**.
- [141] A. Krieger-Liszkay, *J. Exp. Bot.* **2004**, *56*(411), 337–346.
- [142] E. Hideg, C. Spetea, I. Vass, *Photosynth. Res.* **1994**, *39*(2), 191–199.
- [143] A. N. Macpherson, A. Telfer, J. Barber, T. Truscott, *Biochim. Biophys. Acta Bioenerg.* **1993**, *1143*(3), 301–309.
- [144] H. Kautsky, H. de Bruijn, *Sci. Nat.* **1931**, *52*, 1043.
- [145] E. Skovsen, J. W. Snyder, J. D. C. Lambert, P. R. Ogilby, *J. Phys. Chem. B* **2005**, *109*(18), 8570–8573.
- [146] S. Hatz, J. D. C. Lambert, P. R. Ogilby, *Photochem. Photobiol. Sci.* **2007**, *6*(10), 1106–1116.
- [147] E. Hideg, K. Ogawa, T. Kálai, K. Hideg, *Physiol. Plant.* **2001**, *112*(1), 10–14.
- [148] E. Hideg, C. Barta, T. Kálai, I. Vass, K. Hideg, K. Asada, *Plant Cell Physiol.* **2002**, *43*(10), 1154–1164.
- [149] N. Adir, H. Zer, S. Shochat, I. Ohad, *Photosynth. Res.* **2003**, *76*(1), 343–370.
- [150] E. Tyystjärvi, *Coord. Chem. Rev.* **2008**, *252*(3), 361–376.
- [151] A. Krieger-Liszkay, C. Fufezan, A. Trebst, *Photosynth. Res.* **2008**, *98*, 551–564.
- [152] A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education, Essex, **2001**.
- [153] A. Pandya, M. J. Howard, M. Zloh, P. A. Dalby, *Pharmaceutics* **2018**, *10*(4).
- [154] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, B. Towles, *Millisecond-Scale Molecular Dynamics Simulations on Anton*, Association for Computing Machinery, New York, NY, USA, **2009**.

- [155] T. A. Halgren, W. Damm, *Curr. Opin. Struct. Biol.* **2001**, *11*(2), 236–242.
- [156] J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, R. A. Friesner, *J. Chem. Phys.* **1999**, *110*(2), 741–754.
- [157] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*(2), 227–249.
- [158] J. Kostal, *Adv. Mol. Toxicol.* **2016**, *10*, 139–186.
- [159] H. Hu, Z. Lu, W. Yang, *J. Chem. Theory Comput.* **2007**, *3*(2), 390–406.
- [160] L. Shen, J. Wu, W. Yang, *J. Chem. Theory Comput.* **2016**, *12*(10), 4934–4946.
- [161] L. Shen, W. Yang, *J. Chem. Theory Comput.* **2018**, *14*(3), 1442–1455.
- [162] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*(12), 1463–1472.
- [163] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*(3), 327–341.
- [164] *Physikalische Blätter* **1948**, *4*(4), 146–151.
- [165] J. Reinhold, *Quantentheorie der Moleküle: Eine Einführung*, 5. ed., Springer Fachmedien Wiesbaden, Wiesbaden.
- [166] M. Born, R. Oppenheimer, *Ann. Phys.* **1927**, *389*, 457–484.
- [167] D. R. Yarkony, *Rev. Mod. Phys.* **1996**, *68*, 985–1013.
- [168] G. Wedler, *Lehrbuch der physikalischen Chemie*, 6. ed., Wiley-VCH, Weinheim, **2012**.
- [169] C. C. J. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [170] H. Stoll, *Phys. Rev.* **1992**, 6700–6704.
- [171] G. G. Hall, J. E. Lennard-Jones, *Proc. R. Soc. A* **1951**, *205*(1083), 541–552.
- [172] P. Löwdin, *Phys. Rev.* **1955**, *97*, 1509–1520.
- [173] F. J. Holguín-Gallego, R. Chávez-Calvillo, M. García-Revilla, E. Francisco, M. Pendás, T. Rocha-Rinza, *J. Comput. Chem.* **2016**, *37*(19), 1753–1765.
- [174] D. Haase, E. Ruch, *Theor. Chem. Acc.* **1973**, *29*, 235–246.
- [175] P. Hohnberg, W. Kohn, *Phys. Rev.* **1964**, *136*(3B), 864–871.
- [176] M. Levy, *Phys. Rev.* **1982**, *26*(A), 1200–1208.
- [177] J. P. Perdew, K. Schmidt, *AIP Conf. Proc* **2001**, *577*(1), 1–20.
- [178] W. Kohn, L. Sham, *Phys. Rev.* **1965**, *140*(A), 1133–1138.
- [179] K. Burke, L. O. Wagner, *Int. J. Quantum Chem.* **2013**, *113*(2), 96–101.

- [180] J. Perdew, Y. Wang, *Phys. Rev.* **1992**, *23*(45B), 13244–13249.
- [181] J. P. Perdew, J. Tao, V. N. Staroverov, G. E. Scuseria, *J. Chem. Phys.* **2004**, *120*(15), 6898–6911.
- [182] A. D. Becke, *J. Chem. Phys.* **1993**, *98*(7), 5648–5652.
- [183] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
- [184] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*(8), 1200–1211.
- [185] A. D. Becke, *J. Chem. Phys.* **2014**, *140*, 18A301.
- [186] H. S. Yu, S. L. Li, D. G. Truhlar, *J. Chem. Phys.* **2016**, *145*, 130901.
- [187] D. J. Tozer, N. C. Handy, *J. Chem. Phys.* **1998**, *108*(6), 2545–2555.
- [188] R. Bauernschmitt, R. Ahlrichs, *Chem. Phys. Lett.* **1996**, *256*(4), 454–464.
- [189] D. P. Chong, *Recent Advances in Density Functional Methods*, World Scientific, **1995**, p. 155–192.
- [190] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.* **2004**, *393*(1), 51–57.
- [191] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, K. Hirao, *J. Chem. Phys.* **2004**, *120*(18), 8425–8433.
- [192] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.* **2004**, *393*(1-3), 51–57.
- [193] E. Runge, E. K. U. Gross, *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- [194] P. Ehrenfest, *J. Phys.* **1927**, *45*, 455–457.
- [195] R. van Leeuwen, *Phys. Rev. Lett.* **1999**, *82*, 3863–3866.
- [196] L. Chayes, J. T. and Chayes, M. B. Ruskai, *J. Stat. Phys.* **1985**, *38*, 497–518.
- [197] M. Levy, *Proc. Natl. Acad. Sci.* **1979**, *76*(12), 6062–6065.
- [198] N. T. Maitra, T. N. Todorov, C. Woodward, K. Burke, *Phys. Rev. A* **2010**, *81*, 042525–042532.
- [199] V. A. Azov, *Tetrahedron Lett.* **2016**, *57*(49), 5416–5425.
- [200] H. Hettema, H. J. A. Jensen, P. Jørgensen, J. Olsen, *J. Chem. Phys.* **1992**, *97*(2), 1174–1190.
- [201] L. Frediani, H. Ågren, L. Ferrighi, K. Ruud, *J. Chem. Phys.* **2005**, *123*(14), 144117.
- [202] S. Hirata, M. Head-Gordon, *Chem. Phys. Lett.* **1999**, *314*(3), 291–299.
- [203] H.-J. Kunze, *Introduction to Plasma Spectroscopy*, Springer Berlin, Heidelberg, **2009**, p. 153–178.
- [204] E. Fermi, *Nuclear Physics*, University of Chicago Press, Chicago, IL, **1950**.

- [205] P. A. M. Dirac, N. H. D. Bohr, *Proc. R. Soc. A* **1927**, *114*(767), 243–265.
- [206] G. Herzberg, E. Teller, *Z. fur Phys. Chem.* **1933**, *21B*(1), 410–446.
- [207] W. S. Struve, *Fundamentals of molecular spectroscopy*, Wiley-Interscience, **1989**.
- [208] H. H. Jaffe, A. L. Miller, *J. Chem. Educ.* **1966**, *43*(9), 469–473.
- [209] B. P. Krueger, G. D. Scholes, R. Jimenez, G. R. Fleming, *J. Phys. Chem. B* **1998**, *102*(12), 2284–2292.
- [210] G. D. Scholes, I. R. Gould, R. J. Cogdell, G. R. Fleming, *J. Phys. Chem. B* **1999**, *103*(13), 2543–2553.
- [211] B. H. Besler, J. Merz, M. Kenneth, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*(4), 431–439.
- [212] K. Hinsien, B. Roux, *J. Comput. Chem.* **1997**, *18*(3), 368–380.
- [213] R. O. Dror, T. J. Mildorf, D. Hilger, A. Manglik, D. W. Borhani, D. H. Arlow, A. Philippsen, N. Villanueva, Z. Yang, M. T. Lerch, W. L. Hubbell, B. K. Kobilka, R. K. Sunahara, D. E. Shaw, *Science* **2015**, *348*(6241), 1361–1365.
- [214] G. Dawe, M. Musgaard, M. Aurousseau, N. Nayeem, T. Green, P. Biggin, D. Bowie, *Neuron* **2016**, *89*(6), 1264–1276.
- [215] N. R. Latorraca, J. K. Wang, B. Bauer, R. J. L. Townshend, S. A. Hollingsworth, J. E. Olivieri, H. E. Xu, M. E. Sommer, R. O. Dror, *Nature* **2018**, *557*(7705), 452–456.
- [216] M. Ghasemi, A. Shafiei, J. Foroozesh, *Adv. Colloid Interface Sci.* **2022**, *300*, 102594.
- [217] J. R. Hill, J. Sauer, *J. Phys. Chem.* **1994**, *98*(4), 1238–1244.
- [218] T. K. Firman, C. R. Landis, *J. Am. Chem. Soc.* **2001**, *123*(47), 11728–11742.
- [219] Z. Tao, P. T. Cummings, *Mol. Simul.* **2007**, *33*(15), 1255–1260.
- [220] Y. Zhang, *J. Chem. Phys.* **2005**, *122*(2), 024114.
- [221] K. Nam, J. Gao, D. M. York, *J. Chem. Phys.* **2005**, *1*(1), 2–13.
- [222] H. Lin, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*(17), 3991–4004.
- [223] H. M. Senn, W. Thiel, *Angew. Chem., Int. Ed.* **2009**, *48*(7), 1198–1229.
- [224] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*(6), 700–733.
- [225] F. Maseras, K. Morokuma, *J. Comput. Chem.* **1995**, *16*(9), 1170–1179.
- [226] T. Vreven, B. Mennucci, C. O. da Silva, K. Morokuma, J. Tomasi, *J. Chem. Phys.* **2001**, *115*(1), 62–72.

- [227] R. Catlow, P. Burke, J. Goodfellow, D. Tildesley, M. Wilson, K. Morokuma, *Philos. Trans. Royal Soc. A* **2002**, *360*(1795), 1149–1164.
- [228] I. Antes, W. Thiel, *J. Phys. Chem. A* **1999**, *103*(46), 9290–9295.
- [229] Y. Zhang, T.-S. Lee, W. Yang, *J. Chem. Phys.* **1999**, *110*(1), 46–54.
- [230] G. A. DiLabio, M. M. Hurley, P. A. Christiansen, *J. Chem. Phys.* **2002**, *116*(22), 9578–9584.
- [231] G. G. Ferenczy, J.-L. Rivail, P. R. Surján, G. Náray-Szabó, *J. Comput. Chem.* **1992**, *13*(7), 830–837.
- [232] X. Assfeld, J.-L. Rivail, *Chem. Phys. Lett.* **1996**, *263*(1), 100–106.
- [233] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. G. Ferenczy, *J. Comput. Chem.* **1994**, *15*(3), 269–282.
- [234] J. Gao, P. Amara, C. Alhambra, M. J. Field, *J. Phys. Chem. A* **1998**, *102*(24), 4714–4721.
- [235] J. Pu, J. Gao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*(25), 5454–5463.
- [236] J. Pu, J. Gao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*(4), 632–650.
- [237] R. Croce, H. Van Amerongen, *Photosynth. Res.* **2013**, *116*, 153–166.
- [238] J. J. J. Roden, D. I. G. Bennett, K. B. Whaley, *J. Chem. Phys.* **2016**, *144*(24), 245101.
- [239] D. Montepietra, M. Bellingeri, A. M. Ross, F. Scotognella, D. Cassi, *J. R. Soc. Interface* **2020**, *17*(172), 20200813.
- [240] R. C. Bugos, H. Y. Yamamoto, *Proc. Natl. Acad. Sci.* **1996**, *93*(13), 6320–6325.
- [241] F. Bouvier, A. d’Harlingue, P. Hugueney, E. Marin, A. Marion-Poll, B. Camara, *J. Biol. Chem.* **1996**, *271*(46), 28861–28867.
- [242] U. Schreiber, C. Klughammer, G. Schansker, *Photosynth. Res.* **2019**, *142*(1), 35–50.
- [243] T. Polívka, V. Sundström, *Chem. Rev.* **2004**, *104*(4), 2021–2071.
- [244] J. P. Götze, B. Karasulu, W. Thiel, *J. Chem. Phys.* **2013**, *139*(23), 234108.
- [245] J. P. Götze, B. Karasulu, M. Patil, W. Thiel, *Biochim. Biophys. Acta Bioenerg.* **2015**, *1847*(12), 1509–1517.
- [246] P. Xu, V. U. Chukhutsina, W. J. Nawrocki, G. Schansker, L. W. Bielczynski, Y. Lu, D. Karcher, R. Bock, R. Croce, *eLife* **2020**, *9*, e58984.
- [247] J. P. Götze, H. Lokstein, *bioRxiv* (doi: 10.1101/2023.01.26.525641) **2023**.
- [248] J. P. Götze, H. Lokstein, *bioRxiv* (doi: 10.1101/2023.01.26.525634) **2023**.

-
- [249] S. Jurinovich, L. Viani, I. G. Prandi, T. Renger, B. Mennucci, *Phys. Chem. Chem. Phys.* **2015**, *17*(22), 14405–14416.
- [250] D. Casares, P. V. Escribá, C. A. Rosselló, *Int. J. Mol. Sci.* **2019**, *20*(9).