Energy Transfer in Bacterial Photosynthesis Resolved by Machine Learning enhanced Multiscale Simulations

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

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

genehmigte

DISSERTATION

von

M.Sc. David Sebastian Hoffmann

aus Heilbronn

1. Referent: Prof. Dr. Marcus Elstner

2. Referent: Priv.-Doz. Dr. Sebastian Höfener

Tag der mündlichen Prüfung: 29. Oktober 2025

Abstract

The central dogma of computational chemistry - the balance between accuracy, system size, and computational cost - has been reshaped by machine learning (ML) models. These models enable us to address questions that were inaccessible only a few years ago. In this thesis, ML predictions of large proteins and a smaller chlorophyll molecule were integrated into established multiscale methods to provide new atomistic insights into receptors and photosynthetic proteins.

First, AlphaFold2, the Nobel Prize-winning protein prediction tool, was used to study the interspecies signaling between nematode worms and fungi. Molecular docking studies of nematode-specific pheromones and AlphaFold2-predicted receptors revealed the role of two short sequence motifs that are conserved in both organisms. These results provided an atomistic explanation for the phenotypic experiments. Furthermore, the results demonstrated how an iterative loop of predictions validated by experiments and experiments guided by predictions can lead to highly productive collaborations.

The main part of this work focused on bacteriochlorophyll a (BChl a), the key pigment in photosynthetic green and purple bacteria. Its quantum mechanical properties such as excitation energies and excitonic couplings, govern light absorption and energy transfer. In light-harvesting complexes, the protein environment fine-tunes these properties so that multiple BChl a molecules form collective quantum states, optimizing the efficiency of energy transfer in form of excitons. Neural networks can predict the required properties of individual BChl a pigments significantly faster than any quantum mechanical calculation, enabling the description of the collective properties of BChl a molecules.

The multiscale method was benchmarked on the comparably small Fenna–Matthews–Olson (FMO) complex. It reproduced experimental transfer times and absorption spectra under both cryogenic and physiological conditions. Based on these findings, exciton dynamics in the light-harvesting complex 2 (LH2) from purple bacteria was investigated. The simulations reproduced experimentally determined properties such as the absorption spectrum or the delocalization of the exciton among the pigments.

However, the ML-accelerated simulations showed that these experiments and interpretations are misleading. The pigments' excitation energies and couplings are rarely, if ever, arranged in a way that leads to the experimentally determined time-averaged properties. A time-dependent analysis of excitonic properties revealed a transfer mechanism in one subsystem of LH2 known as "transient localization theory", which was originally formulated to describe highly efficient transfer in organic semiconductors. This transfer regime is only possible due to the low reorganization energy of BChl a and the specific protein environment of that subsystem, which yields negligible additional reorganization energy.

Such insights not only deepen the fundamental understanding of energy transfer in photosynthesis, but also open possibilities for the rational design of photosynthetically optimized organisms.

Zusammenfassung

Das zentrale Dogma der computergestützten Chemie, dass sich Gleichgewicht, Systemgröße und Genauigkeit gegenseitig bedingen, wurde durch Modelle des maschinellen Lernens (ML) neu definiert. In dieser Arbeit wurden ML-Vorhersagen für große Proteine und ein kleineres Chlorophyllmolekül in etablierte Multiskalenansätze integriert, um auf atomarer Ebene neue Einblicke in Rezeptoren und photosynthetische Proteine zu gewinnen.

Zunächst wurde das mit dem Nobelpreis ausgezeichneten AlphaFold2 Proteinvorhersage-Tool verwendet, um die Signalübertragung zwischen Fadenwürmern und Pilzen zu untersuchen. Docking-Studien von Pheromonen und Rezeptoren enthüllten die Rolle von zwei in beiden Organismen konservierten Sequenzmotiven. Diese Ergebnisse lieferten eine atomistische Erklärung für die phänotypischen Experimente. Darüber hinaus zeigten die Ergebnisse, wie ein iterativer Abgleich zwischen Experimenten und Vorhersagen zu einer produktiven Zusammenarbeit führen kann.

Der Hauptteil dieser Arbeit konzentrierte sich auf Bakteriochlorophyll a (BChl a), das am häufigsten vorkommende Pigment in photosynthetischen Bakterien. Seine quantenmechanischen Eigenschaften, wie Anregungsenergien und exzitonischen Kopplungen, bestimmen seine Lichtabsorption und den Energietransfer. In Lichtsammelkomplexen moduliert die Proteinumgebung diese Eigenschaften so, dass mehrere BChl a-Moleküle kollektive Zustände bilden, was zu einem optimierten Energietransfer in Form von Exzitonen führt. Neuronale Netze können die erforderlichen Eigenschaften einzelner BChl a-Pigmente deutlich schneller vorhersagen als jede quantenmechanische Berechnung, wodurch die Simulation der kollektiven Eigenschaften von BChl a-Molekülen ermöglicht wird.

Die Methode wurde am vergleichsweise kleinen Fenna-Matthews-Olson-Komplex (FMO) getestet. Sie reproduzierte experimentelle Transferzeiten und Absorptionsspektren sowohl unter kryogenen als auch unter physiologischen Bedingungen. Basierend auf diesen Ergebnissen wurde die Exzitondynamik im Lichtsammelkomplex 2 (LH2) aus Purpurbakterien untersucht. Die Simulationen reproduzierten experimentell gemessene Eigenschaften, wie das Absorptionsspektrum oder die Delokalisierung des Exzitons zwischen den Pigmenten.

Die Simulationen zeigten jedoch auch, dass diese Experimente und Interpretationen irreführend sind. Die Anregungsenergien und Kopplungen der Pigmente sind selten derart angeordnet, dass sie zu den experimentell ermittelten Eigenschaften führen. Eine zeitaufgelöste Analyse des Exzitons ergab einen Transfermechanismus in einem Teilsystem von LH2, der als „Transient Localization Theory“ bekannt ist und ursprünglich für den effizienten transfer in organischen Halbleituern formuliert wurde. Dieses Transferregime ist nur aufgrund der geringen Reorganisationsenergie von BChl a und der spezifischen Proteinumgebung möglich.

Derartige Erkenntnisse vertiefen nicht nur das grundlegende Verständnis von der Energieübertragung in der Photosynthese, sondern eröffnen auch Möglichkeiten für die rationale Entwicklung photosynthetisch optimierter Organismen.

List of Abbreviations

ADF	AutoDockFR suite
AF2	AlphaFold2
ANN	Artificial Neural Network
AO	Atomic Orbital
ATP	Adenosine Triphosphate
BOA	Born–Oppenheimer Approximation
BChl a	Bacteriochlorophyll a
ChEBI	Chemical Entities of Biological Interest
DFT	Density Functional Theory
DFTB	Density Functional Tight Binding
ECL	Extracellular Loop
EET	Excitation Energy Transfer
ESP	Electrostatic Potential
FC	Franck–Condon
FCE	Full Cumulant Expansion
FMO	Fenna–Matthews–Olson
FRET	Förster Resonance Energy Transfer
GPCR	G-Protein–Coupled Receptor
HF	Hartree–Fock
IPR	Inverse Participation Ratio
JDS	Jang–Dempster–Silbey delocalization measure
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LHC	Light-Harvesting Complex
LH2	Light-Harvesting Complex 2
LC-DFTB	Long-Range Corrected DFTB
MAE	Mean Absolute Error
MD	Molecular Dynamics
ML	Machine Learning
MM	Molecular Mechanics
MSE	Mean Squared Error
MO	Molecular Orbital
NAMD	Non-Adiabatic Molecular Dynamics
NN	Neural Network
OSC	Organic Semiconductor
PDB	Protein Data Bank
PES	Potential Energy Surface

PPC	Protein–Pigment Complex
plDDT	Predicted Local Distance Difference Test
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
RMSD	Root Mean Square Deviation
SCC	Self-Consistent Charge
SCF	Self-Consistent Field
SH	Surface Hopping
TFSSH	Tully's Fewest Switches Surface Hopping
TDSE	Time-dependent Schrödinger Equation
TM	Transmembrane helix/domain
TrEsp	Transition Charges from Electrostatic Potentials
VEE	Vertical Excitation Energy

Contents

1. Introduction

For centuries, the research field of biochemistry has been dominated by empirical approaches. Industrial chemistry in the 19th and the discovery of quantum mechanics (QM) in the beginning of the 20th century paved the way for modern chemistry. Since the development of computers at the end of the 20th century computational chemistry has emerged.

Its central dilemma is the balance between accuracy and computational cost.[55] With increasing molecular system size, the number of interactions between atoms grows rapidly. In addition, the timescale of the motion of interest can extend from femto- to milliseconds, increasing the cost of molecular dynamics (MD) simulations steeply with system size. A wide spectrum of methods have been developed to balance accuracy with system size at a given cost. Accurate *ab initio* calculations based on first principles are only affordable for systems with tens of atoms. In semi-empirical QM methods, precalculated and fitted parameters are applied to approximately solve the Schrödinger equation. Molecular mechanics (MM) cannot describe quantum mechanical processes, but can describe motion and dynamics of thousands of atoms. Multiscale QM/MM methods split the molecular system into accurate QM and approximate MM regions. Cheap scoring functions can be used for large-scale screening, but work mostly empirically. Most recently, a new peak of empiricism has been reached by data-driven and machine learning (ML) approaches.[139] They make use of the fact, that the calculations of molecular properties are highly repetitive and the required training data can be generated on demand.

Advances in algorithms, high performance computing (HPC) infrastructures, and graphic processing units (GPUs) have pushed MD simulations by orders of magnitude from polypeptides with thousand atoms in 1977[183] up to entire cell organelles with 100 million atoms.[265] A comparable leap is represented by the rise of ML, which reshapes the balance between accuracy, system size and computational cost by further magnitudes.

One example of a large system with low accuracy is the determination of protein structures. Traditionally, such determination has required crystallographic experiments with inconsistent successes. For many years, homology modeling was the method of choice, but its accuracy depended strongly on the quality and availability of template structures.[301,212] Although MD simulations can refine homology models, they remain too computationally expensive for the long time scales relevant to proteins. The Nobel Prize in Chemistry in 2024[258] was in part awarded to the developers of the ML protein prediction tool AlphaFold2 (AF2).[135] AF2 can generate protein structures with near-crystallographic accuracy within a few hours, clearly outperforming any physics-based prediction method.[151]

Another example is the interaction of molecules and light, which can only be calculated with accurate QM methods. In contrast to first-principle QM, ML models still struggle to predict properties of arbitrary molecules.[147,13] However, when trained on a large dataset of a single molecule, ML models can predict variations of the property of interest nearly as accurate as the reference method but several orders of magnitude faster. This makes them suitable for simulating energy transfer between molecules, which is particularly relevant for studying photosynthetic pigment–protein complexes.

The two examples represent opposite extremes of the accuracy-size-cost triangle. Yet both demonstrate how ML approaches reshape this balance, enabling computational chemistry to gain mechanistic insights at the atomistic level on questions that seemed inaccessible a few years ago. This work addresses both aspects: the application of AF2 predicted protein structures for protein-ligand docking, and ML-accelerated simulations of energy transfer in light-harvesting complexes.

This thesis is organized as follows: In the first chapters, the theoretical background and all computational methods, that are relevant for this work, are presented.

In chapter 6, AF2 is applied to study the molecular reason for inter-species signaling between the predatory fungus *Arthrobotrys flagrans* and its nematode prey, *Caenorhabditis elegans*. G-protein-coupled receptors that can sense nematode-specific ascaroside pheromones have been found in both organisms. Although sequence comparisons have ruled out horizontal gene transfer as the underlying reason for fungal receptor activity, two short conserved motifs (RFAF and SLIL/SLIY) were identified in both receptors. The central question is whether these motifs directly participate in ligand binding or in receptor signaling. To investigate this question, molecular docking is applied in combination with AF2-predicted receptor structures to analyze the receptor-ligand interactions. In this way, the study links sequence analysis and experimental results.

In chapter 7, spectroscopic properties of Bacteriochlorophyll a (BChl a) are analyzed. It is the most widely distributed pigment in photosynthetic green and purple bacteria. In photosynthetic proteins, the excitation energy of BChl a is tuned by geometrical and electrostatic influences. To study photosynthesis with the semi-empirical DFTB method, its sensitivity to these influences has to be benchmarked. The calculated solvatochromic shift, i.e., the shift of the absorption peak with respect to different solvents, is analyzed and partitioned into geometric and electrostatic contributions. Thereby, two methods of calculating absorption spectra are compared.

The generated knowledge is used in chapter 8, which addresses the simulation of seven BChl a molecules in the Fenna-Matthews-Olson (FMO) complex of the green sulfur bacterium *Chlorobaculum tepidum*. Although numerous time-resolved spectroscopic experiments have been performed on the FMO complex, it is currently not possible to obtain a pigment-wise resolution *a priori* due to the collective excitation of the seven BChl a pigments in the form of excitons. The excitons can occupy different eigenstates of the system Hamiltonian (adiabatic states), which are localized or delocalized among the pigments. To benchmark the ML-accelerated NAMD method on photosynthetic complexes, MD and

NAMD simulations with different starting conditions are performed and analyzed at both cryogenic and room temperature.

Chapter 9 extends the analysis to exciton transfer in the light-harvesting complex 2 (LH2) from the purple bacterium *Rhodospirillum molischianum*. This complex contains 24 BChl a molecules arranged in two symmetric rings (B800 and B850). The absorption spectrum and the exciton transfer within each ring and in between these rings are analyzed. The properties that cannot be obtained experimentally, such as fluctuations of the adiabatic surfaces, the contributions of individual pigments, and the mechanisms that enable efficient transfer are focused particularly.

In the last chapter, chapter 10, the most important findings are summarized, open research questions are discussed, and an outlook to future work is given.

2. Calculation of Energies

The basis for the determination of molecular properties such as geometries, forces, binding affinities, or spectral properties is the energy of the molecule of interest. This chapter outlines several methods for calculating energies. These methods are based on either classical or quantum mechanics, vary in accuracy and computational cost, and are therefore suitable for different applications.

2.1. Molecular Mechanics

A simple and intuitive way to describe the potential energy of a molecule is Molecular Mechanics (MM). This efficient approach is especially appealing for large biomolecules, such as membranes or proteins. The approach involves conceptualizing atoms as particles and employing classical mechanics to calculate their interaction within a "ball and spring" model. While the all-atom approach considers each atom as a particle, in so-called coarse-grained models, each particle is larger and consists of a group of atoms. In either scenario, electrons are not explicitly included, and each particle is assigned a set of properties, such as radius or partial charge. Constants for bonds and angles are assigned to pairs or groups of particles using predefined combination rules. The potential energy of the system then expressed as an analytic function of the particle's coordinates.[55,132]

2.1.1. Force Fields

This analytical function relies on parameterized constants assigned to the particles. A specific set of such parameters is called a force field. In this work, only all-atom force fields are used, meaning that each atom is represented as an individual particle. The central idea is to classify all atoms of a molecule into defined atom types, with identical parameters applied to atoms of the same type. This approach is based on the assumption that similar atom types, like aliphatic carbon atoms, share similar properties, like bond distances, regardless of the chemical environment. This approach minimizes the number of necessary parameters. For example, in the Chemistry at Harvard Macromolecular Mechanics (CHARMM) all-atom force field used in this work, there are six different atom types for hydrogen.[29,28] The parameters of all the different atom types implicitly include quantum effects. They have been derived from experiments or *ab initio* calculations and have been parameterized to reproduce certain thermodynamic, kinetic, or electrochemical

properties. Consequently, many different force fields have been developed for different types of molecules, e.g. small molecules, proteins, lipids, or DNA.

The parameters express the potential energy of a molecule as the sum of the bonded and non-bonded interactions. The former include the covalent bonds, angles, and dihedral (torsion) angles between atoms. Non-bonded interactions are described by Van der Waals and electrostatic interactions, often modeled by the Lennard-Jones and Coulomb potentials, respectively. The total potential energy of an all-atom force field can be described as follows:

$$V = \underbrace{\frac{1}{2}\sum_i k_i(b_i - b_i^0)^2}_{\text{bonds}} + \underbrace{\frac{1}{2}\sum_j k_j^\theta(\theta_j - \theta_j^0)^2}_{\text{angles}} + \underbrace{\frac{1}{2}\sum_k \sum_n k_{\phi,n}[1 + cos(n\phi + \delta_n)]}_{\text{dihedrals}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\left[\underbrace{4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)}_{\text{Lennard-Jones}} + \underbrace{\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}}_{\text{Coulomb}}\right]. \tag{2.1}$$

The potential energy of a covalent bond or angle is approximated by a harmonic potential around the equilibrium distance (angle) b_i^0 (θ_j^0) with a force constant k_i (k_j^θ). Dihedral angles between four covalently bonded atoms are described by cosine functions with integer periodicity (n=1, 360°, n=2: 180°, n=3: 120°), the corresponding phases δ_n and the amplitudes $k_{\phi,n}$. The non-bonded Van der Waals interactions can be modeled in several ways, such as the Morse potential or the Buckingham potential. The Lennard-Jones potential is a widely used, computationally efficient model. It is zero for the interatomic distance σ and has a depth ϵ of the potential well. It is also referred to as the 12-6 potential due to the steep increase of the short-range Pauli repulsion and the moderate increase of the long-range London dispersion. The Coulomb term models the electrostatic interaction between two partial charges q_i and q_j at distance r_{ij} with vacuum permittivity ϵ_0.

The parameterized force fields can be used to simulate biomolecules with hundreds of thousands of atoms over µs time scales. However, force fields are limited in their ability to describe certain properties from the outset: The parameters of one atom type are not affected by the environment. However, the properties of a molecule change with respect to its environment. This effect can be seen e.g. for ligands in polar or unpolar binding pockets[213] or in the famous TIP3P water model used in this work, which was parameterized to describe the properties of a water molecule in water and is therefore overpolarized for water in unpolar environments.[163] Another drawback of force fields is that electrons are strictly assigned to nuclei in the form of partial charges. Therefore, the force field cannot describe electronic properties such as bond formation or breaking, charge transfer, or photo-physical processes. In such cases, a more accurate quantum mechanical description of the molecular system is necessary.[55,132]

2.2. Scoring Functions

Scoring functions are an even faster variant for calculating energies than force fields. As with them, all atoms in the system have a fixed type, and their properties do not change. The fundamental difference is that force fields represent atoms as balls connected by springs within the framework of molecular mechanics, whereas scoring functions are empirically fitted, do not always follow first-principles relations. Therefore, they cannot be considered molecular mechanics. Moreover, force fields are used to approximately calculate the total energy of the system, while scoring functions are used to estimate chemical potentials and ultimately the free energy of binding. The profile and minima of the hypersurfaces for total energy and free energy differ significantly. Additionally, the temperature-dependent entropic factor complicates the calculation of the free energy, which is why scoring functions can only be considered as a crude approximation to the free energy.[102]

In most cases, scoring functions are used to compare the binding affinity of thousands of geometric configurations of the same molecules. Precise free-energy simulations can accurately replicate experimental values; however, they require substantial computational resources.[217] Therefore, the primary objective of scoring functions is often to differentiate between potentially native configurations and the others. This results in many false positives among the highest-ranked structures. In many cases, the native pose - defined as the structure that has been experimentally measured - is among the top hundred or top ten structures with the highest score.[102]

Due to these differing objectives and target values (ΔG instead of E_{tot}), scoring functions are not physically exact. Physically motivated, empirical approximations such as the Lennard-Jones potential (eq. 2.1) are often additionally approximated. In many cases, the scoring functions are so simplified that they do not provide a physically meaningful description of the system. However, they are legitimized solely by their performance.[289]

A plethora of scoring functions exists,[83,269,200,137,279] but they all share the same fundamental idea. The total score, c, is calculated as a weighted sum of single contributions:

$$c = \sum_{i<j} \sum_{n} \lambda_n f_n(r_{ij}) \,. \tag{2.2}$$

In this process, the distances r_{ij} between atoms i and j are calculated. Then, a set of n functions with their weighting factor, λ_n, is computed to express different properties. Interactions such as steric hindrance, counts of hydrogen bonds, van der Waals, and electrostatic interactions are often considered. In biochemical applications, a small molecule (ligand) is often bound to a large molecule, such as DNA or protein (receptor). In this case, the functions f_n can additionally handle properties like geometric complementarity between bound and unbound receptors, intermolecular overlap and clashes, buried charges, buried nonpolar surface area, or total buried surface area.[102]

2.3. Quantum Mechanics

Quantum chemistry is the branch of chemistry that involves the application of quantum mechanics (QM) to the description of molecular systems. Since the discovery of the photoelectric effect and wave-particle duality in the early 20th century, quantum mechanics has undergone significant progress. Together with the development of computational capabilities in recent decades, the accurate description of small molecules has become attainable. This section provides an overview of the fundamentals and the methods used in this work.

2.3.1. Wave Function Based Methods

The key concept in quantum mechanics is the wave function. It has no physical meaning, but describes the state of a quantum system in full detail. Its square, integrated over a certain spatial and temporal area, is the probability density of finding the system in a certain state. Other physical properties like momentum, position, or kinetic energy, are extracted by operators that act on the wave function. To calculate the average measurement of an observable A, i.e. the expectation value $\langle \hat{A} \rangle$, of any operator $\hat{A}$, the complex conjugated wave function must be multiplied with the operator acting on the wave function and integrated over the whole space:

$$\langle \hat{A} \rangle = \int \int \int \Psi^* \hat{A} \, \Psi \, dx \, dy \, dz = \langle \Psi | \hat{A} | \Psi \rangle \, . \tag{2.3}$$

The Hamilton operator $\hat{H}$ is the key operator in quantum mechanics, because its observable is the total energy of the system.

$$E = \langle \Psi | \hat{H} | \Psi \rangle \tag{2.4}$$

The temporal development of a quantum system is described by the time-dependent Schrödinger equation (TDSE)

$$\hat{H} | \Psi(\vec{r}, \vec{R}, t) \rangle = i\hbar \frac{\partial}{\partial t} | \Psi(\vec{r}, \vec{R}, t) \rangle \tag{2.5}$$

with the wave function Ψ depending on electronic coordinates $\vec{r}$, nuclear coordinates $\vec{R}$ and time t. i is the imaginary unit and $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant. If the Hamiltonian of the system is not explicitly time-dependent, the time variables can be separated from the spatial variables. In this case, the energy of the wave function in the form of a standing wave can be computed using the time-independent, or stationary, Schrödinger equation:

$$\hat{H} | \Psi(\vec{r}, \vec{R}) \rangle = E | \Psi(\vec{r}, \vec{R}) \rangle \, . \tag{2.6}$$

This eigenvalue problem is fundamental to most quantum chemical considerations, because it determines the eigenfunctions or eigenvectors $\Psi(\vec{r}, \vec{R})$ that are necessary for extracting other properties of the quantum system. The corresponding set of discrete energetic eigenvalues E includes the energies of the ground state (E_0), as well as the first, second, and higher excited states and is directly linked to the Hamilton operator. For molecular systems, this operator includes terms that describe the kinetic energies $\hat{T}$ of nuclei and electrons and the potential $\hat{V}$ between them. For a molecule with N_e electrons and N_n nuclei with their corresponding nuclear mass M and nuclear charge Z, it reads in atomic units

$$\begin{aligned}\hat{H} &= \hat{T}_e + \hat{T}_n + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \\ &= -\frac{1}{2}\sum_{i}^{N_e}\nabla_i^2 - \frac{1}{2}\sum_{A}^{N_n}\frac{1}{M_A}\nabla_A^2 - \sum_{i}^{N_e}\sum_{A}^{N_n}\frac{Z_A}{r_{iA}} + \sum_{i<j}^{N_e}\frac{1}{r_{ij}} + \sum_{A<B}^{N_n}\frac{Z_A Z_B}{r_{AB}}\end{aligned} \tag{2.7}$$

with distances $r_{iA} = |\vec{r}_i - \vec{R}_A|$ and the Laplace operator $\nabla^2 = \Delta$.

In order to reduce the degrees of freedom in the time-dependent partial differential equation and in the time-independent eigenvalue problem, the Born-Oppenheimer approximation (BOA)[26] is frequently applied. It is based on the significant mass and velocity differences between electrons and nuclei, and it approximates electronic positions to be instantaneously adjusted to small nuclear displacements. In response, the electronic and nuclear motions are decoupled and the nuclear coordinates are incorporated into the electronic wave function ψ_{el} as parameters.

$$\Psi_{\text{total}}(\vec{r}, \vec{R}) = \psi_{\text{el}}(\vec{r}; \vec{R})\phi_{\text{nuc}}(\vec{R}) \tag{2.8}$$

In the electronic Hamiltonian under the BOA, the nuclear kinetic energy is neglected and the nuclei-nuclei repulsion is treated as a constant.

$$\hat{H}_e = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} \tag{2.9}$$

Solving the time-independent Schrödinger equation (eq. 2.6) for stationary nuclei under the BOA leads to the potential energy surface (PES), which is a hypersurface in form of a function of nuclear coordinates. The value of the PES function is the expectation value of the electronic Hamilton operator (eq. 2.9) that acts on the electronic wave function.

In order to find an appropriate electronic wave function, an initial function is guessed and iteratively optimized. The variational principle (eq. 2.10) states that the minimal energy of all arbitrary wave functions is achieved with the true wave function in the ground state $\Psi_{0,true}$. In this way, the ground state energy of the true wave function $E_{0,true}$ is a lower bound to all trial wave functions and optimization of the wave function is equal to leading to lower energy.[143]

$$\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle = E_{trial} \geq E_{0,true} = \langle \Psi_{0,true} | \hat{H} | \Psi_{0,true} \rangle \tag{2.10}$$

This important relation allows the use of sophisticated minimization algorithms. The most prominent method for such an optimization scheme for a wave function in molecules is the Hartree-Fock method.

Hartree-Fock is based on the ansatz of describing the electronic wave function Ψ as a product of one-electron wave functions $\Phi_i(i)$ (i=1,...,N):

$$\Psi(r_1, r_2, ..., r_N) = \Phi_1(1)\Phi_2(2)...\Phi_N(N) \ . \tag{2.11}$$

An antisymmetrized product of these one-electron wave functions is used in form of a Slater determinant, to satisfy the Pauli Principle:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Phi_1(1) & \Phi_2(1) & \cdots & \Phi_N(1) \\ \Phi_1(2) & \Phi_2(2) & \cdots & \Phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(N) & \Phi_2(N) & \cdots & \Phi_N(N) \end{vmatrix} . \tag{2.12}$$

In the Slater determinant, each row corresponds to the coordinates of one electron, and each column corresponds to a one-electron wave-functions Φ_i, also called molecular orbital (MO). To find the best set of orthonormal ($\langle\Phi_i|\Phi_j\rangle = \delta_{ij}$) MOs, that minimize the expectation value of the Hamiltonian according to the variational principle, the HF equations are applied. Using the Lagrangian multipliers ϵ_i that can be interpreted as orbital energies, the HF equations read:

$$\hat{f}_i\Phi_i = \epsilon_i\Phi_i \ . \tag{2.13}$$

The Fock operator $\hat{f}$ serves as an effective one-electron operator and consists of the one-electron operator $\hat{h}$ for kinetic energy and nuclear potential, and the two-electron operators $\hat{J}$ and $\hat{K}$, for Coulomb and exchange interactions, respectively.

$$\hat{f}_i = \hat{h}_i + \sum_{j\neq i}^{N} \left(\hat{J}_j(i) - \hat{K}_j(i)\right) \tag{2.14}$$

$$\hat{h}_i = -\frac{1}{2}\Delta_i - \sum_{A}^{M} \frac{Z_A}{r_{iA}} \tag{2.15}$$

$$\hat{J}_j(i)|\Phi_i(i)\rangle = \langle\Phi_j(j)|\frac{1}{r_{ij}}|\Phi_j(j)\rangle\,|\Phi_i(i)\rangle \tag{2.16}$$

$$\hat{K}_j(i)|\Phi_i(i)\rangle = \langle\Phi_j(j)|\frac{1}{r_{ij}}|\Phi_i(j)\rangle\,|\Phi_j(i)\rangle \tag{2.17}$$

The one-electron operator $\hat{h}_i$ includes the kinetic energy of the electron and the external potential of all N_{nuc} nuclei. The first two-electron operator $\hat{J}$ defines the classical electron-electron repulsion between the electron i and another electron j. The second two-electron operator $\hat{K}_j$ has no classical analogue and defines the exchange energy due to the antisymmetry of the Slater determinant.

Because any electron is treated within an averaged potential of the nuclei and all other electrons, the HF equations treat electrons in an averaged field. Therefore, electron correlation cannot be considered. Moreover, the effective potential for one electron is

defined by the MOs of the other electrons. Therefore, the HF equations have to be solved iteratively. Based on the variational principle, a set of initial trial wave functions is created, the resulting Hamiltonians (Fock-operators) are built and the trial wave functions are optimized to yield new Fock-operators for the next cycle. This process is repeated until a convergence criterion is reached meaning that the wave function is consistent with the mean field it generates (self consistent field, SCF).[131]

In order to solve the HF equations, the MOs Φ are constructed as linear combinations of atomic orbitals χ (LCAO):

$$\Phi_i = \sum_{\alpha=1}^{n} c_{i\alpha} |\chi_\alpha\rangle \tag{2.18}$$

The expansion coefficients c combine the set of n atomic orbitals, which is called basis set. Slater type orbitals (STO) and Gaussian type orbitals (GTO) exist for various orbitals (1s, 2s, $1p_x$, $1p_y$,...). The accuracy and the computational cost rises for a rising number of basis functions.

The LCAO ansatz enables the formulation of the HF equations in form of the Roothan-Hall equations:

$$\mathbf{FC} = \mathbf{SCE} \, . \tag{2.19}$$

This matrix equation contains the Fock matrix elements

$$F_{\alpha\beta} = \langle \chi_\alpha | \hat{F} | \chi_\beta \rangle \, , \tag{2.20}$$

the overlap matrix elements

$$S_{\alpha\beta} = \langle \chi_\alpha | \chi_\beta \rangle \, , \tag{2.21}$$

the matrix **C** with the expansion coefficients $c_{i\alpha}$, and the diagonal matrix **E** of orbital energies (ϵ_i). The Fock matrix elements contain Fock operators, where the two-electron integrals are weighted by the density matrix, to calculate repulsion and exchange interactions with other orbitals, proportional to their occupations.

While the primary concept appears straightforward, the computational cost for molecular systems containing tens or hundreds of electrons is significant. The two-electron integrals are the rate limiting step, so the computational cost scales with M^4, where M is the number of basis functions.[131, 55] Higher order wave function based methods such as Configurational Interaction (CI) or Coupled Cluster (CC) can account for electron correlation and can yield highly accurate results; however, their computational scaling is even steeper, restricting them to small systems.

2.3.2. Density Functional Theory

A more efficient, though approximate, method is the Density Functional Theory (DFT). For N electrons, the basis is not a 4N-dimensional wave function (three spatial coordinates and one spin per each electron), but rather the three-dimensional electron density

$$\rho(\vec{r}) = |\Phi(\vec{r})|^2 \, . \tag{2.22}$$

This electron density has cusps (maxima) in the vicinity of nuclei which are treated as point charges. Higher cusps mean higher nuclear charge. The number of electrons is obtained by integrating over the whole three-dimensional space $\vec{r}$:

$$N = \int \rho(\vec{r})\, d\vec{r}\,. \tag{2.23}$$

The relation between the so formulated electron density and the molecular energy was proven by Hohenberg and Kohn[108] in form of two theorems:

$$E_0 = E_0[\rho_0(\vec{r})] \quad \text{and} \quad E_0 \leq E[\rho_{\text{trial}}(\vec{r})] \tag{2.24}$$

The first Hohenberg-Kohn theorem (eq. 2.24, left) states that the system-specific external potential $V_{ext}(\vec{r})$ is an unambiguous functional of the ground state electron density. Therefore, the electron density uniquely determines all ground state properties of the system, like the ground state energy E_0. The second theorem is comparable to the variational principle and defines the true ground state electron density of a system leading to a global minimum of energy with respect to all trial electron densities (eq. 2.24, right).

The expression of the energy as a functional of the electron densities within the Born-Oppenheimer approximation can be expressed similar to equation 2.9:

$$E[\rho] = T_e[\rho] + V_{ee}[\rho] + V_{en}[\rho] \tag{2.25}$$

It consists of the kinetic energy of the electrons $T_e[\rho]$, the electron-electron repulsion $V_{ee}[\rho]$ and the electron-nuclei attraction $V_{en}[\rho]$. Whereas the calculation of the last term is straightforward via the Coulomb potentials from each nucleus, the first two terms are difficult to express as a functional of the electronic density alone. The electron-electron repulsion V_ee can be split into a classical Coulomb term J, which can be calculated exactly, and a non-classical contribution that arises from self-interaction, exchange and correlation effects (E_{XC}), which is absorbed into a correction term. To calculate the kinetic energy of the electrons T_e, Kohn and Sham made a great progress in 1965.[146] They considered the kinetic energy T_{ref} in a reference system with non-interacting electrons. The reference system is exposed to an effective potential which is is set such that the electron density of the real interacting system is reproduced by the reference system. The difference between the true kinetic energy and T_{ref} is incorporated into the same correction term E_{XC}. With the correction term E_{XC}, equation 2.25 can be rewritten as[143]:

$$\begin{aligned} E[\rho] &= T_{\text{ref}}[\rho] + J[\rho] + E_{XC}[\rho] + V_{en}[\rho] \\ &= T_{\text{ref}}[\rho] + \frac{1}{2}\iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}\, d\vec{r}_1 d\vec{r}_2 + E_{XC}[\rho] + \int V_{en}\,\rho(\vec{r})\, d\vec{r} \end{aligned} \tag{2.26}$$

The kinetic energy T_{ref} can be computed from the reintroduced Kohn-Sham orbitals that express the density for an N-particle system as

$$\rho(\vec{r}) = \sum_i^N |\phi_i(\vec{r})|^2\,. \tag{2.27}$$

In terms of orbitals, equation 2.26 can be rewritten as[143]:

$$E[\rho] = -\frac{1}{2}\sum_{i}^{N}\langle\phi_i|\Delta|\phi_i\rangle + \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}\iint |\phi_i(\vec{r}_1)|^2\frac{1}{|\vec{r}_1-\vec{r}_2|}|\phi_j(\vec{r}_2)|^2\,d\vec{r}_1 d\vec{r}_2 + E_{XC}[\rho] - \sum_{i}^{N}\int\sum_{A}^{N_n}\frac{Z_A}{|\vec{r}_1-\vec{r}_A|}|\phi_i(\vec{r}_1)|^2\,d\vec{r}_1 \tag{2.28}$$

Keeping the orbitals orthonormal while using the variational principle yields the one-electron Kohn-Sham equation that expresses the orbital energy ϵ_i as

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}(\vec{r})\right)\phi_i = \varepsilon_i\phi_i\,. \tag{2.29}$$

The effective potential V_{eff} depends on the Hartree term, the exchange-correlation term and the external potential:

$$V_{\text{eff}}(\vec{r}_1) = \int\frac{\rho(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}\,d\vec{r}_2 + V_{\text{XC}}(\vec{r}_1) - \sum_{A}^{N_n}\frac{Z_A}{|\vec{r}_1-\vec{r}_A|}\,. \tag{2.30}$$

The first two (Hartree and exchange-correlation) in Equation 2.30 depend themselves on the density. Therefore, the Kohn-Sham eigenvalue problem (eq. 2.29) must be calculated self-consistently.

If the exchange-correlation correction term $V_{XC}(\vec{r}) = \frac{\delta E_{XC}[\rho]}{\delta\rho(\vec{r})}$ would be known, the ground state energies and orbitals could be exactly calculated with the aforementioned equations. However, there is no analytical expression for the exchange-correlation (XC) functional. Therefore, most of the development in the field of DFT is based on finding new XC-functionals that exhibit a more accurate or computationally more efficient approximation.[143] Up to date, a variety of DFT functionals with different levels of accuracy and computational cost have been identified. They can be organized into a hierarchical structure known as Jacob's Ladder. The *Local Density Approximation* (LDA) types are the simplest functionals. As Kohn and Sham previously outlined,[146] these models are based on the assumption of a homogeneous electron gas, which is characterized by a uniformly distributed electron density throughout the entire space. Its application is primarily limited to metals or other systems that exhibit an equal distribution of valence electrons. The next type of functionals, the *Generalized Gradient Approximation* (GGA) type functionals, additionally include the gradient of the electron density to account for a non-homogeneous electron density:

$$E_{\text{XC}}^{\text{GGA}}[\rho] = \int\rho(\vec{r})\epsilon_{xc}[\rho(\vec{r}),\nabla\rho(\vec{r})]\,d\vec{r}\,. \tag{2.31}$$

Since both LDA and GGA functionals exhibit a self-interaction error, hybrid functionals, that include a certain amount of HF exchange, were developed. While the computation of exchange energy obtained via the Slater determinant in HF is formally exact, DFT only approximates it through the exchange–correlation functional. Conversely, correlation

energy is not addressed in HF theory. Hybrid functionals, such as PBE0 and B3LYP, have gained significant popularity because they strike a good balance between accuracy and computational cost.

However, the asymptotic behavior of the exchange-correlation potential V_{xc} is poorly described by LDA, GGA and standard hybrid functionals, leading to poor calculations for long-range charge-transfer or delocalized electrons. Therefore, range-separated hybrid (RSH) functionals were developed. They apply a distance-dependent amount of HF exchange contribution, based on the error function

$$\frac{1}{r_{12}} = \frac{1 - \mathrm{erf}(\omega r_{12})}{r_{12}} + \frac{\mathrm{erf}(\omega r_{12})}{r_{12}} \,. \tag{2.32}$$

Thus, the electron–electron interaction is divided into short-range and long-range components, which are primarily described with DFT and HF, respectively. Range-separated versions of B3LYP, such as CAM-B3LYP, or other RSH functionals like ωB97-X have been developed and generally offer high accuracy for a wide range of systems.

Despite the advancements in the field of DFT, the computational costs of conventional Kohn-Sham DFT (eq. 2.29) scales as $O(N^3)$ with the number of atoms N and the two-electron integrals for the exchange term in hybrid functionals lead to a scaling of approximately $O(N^4)$.[10] RSH-functionals already include empirically derived parameters in fitting the coefficients of the contributions to the hybrid E_{XC} functionals (3 for B3LYP) or of the range-separation behavior. In order to study molecular systems with approximately 100 atoms, it is necessary to employ a faster and more empirical version of DFT, such as the semi-empirical DFTB method.

2.3.3. Density Functional Tight Binding

Density functional tight binding (DFTB) employs the assumption from solid state physics that core electrons are tightly bound to the nuclei and only valence electrons are needed to describe molecular physicochemical properties. Besides the tight binding approach, this semi-empirical version of DFT incorporates additional approximations that lead to a speedup of approximately three orders of magnitude compared to DFT.[74,149]

Based on the tight binding formalism, the electron density $\rho(\vec{r})$ is defined as

$$\rho(\vec{r}) = \rho_0(\vec{r}) + \delta\rho(\vec{r}) \tag{2.33}$$

with a neutral atomic reference density $\rho_0(\vec{r})$ and the fluctuation $\delta\rho(\vec{r})$. With this ansatz, the total energy in DFTB is expressed as a Taylor expansion of the DFT energy (eq. 2.26) around the reference density.

$$\begin{aligned} E^{DFTB}[\rho_0+\delta\rho] = &\sum_i n_i \left\langle \Psi_i \left| -\frac{\nabla^2}{2} + V_{\mathrm{ne}} + \int \frac{\rho_{0'}}{|\vec{r}-\vec{r}\,'|}\, d\vec{r}\,' + V_{\mathrm{XC}}[\rho_0] \right| \Psi_i \right\rangle \\ &- \frac{1}{2}\iint \frac{\rho_0\rho_{0'}}{|\vec{r}-\vec{r}\,'|}\, d\vec{r}\, d\vec{r}\,' - \int V_{\mathrm{XC}}[\rho_0]\rho_0\, d\vec{r} + E_{\mathrm{XC}}[\rho_0] + E_{\mathrm{nn}} \\ &+ \frac{1}{2}\iint \left(\frac{1}{|\vec{r}-\vec{r}\,'|} + \left.\frac{\delta^2 E_{\mathrm{XC}}[\rho]}{\delta\rho\,\delta\rho'}\right|_{\rho_0,\rho_0'} \right) \delta\rho\,\delta\rho'\, d\vec{r}\, d\vec{r}\,' \\ &+ \frac{1}{6}\iiint \left.\frac{\delta^3 E_{\mathrm{XC}}[\rho]}{\delta\rho\,\delta\rho'\,\delta\rho''}\right|_{\rho_0,\rho_0',\rho_0''} \delta\rho\,\delta\rho'\,\delta\rho''\, d\vec{r}\, d\vec{r}\,'\, d\vec{r}\,'' + \ldots \\ &= E_0 + E_1 + E_2 + E_3 + \ldots \end{aligned} \tag{2.34}$$

Taking into account the different terms of the expansion leads to different levels of accuracy and computational cost. Standard DFTB (DFTB1, non-self-consistent DFTB)[222,260] considers the first order of the Taylor expansion. Truncating after the second term leads to self-consistent charge DFTB (DFTB2, SCC-DFTB)[76] and DFTB3[311] additionally considers the third term. In the LCAO approach, similar to Equation 2.18, the density is separated into atomic contributions: $\rho_0 = \sum_i |\phi(\vec{r})|^2 = \sum_a \rho_a(\vec{r})$. A minimal basis set with one Slater-type function per valence orbital is used. The total energy of the DFTB1 (zeroth and first order in Equation 2.34 can be expressed as:

$$E_{\mathrm{DFTB1}} = E_0 + E_1 = E_{\mathrm{Rep}} + E_{H^0}\,. \tag{2.35}$$

The first term can be expressed in terms of the eigenvalues ϵ_i of the Kohn-Sham equations as

$$E_{H^0} = \sum_{i=1}^{\mathrm{occ}} \langle\phi_i|\hat{H}_0|\phi_i\rangle = \sum_{i=1}^{\mathrm{occ}} \sum_\mu \sum_\nu c_{\mu i} c_{\nu i} H^0_{\mu\nu} = \sum_{i=1}^{\mathrm{occ}} n_i \epsilon_i\,. \tag{2.36}$$

with orbitals ϕ obtained from the minimal basis set and the occupation number n_i. The diagonal elements $H^0_{\mu\mu}$ are obtained from single atom DFT calculations using the PBE functional.[273,219] The precalculated off-diagonal matrix elements $H^0_{\mu\nu}$, which are obtained by a two-center approximation, are listed in the Slater-Koster files for different internuclear distances. The same strategy is applied for the overlap integrals $S_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle$, so that the Kohn-Sham equations

$$\sum_\nu c_{\nu i}\left(H^0_{\mu\nu} - \varepsilon_i S_{\mu\nu}\right) = 0 \tag{2.37}$$

can be solved without iterative computations. The second line in equation 2.34, the repulsive energy E_{Rep}, depends only on the reference density ρ_0 and is approximated as a sum of pairwise potentials:

$$E_{\mathrm{rep}} = \frac{1}{2}\sum_{ab} V^{\mathrm{rep}}_{ab}(\rho^0_a, \rho^0_b, R_{ab})\,. \tag{2.38}$$

The representation of these potentials as spline functions are fitted to theoretical or experimental reference data.[75]

Although the aforementioned procedure with precalculated integrals is computational efficient, charge transfer or properties of heteroatomic molecules cannot be accurately described without fluctuating electron densities. In DFTB2, the third line in equation 2.34 is approximated as:

$$E_{\mathrm{DFTB2}} = E_{\mathrm{DFTB1}} + \frac{1}{2}\sum_{ab}\gamma_{ab}\,\Delta q_a \Delta q_b\,. \tag{2.39}$$

The gamma function approaches the coulomb interaction $\frac{1}{r_{ab}}$ between two partial charges for large distances r_{ab}. The self repulsion γ_{aa} is described with the Hubbard parameter U_α, which expresses the chemical hardness via the second derivative of the total energy with respect to the occupation number. The charge fluctuation Δq_i of atom i is calculated as

$$\Delta q_a = q_a - q_a^0 \tag{2.40}$$

with q_a obtained from a Mulliken population analysis via the MO coefficients and q_a^0 being the number of valence electrons of the neutral atom a.

Unlike in DFTB1, iterative calculations must be performed to calculate the expansion coefficients for the orbitals ϕ with the Kohn-Sham equations (eq. 2.37) via the electronic Hamiltonian. This, in turn, depends on the atomic charges. Therefore, in DFTB2, an initial guess of the charges is made, which is used to calculate the Hamiltonian. The Hamiltonian is then used to update the charges. This procedure is repeated until convergence, which is why DFTB2 is also called the *self-consistent-charge* (SCC) DFTB.

Adding the last term in equation 2.34 results in DFTB3, which accounts for the third-order response of an atom's energy to charge fluctuations. Essentially, this means that its polarizability is not constant and can change depending on the charge fluctuation of the atom. This is expressed through the derivative of the Hubbard parameter with respect to charge fluctuation. With this extension, charged or highly polarized molecules (hydrogen bonds) can be described with greater precision than DFTB2. The total energy reads[311]:

$$E_{\mathrm{DFTB3}} = E_{\mathrm{DFTB2}} + \frac{1}{3}\sum_{a,b}\Delta q_a^2 \Delta q_b \left.\frac{\partial \gamma_{ab}}{\partial q_b}\right|_{q_b^0} \tag{2.41}$$

With the approximations mentioned, the fluctuating electron densities around a reference density can be calculated efficiently. However, as described above for DFT, LDA and GGA functionals suffer from the self-interaction error, leading to incorrect asymptotic behavior for electron-electron interactions. As DFTB is parameterized with a GGA functional, its exchange-correlation potential V_{XC} has the same incorrect asymptotic decay. This can potentially lead to wrong descriptions of long-range charge transfer interactions, excited states, or the delocalization of electrons. To circumvent this problem, long-range corrected (LC) functionals apply a range-separation parameter ω. In analogy to the range-separated

hybrid functionals in DFT (Equation 2.32, the Coulomb operator $\frac{1}{r_{12}} = \frac{1}{|\vec{r}_1 - \vec{r}_2|}$ is separated into a short-range and long-range component:

$$\frac{1}{r_{12}} = \underbrace{\frac{1 - \mathrm{erf}(\omega r_{12})}{r_{12}}}_{\text{short-range}} + \underbrace{\frac{\mathrm{erf}(\omega r_{12})}{r_{12}}}_{\text{long-range}} . \tag{2.42}$$

To balance the fraction of HF and DFT for the short-range and long-range exchange components, the range-separation parameter ω is calibrated using experimental data or higher-level theoretical calculations. The exchange-correlation energy can be written as

$$E_{\mathrm{XC}} = E_{\mathrm{C}}^{\mathrm{GGA}} + E_{\mathrm{X}}^{\omega,\mathrm{LDA}} + E_{\mathrm{X}}^{\omega,\mathrm{HF}} \tag{2.43}$$

with the correlation contribution calculated with the GGA functional PBE[219], the short-range exchange obtained from the LDA functional and the long-range exchange described by the HF exchange.

To describe excited state properties, the time-dependent variant TD-DFTB is applied which is based on linear response theory. According to this theory, the interaction between the molecule and an electromagnetic field is considered as a small perturbation of the first order. This results in a molecule response in form of a transition from the ground to the excited state, that can be described linearly in time. In this manner, the excited state is not calculated directly but rather based on the results of a ground state DFTB calculation. The density response of this calculation peaks at certain frequencies which are interpreted as excitation energies. The time-dependent linear response scheme can be expressed using the pseudo-eigenvalue Casida equation:[36]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} . \tag{2.44}$$

The excitation energies Ω are the eigenvalues of this equation, whereas the eigenvectors $\mathbf{X}$ and $\mathbf{Y}$ determine the corresponding transition densities and oscillator strengths, respectively. The matrix elements $\mathbf{A}$ and $\mathbf{B}$ are calculated from the occupied and virtual Kohn-Sham orbitals interacting with the coupling matrix that represents the linear response of the Hamiltonian via the exchange-correlation kernel.[149,198]

2.4. Quantum Mechanics/Molecular Mechanics

Despite the use of the aforementioned semi-empirical DFTB, the calculation of large molecular systems on long time scales are not affordable. The presented force fields are efficient but less accurate. They do not consider changes in the electronic structure, which precludes the study of phenomenona like bond breaking or charge transfer. The combination of both methods is called Quantum Mechanics/Molecular Mechanics (QM/MM) and is a valuable tool for studying large systems over extended periods of time.

The method is applicable when the system under study can be partitioned into a a QM and an MM region. The total energy of the system is expressed as

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \tag{2.45}$$

with E_{QM} and E_{MM} obtained from the QM method of choice and the force field, respectively. The coupling term $E_{\text{QM/MM}}$ quantifies the interaction between the two regions. In the absence of a covalent bond between the QM and the MM region, the coupling term only includes non-bonding interactions.

The most straightforward approach to describe the non-bonding interactions is mechanical embedding. It includes the determination of partial charges of the QM atoms via a population analysis, such as Mulliken analysis. The parameters for the Lennard-Jones potential (eq. 2.1) are derived from the force field-specific combination rule between atom types. The interaction $E_{\text{QM/MM}}$ is described by the force field. To allow the electron density of the QM region to be polarized by the partial charges of the MM region, the electrostatic potential evoked from the MM region must be included in the QM calculation. This defines the electrostatic embedding scheme and the interaction $E_{\text{QM/MM}}$ can be expressed as:

$$E_{\text{QM/MM}} = \sum_{\alpha}^{N_{QM}} \sum_{m}^{N_{MM}} \left[\frac{Z_\alpha q_m}{r_{\alpha m}} + 4\epsilon_{\alpha m} \left(\left(\frac{\sigma_{\alpha m}}{r_{\alpha m}} \right)^{12} - \left(\frac{\sigma_{\alpha m}}{r_{\alpha m}} \right)^{6} \right) \right] - \sum_{i}^{n_{QM}} \sum_{m}^{N_{MM}} \frac{q_m}{r_{im}} \,. \tag{2.46}$$

The indices α and m run over the nuclei of the QM and the MM zone, respectively. Z_α is the atomic number of the QM nuclei, whereas q_m is the partial charge of the MM atoms. The parameters ϵ and σ stem from a classical force field. The electrons in the QM region are treated with the index i and are polarized by the partial charges of the MM region. The electrostatic embedding scheme is an effective solution that strikes a balance between computational cost and accuracy and is also applied in this work.[262]

In polarized embedding, the electron distribution of the QM region polarizes the MM region, which leads to computational costly iterative calculations of both charge distributions. Sophisticated algorithms are needed to prevent from the polarization catastrophe.[101]

In all embedding schemes, careful attention must be given to treat the covalent bonds between the QM and MM region. Applying a cutoff between two bonded atoms would result in unpaired electrons in the QM region. One technical solution is to saturate the QM molecule with a hydrogen atom (linker atom) at a typical distance. QM calculations are performed in the presence of this linker atom, whereas the force field energy is evaluated for unsaturated molecules. In most cases, the QM-MM boundary is between two carbon atoms. In the MM region, the initial carbon atom is typically set to have a zero charge. The charge is then redistributed to the neighboring MM atoms. This is done to prevent the QM zone from becoming overpolarized by the nearby MM atom.[262]

3. Molecular Dynamics Simulations

The previous section introduced methods for calculating molecular energies, ranging from inaccurate and computationally efficient to highly accurate and costly. This chapter deals with the motion of molecules. In the first subchapter, molecular geometries are optimized without considering the time dimension. Then, nuclei are moved in time, leading to molecular dynamics simulations. The last subchapter focuses on non-adiabatic molecular dynamics to analyze the dynamics of electrons via the time-dependent Schrödinger equation.

3.1. Geometry Optimization

There are numerous reasons to optimize a molecular structure. A small molecule can be optimized with great precision to extract information such as bond lengths or dipoles. In contrast, the experimentally determined X-ray structure of a large biomolecule must be optimized to avoid issues like atom clashes or overstretched bonds. In any case, a static minimum on the potential energy surface (PES) is searched. Finding the optimal geometry of a molecule is equivalent to minimizing the function of the potential energy $V(\vec{r})$ with the Cartesian coordinates $\vec{r} = (\vec{r}_1, \vec{r}_2, ..., \vec{r}_n) = (x_1, y_1, z_1, x_2, ..., z_n)$ of all n atoms.

Two algorithms are frequently employed in the field of geometry optimization to minimize the multidimensional hyperplane. Both methods depend on the calculation of the gradient vector $\vec{g}(\vec{r}) = \frac{dV(\vec{r})}{d\vec{r}}$. This gradient can be efficiently calculated using the force fields outlined in chapter 2.1.1. The gradient of quantum mechanical potential energies can be calculated via expensive analytical methods or by performing numerical calculations, depending on the method applied.[132]

The first algorithm is called gradient descent, which calculates the step $\vec{d}$ as the energy gradient $\vec{g} = \nabla V$ multiplied by the step size η. In each step i the coordinates are updated as:

$$\vec{r}_{i+1} = \vec{r}_i - \vec{d}(\vec{r}_i) = \vec{r}_i - \eta\vec{g}(\vec{r}_i)\,. \tag{3.1}$$

In this manner, all atoms are moved in the opposite direction of the gradient. This straightforward approach requires minimal memory and is highly computationally efficient. However, the multidimensional hyperplane of the molecular PES is typically full of shallow valleys. In this scenario, gradient descent suffers from oscillating around local minima and can also become trapped. A simple improvement is to vary η_n at each step, based on a one-dimensional minimization (line search) along the negative gradient direction. This

algorithm, called steepest descent, yields fewer total iterations at the cost of additional energy calculations per step. However, both algorithms suffer from "zig-zag" convergence in narrow, curved valleys.

Therefore, the steepest descent algorithm is frequently used as an initial optimization routine, with the more accurate conjugate gradient algorithm used to complement it. This algorithm calculates the step $\vec{d}_i$ as a combination of the current negative gradient an the previous search direction $\vec{d}_{i-1}$:

$$\vec{d}_i = -\vec{g}_i + \lambda_i \vec{d}_{i-1} \,. \tag{3.2}$$

The additional parameter λ_i can be calculated using the Polak–Ribière update, among others:

$$\lambda_i = \frac{\vec{g}_i^t(\vec{g}_i - \vec{g}_{i-1})}{\vec{g}_{i-1}^t \vec{g}_{i-1}} \tag{3.3}$$

A line search along $\vec{d}_i$ then determines the optimal step length. Due to the additional calculation and memory consumption of the previous step, the conjugate gradient algorithm is slower than the steepest descent optimization. It is considered to be more accurate and does converge faster for nearly quadratic surfaces.[132]

3.2. Adiabatic Molecular Dynamics

Up to this point, the geometries and energies have been calculated without taking the time domain into consideration. Incorporating the time element involves examining the dynamics of the molecular system. In principle, molecular dynamics (MD) simulations require costly solutions to the time-dependent Schrödinger equation.

Three key approximations enable the simulations of large molecules on long time scales. First, the Born-Oppenheimer approximation (chapter 2.3.1) is assumed to be valid, allowing to separate the dynamics of nuclei and electrons. Therefore, nuclei are propagated on a single PES (an adiabatic surface), usually the one of the ground state. Second, the force fields as described in chapter 2.1.1 are parameterized precisely for this ground state PES, resulting in a rapid calculation of the potential energy. Its negative gradient, the force

$$\vec{F}_i = -\nabla_i V = -\frac{\partial V(\vec{r})}{\partial \vec{r}_i} \tag{3.4}$$

can be efficiently calculated for each atom i due to the harmonic nature of the force fields.

Third, nuclei are treated classically which means that Newton's second law is applied as the equation of motion:

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} \,. \tag{3.5}$$

If costly QM calculations are used in place of force fields, forces for ground or excited states can be obtained. In QM/MM simulations, the total force of the system can be obtained in analogy to Equation 2.45 as:

$$F_{\text{tot}} = F_{\text{QM}} + F_{\text{MM}} + F_{\text{QM/MM}} \tag{3.6}$$

Nevertheless, the nuclei are classicaly moved on one PES via Newton's second law.

For newton's differential equation, no analytical solution is available for multi-atomic systems. However, with the acceleration a_i of each atom in each x, y and z direction, the time-dependent motion of the nuclei can be solved numerically. Integrators like the verlet, the velocity verlet, or the leap frog algorithm are essential to balance numerical accuracy with computational efficiency. The leap frog algorithm was used in this work and is based on a Taylor expansion of the coordinates around their position vector:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \frac{d\vec{r}}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{r}}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3\vec{r}}{dt^3}\Delta t^3 + \dots \tag{3.7}$$

It is correct to the third order and expresses new coordinates and velocities based on the previous ones and the time step Δt:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t + \frac{\Delta t}{2})\Delta t \tag{3.8}$$

$$\vec{v}(t + \frac{\Delta t}{2}) = \vec{v}(t - \frac{\Delta t}{2}) + \vec{a}(t)\Delta t \tag{3.9}$$

The initial coordinates and velocities for the first step are obtained from the geometry optimization as described above and from the Maxwell-Boltzmann distribution, respectively. Based on the initial conditions, new sets of coordinates are computed with the integrator, building up the so called trajectory, which contains the time-dependent movements of all the nuclei in the system.

For a simulation of length t, the described integrator has to be applied $t/\Delta t$ times to provide the trajectory of nuclear movements. As the error of the numerical integration depends on the time step Δt, its value must be carefully chosen to balance computational efficiency and accuracy. In practice, stable simulations are obtained by setting the time steps to one tenth of the fastest periodic motion in the system. Therefore, the commonly applied time step of 1 fs is justified by the vibrations of bonded hydrogen atoms that oscillate with a period of approximately 10 fs. Techniques like restraining bonds to hydrogen atoms allow a longer time step, resulting in increased computational efficiency.[132]

3.3. Non-Adiabatic Molecular Dynamics

Adiabatic MD simulations rely on the Born-Oppenheimer approximation, which is valid for near-equilibrium dynamics, where the PES of interest is well separated from the others. This is usually the case for ground state conformational changes. For the excited state,

it is more likely that PES from different electronic states come close to each other or even intersect. In these (avoided) crossing regions the PES couple, the molecular system doesn't evolve on a single PES anymore, but is in a superposition of two or more surfaces. Transitions between the adiabatic surfaces must be considered and simulations can't be restricted any longer to one adiabatic surface.

Describing the nuclear wavepacket quantum mechanically is possible with methods like the multi-configurational time-dependent Hartree (MCTDH)[306], but the high computational cost negates the study of large molecules and long time scales. Semi-classical approaches like ab initio multiple spawning (AIMS) are still computationally demanding for large systems like protein complexes. They represent the nuclear wavepacket with linear combinations of Gaussian functions along classical trajectories. The gain of computational efficiency by the approximative treatment of the nuclear wavefunction is partly compensated by the necessity to calculate a swarm of trajectories.[62]

Mixed Quantum-Classical Dynamics (MQCD) approaches like Ehrenfest dynamics or trajectory surface hopping (TSH) are an additional approximation. In the TSH approach, the wavepacket is simulated through a swarm of non-interacting, independent, classical trajectories. A fewest switches formulation was developed by Tully,[290] Tully's fewest switches surface hopping (TFSSH). It is has been used in this work and is the method of choice to study non-adiabatic dynamics, due to its good results in different fields and its simplicity and efficient implementation.[53]

3.3.1. Surface Hopping for Exciton Transfer

A molecular excitation in form of a transition e.g. from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) (figure 4.1) leads to a positively charged hole in the lower shell orbitals. The negatively charged excited electron and the corresponding hole can be considered as a quasiparticle which is bound by electrostatic Coulomb interactions. Depending on the (de)localization on single pigments, Wannier-Mott (very delocalized), charge transfer (delocalized on two sites) and Frenkel (localized) excitons can be distinguished.

Transfer of excitation energy in form of excitons can be described in two limits. In the first picture, where the donor gets de-excited and the acceptor gets excited, the coherent transfer is mediated by long-range interactions. Förster theory can describe this process taking the overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor as a critical parameter. The other limit is a fully localized hopping between neighboring sites. In this Dexter mechanism the excited electron of the acceptor is exchanged with a ground state electron of the donor. This transfer requires spatial overlap of the wave functions (or orbitals) and can be described by Redfield theory.[127]

TSH can describe both the Förster and Dexter transfer and other real life scenarios which are in between the two limits. However, several approximations are made that are detailed in the following.

To model the transfer dynamics of a Frenkel exciton in a multi-site system — whether these sites represent pigment molecules, molecular units, or lattice positions in a crystal — a coarse-grained description of the electronic structure can be obtained by the Frenkel Hamiltonian. For N sites in the system, it reads:

$$H^{\text{Frenkel}} = \sum_{i}^{N} \epsilon_i |i\rangle\langle i| + \sum_{i \neq j}^{N} V_{ij} |i\rangle\langle j| \,. \tag{3.10}$$

The site energies ϵ_i describe the transition from the pigments' ground state to the excited state of interest. This approximation is valid when a dominant transition (such as the HOMO-LUMO transition) doesn't compete with other excitations, such as charge transfer excitations between sites. The electronic coupling, the interaction between the pigments, are represented by the second term in equation 3.10. Its magnitude, relative to the phonon-interaction with the environment, determines the (de)localization of the exciton. Therefore, it is a key quantity to determine the rate and mechanism of the transfer. When the distances between sites are large enough, exchange interactions and overlap between orbitals can be neglected. Then, the coupling is dominated by Coulomb interactions between transition densities ρ^{tr}.

$$V_{ij} = \iint \frac{\rho_i^{\text{tr}^*}(\vec{r})\, \rho_j^{\text{tr}}(\vec{r}')}{|\vec{r} - \vec{r}'|} \, d\vec{r}\, d\vec{r}' \tag{3.11}$$

For a molecular transition from the ground state (0) to an excited state (e) the transition density can be described with the density operator $\rho(\vec{r})$ as

$$\rho^{\text{tr}} = \langle \Psi_0 | \rho(\vec{r}) | \Psi_e \rangle \,. \tag{3.12}$$

Just like transition dipoles, it does not have a classical analog. As the accurate calculation of the transition densities and their interaction is computationally demanding, many approximations exist. In the simplest but very commonly used approximation the transition densities are expanded up to dipolar terms, leading to the point transition dipole (PDA) approximation:

$$V_{ij}^{\text{PDA}} = \frac{\mu_i \mu_j}{r_{ij}^3} - \frac{3(\mu_i r_{ij})(\mu_j r_{ij})}{r_{ij}^5} \,, \tag{3.13}$$

with the transition dipole moments μ and the distances r_{ij} between two sites. Other approximations like the transition charges from electrostatic potentials (TrEsp) approach or the transition density cube (TDC) method are considered to be more accurate than the PDA approximation.[127] Another approximation, which has been applied in this work, is the Coulomb coupling between atomic transition charges Q that can be obtained via a Mulliken population analysis.[221] With atoms a and b of sites i and j, respectively, the Coulomb coupling reads:

$$V_{ij}^{\text{Coul}} = \sum_{a \in i, b \in j} Q_a Q_b \zeta_{ab}(|\vec{r}_a - \vec{r}_b|) \,, \tag{3.14}$$

with the function ζ being defined based on atomic orbitals Φ as:[221]

$$\zeta_{ab}(|\vec{r}_a - \vec{r}_b|) = \iint d\vec{r} d\vec{r}' \, \frac{\Phi_a(\vec{r}) \Phi_b(\vec{r}')}{|\vec{r} - \vec{r}'|} \,. \tag{3.15}$$

With the electronic couplings between the sites and the site energies, the Frenkel Hamiltonian (eq. 3.10 is constructed in the site basis, or diabatic basis. The wave function of the exciton can be described as a linear combination of local excitations:

$$\Psi^{\text{tot}} = \sum_i c_i |\psi_i\rangle \,. \tag{3.16}$$

Here, the expansion coefficients c_i of the diabatic wave function ψ_i of the sites express the contribution of each local excitation to the single excitation in the multi-site system. Their squares $|c_i|^2$ are called the diabatic populations of the exciton. Among the infinite numbers of basis vectors to describe the exciton in the system, the site basis is the most intuitive representation.

Diagonalizing the diabatic Frenkel Hamiltonian yields the adiabatic Hamiltonian with the energies E_m of the respective m adiabatic surfaces on the main diagonal:

$$H^{\text{diab}} = \begin{bmatrix} \epsilon_1 & V_{12} \\ V_{12} & \epsilon_2 \end{bmatrix} \quad \rightarrow \quad H^{\text{adiab}} = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} . \tag{3.17}$$

For the special case of two sites having the same site energy, the adiabatic energies are separated by twice the coupling interaction: $E_{1,2} = \epsilon \pm V$. The diabatic-to-adiabatic transformation matrix U, which is calculated from the diagonalization

$$H^{\text{adiab}} = U^{-1} H^{\text{diab}} U \,, \tag{3.18}$$

contains the diabatic occupations b_{im}, which are the expansion coefficients for the m^{th} adiabatic surface with respect to the i^{th} pigment (diabatic site) in the system. Thus, the excitonic wave function Ψ^{tot} can be expressed in the eigenbasis (adiabatic basis) Ψ_m or in the site basis (diabatic basis)ψ_i

$$\Psi^{\text{tot}} = \sum_m a_m \Psi_m = \sum_m \sum_i b_{im} \psi_i = \sum_i c_i \psi_i \,, \tag{3.19}$$

with the adiabatic populations, diabatic occupations and diabatic population a, b and c, respectively. All three representations have their own use case and can be transformed into each other with the transformation matrix U.

During the simulation of the exciton transfer, the exciton is treated quantum mechanically. Inserting equation 3.19 into the TDSE leads to the evolution of the expansion coefficients c_i:

$$\dot{a}_i = -i \sum_j a_j H_{ij} - \sum_j \langle \psi_i | \dot{\psi}_j \rangle \,. \tag{3.20}$$

For large distance between the sites, the term on the right hand side is typically neglected.[148,105]

By contrast, the nuclei are propagated classically on one particular adiabatic surface (active surface) using newton's second law (eq.3.5). Intersections of adiabatic surfaces (trivial crossings) can be captured by state tracking algorithms.[307] Transitions between surfaces

are the key concept in surface hopping methods. The probability for a hop between two surfaces n and m is described as

$$P_{m\to n}^{\mathrm{FSSH}} = \max\left\{0, \frac{2\,\vec{d}_{mn}\,\mathrm{Re}\left(a_n^* a_m\right)}{|a_m|^2}\,\Delta t_{\mathrm{hop}}\right\} \tag{3.21}$$

with the nonadiabatic coupling vector $\vec{d}_{mn} = \langle\Psi_i|\partial_t|\Psi_j\rangle$, the adiabatic expansion coefficients a (Re denotes the real part of a complex number) and the time step between hopping attempts Δt_{hop}.[148] The stochastic nature of the process is emulated with a random number that is generated for any hopping attempt and compared with the hopping probability to determine whether a hop takes place or is rejected.

When a hop takes place, the energy of the adiabatic surface changes abruptly. To ensure energy conservation and the correct Boltzmann statistics, the momenta of the nuclei are scaled along the direction of the nonadiabatic coupling vector. Alternatively, the hopping probability $P_{m\to n}$ itself can be multiplied with the Boltzmann factor g leading to the computationally efficient Boltzmann-corrected FSSH (BC-FSSH) method.

$$g_{mn}^{\mathrm{BC}} = \begin{cases} \exp\left(-\dfrac{E_n - E_m}{k_B T}\right) & \text{if } E_n > E_m \quad \text{(upwards hop)} \\ 1 & \text{if } E_n \leq E_m \quad \text{(downwards hop)} \end{cases} \tag{3.22}$$

So far, calculations based on the molecular properties at given geometries were presented, that determine the dynamics of the exciton. The feedback, i.e. the effect of a hopped exciton on the nuclei in the system is the key to describe the interdependent movements of nuclei and electrons.

As described in chapter 2.4 the total energy of the system E_{tot} can be subdivided into contributions from the QM and MM regions and from the interaction between these two: $E_{\mathrm{tot}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}$. In the present work, a so-called subtractive QM/MM scheme is applied. Here, the force field is applied to all atoms in the system and adjustments are made to account for the QM nature of the process:

$$E_{\mathrm{tot}} = E_{\mathrm{MM}} + \Delta E_{\mathrm{QM}}^{\mathrm{ex}} + \Delta E_{\mathrm{QM/MM}} \tag{3.23}$$

$\Delta E_{\mathrm{QM}}^{\mathrm{ex}} = E_{\mathrm{ex}} - E_{\mathrm{ground}}$ is the energy difference of the ground and excited states in the QM region and $\Delta E_{\mathrm{QM/MM}}$ quantifies the effect of excitations to the interaction of the QM region with its surrounding MM region.

If an exciton hops from one pigment to the other, the acceptor gets excited and is out of the minimum of the PES of the excited state (figure 4.1). During the relaxation into the new minimum, the molecule looses an energy called relaxation or reorganization energy. It can be subdivided into outer (effect of the solvation shells) and inner (rearrangement of the QM molecule) reorganization energy. If explicit forces from the excited state PES are not available, the effect of internal reorganization can be described by a phenomenological equation:[152] The internal reorganization energy λ_{vac} is calculated once with a high level

theory and gets subtracted from the site energies of all pigments according to their population:

$$H_{ii} = H_{ii}^0 - |c_i|^2 \lambda_{vac} \tag{3.24}$$

This implicit relaxation scheme incorporates the effect of $\Delta E_{\mathrm{QM}}^{\mathrm{ex}}$.

To calculate $\Delta E_{\mathrm{QM/MM}}$, one can assume that van der Waals interactions are independent of the electronic state. Given that the force field already covers electrostatic and van der Waals interactions, the effect of excited states shrinks down to an electrostatic interaction of altered partial charges in the QM zone. Therefore, in every step of the NAMD simulation, the force field charges of the QM zones can be updated according to their occupation:

$$q_{ai} = q_{ai}^0 + |b_{im}|^2 \Delta q_a \tag{3.25}$$

Here, the partial charge of atom a in pigment i is modified with the charge difference between excited and ground state Δq, weighted with the diabatic occupation b_{im} (eq. 3.19) of the active surface m.

The coherence length or delocalization length estimates the number of sites on which the exciton is delocalized. There are numerous ways to quantify the delocalization length, including the occupations of the pigments or the autocorrelation function of the wave function[38].

Two quantities that only rely on the diabatic occupations of a certain adiabatic state m are the inverse participation ratio (IPR)

$$\mathrm{IPR} = \frac{1}{\sum_i |b_{im}|^4} \tag{3.26}$$

and an alternative quantity suggested by Jang, Dempster and Silbey in 2001[125] (JDS)

$$\mathrm{JDS} = \sum_n \min\{1, 2N_c |b_{in}|^2\}| \,. \tag{3.27}$$

4. Absorption Spectra

4.1. Molecules and Light

Light is a paragon of the wave-particle duality, being both an electromagnetic wave and a stream of discrete particles called photons. Photons are massless quantum particles which carry both energy and momentum. While the intensity of the electromagnetic wave is proportional to the probability density (or number) of photons, its frequency determines the energy of each single photon. According to Planck's relation, this energy is given by

$$E = h\nu \tag{4.1}$$

where h is Planck's constant and ν is the frequency of the wave. Because the speed of light c is constant in a given medium, the frequency ν of the electromagnetic wave is unambiguously related to the wavelength λ and the wavenumber $\tilde{\nu}$ by

$$\nu = \frac{c}{\lambda} = c\tilde{\nu} \ . \tag{4.2}$$

The visible spectrum of light ranges from high energetic violet light (λ =420 nm, E≈2.95 eV) to low energetic red light (λ =700 nm, E≈1.8 eV).

The quantum mechanical nature of atoms and molecules leads to discrete stationary states, which are the solutions of the stationary Schrödinger equation (Equation 2.6). The solutions in form of eigenvalues and eigenfunctions are characterized by quantum numbers that define the allowed electronic, vibrational, and rotational energy levels. Beyond the ground states with quantum number 0, molecules have an infinite number of excited states. All quantum degrees of freedom (electronic, vibrational, rotational) can be excited independently, resulting in a rich spectrum of possible stationary states.

When the energy difference between one of the possible states and the current state of the molecule matches the energy of a photon that encounters the molecule, this photon can be absorbed and the molecule gets excited. However, only certain transitions are allowed based on selection rules derived from conservation laws, like the conservation of angular momentum.

The probability that a photon gets absorbed by the molecule depends on the transition dipole moment μ of the molecule, defined as:

$$\hat{\mu} = -e \sum_i \vec{r}_i \tag{4.3}$$

with the elementary charge e and the vectors $\vec{r}_i$ connecting the center of charge with the respective electrons. The squared transition dipole's expectation value is proportional to the oscillator strength which determines the probability of a transition and therefore, the experimentally measured intensity of absorption. In atomic units, the oscillator strength f_{ge} for a transition from the electronic ground state g to an excited state e is defined as

$$f_{ge} = \frac{2}{3}\Delta E_{ge}|\vec{\mu}_{ge}|^2 \ , \tag{4.4}$$

with the energy gap ΔE_{ge} between the two states.

Under the Born–Oppenheimer approximation, the total molecular wavefunction Ψ_{tot} is factorized into an electronic part ψ and a vibrational part ϕ. The expectation value of the transition dipole moment becomes:

$$\begin{aligned}
\langle\hat{\mu}_{ge}\rangle &= \langle\Psi^g_{\text{tot}}|\hat{\mu}|\Psi^e_{\text{tot}}\rangle \\
&= \langle\psi^g\phi^g\ |\hat{\mu}|\ \psi^e\phi^e\rangle \\
&= \langle\psi^g(r;R)\phi^g(R)\ |\hat{\mu}(r)|\ \psi^e(r;R)\phi^e(R)\rangle \\
&= \underbrace{\langle\psi^g|\hat{\mu}|\psi^e\rangle}_{\mu_{\text{el}}} * \underbrace{\langle\phi^g|\phi^e\rangle}_{\text{S}}
\end{aligned} \tag{4.5}$$

Here, μ_{el} is the electronic transition dipole, also known as extinction coefficient. Its calculation is based on solving Casida's equation (Equation 2.44) which is necessarily performed in TD-DFT calculations. Therefore, the electronic transition dipole and its resulting oscillator strength is written out by many QM software by default. The second term in Equation 4.5, S, is the vibrational overlap integral. Since the dipole operator acts only on electronic coordinates, the vibrational overlap can be factored out. Its squared value $|\text{S}|^2$ is called Franck-Condon factor. In contrast to μ_{el}, evaluating Franck–Condon factors requires costly geometry optimization and normal mode analysis in both ground and excited states.[240]

Due to the large difference in energy scales, the electronic transition defines the position of the absorption peak and the vibrational levels together with the Franck-Condon factors modulate the vibronic fine structure of the spectrum in form of vibrational side bands. A schematic representation of the absorption spectrum consisting of vibrational side bands is shown in Figure 4.1. Here, the nuclear overlap integrals $\langle\Phi|\Phi\rangle$ dictate the intensities of the corresponding absorption peaks.

Under the Born–Oppenheimer approximation (see section 2.3.1), the large mass difference between nuclei and electrons allows the separation of nuclear and electronic motion. The Franck–Condon approximation similarly assumes that the nuclear positions are effectively frozen during electronic transitions. Therefore, in Figure 4.1 these transitions are vertical on the potential energy surfaces. Each vibrational state can be subdivided into rotational states. They are typically studied using infrared (IR) spectroscopy and involve energies on the order of a few meV - significantly less than the ≈2 eV required for electronic excitations in the UV/visible spectrum. Nevertheless, upon electronic excitation, vibrational quantum numbers often change, leading to coupled electronic-vibrational (vibronic) transitions.

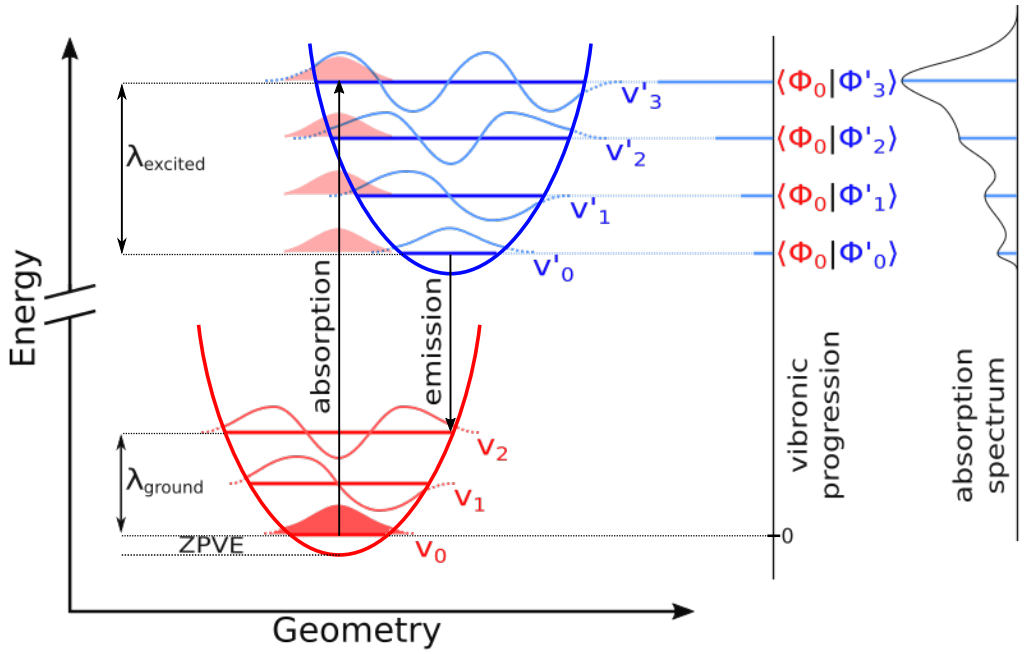


Fig. 4.1.: Scheme of harmonically approximated potential energy surfaces of the electronic ground (red) and the first electronic excited (blue) state. Vibrational energy levels and their corresponding nuclear wave functions are denoted as v_n and Φ_n, respectively. The zero-point vibrational energy (ZPVE) is only shown for the ground state for clarity. The vibronic progression starting from the vibrational ground state leads to sharp peaks whose intensities are related to the nuclear overlaps $\langle\Phi|\Phi\rangle$. The absorption spectrum is formed by the broadened and overlapping peaks. Transitions from higher vibrational states of the electronic ground state ("hot bands") are not shown for clarity. Reorganization energies λ for ground and excited states are shown and discussed in detail in the next sections.

Beyond the vibronic structure, additional broadening mechanisms result in the overlapping peaks in Figure 4.1 that are typically observed in experimental absorption spectra. These broadening effects are classified as *homogeneous* and *inhomogeneous* broadening. The former affects all molecules in an ensemble identically and includes effects such as Doppler broadening due to thermal motion, natural or lifetime broadening due to Heisenberg's uncertainty principle, collisional broadening, and thermal geometry fluctuations. These effects are generally modeled using exponential decay functions, leading to Lorentzian line shapes due to the Fourier transform relation.

In contrast, *inhomogeneous* broadening arises from different molecular environments across the ensemble, such as different protein binding pockets or the presence of different molecules or isomers. When not specifically computed, this effect is often approximated using a Gaussian distribution, assuming a statistical distribution of excitation energies.

4.2. Calculating Absorption Spectra

Based on the theory of broadening and electronic transitions outlined above, the most straightforward approach to simulating a molecular absorption spectrum is to optimize

the ground-state geometry and compute the vertical excitation energy (VEE), i.e., the electronic energy gap ΔE between electronic ground and excited states. The resulting transition is typically broadened with a Lorentzian or Gaussian function.

In this approximation, the absorption spectrum $\sigma(\nu)$ as a function of the photon frequency ν consists of a single Dirac delta peak. Its intensity is proportional to the electronic oscillator strength f_{ge}, which depends on the squared transition dipole moment μ_{el} (Eq. 4.3). The peak is positioned at the excitation energy ΔE_{ge} evaluated at the ground-state minimum geometry R_0^g:

$$\sigma_{\text{VEE}}(\nu) = |\mu_{\text{el}}^2|\delta(\nu - \Delta E_{ge}(R_g^0)) \tag{4.6}$$

The electronic energy gaps ΔE are computed from the minimum of the ground state PES to the excited state PES at the given geometry. Thus, the zero-point vibrational energy (ZPVE, see Figure 4.1) of the ground state is neglected and the energy of an artificial excited state, most often between two allowed vibrational eigenstates (between v'$_1$ and v'$_2$ in Figure 4.1), is considered. While the energy gap between two physically non-existing states is calculated, the VEE method still reliably captures relative trends across molecules due to the different orders of magnitude of electronic and vibrational transitions. Since geometry optimizations are typically performed in implicit solvent, specific solvent interactions, like hydrogen bonding or protein binding pockets, are not included.

To explicitly account for these broadening effects, the so-called ensemble method samples a distribution of nuclear configurations and performs VEE calculations for each geometry. The resulting spectrum is constructed as a histogram of transition energies weighted with the corresponding electronic oscillator strengths from a set of N geometries:

$$\sigma_{\text{Ensemble}}(\nu) = \sum_{i}^{N} |\mu_{\text{el},i}|^2 \delta(\nu - \Delta E_{ge}(R_i)) \tag{4.7}$$

The set of excitation energies is uncorrelated, therefore, the ensemble approach neglects the temporal correlation between configurations and energy gaps. The nuclear geometries are assumed to be statistically independent. Accordingly, the ensemble can be generated via Monte Carlo sampling, Wigner sampling, or classical molecular dynamics (MD) simulations. If force-fields are applied for the sampling, the geometry mismatch between the methods for sampling and VEE calculations are a possible error source.

Both the VEE and ensemble methods only use the electronic dipoles and neglect the vibronic structure. They omit the effects of zero-point vibrational energy and cannot capture vibrational side bands that appear in experimental spectra. To address this, the Franck–Condon method incorporates vibronic states by computing the energy gaps between all vibrational states ω of the ground and excited electronic states. These transitions are weighted by their corresponding Franck–Condon factors $|\langle\omega_g|\omega_e\rangle|^2$ and the population of the vibrational levels in the ground state, represented by the ground-state vibrational density matrix ρ_g:

$$\sigma_{\text{FC}}(\nu) = \sum_{\omega_g} \rho_g \sum_{\omega_e} |\langle\omega_g|\omega_e\rangle|^2 \delta(\nu - \Delta E_{\omega_g,\omega_e}) \tag{4.8}$$

Although the FC method can reproduce vibronic features, it is computationally costly for large molecules due to the need for geometry optimizations and normal mode analyses in both the ground and excited states. For the same reason, it is restricted to implicit solvent models and typically cannot address explicit environments, like protein pockets.

A combination of the ensemble and the Franck-Condon method, called zero temperature Franck-Condon (E-ZTFC) method has recently been introduced.[316] It is motivated by the idea of explicitly sampling the specific solvent or protein environment of the chromophore and performing geometry optimizations in frozen environment, thereby combining the positive effects (explicit solvent and vibronic states) of both methods. In order to reduce the computational costs, the density matrix of the electronic ground state ρ_g is approximated as the low temperature limit, i.e. the vibrational ground state $|0\rangle$ of the electronic ground state:

$$\sigma_{\text{E-ZTFC}}(\nu) = \sum_{i}^{N_{\text{frames}}} \sum_{\omega_e} |0\rangle\langle 0||\langle \omega_g^0|\omega_e\rangle|^2 \delta(\nu - \Delta E_{\omega_g^0,\omega_e}) \tag{4.9}$$

While the E-ZTFC method offers a promising compromise between accuracy and computational costs, it suffers from calculating the vibrational overlap integrals for large and anharmonic systems, and breaks down for strong solute-solvent interactions that couple to the vibrational modes, such as proton sharing.[317]

The methods discussed thus far are primarily time-independent. Even though the Franck–Condon method incorporates dynamic information about the nuclei in form of vibrational structure, it does not explicitly consider the temporal evolution of molecular geometries. However, time-dependent methods that link spectral features to dynamical fluctuations of molecular properties, are of great interest. From a fundamental standpoint, the absorption spectrum can be obtained via the Fourier transform of the autocorrelation function of the transition dipole moment. The time-dependent linear response function $\chi(t)$ captures how the dipole responds to an external perturbation (the photon), and the spectrum $\sigma(\nu)$ is given by the Fourier-transformation of the fluctuations of the transition dipole (Equation. 4.10).

$$\sigma(\nu) = \alpha(\nu) \int_{-\infty}^{+\infty} dt\, e^{i\nu t}\, \chi(t), \quad \text{with} \quad \chi(t) = C_\mu(t) = \langle \vec{\mu}(t)\vec{\mu}(0)\rangle \tag{4.10}$$

The prefactor $\alpha(\nu)$ depends on the precise definition of $\sigma(\nu)$ and is often set to unity.[317] The function $C_\mu(t)$ is the autocorrelation function of the transition dipole and connects time-domain fluctuations of molecular properties to their frequency-domain spectral features.

A particularly promising approach within this time-dependent regime is the cumulant expansion method. Based on linear response theory, this method expands the logarithm of the transition dipole fluctuations in terms of cumulants of the energy gap fluctuations.[170] Within this formalism, the response function $\chi(t)$ from Eq. 4.10 can be approximated as:

$$\chi(t) \approx |\mu_{ge}|^2 e^{-i\Delta E_{ge}^{\text{av}} t} \exp[-G_\infty(t)] \tag{4.11}$$

where μ_{ge} is the transition dipole moment between ground and excited states, $\Delta E_{e,g}^{\text{av}}$ is the time-averaged vertical excitation energy, and the exponential damping term $G_\infty(t)$ is the

full line shape function, which captures the effect of nuclear motion on the absorption line shape. It is expanded in a cumulant series with

$$G_m(t) = \sum_{n=2}^{m} g_n(t) \tag{4.12}$$

being the m^{th} order approximation to the full line shape function $G_\infty(t)$ and $g_n(t)$ denotes the n^{th}-order cumulant term. Whereas the first-order term vanishes for systems at equilibrium, the leading contribution arises from the second-order term, which is expressed in terms of energy gap fluctuations:

$$g_2(t) = \frac{1}{\pi} \int_0^\infty d\omega \, \frac{J(\omega)}{\omega^2} \left[\coth\left(\frac{\beta\omega}{2}\right) [1 - \cos(\omega t)] - i \, [\sin(\omega t) - \omega t] \right] \tag{4.13}$$

Here, $\beta = 1/k_B T$ is the inverse thermal energy and covers the temperature dependence of the spectrum. This aspect together with the vibrational frequencies ω will be discussed in more detail in the following section. If the energy gap fluctuations obey Gaussian statistics, then $g_n(t) = 0$ for $n \geq 3$ and the second-order approximation becomes exact for calculating absorption spectra.[191] This condition is reasonably satisfied for weakly anharmonic systems. The second-order line shape function depends on the spectral density $J(\omega)$.

4.3. Spectral Density

This is a frequency-dependent function that quantifies the system-bath coupling in form of vibrational modes that modulate the energy gap.[182] Experimentally, $J(\omega)$ can be determined via fluorescence line-narrowing spectroscopy, where peak intensities are proportional to the Huang–Rhys factors that describe the electron–phonon coupling strength.

The frequencies ω in the spectral density correspond to the vibrational eigenmodes of the system. These are the same modes that appear in the Franck–Condon treatment of vibronic transitions (Equation 4.8 and Figure 4.1). Each delta peak in the absorption spectrum, of the form

$$\nu = \Delta E_{\omega_g,\omega_e} = \Delta E_{00} + \sum_k n_k \omega_k \ , \tag{4.14}$$

corresponds to the absorption of a photon whose energy differs from the 0-0 transition (ΔE_{00}) by one or more quanta n_k of mode k. Peaks in $J(\omega)$ thus appear at the vibrational mode frequencies ω_k.

The vibrational levels in this context are typically modeled in a harmonic way, so that the potential energy surfaces of the ground and excited states are approximated as parabolic (Figure 4.1). Therefore, the vibrational eigenenergies are considered to be $E_n = \hbar\omega(n + \frac{1}{1})$

with n being the vibrational quantum numbers (n=0,1,2,...). However, for the anharmonic Morse potential, the energy levels are given by

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) - \hbar\omega\alpha\left(n + \frac{1}{2}\right)^2 \tag{4.15}$$

with the anharmonicity constant α. For full anharmonic systems, there are no analytic solutions for the vibrational eigenstates, which limits the calculation of spectral density and accordingly the robustness of the FCE approach for strong anharmonic systems.[318]

Theoretically, the spectral density can be expressed via the Wiener–Khinchin theorem, which relates the Fourier transform of a time correlation function to its power spectral density. The quantum mechanical expression for the spectral density is:

$$J(\omega) = i\theta(\omega)\int dt\, e^{i\omega t}\,\mathrm{Im}\, C_{\delta\Delta E}(t)\ , \tag{4.16}$$

with θ being the Heavside step function and

$$C_{\delta\Delta E}(t) = \langle \delta\Delta E(\hat{q}, t)\delta\Delta E(\hat{q}, 0)\rangle \tag{4.17}$$

being the time correlation function of energy gap fluctuations with the fluctuation $\delta\Delta E(\hat{q}, t) = \Delta E(\hat{q}, t) - \langle\Delta E\rangle$ and the nuclear coordinates $\hat{q}$ treated as quantum operators.

Because the exact time correlation function can only be calculated for simple Brownian Oscillator type models, numerous classical versions exist which are based on Kubo-transformations of the quantum correlation functions.[192] In practice, the spectral density is frequently computed using the cosine Fourier transform of the classical energy gap autocorrelation function.[209]

$$J(\omega) = \frac{\beta\omega}{\pi}\int_0^\infty dt\, C^{\mathrm{cl}}_{\delta\Delta E}(t)\cos(\omega t) \tag{4.18}$$

Here, $\beta = 1/k_B T$ is the inverse thermal energy. This prefactor is crucial within the harmonic approximation to compensate for the thermal distribution inherent in the classical energy gap correlation function $C^{\mathrm{cl}}_{\delta\Delta E}(t)$. The resulting spectral density is considered to be temperature-independent[293] and can be reused across different temperatures.

The classical correlation function can be obtained from a molecular dynamics trajectory with N frames as[63]:

$$C^{\mathrm{cl}}_{\delta\Delta E}(t_j) = \frac{1}{N-j}\sum_{k=1}^{N-j}\Delta E(t_j + t_k)\Delta E(t_k) \tag{4.19}$$

Integrating the spectral density obtained from Equation 4.18 yields the reorganization energy λ.[21]

4.4. Reorganization Energy

The reorganization energy λ is a fundamental quantity that quantifies the nuclear response to an electronic transition. It originates from the vertical electronic transitions and the displaced PESs of the ground and excited states. After the absorption or emission of a photon, the nuclear geometries are typically far from the minimum of the new PES (see Figure 4.1). λ then quantifies the energy required to reorganize the nuclear geometry of a molecule and its environment into the new minimum geometry. It can be subdivided into outer (effect of the solvation shells) and inner (rearrangement of the molecule) reorganization energy. This quantity plays a key role in the theoretical treatment of both absorption spectra, and charge or exciton transfer.

For transfer processes, it has a significant impact on the dynamics. In the Marcus rate equation

$$k = \frac{2\pi}{\hbar}|V|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} \exp\left[-\frac{(\Delta G^\circ + \lambda)^2}{4\lambda k_B T}\right] \quad , \tag{4.20}$$

that includes the electronic coupling V between donor and acceptor and the Gibbs free energy change ΔG° of the reaction, the reorganization energy λ enters the exponent quadratically. Thus, the rate k is highly sensitive to even moderate changes in λ. In the surface hopping simulations used in this work, the reorganization energy effectively reduces the energy gap of the pigments via implicit relaxation (see subsection 3.3.1).

In the context of absorption spectra, the molecular property λ is observed as broadened absorption peaks. Larger values of λ imply stronger coupling of electronic transitions to nuclear motion. In particular, the greater the reorganization energy, the more displaced the ground and excited state PESs are, resulting in a broader distribution of vibronic transitions. In the ensemble method (section 4.2), λ correlates with the spread in excitation energies across geometric fluctuations. In the cumulant expansion framework, λ determines the width of the second-order line shape function $g_2(t)$ (Equation 4.13), which controls the damping of the dipole autocorrelation function and thus finally the spectral broadening. In summary, the reorganization energy connects geometry fluctuations to the energy gap fluctuations, thereby bridging microscopic nuclear dynamics with macroscopic spectral line shapes.

Besides calculating this critical parameter from the spectral density (λ_{spd}), the reorganization energy can also be obtained from the variance of the excitation energy distribution (λ_{var}), assuming classical Gaussian statistics. Experimentally, it is determined by the Stokes shift (λ_{stokes}), which is defined as the difference between the absorption and emission peak energies. The three formulations read:

$$\begin{aligned} \lambda_{\text{spd}} &= \frac{2}{\pi}\int_0^\infty d\omega\, \frac{J(\omega)}{\omega} \\ \lambda_{\text{stokes}} &= E_{\text{absorption}} - E_{\text{emission}} \\ \lambda_{\text{var}} &= \frac{\sigma^2_{\Delta E}}{2kT} \end{aligned} \tag{4.21}$$

Under the harmonic approximation, where both the ground and excited state PESs are assumed to be parabolic and of equal curvature, all three definitions are equivalent:[21]

$$\lambda_{\text{spd}} = \lambda_{\text{stokes}} = \lambda_{\text{var}} \tag{4.22}$$

However, in realistic systems, the three formulations can yield differing results. The ground and excited states reorganization energies shown in Figure 4.1 can strongly deviate from each other. The stokes shift being the sum of λ_{ground} and λ_{excited} can be calculated with the so called 4-point scheme via geometry optimization in ground and excited states and subsequent energy evaluations of each structure in both states. Due to the geometry optimization in the excited state, this approach is demanding and error prone, particularly for large molecules. For the Bacteriochlorophyll molecules studied in this work, reported values for λ vary significantly from 280 $\text{cm}^{-1}$to more than 10,000 cm^{-1}depending on the DFT functional and basis set.[229]

4.5. Absorption Spectra in Coupled Systems

The absorption spectrum of a multichromophoric system, like light harvesting protein-pigment complexes, can be calculated in analogy to the absorption spectrum of single molecules. However, the coupling between the pigments give rise to delocalized adiabatic (excitonic) states (subsection 3.3.1). For n pigments giving rise to n adiabatic eigenstates, the ensemble method (Equation 4.7) includes all n states per snapshot of the trajectory. The transition dipoles of an adiabatic state m can be obtained by a linear combination with the diabatic populations c_i^m per pigment i (Equation 3.19) and the pigment specific dipoles μ_i:

$$\mu_m = \sum_i c_i^m \mu_i \quad . \tag{4.23}$$

Just as for isolated molecules, the time-independent ensemble averages of energy gaps and transition dipoles are insufficient to reproduce experimental absorption spectra of excitonic systems.[45]

To address the time-dependency of the energy gap fluctuation, the cumulant expansion method is extended to coupled systems. The absorption spectrum is computed as a sum over all excitonic states m with their corresponding transition dipole μ_m, excitonic energy gap ν_m, and a phenomenological lifetime τ_K. The expression reads:

$$\sigma(\nu) \propto \nu \sum_m |\mu_m|^2 \int_{-\infty}^{\infty} e^{-i(\nu_m - \nu)t - g_m(t) - t/\tau_K} \, dt \tag{4.24}$$

The line shape function g_m can be calculated comparable to the transition dipoles (Equation 4.23) with the excitonic coefficients, namely,

$$g_m(t) = \sum_i |c_i^m|^4 \, g_i(t) \quad , \tag{4.25}$$

with $g_i(t)$ being the line shape function of chromophore i. Thus, the fluctuations are assumed to be uncorrelated between pigments and each chromophore's spectral density contributes independently to the excitonic line shape.[45]

In summary, the description of absorption in multichromophoric systems requires both an accurate dynamical treatment of nuclear geometries (sampling) and a precise excitonic representation. The excessive benchmarking of TD-LC-DFTB2 for site energies and couplings[25], together with the parameters DFTB3-3OB-f developed for vibrational properties,[92] were finally evaluated in this work in form of absorption spectra.

5. Machine Learning

After exploring the physical aspects of energies and geometries of molecules, this chapter introduces data-driven methods for predicting energies and geometries. Machine learning has changed our daily lives and continues to shape the work of theoretical chemists, particularly in the light of the Nobel Prize awarded in 2024 for machine-learned protein structure prediction. The chapter begins with an introduction to machine learning, followed by an examination of neural networks, which are the core technology of many ML applications. The third subchapter presents special features in the treatment of molecules, while the final subchapter provides an overview of AlphaFold 2 (AF2).

5.1. Foundations of Machine Learning

Machine learning (ML) is based on the idea that the relationship between arbitrary data can be described by a function that maps a set of descriptive variables (features, input, x) to a set of response variables (output, targets, labels, y).

$$\mathbf{y} = f(\mathbf{x}) \tag{5.1}$$

Rather than directly finding the exact and often sophisticated function, ML models use a data-driven approach in which the data is used to iteratively optimize the parameters of a predefined calculation specification. In many cases, this calculation specification is simple, but long, such as a certain type of extensive linear combination. After the training, the hidden function f is approximated with the optimized parameters w_i of the predefined calculation specification, and the trained model can now be applied to unknown data.

$$\mathbf{y} = f(\mathbf{x}) \approx g(\mathbf{x}, w_i) \tag{5.2}$$

The large field of ML can be divided into two main concepts: unsupervised and supervised ML. The former deals with data sets that don't have (known) target values. Therefore, the function $f(x)$ is approximated to discover patterns or relationships in data without known output labels. In the supervised learning approach, which was also applied in this work, the parameters w_i are adjusted to express the known targets.

Supervised learning is divided into regression and classification tasks. In either case, all available data is split into a training, a validation, and a test dataset. During the training, the goodness of a model's prediction must be quantified in order to adjust the parameters w_i in a certain direction. This so-called loss function, or cost function, is partially derived

to obtain gradients for all parameters. Commonly used loss functions in regression tasks are the mean absolute error (MAE) or the mean squared error (MSE). With N data points, they are based on the difference between the known reference data points y_i^{ref} and the predictions of the model y_i^{pred}.

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{i=1}^{N} \left| y_i^{\text{ref}} - y_i^{\text{pred}} \right| \\ \text{MSE} &= \frac{1}{N} \sum_{i=1}^{N} \left(y_i^{\text{ref}} - y_i^{\text{pred}} \right)^2 \end{aligned} \tag{5.3}$$

While the trainable parameters are changed iteratively, the loss function itself must be selected in advance and remains constant throughout the training process. These parameters, known as hyperparameters, define the concrete architecture of the models. Other examples of hyperparameters are the number of trainable parameters, the number of epochs that all training data is used for training, or the learning rate, which determines the amount of adjustment of a parameter with respect to its gradient. During a hyperparameter search, the validation set is used to compare models with different architectures. In addition, the validation set can be used as a diagnostic tool during the training, preventing the model from overfitting, which occurs when the parameters are adjusted in a way that predicts the training data too closely, making it unable to generalize to new, unseen data points.

Once the training is complete due to a given convergence criterion or a maximum number of epochs, the model is evaluated on the test set. These data points have been excluded entirely during model selection and training. Therefore, the test data reflects real-life scenarios, and several metrics exist to quantify the model's accuracy. In the context of regression tasks, the coefficient of determination (R^2) is a frequently utilized metric.

$$R^2 = 1 - \frac{\sum_i^N \left(y_i^{\text{ref}} - y_i^{\text{pred}} \right)^2}{\sum_i^N \left(y_i^{\text{ref}} - \bar{y}^{\text{ref}} \right)^2} \tag{5.4}$$

The neural network (NN) model is a versatile architecture suitable for both simple and complex scenarios in both classification and regression tasks.

5.2. Neural Networks

In many chemical tasks, artificial neural networks (ANNs), also known as NNs, are the preferred model due to their high flexibility, predictive accuracy, and computational efficiency once trained. The core of NNs are the perceptrons. Their development can be traced back to the work of Frank Rosenblatt and others in the 1950s and 1960s.[199] A perceptron is capable of producing one binary output from multiple binary inputs.

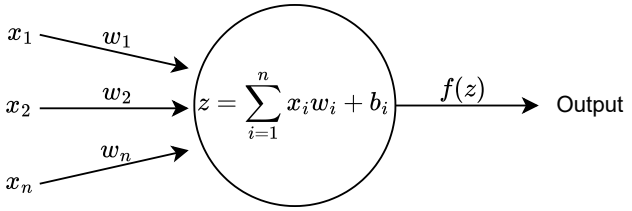


Fig. 5.1.: Scheme of a perceptron that produces one output based on a set of n inputs. A function f is applied to the linear combination of all inputs x_i weighted by their respective weights w_i and biases b_i.

Based on one data point in the training set, the error of one weight w_i to the prediction can be calculated directly via the chain rule. With the frequently used squared loss function $L = \frac{1}{2}(y^{\text{pred}} - y^{\text{ref}})^2 = \frac{1}{2}(f(z) - y^{\text{ref}})^2$, the partial derivative reads:

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial f(z)} \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial w_i} = (f(z) - y^{\text{ref}}) \frac{\partial f(z)}{\partial z} x_i \,. \tag{5.5}$$

The first activation functions $f(z)$ were simple step functions using a threshold.

$$\text{output} = \begin{cases} 0, & \text{if } \sum_j w_j x_j \leq \text{threshold} \\ 1, & \text{if } \sum_j w_j x_j > \text{threshold} \end{cases} \tag{5.6}$$

The perceptron's activation functions are analogous to the firing or non-firing of neurons in the brain. As demonstrated in equation 5.5, the activation function is a bottleneck in building the partial derivative of the loss with respect to the weights. Therefore, they are chosen to be easily derivable, such as linear, sigmoidal, ReLU, Tanh, or SoftMax functions.[263]

Connecting many neurons in series and with each other, all linked by weights, leads to the main architecture of the NN (Figure 5.2).

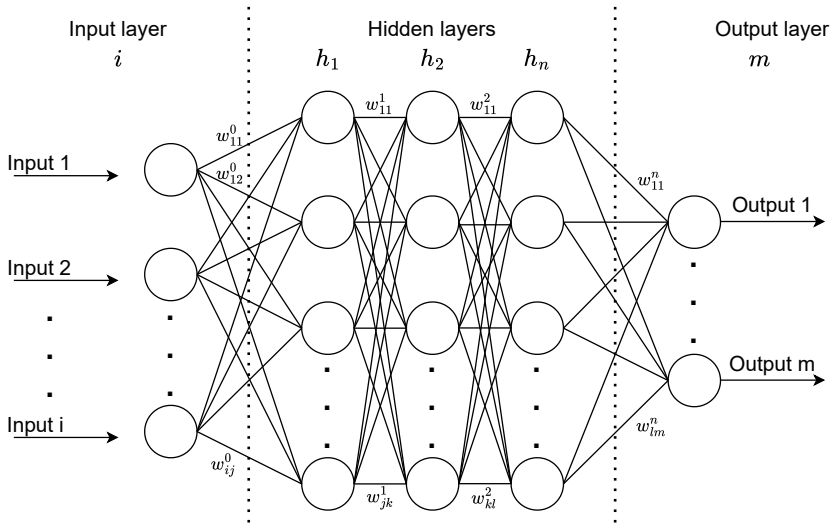


Fig. 5.2.: Schematic architecture of a fully connected neural network. The n binary inputs are fed to the neurons in the input layer. The neurons in the first hidden layer h_1 linear combine all inputs with the respective weights w_ij, apply an activation function and feed the output to the next neurons. The output layer m reads out the accumulated values, often with a different activation function.

Adding one neuron per layer with the output one, enables the inclusion of the biases into the weight matrices w^k_{ij}. These matrices represent all connections between neurons in neighboring layers. The weights and biases are the central property of the NN, and they are used to create the final outputs from the inputs via nested linear combinations and activation functions (feed forward). During the training, the reverse operation is performed, known as backpropagation. For each training point, the partial derivatives of all weights are calculated as shown in Equation 5.5. This calculation takes into account the outputs of all neurons in subsequent layers. Once all training data points have been completed, the initial epoch is concluded and the process restarts from the beginning until convergence criteria or the predefined maximum number of epochs is reached.

In addition to the learning rate, convergence criteria, and number of epochs, important hyperparameters for NNs include the number of hidden layers, the number of neurons per hidden layer, and the activation function.

5.3. Representation of Molecules

The representation of the molecule can be regarded as another hyperparameter. However, it defines the architecture of the NN to a large degree. Simple Cartesian coordinates are not ideal as input features because symmetry operations, such as translation, rotation, reflection, and permutation of atoms, result in different inputs. Therefore, the relationship between input and output is ambiguous and a very large training data set would be necessary, to learn all possible geometries. The descriptors are not invariant with respect

to the symmetry. A commonly used molecular representation is the Coulomb matrix (equation 5.7), which is rotational and translational invariant.

$$X_{ij} = \begin{cases} 0.5\, Z_i^{2.4}, & \text{if } i = j \\ \frac{Z_i Z_j}{r_{ij}}, & \text{if } i \neq j \end{cases} \tag{5.7}$$

The atomic number Z_i represents the nuclear charge and the distance r_{ij} leads to the Coulomb interaction between two atoms. If the atomic number and the diagonal entries are neglected, the Coulomb matrix reduces to the inverse distance matrix.[12]

Alternative molecular representations, such as Behler–Parrinello symmetry functions[17] or graph-based descriptors,[232] are capable of capturing the local chemical environment of atoms, as well as bonds and angles between them. These representations are invariant regarding translation, rotation, and even permutation. The underlying models are more complex, robust and physically meaningful. However, the high number of trainable parameters result in longer training times.

For macromolecules such as proteins and DNA, the atomic-level representation can easily overwhelm the feature space. Therefore, coarse-grained representations are employed, such as one-hot encoding or descriptors derived from sequence alignment methods (e.g., multiple sequence alignments, MSA). The MSA feature is applied in the novel protein structure prediction NN AlphaFold 2.[135]

5.4. Protein Prediction with Alphafold

The sequence-structure-function paradigm, one of the central dogmas in structural biology, states that the function of a protein is determined by its structure, which in turn is determined by its sequence. To understand their functions, plenty of labor and cost intensive X-ray crystallographic or electron microscopic work on proteins has been performed over the last decades. This work has resulted in the creation of more than 100,000 experimental structures, which are stored in the protein data bank (PDB).[223] However, more than 100 million protein sequences are available in the Universal Protein Resource database (UniProt).[52]

To address this gap, protein structure prediction focuses on how a protein's sequence determines its fully folded three-dimensional structure. According to the sequence-structure-function paradigm, the majority of proteins can be described with the assumption that similar (homologous) protein sequences lead to similar structures and functions. This assumption is valid for most, but not all, proteins.[144] Therefore, over the past several years, homology modeling in combination with geometry optimization (Chapter 3.1) via force fields (Chapter 2.1.1) has been the method of choice.

Every two years, different protein prediction software can compete with each other in the Critical Assessment of protein Structure Prediction (CASP). In the 14$^{\text{th}}$ CASP[151] (2020), AlphaFold 2 has significantly outperformed other methods. The software, that was

initially developed by the London-based company DeepMind and is now part of Google's Alphabet Inc., has predicted structures with quality approaching that of experimental determination. AF2 has been published[135] and its source code and hundreds of thousand protein predictions have been made public.[280] Since then, it has been heavily used by researchers and its predictive power was affirmed several times. For entries in the PDB, that were not part of the training set, the median backbone r.m.s.d_{95} (Cα root-mean-square deviation at 95% residue coverage) between the AF2 prediction and the experimental structure is 1.46 Å[284] which is smaller than the resolution of many X-ray studies.

The main idea of AF2 is to solve a graph problem in the three-dimensional space with nodes being residues and edges are not solely defined by covalent bonds but by spatial proximity. In this revolutionary approach, the covalently bonded chain structure is broken. An initial hypothesis of the structure gets iteratively refined.

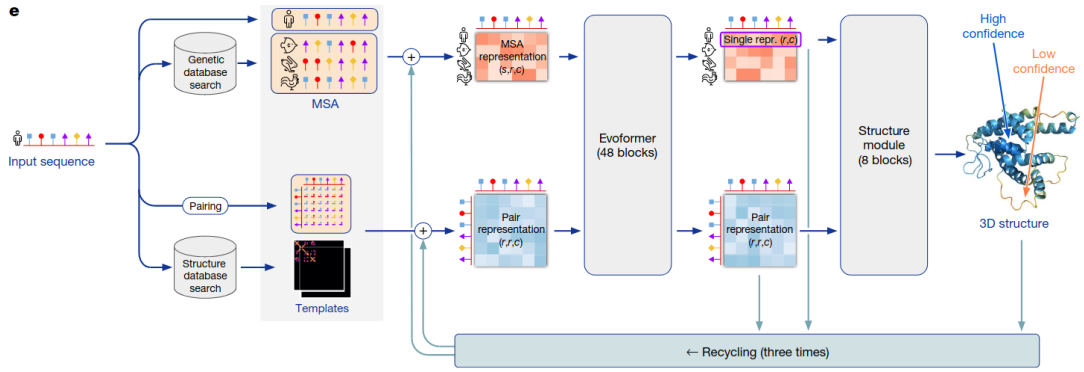


Fig. 5.3.: Schematic workflow of the AlphaFold2 algorithm. Reproduced from Jumper et al. (2021)[135], licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). No changes were made.

In detail, the feed forward pass encompasses three main modules[135] which are shown in Figure 5.3: At first, a Multiple Sequence Alignment (MSA) on many databases is performed to create an N_{seq} x N_{res} array with the number of sequences N_{seq} and the number of residues N_{res}.

Under the assumption that correlated residues are in proximity, the MSA input is transformed into a two dimensional distance matrix, the "pair representation". It is, in essence, a N_{res} x N_{res} matrix describing the likelihood of two amino acids being in contact with each other.

In the second module, a novel neural network called Evoformer processes the MSA and the pair representation in multiple layers. An improved MSA model leads to a more precise representation of the molecular geometry, which, in turn, refines the MSA. This way, both representations exchange information with each other and are iteratively refined. The pair representation is updated via triangle-multiplicative updates to capture the geometric constraints of three residues. The refinement is repeated until a predetermined number of cycles is reached (48 iterations in the published AF2). The key novelty in this block are the learned pair representation from MSA embedding and the simultaneous refinement of amino acids with disregarding covalent bonds.

The third module, called the structure module, maps the refined MSA and pair representation to a three dimensional structure of the protein. For each residue, a local coordinate system ("frame") is created with the origin being the coordinates of the Cα atom, and a Gram–Schmidt process of the vectors connecting the Cα atom with the other two backbone atoms, N and C. The location of every atom in or around a residue is then expressed relative to that residue's local frame, making all subsequent comparisons invariant to global rigid-body motions. Regarding the orientation of the frames relative to the global structure, all frames are initialized as zero translation and identity rotation. The per-residue frames are then iteratively refined in eight sequential blocks to obtain a set of N_{res} independent translations and rotations. Inside each block, an Invariant Point Attention (IPA) layer uses the pair and MSA representation to predict equivariant updates to each frame's rotation and translation. At the end, backbone torsion angles are predicted and side-chain conformers are placed.

Uniquely, AF2 directly outputs Cartesian coordinates in a single end-to-end pass, without any physics-based minimization based on force-fields. However, a final relaxation of the structure using the Amber force field[112] and a gradient descent relaxation (Chapter 3.1) is performed to remove steric clashes.

Prior to this final relaxation, a recycling procedure is applied. The newly predicted 3D coordinates and the refined MSA and pair representations are fed back, repeating the last two modules (Evoformer and structure module) with additionally using the latest structural prediction. The recycling typically converges in few iterations, using three in the original publication.

The training set of AF2 contained all structures in the PDB downloaded at April 2018, consisting of >100,000 protein structures. Training was performed end-to-end with a combined loss function that includes the Frame-Aligned Point Error (FAPE) and many auxiliary losses (Equation 5.8). A slightly deviating loss function that also includes violations for bond-lengths was used for fine-tuning.

$$L = 0.5L_{\text{FAPE}} + 0.5L_{\text{aux}} + 0.3L_{\text{dist}} + 2.0L_{\text{msa}} + 0.01L_{\text{conf}} \tag{5.8}$$

FAPE is summed over all residue-specific frames and measures coordinate deviations for each frame, after aligning the predicted structure to the true structure. This way, the loss becomes invariant to global rotations and translations and the location of all neighboring atoms are evaluated relative to each residue of interest. The auxiliary loss L_{aux} also applies to the structure module. It employs cross-entropy loss functions for torsion angles of the backbone and the side-chain. Regarding the pair representation (distogram) and the MSA prediction, L_{dist} and L_{msa} are both evaluated as an averaged cross-entropy loss. The confidence loss L_{conf} is based on the predicted local-distance difference test (pLDDT).

Besides the structure, AF2 outputs a confidence measure, the pLDDT. A separate network has been trained to predict its own prediction reliability via the Cα local-distance difference test (lDDT-Cα)[178]. It quantifies the agreement of predicted and experimental Cα–Cα distances for residue pairs within a defined cutoff, scoring each residue by the fraction of its local distances that fall within bins of certain tolerance thresholds. This superposition-free

measure is scaled from 0 to 100, providing a robust assessment of local geometry quality for each residue.

The quality of the predicted structures depends on the flexibility and length of the protein. With high pLDDT scores, detailed analysis like enzyme active sites, interfaces between proteins, or binding pockets can be undertaked. The latter has also been applied in this work.

6. Protein Structure Prediction and Molecular Docking

Reprinted in parts with permission from
Xiaodi Hu, David S. Hoffmann, Mai Wang, Lars Schuhmacher, Maria C. Stroe, Birgit Schreckenberger, Marcus Elstner and Reinhard Fischer: GprC of the nematode-trapping fungus Arthrobotrys flagrans activates mitochondria and reprograms fungal cells for nematode hunting
***Nat. Microbiol.* 2024, 9, 1752–1763.**
DOI: 10.1038/s41564-024-01731-9

Author Contributions:
The results presented in this chapter were generated in collaboration. Xiaodi Hu planned and performed most of the experiments and analyzed the data. Mai Wang was responsible for the gene deletions. Lars Schuhmacher, Maria C. Stroe and Birgit Schreckenberger were responsible for cell fractionation. Modeling of the protein structure, optimization of the ligands, molecular docking studies, and binding site analysis were performed by myself.

6.1. Introduction

The nematode *Caenorhabditis elegans* is used as a model organism in many biological studies. The predatory fungus *Arthrobotrys flagrans* can recognize its prey, *C. elegans*, and can subsequently develop hyphae that form round, adhesive traps in which the nematode becomes trapped. The question of how the fungus senses the nematode on a chemical level was answered by the discovery of ascaroside molecules. These nematode-specific pheromones control many developmental processes and are hijacked as signaling molecules by the fungal predator.[31,82] Approximately two hundred ascaroside molecules have been identified as being secreted by nematodes. Many of them have overlapping effects.[116]

Eight ascaroside-sensing receptors have been described in detail in *C. elegans*, including SRBC64 and DAF37.[215] All eight of these proteins are G-protein-coupled receptors (GPCRs), which are among the most largest families of proteins, found in all eukaryotes, and important drug targets in humans.[309] Detailed information on ascaroside-sensing receptors in fungi is lacking. However, the authors of the present study demonstrated that ascaroside number 18 (ascr#18) stimulates hyphal development in *A. flagrans* similarly to the presence of nematodes. To identify the ascaroside receptor in the fungus, the genome of *A.*

flagrans was compared to those of other fungi. This revealed six candidates (GprA-F) that are potentially ascaroside-sensing receptors in *A. flagrans*. Gene knockout experiments revealed that the deletion of GprC had the strongest defect on trap formation.

Because the nematode receptor SRBC64 recognizes ascr#1 but not ascr#5[215] and the same trend was found for trap induction in the fungus *A. oligospora*[114], the authors hypothesized that the fungal receptors should be similar to *C. elegans* SRBC64. While no fungal GPCRs share significant sequence similarities with SRBC64, sequence comparisons between the fungal GprC and *C. elegans* ascaroside-sensing GPCRs revealed two short conserved sequence motifs (RFAF and SLIL/SLIY). This suggests the conservation of ascaroside binding and/or signaling.

To validate this hypothesis, chimeric proteins were expressed. Combining the first four transmembrane (TM) helices of SRBC64 with the last three TM helices of GprC resulted in functional receptors, which rescued the trap defect of the gprC knockout strain. However, using the first four TM helices of DAF37 with the last three TM helices of GprC produced nonfunctional receptors. A comparison of these results with the distribution of S and R motifs in the sequences suggested that both motifs are essential for a functioning receptor.

In summary, the fungus *A. flagrans* senses some of the *C. elegans*-specific ascaroside pheromones. Horizontal gene transfer can be ruled out due to the absence of sequence similarities between the ascaroside-sensing receptors in the two organisms. Based on their experimental data, the authors hypothesized that the R and S motifs are essential for signaling or binding ascaroside molecules.

The goal of this study was to bridge the gap between sequence comparisons and experimental outcomes by computationally analyzing receptor-ligand interactions. The evolutionary evolved structure-function-relationship of the nematode and fungal receptors was studied regarding their similarities and discrepancies in binding ascaroside molecules. Consequently, the hypothesis that the R and S motifs are essential for binding or signaling was investigated.

6.2. Computational Details

Proteins

The protein structures of GprC, SRBC64 and DAF37 were available in the "AlphaFold Protein Structure Database",[294] which is linked to some entries in the UniProt database.[52] The ColabFold web server[187] was used to model mutant receptors and chimeric proteins. The models were generated with models per output, max_recycles and num_ensemble set to 5, 24 and 2, respectively. All protein structures were obtained in the relaxed state using the Amber force field. Long extra- and intracellular loops were removed, resulting in averaged pLDDT scores of above 0.85, which allows for docking studies in principle.[159]

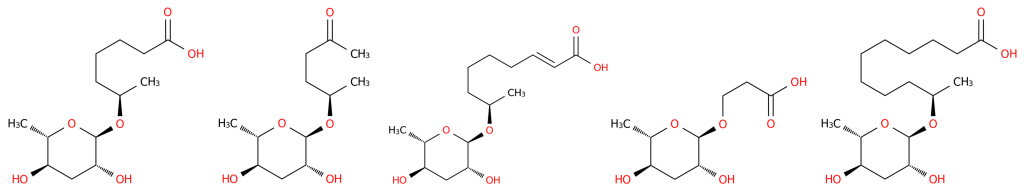


Fig. 6.1.: Ascaroside molecules used in this study. From left to right: ascr#1, ascr#2, ascr#3, ascr#5, ascr#18

Ligands

The ligands ascr#1, #2, #3, #5 and #18 were selected for the study. The first four molecules have been reported to have specific interactions.[215] Ascr#18 was used in experimental work by the authors, because it is commercially available and has also been shown to induce trap formation.[116].

The ligands were obtained from the Chemical Entities of Biological Interest (ChEBI) database.[103] They are shown in Figure 6.1. Their geometries were further optimized using ORCA version 5.0.1.[195,194] on a B3LYP[14,158]/def2-TZVP[303,302] level of theory. Using the TightOpt keyword, the energy and maximum gradient tolerances were set to 1×10^{-6} Ha and 1×10^{-4} Ha, respectively. The def2/J auxiliary basis set was used for the RIJCOSX approximation. Glucose and sucrose, the putative ligands for nutrient sensors, were obtained from the Crystallography Open Database.[99] Their structures were verified in terms of bond lengths and angles by comparing them to other experimental data.[119,310]

Docking

For protein-ligand docking with Autodock Vina version 1.2.3,[289,73] the transmembrane portions of the protein structures were prepared using the "prepare_receptor" tool of the ADFR suite.[230] Using the default settings of the "prepare_ligand" tool in the ADFR suite, all single bonds not included in the ascarylose ring were set to be rotatable. This resulted seven to eleven degrees of freedom for the five ascarosides. The atomic charges of the ligand atoms were determined using the Gasteiger charge model.[89] After processing the pdbqt files with the ADFR suite, docking was performed with a box size of 40 Å around the center of mass of the receptor, an exhaustiveness of 200, and 0.1 Å spacing.

The Autodock Vina scoring function has the same shape as described in Equation 2.2 using a weighted sum of functions. The argument of the functions $d_{ij} = r_{ij} - R_i - R_j$ establishes a relationship between the distance r_{ij} between the atoms i and j, and their van der Waals radii, R_i and R_j.[124] The applied weights are listed in Table 6.1. The first two Gaussian functions and the repulsive penalty for overlapping atoms address steric interactions.

For each protein-ligand complex, the binding affinities and structures of the first five binding poses were used for further analysis. The structures were analyzed and visualized using VMD.[118]

Tab. 6.1.: Weights and terms of the scoring function implemented in Autodock vina.

Weight	Term
-0.0356	$gauss_1$
-0.00516	$gauss_2$
0.840	Repulsion
-0.0351	Hydrophobic
-0.587	Hydrogen bonding
0.0585	N_{rot}

6.3. Results and Discussion

The fungal receptor GprC was analyzed together with the two nematode receptors SRBC64 and DAF37. There is minimal sequence similarity between them (see Table A.2). Structurally, all three receptors are classical GPCRs with seven TM domains, three intracellular and three extracellular loops. The transmembrane domains TM2 and TM3 are longer and more tilted in DAF37 compared to GprC or SRBC64. In addition, the second extracellular loop (ECL2) that connects the two domains TM4 and TM5 is longer in DAF37 compared to the other two receptors. More similarities or discrepancies cannot be easily detected from the protein structure alone.

To compare GPCRs across different families, several generic residue numbering schemes have been introduced.[121] The Ballesteros–Weinstein numbering scheme[11] uses the most conserved residue for each TM helix, individually and denotes it as number 50. For example, Pro2.53 defines a proline residue in TM2, which is three residues closer to the C-terminus than the most conserved residue in TM2. In the three receptors, Asp2.50 was found, which is highly conserved between different GPCRs in many organisms. In most GPCRs, the highly conserved residue in TM6 is a proline, which is present in GprC and SRBC64. DAF37 has a tryptophane instead.

With regard to the most conserved residue in TM6, the length of ECL2, and the shape of TMs 2 and 3, the two receptors GprC and SRBC64 are more similar to each other than to DAF37. In order to address the question of whether this similarity extends to the relevant parts for ligand binding, and to investigate the role of the R- and S-motifs in binding, docking studies were performed.

The docking scores of the five ascaroside molecules did not always align with the trend of receptor activation. SRBC64 was found to be activated by ascarosides #1,#2 and #3, but not by ascr#5,[116] which was correctly reproduced in the docking results (see Table A.1). However, ascr#2 resulted in the lowest docking score for DAF37, yet it was found to activate the receptor. Therefore, the predicted docking scores should be treated with great caution. One potential error source is the formation of more hydrogen bonds or hydrophobic contacts to molecules with more heavy atoms. The scoring function (Equation 2.2) only compensates for the entropic penalty of freezing rotors via dividing the obtained score by $(1 + \omega N_{rot})$ with the weight ω (Table 6.1) and the number of free rotatable bonds

N_{rot}. Additionally, discrepancies between binding affinity and receptor activation may be attributed to induced fit mechanisms, kinetic effects, or entropy contributions from the receptor's conformation or from the displacement of water molecules.

Therefore, it is crucial to avoid overinterpretation and instead prioritize the various binding poses. The selection of the five best binding poses for the five ascaroside ligands resulted in an ensemble of 25 binding poses per receptor. Notably, all ligands occupied a typical binding pocket of GPCRs, in which TM 1 and TM 4 are not involved in binding.[40]

To verify the predicted binding poses, mutagenesis studies were performed on certain amino acids. In GprC, two asparagines in TM2 were selected, separated by a helical loop. The mutation of the more intracellular N2.53 to alanine demonstrated no effect on receptor function, whereas the mutation of N2.57 resulted in a loss of function. In SRBC64, the asparagine N2.60 affected the function of the receptor after mutation, thereby approving the binding pocket prediction. The ensemble of binding poses for the three receptors, together with the positive and negative mutations are shown in Figure 6.2.

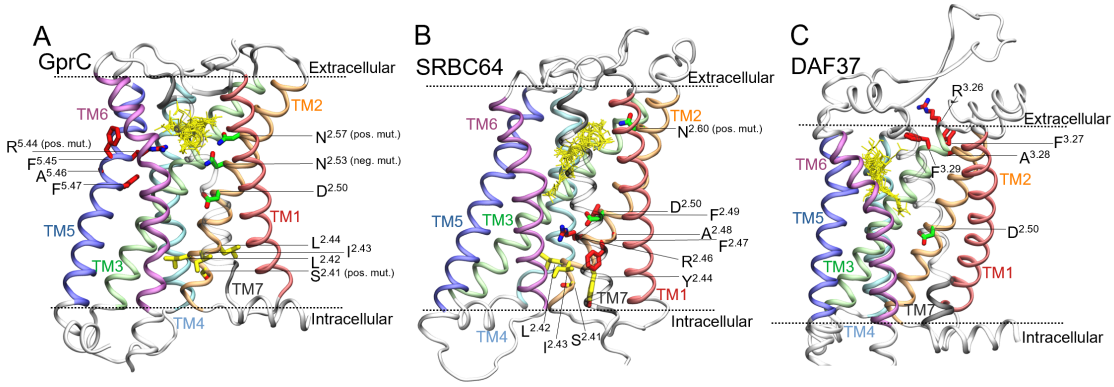


Fig. 6.2.: Twenty-five binding poses of ascarosides #1, #2, #3, #5 and #18 (yellow) docked into different receptor models (A: GprC, B: SRBC-64, C: DAF-37). A highly conserved Asp residue in TM2 is shown in green. Residues that lead to a loss of receptor function upon mutating to alanine are indicated as positive mutation (pos. mut.), whereas N2.53 in Gprc (A) was experimentally proved to be not essential for binding (neg. mut.). Extracellular side on top, intracellular side at the bottom.

As illustrated in Figure 6.2, the common binding site in the extracellular third of the receptor is clearly recognizable. The positions of the two motifs (S and R) are also clearly apparent: The R-motif is present in all three receptors. In GprC, the arginine R6.45 points to the binding site and coordinates ascarosides. Docking ascarosides into the mutant receptor, in which R6.45 is mutated to alanine, has led to a drastic drop in binding affinity (Table A.1). Furthermore, the arginine residue has been experimentally confirmed to be essential for receptor function (Figure 6.2A). In SRBC64 and DAF37, the R-motif is not involved in binding (Figure 6.2b,c). The S-motif is only present in GprC and in SRBC64 and occupies the exact same position (S2.41) in both receptors. It is far from the binding site but has been experimentally confirmed to be essential for the function.

In the model of the chimeric protein DAF37-GprC, which did not rescue the GprC function, the S-motif is missing and a completely different binding site between the TMs 1, 2 and 7

is occupied (Figure 6.3a). The chimeric receptor SRBC64-GprC coordinated ascarosides between TM helices 3, 5, 6 and 7 (Figure 6.3b). Since the S-motif is in the same relative position (S2.41) in the rescued chimeric receptor, we hypothesized that it is essential for signaling.

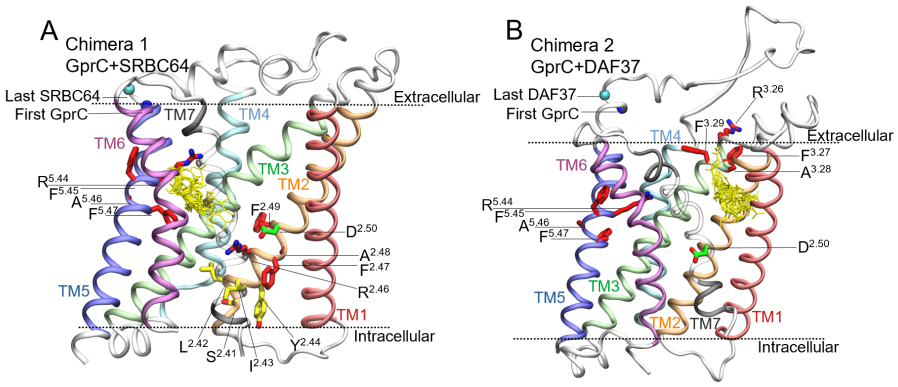


Fig. 6.3.: Twenty-five binding poses of ascarosides #1, #2, #3, #5 and #18 (yellow) docked into chimeric receptor models (A: Functioning GprC+SRBC64, B: Inactive GprC+DAF37). The chimeric proteins contain the N-terminal part with the first four TM helices and the extracellular loop connecting TM4 and TM5 from SRBC64 or DAF37 and the C-terminal part containing the helices TM5, TM6, and TM7 from GprC. The last residue of the first part and the first residue of the GprC part is shown in bright and dark blue, respectively. A highly conserved Asp residue in TM2 is shown in green. Extracellular side on top, intracellular side at the bottom.

The finding of the identical location of the S-motif in both SRBC64 and GprC and the fact that the chimeric protein SRBC64-GprC builds a functioning receptor complements the detected similarities between SRBC64 and GprC described above. This has lead to the hypothesize that the binding pockets of both receptors are also similar.

To compare them with each other, information about coordinating residues was extracted with the Ballesteros-Weinheim scheme. The classical orthosteric binding pocket of GPCRs[40] was occupied by the ascarosides. The relative positions of all residues that have any atom within 3.5 Å of any of the 25 ascaroside binding poses were identified. No trends of motifs of the same chemical nature in the same sequential position were observed, apart from aromatic residues in ECL2, a polar residue in 2.60 and a basic residue in 2.61 or 2.63 (see appendix, Table A.3).

Given the potential variations in length and tilt of the helices, the sequential analysis was complemented with a structural analysis. In Figure 6.4 one binding pose of an ascaroside is shown exemplarily together with the alpha carbons of residues that are involved in binding.

The structural analysis confirmed the absence of trends in motifs of the same chemical nature in similar positions. The binding pockets of SRBC64 and GprC are more similar to each other than they are to the binding pocket of DAF37 with regard to TM2 and TM5, but rarely TM6 and TM7 involved in binding. The residues in TM2 and TM5 do not align with each other between GprC and SRBC64.

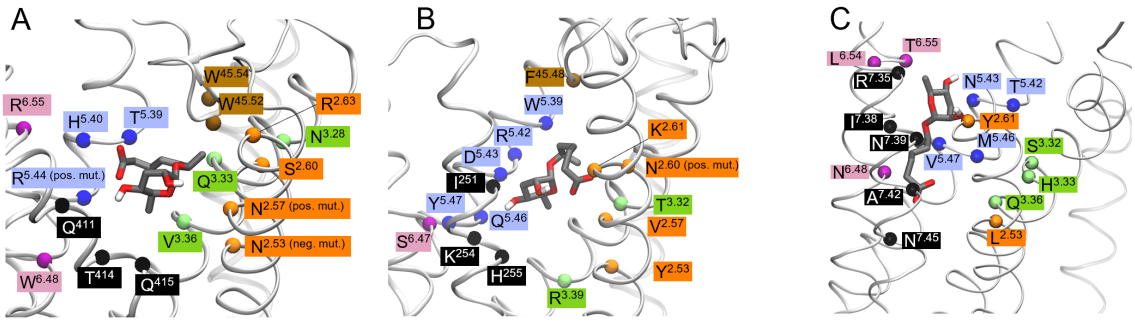


Fig. 6.4.: Binding pose of ascaroside #1 with the Cα atom of all residues that have any atom within 3.5 Å not considering backbone atoms. The receptors GprC (A), SRBC64 (B) and DAF37 (C) are shown.

6.4. Conclusion and Outlook

The central aim of this study was to investigate how the fungus *A. flagrans* senses the nematode-specific pheromones emitted by *C. elegans*. Sequence analysis revealed a lack of matching amino acids, ruling out the possibility of horizontal gene transfer. However, experimental data suggested that the nematode receptor SRBC64 is very similar to the fungus receptor GprC in the role of trap induction, suggesting the presence of conserved sequence motifs in critical structural regions. It was hypothesized that the RFAF and SLIL/SLIY motifs (R and S motifs) are involved in binding or signaling.

The AlphaFold2 protein predictions for the two receptors SRBC64 and GprC were analyzed, revealing numerous structural similarities. The S motif is localized at the same relative position in both receptors. A prediction of the binding pocket was made using computational docking, which was confirmed experimentally. An analysis of the two binding pockets revealed that the S motif in both receptors is not involved in coordinating ascaroside molecules. The R motif in GprC has been shown to bind ascarosides, which partially confirms the initial hypothesis. A comparison of the two binding pockets indicated that their two structures do not contain any easily identifiable trends regarding to similar amino acids in analogous positions. Despite the many similarities between the two receptors, their relevant regions, i.e., the binding pockets, differ, suggesting millions of years of convergent evolution.

The present study's workflow included structural analysis based on experimental and sequential data, followed by binding site prediction, which was verified experimentally, and finally followed by a detailed analysis of the binding pockets. Without experimental approvals, docking studies alone would not have yielded meaningful results due to the following drawbacks.

The primary disadvantage is that the scoring function, as outlined in Chapter 2.2, is merely a rough approximation with empirical weights. The flexible docking approach involves the treatment of some side chains as freely rotatable, thereby enhancing accuracy.[159] MD simulation can be used to create snapshots of the receptors, enabling the calculation of averages from multiple docking computations.[6] Additionally, MD simulation can be

used to refine the docked structure, account for induced fit, or displaced water molecules. Furthermore, costly enhanced sampling methods, like metadynamics, can be employed to accurately calculate free energy differences.

A subsequent collaborative research project is currently underway with Xiaodi Hu focusing on the GprB receptor.

7. Spectroscopic Properties of Bacteriochlorophyll a

7.1. Introduction

Bacteriochlorophyll a (BChl a) is the most widely distributed pigment in photosynthetic bacteria. It is present in purple, green sulfur, as well as green non-sulfur bacteria and plays a central role in both light harvesting and reaction centers.[56] In bacterial light-harvesting complexes, the excitation energies of BChl a molecules are tuned by the protein binding pockets directly through electrostatic effects, and indirectly by changing the geometry of BChl a.[45] In these complexes, even small changes of excitation energies can strongly affect the energy transfer.[272]

BChl a is an aromatic biomolecule with a bacteriochlorin ring containing 18 π-electrons. Its chemical structure is shown in Figure 7.1. The magnesium (Mg) ion in the center is coordinated by four nitrogen atoms, and in protein or solvent environments one or two additional coordinating atoms along the z-axis of the molecule can lead to both penta- or hexa-coordination.[165,233,229] The long aliphatic chain, the phytyl tail, is esterified to the pyrrole ring IV. Numerous studies have shown, that it influences the excitation energy of BChl a only electrostatically. Therefore, it is usually excluded from the QM zone in computational studies.[165,16,8,63]

Experimental absorption spectra of BChl a show three prominent peaks: the near-ultraviolet Soret band (B band) around 400 nm, the visible Q_x band around 600 nm, and the near-infrared Q_y band around 780 nm.[229,247] The Q bands are defined by the orientation of their transition dipole moments (see Figure 7.1). In BChl a, unlike in chlorophylls or in bacteriochlorophyll b, the Q_x and Q_y bands are separated by several thousand cm^{-1}, which enables unambiguous calculation of the S_0–S_1 transition (Q_y). Since the Q_y transition dominates photosynthetic energy transfer, it is the main focus of this work.[56]

Several geometric factors are known to affect the Q_y transition. There is broad agreement that the 5-fold or 6-fold coordination of the Mg ion has little influence.[233,229] In contrast, the dihedral angle of the acetyl group (atoms 1–4 in Figure 7.1) regulates how much the π orbitals of the C_3-O_4 bond contribute to the aromatic system. An out-of-plane motion destabilizes the excited states more than the ground state, leading to a blue shift of both Q_y and Q_x.[7,165] Another geometric effect arises from ring curvature. While DFTB geometry optimization maintains the planarity of the bacteriochlorin ring, the long-range corrected DFT functional CAM-B3LYP leads to a bowl shape and a red shift of excitation

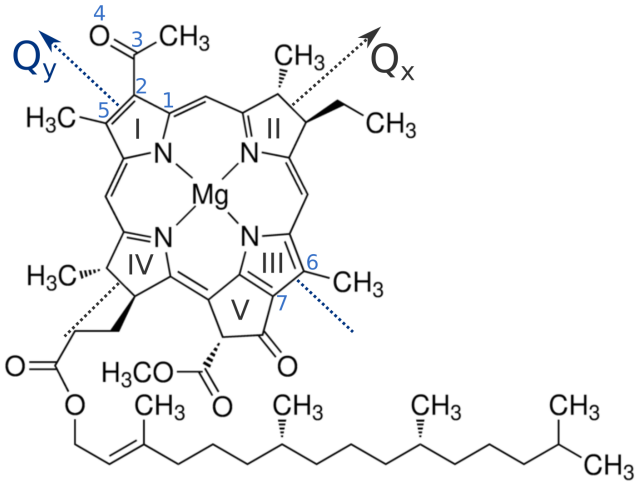


Fig. 7.1.: Chemical structure of BChl a. The molecular axes have been defined by the scientific community so that the z-axis is perpendicular to the molecular plane. The x- and y- axes are indicated with an arrow that shows the transition dipole moments of the Q_x and the more important Q_y excitation. Atoms that are relevant for the present analysis were numbered from 1 to 7. The aliphatic chain at the bottom of the molecule (phytyl tail) has been excluded from all energy calculations.

energies.[25] This red shift is attributed to the curvature of the ring destabilizing the ground state more than the excited state.[8] In addition, as for all conjugated molecules, the bond length alternation (BLA), i.e. the difference of neighboring double and single bond lengths, correlates with the excitation energy.[25]

In addition to intrinsic geometry, the chemical environment of BChl a affects its excitation energy. The electrostatic potential (ESP) induced by surrounding molecules act on BChl a and tune its excitation energy. The electric field has to be in the direction of the Q_y transition dipole moment to significantly affect the S0-S1 transition.[174] Experimentally, the excitation energy of BChl a was found to decrease with increasing solvent polarity, but even more strongly with solvent polarizability.[233] Dispersive effects can be modeled using polarizable force fields[61] or by enlarging the QM region, although the latter also includes charge transfer.[42] Charge transfer in form of protic or aromatic solvents have been observed to red shift the excitation energy of BChl a.[233]

The semi-empirical TD-LC-DFTB2 method has been benchmarked on geometry-optimized BChl a molecules[25] and also applied to BChl a containing protein complexes.[172] However, an analysis of the solvatochromic shift, i.e. the shift of the absorption peak in dependence of the solvent, is missing. An in-depth analysis of the sensitivity of DFTB on the geometrical and electrostatic effects of BChl a is crucial, as even small changes of excitation energies can strongly affect photosynthetic energy transfer.

In order to evaluate this sensitivity of DFTB, BChl a was simulated in various solvents. Ground state geometries of the pigment were sampled either with a force field or with DFTB, and excitation energies were calculated both with and without considering the ESP of the environment. In addition, the effect of explicitly including solvent molecules

in the QM region was investigated. Finally, absorption spectra were computed and the corresponding solvatochromic shifts were analyzed.

7.2. Computational Details

7.2.1. Simulations and Energy Calculations

QM/MM simulations of BChl a were performed with an in-house version of GROMACS[1] and DFTB+.[113] To save computational cost, the phytyl tail of BChl a was cut after the carboxylate ester and saturated with a hydrogen atom. The resulting 85 atoms were treated as the QM zone. The atoms in the phytyl tail and all solvent molecules were treated as the MM zone.

The QM zone was described with the DFTB3 method[312,91] with the 3ob-f parameter set that was optimized for vibrational properties[90] and already shown to reproduce accurate spectral properties of BChl a molecules.[173]

The MM part of the BChl a molecule was described with the CHARMM27 force field[29,28]. For the five solvents, the GAFF-ESP-2012 force field[33] was used. For vacuo simulations a large box size of 100 nm was chosen together with an electrostatic cut-off of 1 nm. The solvent boxes were filled with approximately 1 000 solvent molecules in a cubic box with a border length of 5 nm. The solvent simulations were equilibrated for 50 ps to 300 K using the force field for BChl a and the Nosé-Hoover thermostat with a time constant of 0.5 ps. The 5 ps long NPT equilibration to 1 bar was performed either with DFTB3/3ob-f or with the force field. In both cases, the temperature was controlled as in the NVT equilibration and the c-rescale barostat was used with a compressibility of $4.5\text{x}10^{-5}$ and a time constant of 5 ps. The production runs of 1 ns length or 60 ps length were performed with the same settings as in the NPT equilibration. Snapshots were saved every 100 fs and every 1 fs, leading to 10 000 and 60 000 snapshots, for 1 ns long and 60 ps long simulations, respectively.

Excitation energies on snapshots of corresponding trajectories were calculated using the semi-empirical TD-LC-DFTB2 method with a modified mio parameter set, which has been extensively benchmarked for BChl a molecules.[25,24,272] The software package DFTB+ was used with SCC tolerance set to 10^{-10} electrons and the screening threshold was 10^{-6}. As in the QM/MM simulation, the phytyl tail of BChl a was cut. The phytyl tail and all solvent molecules were considered as MM environment in which the QM zone was electrostatically embedded. For excitation energy calculations, the electrostatic potential (ESP) of the MM zone acting on each atom of the QM zone was calculated with the MDAnalyis package in python.

7.2.2. Geometric Analysis

Geometric properties were analyzed with GROMACS,[1] MDAnalysis,[186] and VMD.[117] Average structures were generated with the *rmsf* command in GROMACS. Dihedral angles were calculated with the *Dihedral* package in MDAnalysis, The curvature along the Q_y transition dipole moment was estimated in terms of the distance between the pyrrole rings I and III. Based on Figure 7.1, the distance (C_2-C_6 + C_5-C_7)/2 was calculated, as performed in Ref. [7, 8].

The displacement of the magnesium (Mg) ion from the BChl plane was quantified by the average distance between Mg and the four planes based on triplets of the four coordinating nitrogen atoms (N_I, N_{II}, N_{III}, N_{IV}). With the cartesian coordinates $\mathbf{r}_{N_I}, \mathbf{r}_{N_{II}}, \mathbf{r}_{N_{III}}, \mathbf{r}_{N_{IV}} \in \mathbb{R}^3$ the unity normal $\mathbf{n}$ of the plane was defined for each of the four triplets (a, b, c) by the cross product

$$\mathbf{n}_{abc} = \frac{(\mathbf{r}_b - \mathbf{r}_a) \times (\mathbf{r}_c - \mathbf{r}_a)}{\left\|(\mathbf{r}_b - \mathbf{r}_a) \times (\mathbf{r}_c - \mathbf{r}_a)\right\|} .$$

The perpendicular distance of Mg to that plane was obtained from the projection

$$d_{abc} = \mathbf{n}_{abc} \cdot (\mathbf{r}_{\mathrm{Mg}} - \mathbf{r}_a) .$$

Because of the arbitrary orientation, absolute distances were taken and averaged over the four planes:

$$d_{\text{Mg-Plane}} = \frac{1}{4} \sum_{(a,b,c)\in\mathcal{T}} \left|d_{abc}\right| .$$

For the bond length alternations, the equilibrium bond lengths of the force field parametrization were used to define single and double bonds. The conjugation is shown in Figure 7.1 and was verified on the gas phase QM/MM simulation. In the aromatic ring which contains 18 π electrons, nine single and nine double bonds were selected. Based on the atom names of the force field, the single bonds were: C4A-CHB, C1B-C2B, C3B-C4B, CHC-C1C, NC-C4C, CHD-C1D, C2D-C3D, C4D-CHA, and C1A-NA. The nine double bonds were: NA-C4A, C1B-CHB, C2B-C3B, C4B-CHC, C1C-NC, C4C-CHD, C1D-C2D, C3D-C4D, and CHA-C1A.

7.2.3. Frequency Analysis

Power spectra were generated with the *velacc* command in GROMACS which computes the velocity autocorrelation function. Its Fourier transform is the vibrational power spectrum. Spectral densities were obtained from a 60 ps long simulation spaced by 1 fs. The classical time-correlation function (Equation 4.19) was applied based on 60 000 excitation energy calculations with electrostatic embedding. Absorption spectra with the cumulant expansion method[170] were generated using a Fortran code published from the university of Pisa.[57,58] The time-averaged excitation energy was obtained from 1 ns long MD simulations.

7.3. Results and Discussion

In this chapter, two research questions were addressed. First, it was investigated whether sampling and subsequent energy calculations with DFTB can model the absorption spectrum of BChl a. Second, the sensitivity of DFTB to different solvents in form the shifted absorption peaks (solvatochromic shift) was analyzed. Both aspects were compared with regard to the BChl a molecule being described with force fields in a classical MM simulation, and with DFTB3/3ob-f in a QM/MM simulation.

To investigate the solvatochromic shift, BChl a has been simulated in vacuum and in 5 different solvents: dimethyl sulfoxide (DMSO), methanol (MeOH), pyridine, quinoline, and triethylamine (TEA). For all QM/MM or MM simulations in gas phase or in solvents, a longer simulation of 1 ns was conducted to sample the relevant phase space, and a shorter simulation of 60 ps was performed to calculate spectral densities.

7.3.1. Simulations in Gas Phase

As a first step, BChl a was simulated for 1 ns in gas phase. The average structures from the 10,000 snapshots are shown in Figure 7.2.

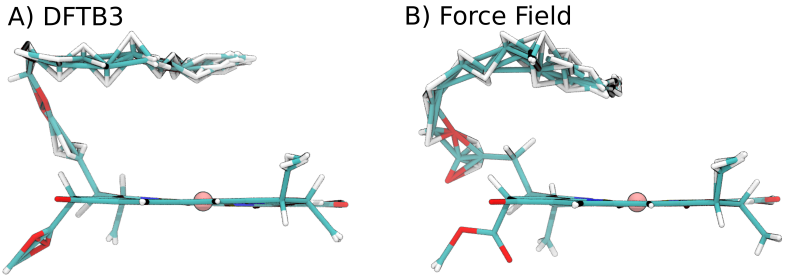


Fig. 7.2.: Average structures from the 1 ns long MD simulation of BChl a in gas phase, sampled with DFTB3/3ob-f (A) or with the force field (B).

Both average structures show that the central magnesium ion, the acetyl group on the right hand side of each molecule and the bacteriochlorin ring itself are in plane, in agreement with geometry optimizations with DFT/B3LYP[25,167] and the crystal structures of light-harvesting complexes.[214,185] However, the average structure from the QM/MM simulation (Fig. 7.2A) is more planar than the one from the MM simulation (Fig. 7.2B) . The phytyl tail shown on top of the structures extends away from the bacteriochlorin ring. In contrast to the relatively rigid ring, the averaged positions of the tail are not chemically meaningful, as averaging over many conformations of flexible groups leads to non physical intermediate structures rather than actual geometries.[150]

Because the averaged ring coordinates were a meaningful description of the bacteriochlorin ring, 10 000 equally spaced snapshots were used to calculate excitation energies. To isolate the geometric effect of the bacteriochlorin ring, the ESP induced from the phytyl tail was

neglected. Figure 7.3 shows the histograms and Gaussian fits of the excitation energies along the 1 ns long simulations.

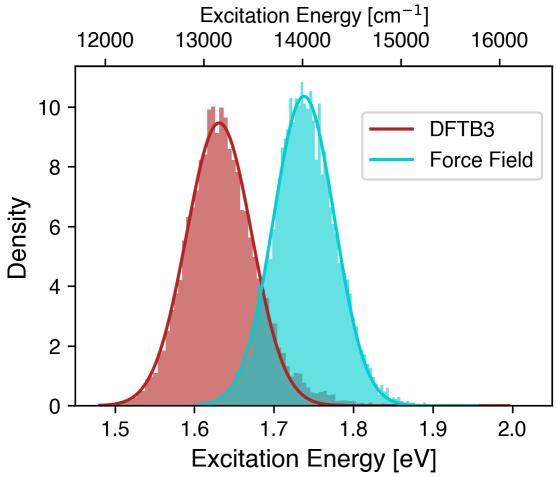


Fig. 7.3.: Histogram of excitation energies of the 1 ns long MD simulation in gas phase sampled with DFTB3/3ob-f (red) or with the force field (blue).

The mean value of the excitation energies of the QM/MM simulation is red shifted by 0.15 eV in comparison to the MM simulation. This is attributed to the geometry mismatch of the underlying potential energy surfaces (PESs). As the minima of the DFTB3 and force field PES do not coincide, the configurations sampled with the force field are biased towards higher energies. This shift has already been observed in a study on BChl a molecules in a light-harvesting complex.[172] Moreover, the choice of the force field, like Amber or Charmm has been shown to significantly influence the calculated excitation energies in BChl a molecules.[299,42]

With a standard deviation of 48 meV, the distribution of excitation energies in the QM/MM simulation is slightly broader than the distribution obtained from the force field (σ = 40 meV). This can be rationalized by the harmonic nature of classical force fields (see section 2.1.1) leading to generally stiffer potentials compared to the more anharmonic QM-based PES. Ultimately this leads to reduced structural fluctuations of BChl a in MM simulations compared to QM/MM sampling.[235]

To elucidate the underlying geometrical cause for the red-shifted and broader distribution of QM/MM sampled excitation energies, four geometric effects were analyzed. Figure 7.4 shows the time series of the four geometric properties along the 1 ns long MD simulation. A scatter plot visualizes correlations between the respective geometric property and the excitation energies within one data set.

The planarity of the ring (Figure 7.4A), as estimated by the distance between the two pyrrole rings I and III, is higher in the QM/MM simulation than in the MM simulation by 0.1 Å. A longer distance could also result from longer bond lengths, which is the case for the QM/MM simulation (see Figure B.1). However, the average structures that are shown in Figure 7.2 indicate a slight bowl shaped ring for the force field simulation, indicating

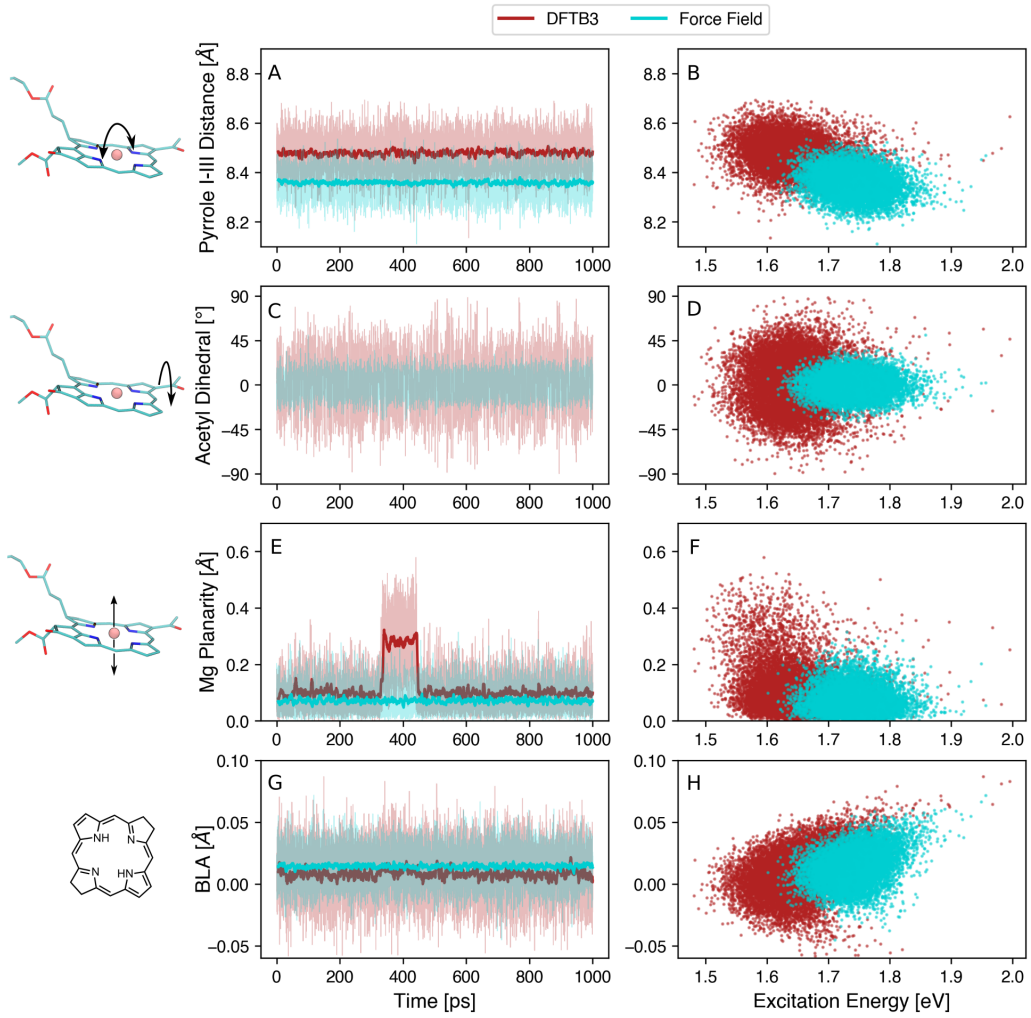


Fig. 7.4.: Geometric effects along the 1 ns long MD simulation performed with DFTB3 (red) or the force field (turquoise). Data is shown in transparent colors, running averages over 50 ps windows are shown in bold colors. For the dihedral angles, no running averages were calculated due to negative values. One point in the scatter plots (right column) refers to one out of the 10 000 snapshots.

that higher pyrrole distances mean more planar structures. In the scatter plot (Panel B) a negative correlation between excitation energy and ring planarity is observed within both data sets. This is in contradiction to excitation energy calculations with TD-LC-DFTB2 on geometry optimized structures with different DFT functionals, in which CAM-B3LYP led to a bowl shape and a smaller excitation energy.[25] Another study based on TD-DFT/PBE0 calculations stated that the ring curvature affects the excitation energies via controlling the shape of the molecular orbitals that are involved in the electronic excitation. Curvatures destabilizing the ground state more than the excited states lead to a red shift of excitation energies.[7,67] However, the same authors also report a blue shift upon ring curvature.[8]

The dihedral angle of the acetyl group regulates the extension of the aromatic system to the two π orbitals of the carbonyl group and is therefore expected to influence the excitation energy. Figure B.2 shows that the orbitals of the carbonyl group are part of the highest

occupied and lowest unoccupied molecular orbitals. Therefore, an effect of the acetyl dihedral angle on the excitation energies is expected. Calculations with TD-DFT/PBE0 showed an effect the Q_y excitation energy by 39 meV upon rotating by 90°.[8] In the present scatter plot in panel D, no correlation between the orientation of the acetyl group and the excitation energies can be observed. It is noteworthy, that DFTB samples between +/- 50° and the force field only samples between +/- 30°, visualizing the high force constant and stiff potentials for dihedral angles in the present force field.

Panels E,F display the effect of the out-of-plane movement of magnesium. In the bacteriochlorin ring, the Mg is coordinated by four nitrogen atoms and it can be 5-fold or 6-fold coordinated dependent on the solvent.[243] TD-DFT calculations as well as spectroscopic experiments showed a negligible effect of the coordination on the excitation energy.[233,229,34] In agreement, the scatter plot (panel F) shows no correlation between the planarity of magnesium and the excitation energy. The time frame of the DFTB3 simulation around 400 ps is a clear outlier in which the Mg has been located out of the ring by almost 0.5 Å. A negative correlation of this part of the trajectory with the excitation energy was observed. However, it was not found that the out-of-plane movement of Mg significantly destabilizes the ground state (see Figure B.3).

At last, the bond length alternation was expected to have the highest influence, as it has been shown to be a strong factor for excitation energies in aromatic systems in general[177,22] and specifically for BChl a.[25] In the scatter plot in panel H a clear correlation of BLA and excitation energies is observed. The fact that for same BLA values the blue scatters of the MM simulation are systematically shifted to higher energies, indicates that another factor influences the excitation energies. Based on the analysis above, this is explained by the planarity of the ring (panel B). The distributions of BLAs in panel G show that DFTB3 samples on average smaller bond length alternations than the force field. In addition, the range of the sampled BLA values is wider for DFTB3, which explains that the histograms of excitation energies (Figure 7.3) are red shifted and broader for DFTB3 than for the force field. Negative BLA values are a consequence of the vibrating carbon-carbon bonds and are rarely reached in MM simulations.

In summary, the bond length alternation has the strongest effect on excitation energies of BChl a, followed by the ring curvature (Distance Pyrrole I-III). The DFT calculated effect of the dihedral angle of the acetyl group[7,8] was not observed in the present study with TD-LC-DFTB2. The location of the magnesium atom effects the excitation energy when it pops out of the ring. For small fluctuations it has no significant effect on the excitation energies. The analysis of geometric influences in BChl a helped to discriminate between geometric and electrostatic effects in simulations of BChl a in five different solvents.

7.3.2. Simulations in Solvents

For photosynthetic pigment protein complexes the electrostatic and geometric tuning from individual binding pockets are essential to accurately model the energy transfer. To benchmark the sensitivity of DFTB on the environment, BChl a was simulated in five

solvents. All solvents were described with force fields. As in the previous subsection, BChl a was sampled either in a QM/MM simulation using DFTB3/3ob-f or in a pure MM simulation using the force field. Experimentally, a solvatochromic shift of 37 meV (300 cm^{-1}) was measured for the five solvents dimethyl-sulfoxide (DMSO), methanol (MeOH), pyridine, quinoline and triethylamine (TEA).[233] In the upcoming subsections, they are colored accordingly from red (quinoline, lowest absorption energy) to blue (TEA, highest absorption energy).

First, the geometric properties were analyzed and compared to the previous findings. The excitation energies of the BChl a molecules were calculated "without ESP", i.e., in the given geometry of the corresponding solvent simulation, while neglecting the ESP from the solvent or from the phytyl tail.

Regarding the magnesium ion, the average structures (Figure B.4) and the running averages of the planarity (Figure B.5) show a pronounced out-of-plane movement of Mg only for some force field simulations (pyridine, DMSO, MeOH). However, the Mg is on average 0.1 Å closer to the ring than in the outlying time interval in the gas phase QM/MM simulation. Therefore, no effect of the planarity of magnesium on the excitation energies in all solvent simulations is expected.

The two most relevant geometric influences from the analysis of gas phase simulations, are bond length alternations and pyrrole I-III distance. Their values along the solvent simulations can be seen in the appendix (Figures B.6 and B.7). The mean values are presented in Table 7.1.

Tab. 7.1.: Mean values of bond length alternation (BLA) and the distance between the two pyrrole rings I and III (Pyrrole I-III) over 1 ns long MD simulations with force fields or DFTB3. Simulations were performed pyridine, quinoline, dimethyl-sulfoxide (DMSO), methanol (MeOH), and triethylamine (TEA). For comparison, the mean value of the simulations in vacuum (Figure 7.3) are also shown (Vac). All values are in Å.

	DFTB3					
	Vac	Quinoline	Pyridine	DMSO	MeOH	TEA
BLA	0.008	0.007	0.010	0.013	0.009	0.008
Pyrrole I-III	8.478	8.479	8.486	8.511	8.505	8.474
	Force Field					
	Vac	Quinoline	Pyridine	DMSO	MeOH	TEA
BLA	0.014	0.014	0.014	0.014	0.014	0.014
Pyrrole I-III	8.358	8.359	8.363	8.355	8.366	8.358

The QM/MM simulations with DFTB3 yield different values for the bond length alternation (BLA). In particular, the BLA obtained for the simulation in DMSO is 50% larger than that of the vacuum simulation. Consequently, the excitation energies of BChl a in DMSO are expected to be the highest. In contrast to the QM/MM simulations, the MM simulations with classical force fields show identical BLA values of 0.014 Å across all solvents. Since BLA correlates more strongly with excitation energy than the planarity of the bacteriochlorin ring, nearly identical excitation energies are expected for all MM simulations. To validate

that expectation, histograms of the excitation energies were created. Even though the weighting by oscillator strengths has a negligible effect (Figure B.8), Gaussian fits of the weighted histograms are presented in Figure 7.5.

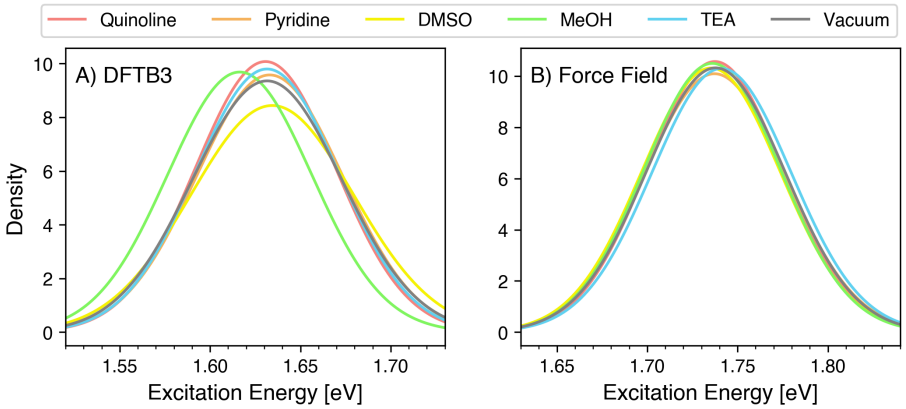


Fig. 7.5.: Excitation energies without ESP. Gaussian fits of excitation energies obtained from 1 ns long QM/MM (A) or MM (B) simulations weighted with the corresponding oscillator strengths. Excitation energies with TD-LC-DFTB2 were calculated on geometries of the BChl a without the phytyl tail and without any electrostatic influence.

The histogram of the QM/MM simulation (Figure 7.5A) reveals a clear outlier: methanol is red-shifted by about 25 meV relative to all other solvents. This deviation cannot be explained by the bond length alternation, the pyrrole I–III distance, or the planarity of the Mg ion, as no anomalies are apparent in the corresponding time series (Figures B.5–B.7). Thus, there must be an additional influence on the geometries beyond the parameters analyzed here. Apart from this exception, all expectations from Table 7.1 were confirmed: in the QM/MM simulations, DMSO, which exhibits the largest BLA, also showed the highest excitation energy, while quinoline, with the lowest BLA, yielded the lowest excitation energy. However, all these differences remain small, below 10 meV. In the MM simulations (Figure 7.5B), the histograms are nearly identical. Methanol appears slightly red-shifted, in consistence with the pyrrole I–III distance of BChl a in methanol being the largest among all solvents.

In summary, apart from the outlier observed for BChl a in QM/MM simulations with methanol, the excitation energies in all solvent simulations were comparable to those of the gas-phase simulation, differing by less than 5 meV. The experimentally determined solvatochromic shift of 37 meV therefore cannot be attributed to geometrical effects. To analyze the role of electrostatic influences, excitation energies were calculated including the ESP of the environment. The histograms are shown in Figure 7.6.

Including the ESP of the environment substantially increased the spread of the mean excitation energies from about 5 meV (geometric contribution only) to roughly 30 meV (Figure 7.6). Methanol remains a clear outlier in the QM/MM simulations, showing an even stronger red shift. In the MM simulations (panel B) all solvents were systematically blue-shifted, which is explained by the geometric mismatch between the force field and DFTB, as outlined in the previous subchapter.

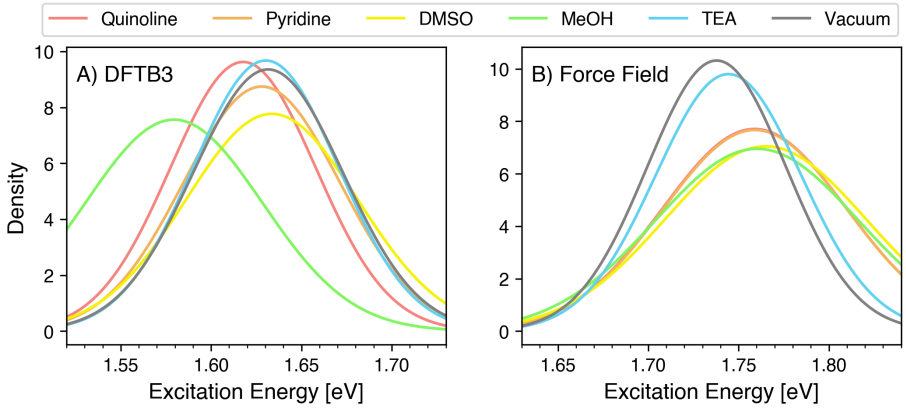


Fig. 7.6.: Excitation energies with ESP. Gaussian fits of excitation energies obtained from 1 ns long QM/MM (A) or MM (B) simulations weighted with the corresponding oscillator strengths. Excitation energies with TD-LC-DFTB2 were calculated on geometries of the BChl a without the phytyl tail. The electrostatic potential of the phytyl tail and of the solvent molecules was considered.

Overall, the solvatochromic shift was not well reproduced. Experimentally, BChl a in TEA has the highest absorption energy among the five solvents, while the excitation energies of the simulations in TEA were on average (QM/MM) or even lower (MM) compared to BChl a in other solvents. Since absorption peaks and mean of excitation energies are not the same, the absorption spectra of BChl a were calculated in the following subsections.

7.3.3. Spectral Densities

The velocity power spectrum is obtained by Fourier transforming the velocity autocorrelation function of the atoms. In the harmonic approximation and at low temperature, its peaks coincide with the vibrational eigenmodes obtained from normal mode analysis. The power spectrum therefore provides access to all vibrational modes of the molecule and is closely related to the infrared (IR) spectrum.[20] In line with IR spectroscopy, frequencies are reported in cm^{-1}. For chlorophylls, prominent modes include the C–H stretching vibrations around 3 000 cm^{-1} (period of 12 fs), the C=O stretching vibrations near 1 700 cm^{-1} (20 fs), and the aromatic C=C vibrations around 1 500 cm^{-1} (25 fs).[176] The region below 1 000 cm^{-1} is generally considered the fingerprint region, arising from bending modes of the bacteriochlorin ring together with C–O, C–N, and C–C single-bond vibrations.[109,37] Peak assignments in this region are often ambiguous. For BChl a, signals around 800 cm^{-1} (40 fs) have been attributed to a mixture of aliphatic C–C and C–N vibrations,[229] whereas slower motions near 300 cm^{-1} (100 fs) were assigned to in-plane deformations of the bacteriochlorin ring.[244]

In contrast, the spectral density is obtained by performing a Fourier transform on the excitation energy autocorrelation function. Thus, the spectral density reveals the frequencies that periodically modulate the excitation energies. It quantifies system-bath coupling. In the gas phase, the bath is considered a set of harmonic oscillators stemming from vibrational modes.[292] Therefore, the spectral density of BChl a quantifies how strongly

the aforementioned vibrational modes of the power spectrum couple to its excitation energy.

Figure 7.7 compares the velocity power spectrum from the QM/MM and MM gas-phase simulations with the corresponding spectral density. At first glance, the two spectra in Figure 7.7 are very different. The power spectra (panel A) display sharp peaks at high, medium, and low frequencies, while the spectral densities (panel B) show broader and redistributed peaks around the medium-frequency region.

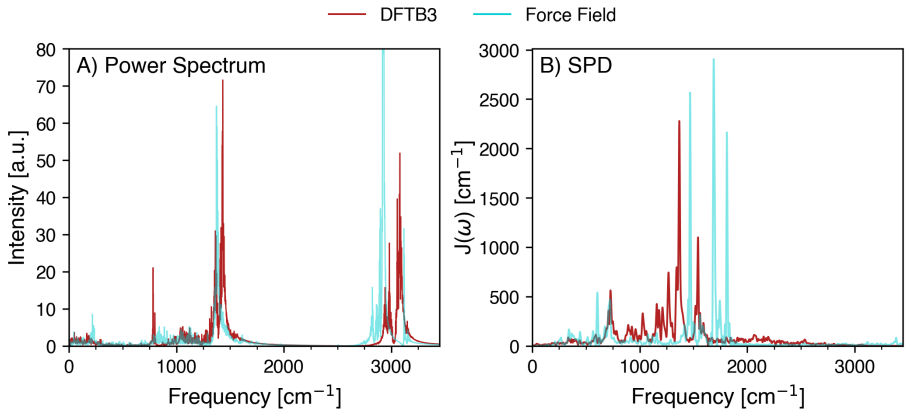


Fig. 7.7.: Frequency analysis of gas-phase simulations sampled with DFTB3 (red) or with the force field (blue). The velocity power spectrum of the 60 ps trajectory (A) is compared with the corresponding spectral density (B).

Conceptually, the spectral density should be regarded as a subset of the power spectrum: the latter reflects all periodic nuclear motions on the ground-state PES, whereas the spectral density highlights only those motions that modulate the excitation energy. This explains why the intense hydrogen vibrations at 3 000 cm^{-1}, clearly visible in the power spectrum, are absent in the spectral density. Negligible effects of the C-H stretching modes on excitation energies are expected, as the HOMO and LUMO orbitals carry hardly any density on the C-H bonds (Figure B.2).

At medium frequencies around 1 500 cm^{-1}, the power spectra are dominated by sharp peaks from aromatic C=C and C–C vibrations. The overlay between DFTB3 and force field simulations is due to the force field being parametrized to yield the correct carbon-carbon stretching frequencies. These modes also appear in the spectral densities because they determine the bond-length alternation, which strongly influences the excitation energy. However, in the spectral density the sharp peaks are smeared and redistributed. This arises because the spectral density is derived from the excitation-energy autocorrelation function. Thus, it depends not only on the vibrational frequencies of the ground state PES, but also on how the gap between ground and excited states changes with those coordinates. For large molecules such as BChl a, the excited-state PES is anharmonic and has different curvatures and eigenmodes than the ground state (Duschinsky rotation). As a result, ground-state vibrations couple to to different modes in the excited states, leading to smeared and broadened peaks.[192,245]

In addition, the peaks for aromatic carbon-carbon vibrations in the spectral densities are shifted to higher frequencies. This shift is explained by the prefactor ω (Equation 4.18) in the definition of the spectral density, which skews intensities towards higher frequencies. For the force field simulations, the mismatch between the force field PES and the DFTB3 PES leads to a sampling around a geometry where the derivatives of the energy gap w.r.t coordinates is steeper. Consequently, force-field-based spectral densities show artificially blue-shifted peaks, which were considered to be a systematic error of force fields.[172] Finally, the bands of the force field-based spectral densities are sharper and more intense than those obtained from the QM/MM simulations. This reflects the stiffer, more harmonic character of the force field PES compared to the more anharmonic DFTB3 surface, as discussed earlier in subsection 7.3.1.

To analyze the effect of solvents on the spectral density, all 60 ps long trajectories in the five solvents were analyzed. Figure 7.8 summarizes the spectral densities obtained from QM/MM and MM simulations with and without considering the electrostatic potential of the environment.

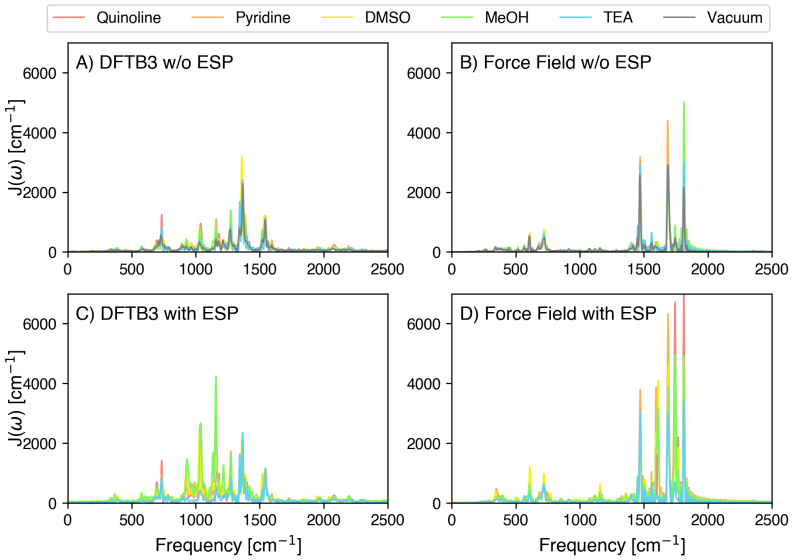


Fig. 7.8.: Spectral densities obtained from 60 ps long simulations of BChl a in different solvents. QM/MM simulations with DFTB3/3ob-f (left column) and MM simulations with force fields (right column) are shown. Excitation energies were calculated without (upper row) and with (bottom row) the influence of the environment. For comparison, the spectral density from the gas-phase simulation (Vacuum) is also shown in the upper row.

The spectral densities from the simulations in solvents show the same general trends, as observed in gas phase: force field simulations lead to sharper peaks that are shifted to higher frequencies compared to the QM/MM simulations. The peaks of the spectral densities from simulations in vacuum and the solvent-based calculations without the ESP are nearly identical (panels A,B). This indicates that the solvent environment does not significantly affect the internal vibrational modes of BChl a that couple to the excitation energy.

When the electrostatic influence of the environment is included in the calculation of excitation energies (panels C,D), the electric field acting along the Q_y transition dipole amplifies the fluctuations of the excitation energy.[174] Because of the close overlap between the gas-phase spectral densities and the solvent-based spectral densities without electrostatic contributions (upper panels), the more intense peaks observed in the lower panels can be attributed solely to solvent–solute electrostatic interactions. In the case of the force fields (panel D), the systematic blue-shift and higher intensities are even more pronounced than for DFTB3, consistent with the PES mismatch discussed above.

A useful technical aspect is illustrated by the comparison of spectral densities obtained from 60 ps simulations with different integration time steps. Using a 3 fs time step instead of 1 fs reduces the computational cost while retaining all relevant spectral features (Figure B.9).

7.3.4. Reorganization Energies

To quantify the effects of the different spectral densities, the reorganization energies were obtained from their normalized integrals. Table 7.2 compares these values with the reorganization energies derived from the variance of the distributions of excitation energies.

Tab. 7.2.: Gas-phase reorganization energies λ in meV (cm^{-1}). λ was calculated based on the normalized integral of the spectral densities (λ_{spd}) from 60 ps long MD simulations shown in Figure 7.7B, or based on the variance of the Gaussian distributions (λ_{var}) from 1 ns long MD simulations shown in Figure 7.3.

	DFTB3	Force Field
λ_{spd}	24 (190)	19 (150)
λ_{var}	38 (310)	30 (240)

For the calculation of the reorganization energy λ_{spd} from the spectral densities, the position of the peaks are eliminated by normalizing the integral with the frequency. Therefore, only the area under the curves are relevant. This explains why the intense, but narrow peaks of the force field based spectral densities result in a low reorganization energy of 19 meV. The aforementioned broader peaks for DFTB3 based spectral densities lead to a higher λ_{spd} of 24 meV.

Regarding the variance obtained reorganization energies (λ_{var}), the sampling with DFTB3 leads to a higher reorganization energy of 38 meV, compared to the force field sampling leading to 30 meV. The issue of anharmonic potentials of DFTB3 leading to a broader distribution of excitation energies was already discussed in subsection 7.3.1.

For both QM/MM and MM simulations, λ_{var} is higher than λ_{spd}. In the harmonic and Gaussian limit of energy-gap fluctuations, these two quantities are expected to be equivalent. As discussed by Blumberger,[21] non-ergodic movements in proteins can produce an almost two times higher value of λ_{var} than the reorganization energy obtained from the stokes shift. Even for the present gas-phase simulations without proteins or solvents,

the energy-gap fluctuations are not perfectly Gaussian reflecting nonlinear coupling of nuclear motions to the excitation energy. Such deviations broaden the variance of the excitation energy but are too short-lived or uncorrelated to proportionally increase the autocorrelation-based spectral density.[192] Moreover, technical limitations arising from finite-time autocorrelation functions and Fourier transforms with windowing and padding tend to underestimate contributions from low-frequency motions. As a result, the spectral densities show reduced intensity in the low-frequency region, which leads to a systematic underestimation of λ_{spd}.

Based on these two arguments, the 19 meV from λ_{spd} with force fields is considered as a lower bond, and the 38 meV from λ_{var} with DFTB3 is considered as an upper bond of the true reorganization energy in gas phase. A corresponding calculation with DFT/B3LYP yielded a value of 35 meV[272], which is in between the calculated values in Table 7.2

The gas-phase reorganization energy of approximately 35 meV is also called the inner-sphere reorganization energy. The contribution of the environment, the outer-sphere reorganization energy is estimated as the difference between calculations performed with and without the electrostatic potential. For λ_{spd} it is just the difference of the obtained values with and without ESP. For λ_{var} however, it is not based on the difference of the variances, but rather calculated as $\sigma(E^{\text{with ESP}} - E^{\text{w/o ESP}})^2$, i.e., the variance of the differences of the excitation energies with and without the ESP.[61] Table 7.3 summarizes the values for the solvent simulations.

Tab. 7.3.: Reorganization energies λ in meV (cm^{-1}) in solvents with, and without (w/o) considering the electrostatic potential (ESP). λ was calculated based on the normalized integral of the spectral densities (λ_{spd}) from 60 ps long MD simulations shown in Figure 7.8, or based on the variance of the Gaussian distributions (λ_{var}) from 1 ns long MD simulations shown in Figures 7.6 and 7.5. Averaged outer-sphere reorganization energies were calculated either as $\lambda_{spd}^{\text{outer-sphere}} = \lambda_{spd}^{\text{with ESP}} - \lambda_{spd}^{\text{wo/ ESP}}$ or based on the variance $\sigma(E^{\text{with ESP}} - E^{\text{w/o ESP}})^2$.

	DFTB3 w/o ESP	DFTB3	Force Field w/o ESP	Force Field
λ_{spd}				
Quinoline	27 (220)	39 (320)	18 (150)	40 (330)
Pyridine	26 (210)	45 (370)	19 (160)	40 (320)
DMSO	27 (220)	47 (380)	19 (150)	41 (340)
MeOH	28 (230)	63 (520)	21 (170)	38 (310)
TEA	22 (180)	22 (180)	19 (160)	21 (170)
Mean value	26 (210)	43 (350)	19 (160)	36 (290)
λ_{var}				
Quinoline	35 (280)	42 (340)	29 (240)	58 (470)
Pyridine	43 (340)	61 (490)	32 (260)	60 (490)
DMSO	56 (450)	87 (700)	30 (240)	70 (570)
MeOH	43 (350)	96 (770)	29 (230)	71 (580)
TEA	40 (320)	41 (330)	31 (250)	35 (280)
Mean value	43 (350)	65 (530)	30 (240)	59 (470)
$\lambda_{spd}^{\text{outer-sphere}}$		17 (140)		17 (140)
$\lambda_{var}^{\text{outer-sphere}}$		16 (130)		18 (150)

Regarding λ_{spd}, the aforementioned negligible influence of the solvent on the geometrical contribution to the spectral density is confirmed by mean values of 26 meV and 19 meV for DFTB3 and the force fields, respectively. These are essentially identical to the values obtained from the gas-phase simulations (Table 7.2), which were 24 meV and 19 meV.

A similar trend is observed for the reorganization energies obtained from the variance of the excitation energy distributions (λ_{var}). Since the histograms of the solvent and gas-phase simulations in Figure 7.5 are very similar, the mean values of 43 meV (DFTB3) and 30 meV (force field) agree well with the corresponding gas-phase values of 38 meV and 30 meV.

When the ESP is included, all reorganization energies increase due to the broader distributions of excitation energies and the more intense peaks in the spectral densities. The outlying QM/MM simulation in methanol yielded reorganization energies almost twice as high as those of the other solvents.

The outer-sphere reorganization energies of 16-18 meV are slightly lower than the inner-sphere contributions obtained in gas-phase simulations (19–38 meV). Interestingly, the outer-sphere reorganization energy from the spectral density ($\lambda_{spd}^{\text{outer-sphere}}$) is 17 meV for both force field and DFTB3. This indicates that the more intense but narrow peaks from

the force-field simulations are exactly balanced by the low but broad peaks obtained from DFTB3 (see Figure 7.8C,D).

In summary, the anharmonic ground state PES of the QM/MM sampling leads to higher reorganization energies compared to the stiffer harmonic surface of force fields. The analysis of reorganization energies and spectral densities is essential for interpreting calculated absorption spectra. The total reorganization energy leads to homogeneous line broadening, while the detailed shape of the spectral density determines the vibronic fine structure.

7.3.5. Absorption Spectra and Solvatochromic Shift

As a first step, the calculated absorption spectra are compared with the experimental measurement. In absorption spectroscopy, wavelengths in nanometers are traditionally used to describe excitation energies. However, the inverse ordering and non-linear scaling of the energy axis is misleading. Therefore, in this work absorption spectra are consistently reported in electronvolts (eV) and wavenumbers (cm^{-1}), in line with the preceding analysis. The spectrum plotted against wavelength (nm) can be found in the appendix in Figure B.10.

Figure 7.9 shows the calculated absorption spectra of BChl a in pyridine. Weighted histograms of excitation energies (ensemble method) are compared with the full cumulant expansion (FCE) approach.

The ensemble spectra in Figure 7.9 (bright colors) are the same histograms for pyridine that were presented in Figure 7.6, shifted to align with the experimental absorption maximum. Both calculated distributions are broader than the experimental peak. The distribution from the force-field simulation (bright blue) is particularly broad, consistent with the PES mismatch and the higher variance-based reorganization energy (λ_{var}) obtained for pyridine (Table 7.3).

The experimental absorption peak is not Gaussian but asymmetric, being truncated in the low-energy side and broader in the high-energy side. Among several inhomogeneous broadening mechanisms, the dominant origin of this skewness in BChl a is vibronic coupling.[15] This reflects the fact that no vibrational states exist below the vibrational ground state, while many states exist above it (see also Figure 4.1 in section 4.1).[291]

These vibronic sidebands are captured by the spectral densities within the harmonic approximation. The second-order cumulant expansion is equivalent to the full cumulant expansion (FCE) and is exact if the excitation-energy distribution is Gaussian.[318] For the QM/MM simulations with DFTB3, the spectral densities in Figure 7.8C exhibit peaks around 1 200 cm^{-1}, giving rise to a vibrational shoulder in the FCE spectrum (dark red) displaced by 1 200 cm^{-1}relative to the main absorption peak. In contrast, the force-field spectral densities (Figure 7.8D) peak around 1 700 cm^{-1}, leading to a vibrational shoulder in the FCE spectrum (dark blue) that is shifted by an additional 500 cm^{-1} towards higher energy.

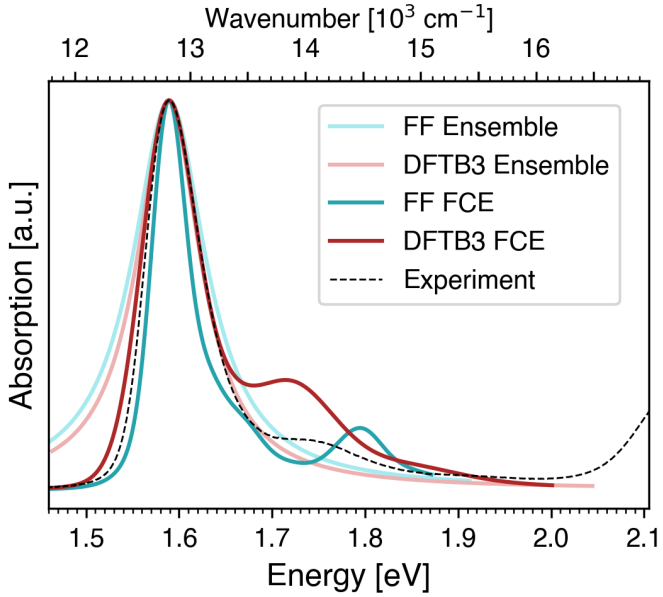


Fig. 7.9.: Absorption spectra obtained from solvent simulations in pyridine either with DFTB3 (red) or force field (blue). Results from the full cumulant expansion (dark colors) are compared to those from the ensemble method (bright colors), where the latter correspond to histograms of excitation energies weighted by oscillator strengths. All calculated spectra were shifted in position and scaled in intensity to match the experimental absorption spectrum (black). The experimental reference spectrum of BChl a in pyridine was taken from Ref. [197]. The Q_x band at 580 nm (2.13 eV) is indicated for comparison with the width of the Q_y peak.

The FCE spectrum obtained from DFTB3 sampling reproduces the experimental spectrum almost exactly: the vibrational shoulder appears at the correct position but with too high intensity. Intense high-frequency shoulders are a known artifact of the FCE method.[205] In contrast, the force-field-based FCE spectrum is too narrow, which is explained with the lower reorganization energy of BChl a in pyridine when sampled with force fields.

This comparison illustrates clearly how the spectral density determines the shape of the absorption spectrum. To assess whether it also influences the relative position of the absorption peak, the ensemble and FCE spectra were compared across all solvents. The comparison is showed in Figure 7.10.

The FCE spectra from the QM/MM simulations (upper row) generally show a single vibrational shoulder on the high-energy side. This reflects strong homogeneous broadening induced by the larger reorganization energies λ_{spd} compared to the force-field simulations (43 meV vs. 36 meV on average). The broadest spectrum was found for MeOH (λ_{spd} = 63 meV), where the vibrational shoulder almost merges with the main peak. MeOH has been an outlier in the distributions of excitation energies. Nevertheless, the pronounced vibronic side band is consistent with experiment, where the shoulder intensity in MeOH was measured to be approximately three times larger than in pyridine.[15] The opposite case is TEA, which yields the weakest spectral densities, resulting in two distinct shoulders.

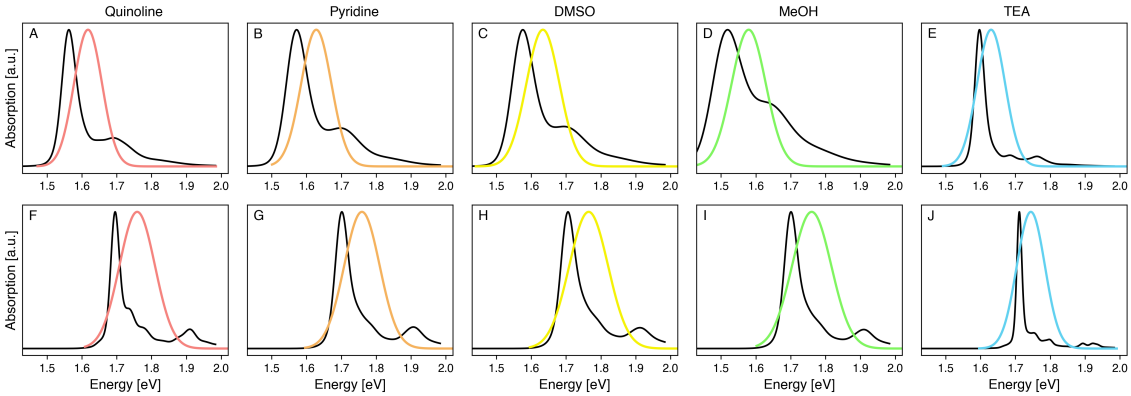


Fig. 7.10.: Absorption spectra obtained from simulations either with DFTB3 (upper row) or force field (lower row) in different solvents: quinoline (A,F), pyridine (B,G), DMSO (C,H), MeOH (D,I), and TEA (E,J). The Full-Cumulant expansion method (black) was compared against the ensemble method with Gaussian fits of the histograms (colors). Positions were not modified, but intensities were scaled to produce the same peak heights.

The FCE spectra from the force-field simulations (lower row) are much sharper in all solvents, consistent with the overall lower spectral densities and smaller reorganization energies.

In all panels of Figure 7.10, the colored Gaussian fits of the excitation-energy distributions are blue-shifted relative to the FCE spectra. This is rationalized by the mean excitation energy being effectively broadened by the spectral density with less weight on the high-energy shoulders. The off-set between the ensemble and FCE peaks vary across the solvents. The resulting implication fot the solvatochromic shift become clearer when the same data is presented differently: Figure 7.11 compares ensemble and FCE spectra for both QM/MM and force-field simulations. Within each data set, the spectra are shifted with a constant, such that the quinoline peak matches the experimental reference.

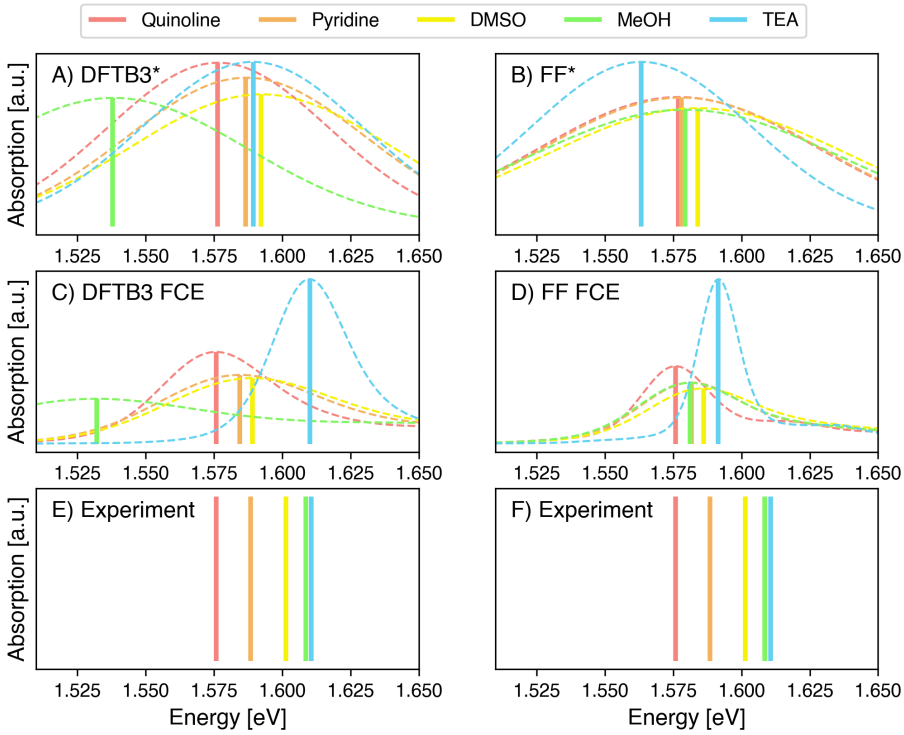


Fig. 7.11.: Maxima of calculated absorption spectra. The weighted histograms from QM/MM simulations with DFTB3 (A) and from force field simulations (B) are denoted with an asterisk. FCE spectra from DFTB3 (C) and force fields (D) are also shown. The same data as shown in Figure 7.10 is presented, but arranged differently. The absorption peaks are represented as vertical bars and the x-axis for each data group was shifted so that the calculated peak of quinoline is aligned to the experimental peak (E, F). Experimental absorption peaks in different solvents were taken from ref. [233].

Figure 7.11 illustrates that, the method used to calculate absorption spectra leads to qualitatively different solvatochromic shifts, even for the same QM/MM or MM simulations.

For the QM/MM simulations with DFTB3, the ensemble method (panel A) yields the following order of peak positions from low to high energy: MeOH, quinoline, pyridine, TEA, DMSO. Using the FCE method (panel C), the spectral density of TEA causes a relative shift of about 25 meV (200 cm^{-1}), changing the order to MeOH, quinoline, pyridine, DMSO, TEA. Apart from the outlier MeOH, the FCE results reproduce the experimental solvatochromic shifts better than the ensemble method.

For the force-field simulations (panels B,D), TEA is likewise shifted by approximately 25 meV. However, both approaches underestimate the experimental solvatochromic shifts, indicating that the electrostatic contributions of the solvents are not captured on the force field PES.

Beyond electrostatics, specific quantum-mechanical effects modulate solvatochromic shifts. Hydrogen bonds, as well as aromatic interactions between BChl a and the solvent were experimentally determined to affect absorption peaks by approximately 12 meV (100 cm^{-1}).[233] A preliminary analysis of charge transfer between MeOH and BChl a yielded a blue shift of only 5 meV (see Figure B.11), thereby not resolving the outlying MeOH energies. In

addition, polarization effects of the solvent environment can be addressed by by polarizable force fields,[259] or by including solvent molecules in the QM region[42].

Because of these specific interactions, experimental solvatochromic shifts are challenging to reproduce accurately. However, the FCE approach improves the agreement with experiment compared to the ensemble method, especially for QM/MM sampling.

7.4. Conclusion and Outlook

In this chapter, excitation energies of the photosynthetic bacteriochlorophyll a (BChl a) pigment were calculated using the semi-empirical TD-LC-DFTB2 method. Although the DFTB method had previously been benchmarked on geometry-optimized BChl a structures,[25] a detailed analysis of its sensitivity to geometric and electrostatic effects was still lacking.

Comparing DFTB3/MM simulations with pure MM simulations revealed that excitation energies are sensitive to the bond-length alternation and ring curvature of BChl a. The well-documented effect of the acetyl dihedral angle[8,165] could not be reproduced. In solvent simulations, electrostatics had a much stronger influence on relative shifts of excitation energies than geometry.

Additionally, the coupling of excitation energies to vibrational modes was analyzed in the form of spectral densities. Geometry mismatch between force fields and DFTB3 resulted in the overestimation of force field-based spectral densities. Based on this comparison, upper and lower bounds were obtained for the reorganization energy λ, a critical parameter for absorption, fluorescence, and electron transfer.[21] Although electrostatics had a much stronger influence on the relative shifts of excitation energies than geometry did, the contributions to λ quantified as inner- and outer-sphere reorganization energies were almost the same for geometries and electrostatics.

The obtained spectral densities were used to calculate absorption spectra using the full cumulant expansion (FCE) method. Different absorption energies in various solvents (solvatochromic shifts) were investigated. The FCE method was found to outperform simple histograms of excitation energies, both in reproducing the spectral shape and in capturing the solvatochromic shift. This reflects the fact that the absorption peak is not simply determined by the mean excitation energy, but rather by the vibronic fine structure of the spectrum. Therefore, accurate spectral densities are essential. In obtaining these, DFTB3 with the ob-f parameter set optimized for vibrational frequencies clearly outperformed the force field. Besides the outlier simulation of methanol, the solvatochromic shift was correctly reproduced only with the combination of DFTB3 and the FCE method.

Whereas the inability of TD-LC-DFTB2 to capture the dihedral effect is considered a methodological limitation, the out-of-plane motion of the magnesium atom in DFTB3 trajectories may be an artifact or a true part of the phase space. Calculations of total energies with higher level theories such as DFT/B3LYP are planned. Future developments

could address specific QM interactions, such as hydrogen bonding or π–π stacking between solvents and BChl a. The reason for the outlier behavior of methanol in the QM/MM simulations could not be identified within the scope of this work, and a more detailed analysis or repeated simulations are planned.

Altogether, this work establishes a pipeline for extracting spectral densities and calculating absorption spectra of single molecules, capable of reproducing the experimentally observed solvatochromic shifts more accurately than previous approaches based on excitation energy histograms.

8. Exciton Transfer in the Fenna–Matthews–Olson Complex at Cryogenic Temperatures

Reprinted in parts from
Monja Sokolov, David S. Hoffmann, Philipp M. Dohmen, Mila Krämer, Sebastian Höfener, Ulrich Kleinekathöfer and Marcus Elstner: Non-adiabatic molecular dynamics simulations provide new insights into the exciton transfer in the Fenna–Matthews–Olson complex
***Phys. Chem. Chem. Phys.* 2024, 26, 19469-19496**
with permission from the PCCP Owner Societies
DOI: 10.1039/D4CP02116A

Author Contributions:
The results presented in this chapter were generated in collaboration. The method has initially been developed for organic semiconductors and has been adjusted to biological systems by Philipp Dohmen and Monja Sokolov. Neural networks were developed by Mila Krämer. The FMO system was obtained from the group of Prof. Ulrich Kleinekathöfer. Monja Sokolov has trained the neural networks for the FMO system and has performed all MD and NAMD simulations at 300 K. All simulations at 77 K and subsequent analyses were performed by myself. In addition, I have performed a hyperparameter search for the environment-sensitive neural network and tested the trained neural network on QM/MM phase spaces. Also, I have performed additional analyses of simulations at 300 K in terms of absorption spectra, adiabatic states, and transfer rates.

8.1. Introduction

The bacteriochlorophyll a (BChl a) molecule, whose spectroscopic properties have been analyzed in the previous section, plays a central role in bacterial photosynthesis.

Comparably to the photosynthesis of plants, bacteria utilize antenna complexes (light-harvesting complexes, LHCs). These are protein-pigment-complexes that collect sunlight energy and transfer it to the reaction centers of the photosynthetic machinery. The *special pair* inside a reaction center initiates the multi-step photosynthesis process through charge

separation. This is essential for generating a proton gradient across the membrane, which ultimately leads to the synthesis of adenosinetriphosphate (ATP).

Green sulfur bacteria, unlike plants, use hydrogen sulfide and carbon dioxide as reactants to produce carbohydrates. They also employ a simplified photosynthetic apparatus that includes chlorosomes in the cytoplasm. These chlorosomes collect sunlight in form of excited carotenoid or BChl a molecules. The chlorosome is attached to reaction centers in the cell membrane via a protein baseplate and Fenna-Matthews-Olson (FMO) complexes. The FMO complexes transfer the collected energy from the chlorosome baseplate to the reaction centers.[43]

The water-soluble and trimeric FMO complex was the first LHC to be crystallized by Fenna and Matthews in 1975 in the organism *Chlorobium limicola*.[81] The structures from other organisms, such as *Prosthecochloris aestuarii* and *Chlorobaculum tepidum*, were resolved at the end of the 20th century.[288,162] Each monomer consists of one protein that coordinates seven BChl a molecules, specifically arranged in distinct binding pockets. An eighth BChl a molecule has been confirmed in 2011.[304] It is located at the interface between monomers and is easily lost during sample preparation and purification. Consequently, it is not considered in many experimental studies and is not included in this work.

The arrangement of the seven BChl molecules in one monomer is shown in Figure 8.1. Due to the orientation of the trimeric FMO complex between the chlorosome baseplate and the reaction centers, the excitation energy transfer (EET) is generally assumed to flow from pigments 1 and 6, which are closest to the baseplate, to pigments 3 and 4, which are closest to the reaction centers.[43]

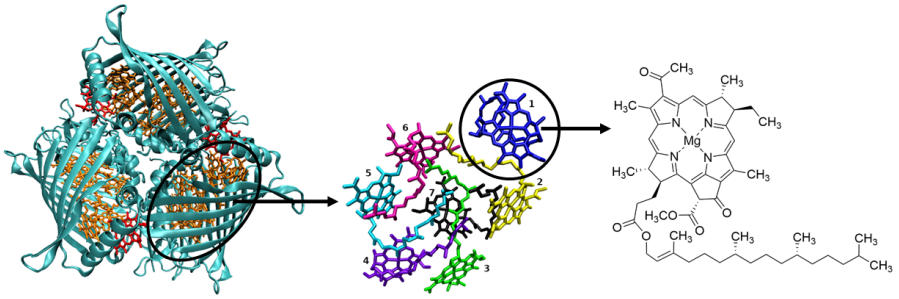


Fig. 8.1.: Arrangement and numbering of the seven BChl a molecules in one monomer of the FMO complex. The chlorosome baseplate is located above of BChl 1 and the reaction centers are located below pigments 3 and 4. Figure adapted from Sokolov et al. (2024)[272]

The FMO complex has been a model system for EET studies for decades due to its water solubility and relatively small size. The EET across the entire complex occurs in the sub-picosecond to picosecond regime with extremely high quantum yields.[297,84,27,281,69,231,106,71]

However, the efficiency of this process also poses a significant challenge in the study of light-harvesting complexes. Energy transfer in these complexes occurs via excitons, which can be localized to a single pigment or be distributed across several pigments. Delocalization is determined by the balance of the relative excitation energies of the

individual pigments (site energies) and their interactions (couplings). However, the reason for its astonishingly efficient transfer is also the main challenge in studying light-harvesting complexes. Energy transfer in these type of complexes occurs via excitons, which can be localized to a single pigment or distributed across several pigments. Delocalization is determined by the balance between relative excitation energies of the individual pigments (site energies) and the interaction between them (couplings). Diagonalizing the matrix with site energies on the main diagonal and couplings on the off-diagonal (Frenkel Hamiltonian, Equation 3.10), yields excitonic states which act as spectroscopically active adiabatic surfaces. Therefore, the aforementioned experiments reveal the transfer between excitonic states, which makes a pigment-wise resolution not directly accessible.

This situation is reflected in the absorption spectrum of FMO at room temperature.[282] Inhomogeneous broadening reduces the seven expected peaks to a single broad band. This indicates that there are minor differences in site energies compared to thermal motion. At cryogenic temperatures, the seven pigments produce three absorption peaks.[248] The peak with the lowest absorption energy is attributed to pigment 3.[282,281,39] The attribution of the remaining pigments to the peaks is ambiguous.[305,4]

Such interpretations of experimental measurements are only possible through computational studies and calculations of site energies and couplings. Minor differences in the site energies of the identical BChl a molecules are attributed to geometric and electrostatic tuning by the protein scaffold. The results of the calculations depend heavily on the quantum chemical method applied to the relatively large BChl a molecules. Furthermore, they depend on the force fields of the protein, the treatment of the environment, and the FMO species.[133,136,244]

The couplings mainly depend on the pigments' distances and orientations, which are predetermined by the protein matrix. Their theoretical predictions are more robust across different methods. Therefore, many studies utilize calculated couplings and fitted site energies to reproduce the absorption spectrum.[104,218,305,3] However, these fitted site energies depend on the initial guess and the optimization algorithm, so they cannot be considered as experimental values.

Many computational studies have compared their calculated site energies and couplings with other values in the literature and have investigated exciton dynamics in the FMO complex using Ehrenfest-type mean-field approaches, density matrix propagation methods such as Redfield theory, or hierarchical equations of motion (HEOM). Thus, there is a large agreement in the scientific community about the pathway of the exciton from pigment 1 or pigment 6 to pigment 3 or 4.[210,98,123,236]

While these methods differ in terms of accuracy and computational cost, they all describe the exciton as a population, rather than following a single exciton in real space. As such, they provide an ensemble-averaged picture of energy transfer, but they do not account for the effect of the exciton on the geometry of the BChl a molecule. Specifically, they neglect the feedback loop involving the excited-state charge redistribution in BChl a, its back-reaction on the protein scaffold via altered electrostatic interactions, and the resulting feedback of the protein on BChl a site energies.

To address this gap, real-time exciton simulations that capture geometric effects, coupling fluctuations, energetic disorder, and exciton dynamics in atomistic detail are highly desirable to elucidate the details of the exciton transfer.

Non-adiabatic simulations, such as surface-hopping methods, can describe the coupled motion of electronic and nuclear degrees of freedom in atomistic detail. To apply this method to the FMO complex, a benchmark of the semi-empirical TD-LC-DFTB2 method on site energies and couplings of BChl a molecules was performed.[25] Subsequently, neural networks were trained to accelerate the calculation by three orders of magnitude.[271] This speed increase is accompanied by the fragmentation of the system in the form of a Frenkel Hamiltonian. The coupled electronic-nuclear degrees of freedom are then propagated using trajectory surface hopping (see section 3.3) with pigment-wise population of the exciton being determined by the Frenkel Hamiltonian.

The exciton's feedback on the nuclear degrees of freedom is modeled using an altered charge distribution, which is scaled by the excitonic populations (Equation 3.24) on the respective pigments. This approach explicitly samples the outer-sphere reorganization energy (see section 4.4). The internal reorganization energy λ of a BChl a molecule in a protein pocket is not known. It is possible to explicitly model this using excited states forces, which are currently unavailable. Therefore, an implicit relaxation scheme is applied. The site energies are reduced proportionally to the excitonic population, using a predefined λ as the scaling parameter. The nuclei are propagated on a single adiabatic potential energy surface (PES), and there is a certain probability for a hop from one PES to another, when the surfaces approach or cross each other. Therefore, the feedback of the protein on the Frenkel Hamiltonian is explicitly taken into account, thus providing a bottom-up framework to analyze exciton dynamics. The main drawbacks of the surface-hopping method are that there is no rigorous derivation from first principles, and swarms of trajectories have to be sampled to yield ensemble averages in order to compare to experimentally revealed observables.

In this study, experimental absorption spectra and measurements of energy transfer were simulated. The temporal evolution of site energies was analyzed in terms of spectral densities and absorption spectra. NAMD simulations at room temperature were analyzed to define appropriate input variables. Simulations at 77 K were utilized to test the ability to reproduce cryogenic experiments and to elucidate the exciton transfer mechanism. Together, these results establish a benchmark for the capacity of atomistic simulations to link microscopic fluctuations with ensemble observables.

8.2. Computational Details

8.2.1. Trajectories at 77 K

In addition to the existing MD simulation of the FMO complex from *C. tepidum* (PDB: 3ENI) at 300 K[25], the same crystal structure was set up by myself and equilibrated at 300 K. The

two structures were simulated by Monja Sokolov at 300 K and are denoted as *set 1* and *set 2*, respectively.

Six additional MD simulations at 77 K were conducted by myself. The first five simulations result from the 300 K MD simulations by cooling down five equally spaced snapshots from the first 10 ns of *set 1*. Cooling down was achieved by performing a 2 ns NVT equilibration with a 1 fs time step and a Nosé-Hoover thermostat. Then, a 25 ns NPT equilibration was performed with a 1 fs integration step, Nosé-Hoover thermostat and Parrinello-Rahman barostat to maintain the target pressure of 1.013 bar. Finally, a productive simulation was performed for 100 ns with a 2 fs time step and hydrogen bond constraints using the LINCS algorithm. The Nosé-Hoover thermostat and the Parrinello-Rahman barostat were used. These five simulations are referred to as *set 3* to *set 7*.

For the sixth simulation at 77 K (*set 8*), the energy minimized crystal structure (PDB: 3ENI) was heated up to 77 K using the same protocol for NVT and NPT as described above. Since the density had not yet converged, two further NPT simulations, first of 20 ns, then of 100 ns, were performed before starting the productive run for 100 ns with a 2 fs time step. No restraints on heavy atoms were required for any of the productive or equilibration runs due to the slow movement of the atoms.

Tab. 8.1.: Summary of simulations that provided the sets of snapshots used in this study.

simulation	description
set 1	20 ns, 300 K, heat up PDB: 3EOJ
set 2	10 ns, 300 K, heat up PDB: 3ENI
set 3	100 ns, 77 K, start from set 1, frame 0.0 ns
set 4	100 ns, 77 K, start from set 1, frame 2.5 ns
set 5	100 ns, 77 K, start from set 1, frame 5.0 ns
set 6	100 ns, 77 K, start from set 1, frame 7.5 ns
set 7	100 ns, 77 K, start from set 1, frame 10 ns
set 8	100 ns, 77 K, heat up PDB: 3ENI

8.2.2. Single Point Calculations and Absorption Spectra

In all calculations of BChl a molecules, their phytyl tail has been excluded from the QM region and replaced by a hydrogen atom. It has been shown that this procedure does not significantly affect their excitation energies[165,54]. Excitation energies of the bright state and the corresponding transition charges were calculated with TD-LC-DFTB2 using the *DFTB+* package with the SCC tolerance set to 10^{-10} electrons. For the site energy calculations, the BChl a was treated with electrostatic embedding, i.e. its chemical environment was included using its electrostatic potential (ESP). The pairwise couplings were determined as Coulomb interactions between the transition densities of the two respective pigments, approximated by Mulliken transition charges from gas phase TD-LC-DFTB2 calculations

of the respective geometries at that moment in time. Both quantities were shown to agree well with DFT and ab initio references.[149,23,270,249]

Two modifications of the obtained Coulomb couplings were considered: The first correction refers to the applied charge model used to compute the Coulomb couplings. The transition dipole moments calculated from these charges ($\approx$ 7.5 D on average) are overestimated compared to the experimental value ($\approx$ 6.1 D[141,3]). Such an error in the charge model has also been observed for TD-DFT/B3LYP, which was previously used to determine transition charges for TrESP couplings.[171,210] Usually the charges are simply scaled to the experimental value.[210,3] For TD-LC-DFTB2, this results in a factor of 0.656 ($\widehat{=}$ factor 0.81 for the transition charges) to correct for the overestimated TD-LC-DFTB2 vacuum couplings. While this first scaling factor corrects the charge model, a second factor is typically applied to account for the polarizability of the medium between two pigments. Scholes et al.[253] found an empirical formula to determine a reasonable scaling factor as a function of the distances between different pigments in different photosynthetic proteins. The resulting scaling factor was between 0.56 and 0.68 for the present system.

Therefore, in the present study, three coupling models are compared: the TD-LC-DFTB2 couplings, the scaled ones (first scaling factor applied) and the scaled and screened ones (first and second scaling factor applied).

Spectral densities were obtained from a 60 ps QM/MM simulation of each individual chromophore at room temperature. The QM zone was described with the DFTB3 method, using the 3OB-f parameters developed for vibrational properties[90] and proven to reproduce spectral densities in very good agreement with experimental results.[172] The classical time-correlation function (Equation 4.19) was applied based on 60 000 site energy calculations with electrostatic embedding.

Absorption spectra with the cumulant expansion method[170] were generated using a Fortran code published from the University of Pisa.[57,58] The time-averaged Hamiltonians were created from the MD simulations at the respective temperature (*set 1* or *set 3*). The time-averaged transition dipoles were obtained from the same MD simulations via rotating the calculated transition dipoles to be aligned to the first snapshot of the trajectory. Vector averages were calculated across all frames in two steps. First, each x, y, and z component was averaged independently. The length of the resulting vector was subsequently scaled to the averaged norm.

8.2.3. Exciton Transfer Simulations

The propagation of the electronic and nuclear degrees of freedom as described in chapter 3.3 is based on Gromacs version 4.6, coupled with an in-house charge/exciton transfer code. A small integration time step of 0.1 fs was used in the MD steps of the NAMD simulations. For every set of input parameters (reorganization energy, couplings) or temperatures, 100 NAMD simulations with a length of 10 ps each were performed. The 100 different initial structures stem from the MD simulations (*set 1* to *set 8*) described above, evenly distributed over the respective trajectories. The simulations were initiated with the exciton assigned

to one pigment, which was either pigment 1 or pigment 6. This is a common practice in literature, e.g.[70,193,172], since these pigments are most likely the entrances of the energy coming from the baseplate.

To obtain the charges for the BChl a molecules in their excited state, the geometries of the BChl a molecules (without phytyl tail) were optimized in the ground and first excited state on a B3LYP[14,158,275]/6-31+G(d,p) level of theory using Gaussian 09.[86] The electron density was calculated, and the partial charges were obtained from a fit to the electrostatic potential on a CAM-B3LYP/def2-TZVP level[308] using the KOALA program[107]. The partial charges of the BChl a molecules for the force field were updated with the difference between the ground and excited state charges, weighted with the occupation.

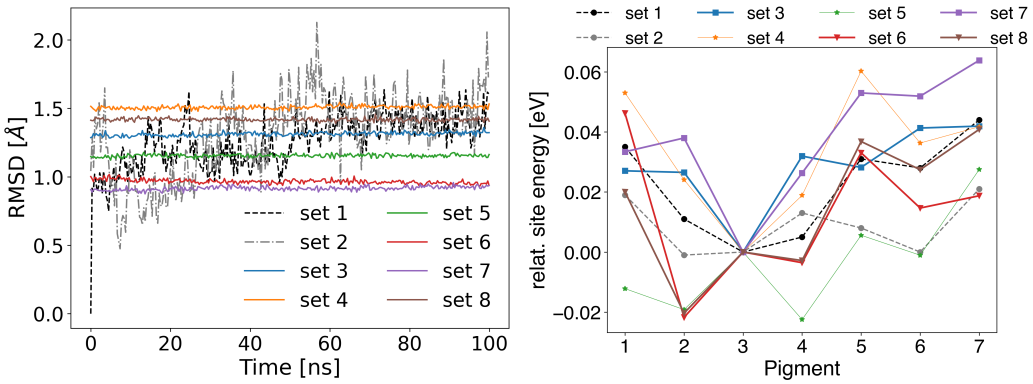
8.3. Results and Discussion

The results section is structured as follows: First, the experimental absorption spectra at two temperatures are reproduced using ground-state MD simulations. Next, swarms of NAMD simulations are analyzed to determine the optimal combination of input parameters for reproducing the experimentally determined time scales of exciton dynamics. Finally, single NAMD trajectories are analyzed to draw conclusions about the transfer mechanism.

8.3.1. MD at 77 K

In order to compare to spectroscopic experiments at cryogenic temperatures, the two room-temperature simulations performed by Monja Sokolov (sets 1-2) were complemented by six independent MD simulations at 77 K (sets 3–8). In accordance with the protocol outlined in section 8.2, five of the simulations (sets 3–7) were obtained by cooling down from a 300 K trajectory, while one trajectory (set 8) was generated by heating up directly from the crystal structure.

To qantify the reduced geometric fluctuations at low temperatures, the root mean square deviation (RMSD) of the FMO protein backbone was monitored over time (Figure 8.2a). In all 77 K simulations the RMSD remains flat, fluctuating below 0.1 Å. This indicates that the proteins are trapped in local minima in all single simulations. The fluctuations are significantly larger in the two simulations at 300 K (sets 1–2), which show RMSD fluctuations of up to 1.5 Å. Comparable low fluctuations are also evident in the time-series of site energies and Coulomb couplings (Appendix Figs. C.1 and C.2), confirming that thermal motions are effectively quenched at 77 K.



(a) RMSD **(b)** Site energies

Fig. 8.2.: The two simulations *set 1* and *set 2* at 300 K (black, gray) are compared with the six simulations sets 3-8 at 77 K (colored). a) Root mean square deviation of the FMO backbone in Å with respect to the starting structure of the simulation of *set 1*. The RMSD of *set 2* was shifted by -1 Å. b) Mean of site energies of the nine sets during the MD simulation.

The averaged site energies of the pigments at 77 K were compared with the ones obtained from the two simulations at 300 K. As illustrated in Figure 8.2b, two clusters were identified in the 77 K simulations. Cluster A (*set 3* and *7*) replicates the observed ordering at 300 K: pigment 3 has the lowest site energy, followed by pigments 2 and 4. In cluster B (*set 6* and *8*), however, pigment 2 has the lowest excitation energy, followed by pigments 3 and 4. The remaining two simulations (*set 4* and *5*) demonstrated inconsistent rankings. Apparently, they have sampled certain local minima of the 300 K phase space and were therefore excluded from further analysis. The sets 3, 6, 7, and 8 were selected as representative low-temperature ensembles for further analysis.

8.3.2. Neural Networks and Phase Space

To carry out NAMD simulations at 77 K, neural networks for site energies and couplings are required. Therefore, the transferability of the NNs across the two temperatures was investigated. The NN, that has been trained on geometries from the 20 ns long MD simulation at 300 K (*set 1*) was applied to the data points of the simulations at 77 K. As demonstrated in Figure 8.3, the mean absolute errors were determined to be approximately 4 meV for site energies and less than 1 meV for couplings. These performance metrics are identical to those on the original 300 K test set, indicating that the 77 K local minima lie within the phase space of the higher-temperature MD. Therefore, no retraining was performed.

In addition, the transferability of the force field-trained neural network to QM/MM geometries was analyzed. Because QM/MM embedding with predicted DFTB3 forces is planned for future works, the NN predictions of geometries along a 60 ps long ground state QM/MM simulation at 300 K were analyzed. This simulation was used to calculate the spectral density. The results are summarized in Figure 8.4.

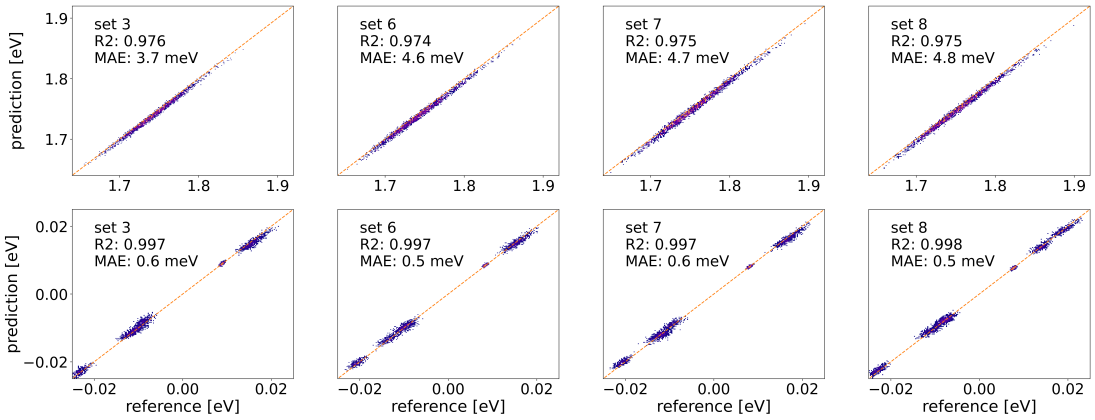


Fig. 8.3.: Scatter plots of the neural network that was trained on 100,000 geometries from 300 K simulations for site energies (top row) and coulomb couplings (bottom row) predicting the respective properties from the four simulation sets at 77 K.

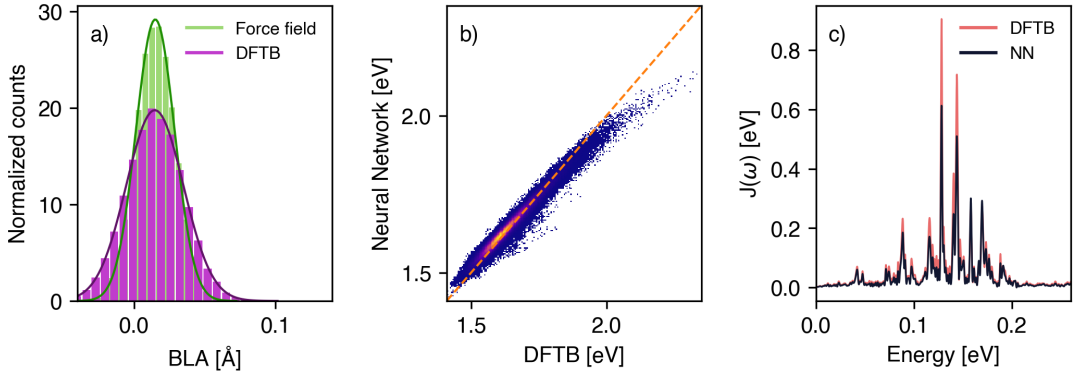


Fig. 8.4.: Assessment of neural network predictions for two phase spaces at 300 K. a) Bond length alternations for seven pigments in the ground state QM/MM simulation with DFTB3/3ob-f (purple), in comparison with the MD simulation (green). b) Scatter plot comparing the site energies of all pigments in the ground state QM/MM simulation, evaluated using the neural network, or using the reference method TD-LC-DFTB2. c) Spectral density of one pigment derived from the ground state QM/MM trajectory. The spectral density obtained from predictions (black) is compared against the one obtained from calculations with TD-LC-DFTB2 (red).

In Figure 8.4a, the bond-length alternations (BLA) of the seven pigments were compared between the QM/MM geometries obtained by DFTB3/3ob-f and the MM geometries of the training data. The bond length alternation is an important geometric measure impacting the site energies of BChl a (see also chapter 7.3.1). The BLA of each individual pigment can be found in the appendix (Figure C.3). The distribution from DFTB is broadened at both ends of the BLA range compared to the stiffer distribution from the force field. This is consistent with generally stiffer potentials of force fields compared to QM methods. Consequently, there is a lack of training data for the extreme high and low BLA values. The effect is illustrated in the scatter plot in Figure 8.4b. It demonstrates the force field network's generally good prediction capability for QM/MM geometries. The performance, as indicated by an R^2 of 0.96 and an MAE of 12 meV, is only marginally worse than that of the original MM test data set (R^2 of 0.99 and MAE of 4 meV). In the high- and low-site

energy regions, the predictions are too high for low site energies and too low for high site energies. This bias towards the mean value is explained by the correlation between bond length alternation and excitation energy for aromatic molecules, as well as the lack of training data for extreme BLA values.

To estimate the effect of the poor predictions in the high- and low-site energy extremes, the spectral density along the 60 ps long QM/MM simulation was utilized. Figure 8.4c compares the spectral density from the NN predictions (black) with the one obtained from the reference method (red). The position of the peaks is strikingly similar, indicating that the temporal evolution of site energies was accurately reproduced. Only the peak intensities of the NN predictions (black) are lower, suggesting that the NN reproduces the frequencies, but underestimates the amplitudes of the site energy fluctuations.

Together, the results demonstrate that the existing neural network trained on force field geometries at 300 K, is sufficiently transferable to the force field geometries of the four simulation sets at 77 K. This finding aligns with the previous subchapter's conclusion that each 77 K simulation samples a local minimum of the room temperature trajectory. Consequently, no retraining is necessary, and NAMD simulations at 77 K can be performed directly. The same network is also capable of predicting site energies from 300 K QM/MM geometries that were sampled with DFTB3/3ob-f. Despite a few outliers in the very high and very low site energy domain, the spectral density was reproduced with minor deviations. The impact of these deviations on the absorption spectra is analyzed in the next chapter.

8.3.3. Absorption Spectrum

In order to assess the impact of site energies, couplings, temperature and spectral density on the absorption spectrum, a series of analysis was performed. In Figure 8.5, the key findings on the absorption properties of the FMO complex at different temperatures are collected. The top row shows the excitation energies of the seven BChl a pigments and the resulting eigenstates for the 100 ns long *set 3* simulation at 77 K (panel A) and the 20 ns long *set 1* simulation at 300 K (panel B). In both cases, 10,000 frames were analyzed.

The leftmost Gaussian fit to the site energy histogram in Figure 8.5A corresponds to pigment 3. Its distribution is clearly distinct from the others, and its significantly reduced mean value is also shown in blue in Figure 8.2b. All other pigments exhibit strong overlapping site energy distributions. The diagonalization of the time-averaged Frenkel Hamiltonian results in a set of adiabatic eigenenergies labeled as "Adiab average 1". Due to excitonic couplings, their splitting of 600 cm^{-1} exceeds that of the bare site energy range. An alternative averaging procedure ("Adiab average 2") represents diagonalization of the Frenkel Hamiltonians in every frame and subsequently averaging the 10,000 eigenstates. These averages yield a significant splitting of 800 cm^{-1}, due to averaging the states with the first, second, etc. highest energy without considering their composition. This non-physical method of averaging will be discussed in more detail in chapter 9. In the present

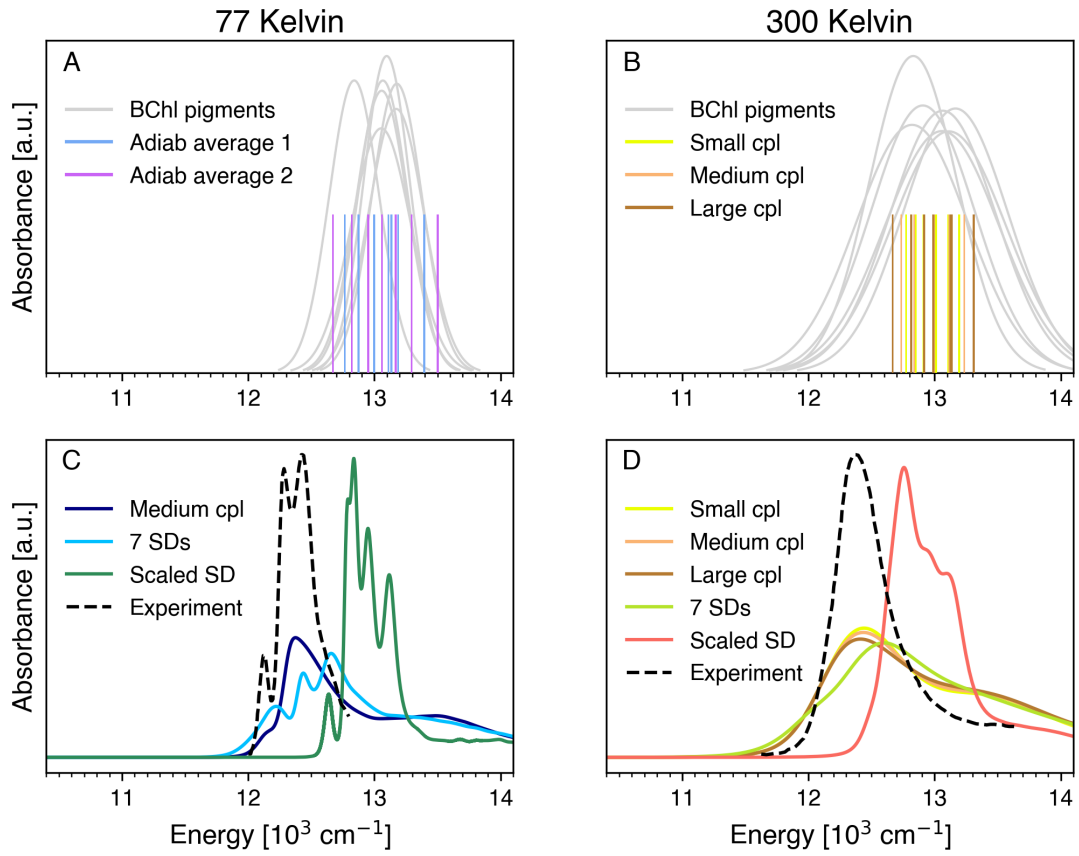


Fig. 8.5.: Absorption properties of FMO at different temperatures. Left: 77 K, *set 3*. Right: 300 K, *set 1*. Top: Site energies of the pigments (gray) and adiabatic energy levels (colors). Adiabatic averages and different couplings (cpl) are described in the text. Bottom: Absorption spectra calculated with the FCE method using several inputs (colors) compared to the experimental absorption spectrum at 77 K[248] and 300 K[282] (black). In all panels, medium couplings and an unscaled averaged spectral density (SD) were applied if not stated otherwise. For the representation of site energies (A, B, gray) as well as for all subsequent calculations (A-D, colors), the excitation energies were shifted by 1,003 cm^{-1} to compensate for the systematic overestimation of TD-LC-DFTB2, as outlined in the computational details. The intensities of all spectra in C and D are scaled to have the same integral.

analysis, only the first procedure, the diagonalization of the time-averaged Hamiltonian, is applied.

Panel B shows broader site energy distributions at 300 K (gray), as expected from the larger nuclear fluctuations. Furthermore, the effect of the couplings to the adiabatic states was explored. As outlined in section 8.2, the raw couplings are scaled with a factor of 0.656 to correct for the overestimation of the transition dipoles from TD-LC-DFTB2. Additionally, the couplings can be scaled further with an empirical distance-dependent screening factor found by Scholes et al.[256] which lies between 0.5 and 0.7 for the pigments in the present system. The raw, scaled, and scaled plus screened couplings are labeled as large, medium, and small, respectively. The corresponding adiabatic states are separated by 600 cm^{-1}, 500 cm^{-1}, and 400 cm^{-1}.

In summary, the effect of couplings to the adiabatic states is significant. The split between the lowest and the highest eigenstate is increased from 400 cm^{-1} to 600 cm^{-1} upon tripling the coupling value. However, when compared to the 600 cm^{-1} split found at 77 K with medium couplings, the effect of an energetically isolated pigment is stronger.

For the calculation of absorption spectra, the system's adiabatic energies are effectively broadened in the full cumulant expansion (FCE) method. Broadening is mediated by the system-bath coupling, which is quantified by the spectral densities of the pigments. In the case of zero spectral densities, the absorption spectrum consists of seven sharp peaks that precisely match the eigenvalues of the time-averaged Frenkel Hamiltonian. Consequently, different absorption spectra were expected for small, medium, and large couplings. However, at 300 K (panel D), the FCE-computed spectrum remains essentially unchanged by the three coupling strengths (brown, orange, yellow curves). All subsequent analysis were conducted using the scaled, but not screened (medium) excitonic couplings.

In addition to calculating the mean spectral density of the seven pigments, an FCE spectrum was created with the seven individual spectral densities (green curve). It also shows minimal deviations from the other three spectra. All of them are significantly broader than the experimental spectrum. The spectral densities are apparently very intense. As a result, the broadening caused by the system-bath coupling overlays the excitonic splitting induced by excitonic coupling. Reducing the mean spectral density by a factor of four results in a spectrum (red curve) whose full-width at half-maximum closely reproduces the experiment. The width matches that of the adiabatic eigenstates (Panel B, orange), but the entire spectrum is shifted to higher energy. This "reorganization-energy shift" is a well-known artifact of FCE modeling in photosynthetic proteins.[205] In practice, this is compensated for by shifting and scaling the calculated spectrum to match the experimental absorption spectrum.[58] For the current spectra, only the site energies were shifted by a fixed value to compensate for the systematic error in TD-LC-DFTB2, and the intensities of the spectra were scaled to yield the same integrals. No further manipulation was performed.

Panel C presents an analogous analysis for 77 K, employing the same set of spectral densities as in panel D, because the temperature effectively cancels in the calculation of classical spectral densities (see section 4.3). With the averaged spectral density, the spectrum is overly too broad (blue curve). It should be noted that FCE spectra in general suffer from too intense high energy side bands[58], which partially explains their broadening. However, the absence of fine structure is a clear identification for too intense spectral densities. The use of seven unscaled densities (cyan curve) reveals three peaks, but the curve is still too wide. After downscaling the mean spectral density, the calculated width aligns with the experimental result. The prominent low-frequency peak is attributed to pigment number 3, in consistency with prior studies.[281,4] However, the experimental double peak could not be reproduced, even with the downscaled spectral density.

The integral of the spectral density yields the reorganization energy λ (Equation 4.21). For the calculated spectral density, it was 486 cm^{-1} (60 meV). Reducing their intensities by a factor of four led to spectra that more closely agree with experiments. The resulting

reorganization energy of only 120 cm^{-1} (15 meV) is lower than most calculations of reorganization energies that employ a higher level of theory[229] and also lower than the value of 35 meV that was calculated by Monja Sokolov with DFT/B3LYP. However, it is higher than estimations of the outer-sphere reorganization energy in the FMO complex.[236,244] This critical parameter will be discussed in more detail at the end of the following subsection.

In summary, the analysis presented in Figure 8.5 shows that the vibronic fine structure of the cryogenic spectrum could be reproduced only to a small extent. At 300 K, the treatment of excitonic couplings resulted in adiabatic eigenstates whose energy split increased from 400 cm^{-1} to 600 cm^{-1} upon increasing the coupling. In contrast to the differing splitting of eigenstates, almost identical absorption spectra were obtained when the raw spectral density was used. A reduction to one-fourth produced a spectrum that closely matches the experimental one at 300 K. From the standpoint of absorption spectra, the downscaled spectral density and the resulting reorganization energy of 120cm^{-1} (15 meV) is a realistic description of the Bacteriochlorophyll a molecule. The interplay between temperature, site energy distribution, reorganization energy, and excitonic coupling will be examined in the next section from a dynamical point of view.

8.3.4. NAMD at 300 K

In order to disentangle the effects of the reorganization energy λ, and the excitonic couplings between pigments, six sets of NAMD simulations were performed. The inner sphere reorganization energy was set to 35 meV (282 cm^{-1}) or 65 meV (524 cm^{-1}), and large, medium, or small couplings were applied for each simulation. In Figure 8.6, the time-evolution of the excitonic population of the seven pigments is shown. All values were obtained by averaging over the swarm of 100 trajectories, each with pigment 1 being initially occupied. The curves are not entirely smooth, but 100 trajectories can be regarded as sufficient to converge the populations.

In all six simulations, the exciton's pathway was consistent (Figure 8.6): it transitioned from pigment 1 (blue) to the neighboring pigment 2 (orange) and then to pigment 3 (green). An almost converged equilibrium state at approximately 1.5 ps has been observed, in which the exciton was primarily located on pigments 3 and 4. A clear trend of the transfer speed can be identified regarding the coupling strength. The speed of the transfer from pigment 1 to pigment 2 was estimated by the intersection of the two curves and illustrated by the width of the gray box. The transfer was fastest for the largest couplings (a, b), intermediate for the medium couplings (c, d) and slowest for the smallest couplings (e, f).

The interpretation of the effect of the reorganization energy is more challenging. At large and medium couplings, the transfer is slightly faster for the higher reorganization energy (b, d) than for the lower one (a, c). However, this trend is reverted for the small couplings where a lower value of λ (e) leads to a faster transfer than the higher one (f). This indicates that the ratio of couplings to reorganization energy within the disordered energy landscape impacts the dynamics in a nontrivial way.

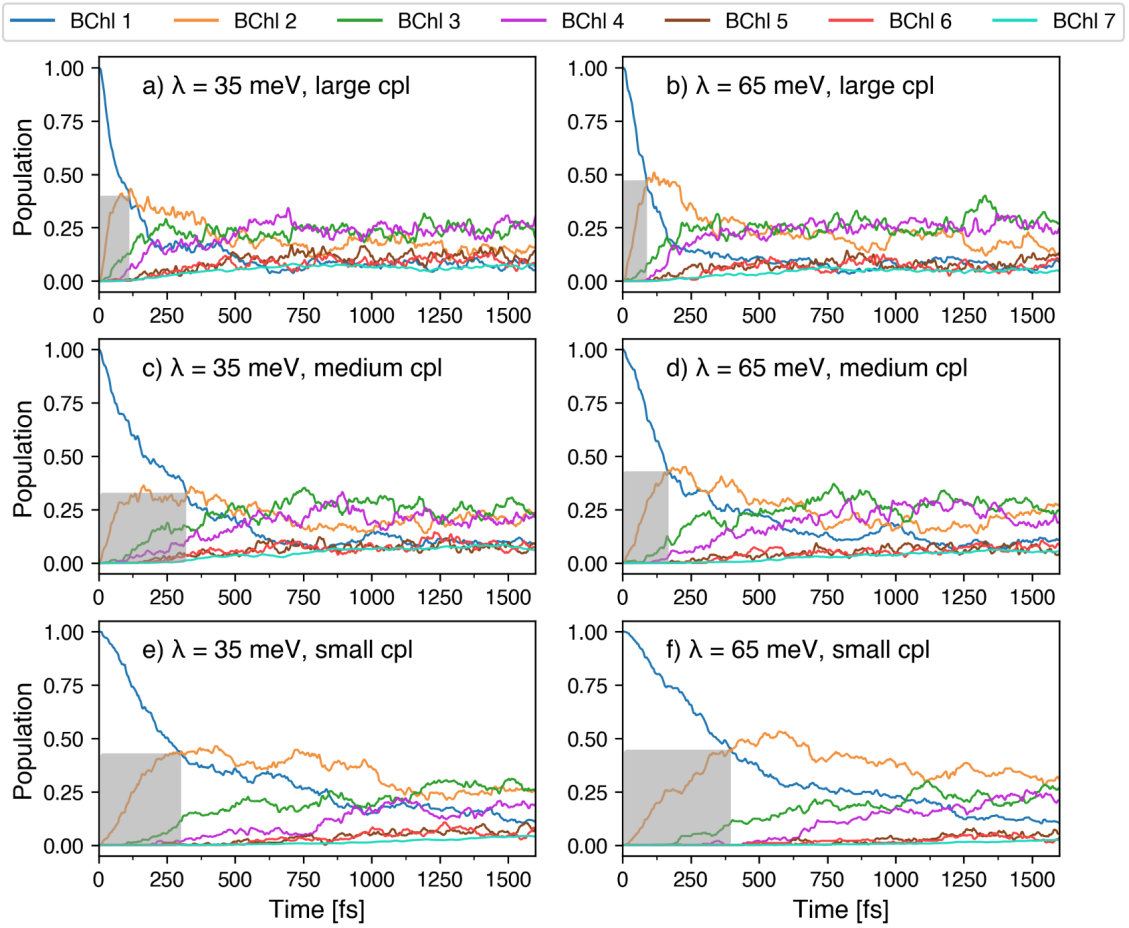


Fig. 8.6.: Diabatic population of the exciton at 300 K for different input settings. All populations were averaged over the swarm of 100 trajectories per input condition. Low reorganization energies of 35 meV (a, c, e) were compared against high reorganization energies of 65 meV (b, d, f). The simulations were run with the raw couplings (large, a, b), scaled couplings (medium, c, d) and scaled plus screened couplings (small, e, f). A gray box from the origin to the intersection of pigment 1 and 2 estimates the speed of the transfer.

A set that is particularly suitable for NAMD simulations has been determined to be medium (scaled but not screened) couplings in conjunction with the reorganization energy of 35 meV (282 cm^{-1}). This decision was made for several reasons:
i) In many computational studies, a transfer time from pigment 1 to pigment 3 of 700 fs up to 1 ps was calculated.[4,193,244] The transfer from pigment 1 to pigment 2 may take approximately 200 fs[72,236] up to 500 fs[244]. Figure 8.6c reproduces these properties.
ii) The selection of scaled, but not screened, couplings is particularly relevant in light of a recent study that calculated the couplings between chlorophylls using the TD-DFT/M062X level of theory in a polarizable force field. The authors determined that the weakening screening effect by the electrostatic medium is exactly compensated by the boosting influence of the induced dipoles in the environment.[46]
iii) The reorganization energy of 35 meV lies within a huge bandwidth of computations: It was suggested to be 15 meV[72], between 9 and 53 meV[244], or between 16 and 83 meV[229]. The experimentally determined reorganization energy in triethylamine was 65 meV.[229]

This value for the pigment in solution serves as an upper bond for pigments in a protein matrix.[157,179]

Reorganization energies are of particular interest, since they can slow down the propagation significantly, as pointed out by Duan et al.[72] Their estimation of about 15 meV is much smaller than the parameter ranges for the inner-sphere reorganization energy between 35 and 65 meV that was considered in this work. However, they reported the reorganization energy applicable for a localized transfer. For delocalized excitons, it is reduced by the population of the pigments (Equation 3.24). For instance, if an exciton is delocalized over two sites, the reorganization energy would be reduced per site by 50%, which already can rationalize the difference between the theoretical values and experimental estimates.

A detailed study about the delocalization was performed on the NAMD simulations at cold temperature.

8.3.5. NAMD at 77 K

Four swarms of NAMD simulations at 77 K were performed with the aforementioned parameter set. The starting structures were taken from the four 100 ns long MD simulations at 77 K, which were categorized as cluster A (sets 3 and 7) and cluster B (sets 6 and 8) at the beginning of this chapter. In Figure 8.7, the time evolution of the averaged excitonic population of the seven pigments is presented. As was the case at 300 K, the curves are sufficiently converged to allow statements about the temporal evolution of the exciton.

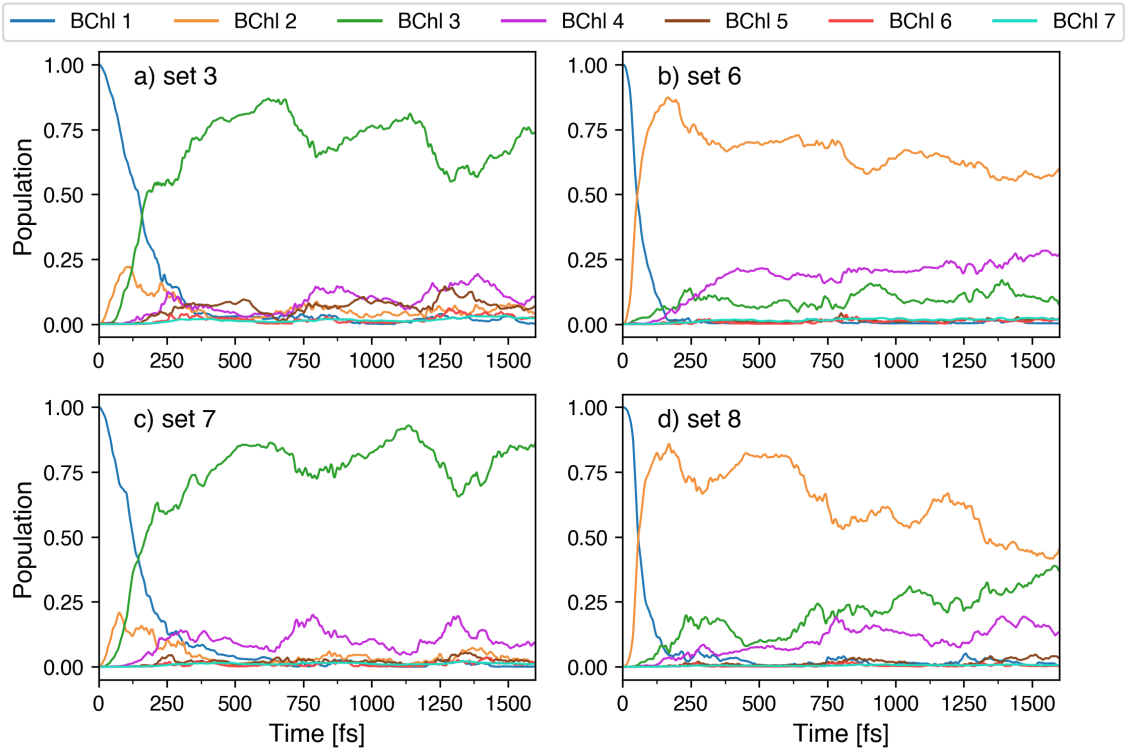


Fig. 8.7.: Diabatic populations at 77 K for the four different sets averaged over the swarm of 100 trajectories per set. All simulations were performed with pigment 1 being initially populated, a reorganization energy of 35 meV, and scaled but not screened couplings.

Cluster A (left panel) displays the same qualitative progression of the exciton compared to the simulations at 300 K (Figure 8.6): pigment 1 is continuously depopulated, pigment 2 serves as an intermediate state as in quasi-equilibrium kinetics, and pigment 3 is the final acceptor. The transfer to the final acceptor is significantly faster at 77 K than at 300 K. In addition, after one picosecond, pigment 3 reaches a three times higher population than at 300 K, indicating localization at the lowest energy pigment. The localization is clearly driven by the site energy of pigment 3. Its mean value was approximately 30 meV lower than the value of the second lowest pigment. At 300 K, this difference was less than 10 meV.

The simulations of cluster B (right panel) showed the same trend as the simulations of cluster A: a faster transfer and a higher localization on the final acceptor than at room temperature. The final acceptor was pigment 2 in this case. The energetic distance between the site energy of pigment 2 and the second lowest pigment was approximately 20 meV, which explains the moderate localization on the final acceptor compared to cluster A.

In all four simulations, a periodic population transfer from and to the highest occupied pigment can be observed over the time span between 250 fs and 1 250 fs. Because quantum effects like long-lived coherences were recently ruled out[35] and the applied surface hopping method employs decoherence corrections, the cause of the wave-like shape of the population plots is assumed to be randomness and lack of convergence.

In summary, two trends were observed. First, the transfer in all NAMD simulations at 77 K is faster than at 300 K. Comparisons to accurate experimental kinetics are not available, due to the excitonic nature of the individual pigments. Moreover, the excitation pulses, that are used in two-dimensional spectroscopic experiments, are on the same time scale as the exciton lifetimes.[281] However, the finding is in qualitative agreement with computational studies by Fleming and Ishizaki[123] and by Gillis.[98] The latter argued that higher temperatures lead to a slower transfer, due to higher fluctuations of site energies and the resulting lower probability of the exciton reaching the energy sink.[98] This is precisely the case in the present study, as the reduced movement of nuclei (Figure 8.2a) leads to a fourfold decrease in fluctuations of site energies at 77 K (Figure C.1). The second observed trend is the localization of the exciton on the final acceptor (energy sink), which is significantly higher at 77 K compared to 300 K. This finding aligns with other studies[98,210,237,242] and is also attributed to the reduced fluctuations at cold temperatures.

The two observations of faster and more localized excitons at 77 K contradict each other, because localized hopping transport is generally slower than delocalized wave-like transport.[257] While the correlation between fluctuations and delocalization appears intuitive, its applicability to the present protein complex remains unclear. In this case, the key factor for delocalization is the interplay of static energetic disorder, size of couplings, and fluctuations of site energies and couplings. Furthermore, the definition of delocalization has a significant impact on the overall analysis, resulting in non-trivial temperature dependencies.[64] Therefore, a detailed analysis of delocalization is necessary.

8.3.6. Delocalization

To quantify the trend regarding higher localization at lower temperature, the inverse participation ratio (IPR) was calculated over all trajectories of each swarm and averaged over the entire length of 10 ps. The result is shown in Table 8.2.

Tab. 8.2.: Inverse participation ratio (IPR) of the exciton over the FMO pigments averaged over 100 trajectories with a length of 10 ps each. The starting structures were taken from the classical MD simulations (Table 8.1). The couplings were scaled by a factor of 0.656 to correct for the overestimation of the transition dipole moment with TD-LC-DFTB2. The reorganization energy λ was set to 35 meV.

	300 K	77 K			
	set 1	set 3	set 6	set 7	set 8
$\lambda = 35$ meV, scaled	1.37	1.34	1.26	1.27	1.23

A clear trend of low temperature leading to more localization can be observed, confirming the local trapping hypothesis. The analysis is in good agreement to Kleinekathöfer et al. who estimated the delocalization length in the FMO monomer to be 1.4 at physiological temperatures.[210] Moreover, the table shows the sensitivity of the IPR measurement. The population plots at 77 K (Figure 8.7) appear to be much more localized than the corresponding plots at 300 K (Figure 8.6). However, the IPR decays only by less than 10% upon reducing the temperature.

Averaging the IPR over the whole simulation length of 10 ps quantifies the steady state delocalization of the exciton in equilibrium. However, the initial few femtoseconds of the transfer process are of particular relevance, which motivates the time-resolved analysis of populations and delocalization. In order to obtain results independent from the different phase spaces of the seven simulation sets, the two swarms of NAMD simulations at 77 K (*set 3*) and 300 K (*set 1*) were complemented by two simulations using the same starting structures respectively, but having pigment 6 instead of pigment 1 initially occupied.

The simulations at 300 K have already proven that the NAMD simulations reproduce the correct path of the exciton, either from pigment 1 via pigment 2 to the final acceptor pigment 3, or from pigment 6 via the pigments 5 and 4 to pigment 3.

As illustrated in Figure 8.8, the pathways are unaffected by temperature variations. Furthermore, the transfer is fastest for 77 K and pigment 6 (d), followed by 300 K and pigment 6 (c), 77 K and pigment 1 (b), and finally 300 K and pigment 1 (a). The finding, that relaxation from site 6 is faster than from site 1 has been reported before.[70] The IPR for the first 100 fs displays the same trend, reinforcing the assertion that a delocalized transfer mechanism is faster than a localized one. The observed range order of the IPR is different from the time-averaged values of the IPR (Table 8.2).

Due to this discrepancy, a steady state delocalization as quantified in some studies[220,189] is not a useful quantity. Instead, the delocalization must be followed in a time-dependent manner to explain the speed of the transfer. This finding is in agreement with a recent study by Rodriguez and Kananenka[237] who systematically analyzed the coherence length dynamics in FMO with hierarchical equations of motions (HEOM). They found that "the coherence length decreases for low temperatures in late stages of the dynamics". However, for the first few hundred femtoseconds, low temperatures lead to a higher delocalization. The authors conclude that there is a "correlation between the coherence length dynamics and the EET process" which is not obvious when only equilibrium values are considered.[237]

Therefore, the time domain of the delocalization must be taken into account to explain the transfer speed. However, when making statements about the transfer mechanism, the time-domain averaged over a swarm of surface hopping trajectories is insufficient. Instead, individual trajectories must be analyzed.

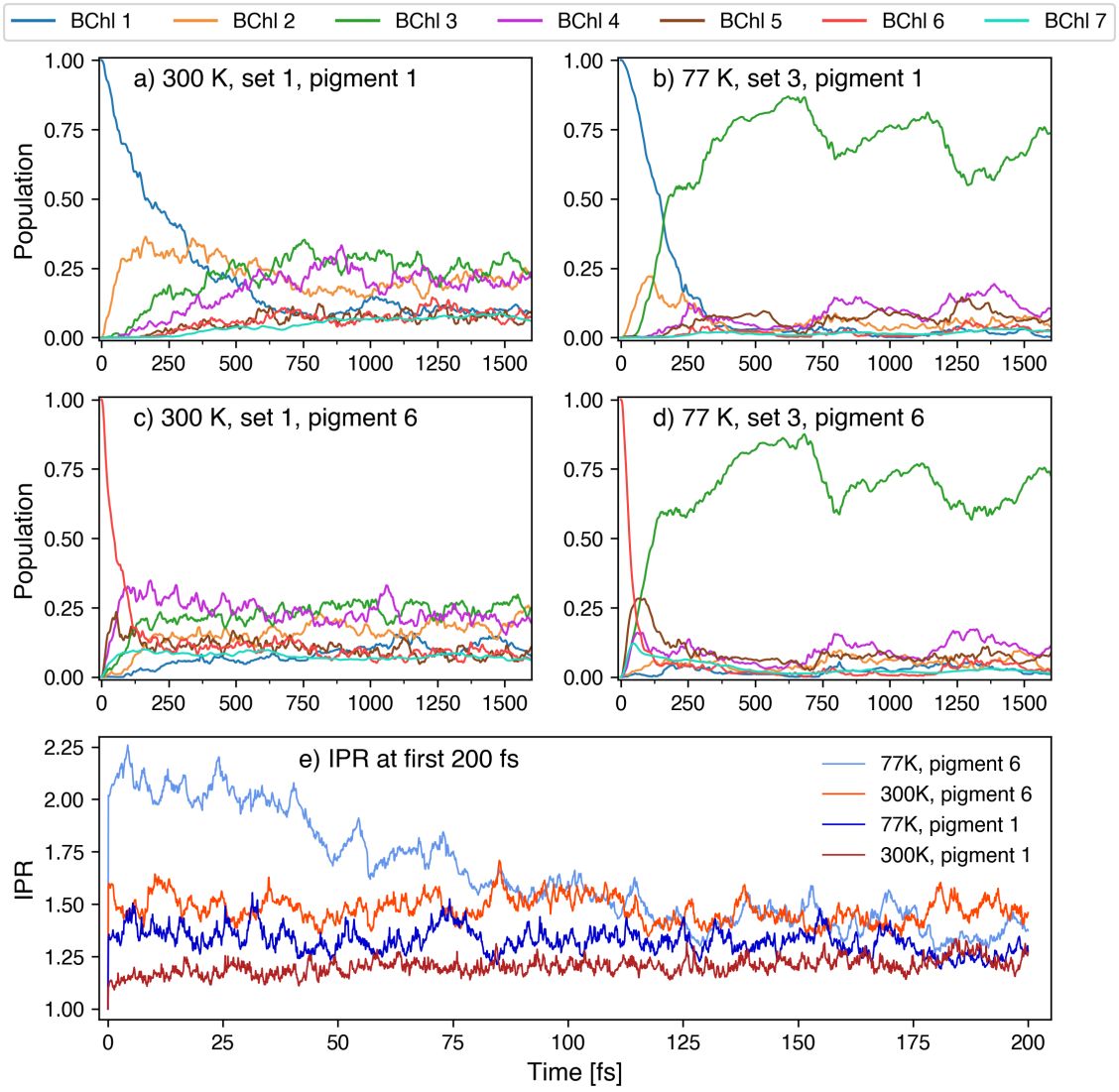


Fig. 8.8.: Averaged diabatic populations of the pigments during the 100 NAMD simulations of *set 1* and *set 3* with $\lambda = 35$ meV, scaled but not screened couplings and the exciton initially localized at pigment 1 or 6. (a) Trajectories of *set 1* with our default settings at 300 K and initial occupation of pigment 1 and (c) pigment 6. (b) Trajectories of *set 3* with the exciton being initially localized at pigment 1 and (d) pigment 6. (e) The IPR of the four simulations shown above, each averaged from 100 NAMD simulations.

8.3.7. Single Trajectory

The two pathways that the exciton takes from the initial pigment (1 or 6) to the final acceptor (pigment 3) have been correctly reproduced with the swarm of 100 NAMD trajectories. In addition, the temporal evolution of delocalization and the relative speed of the EET were successfully reproduced, thereby validating the swarms as an accurate representation of the underlying reality. To make statements about the transfer mechanism, one exemplary trajectory is shown for each of the two starting conditions in Figure 8.9.

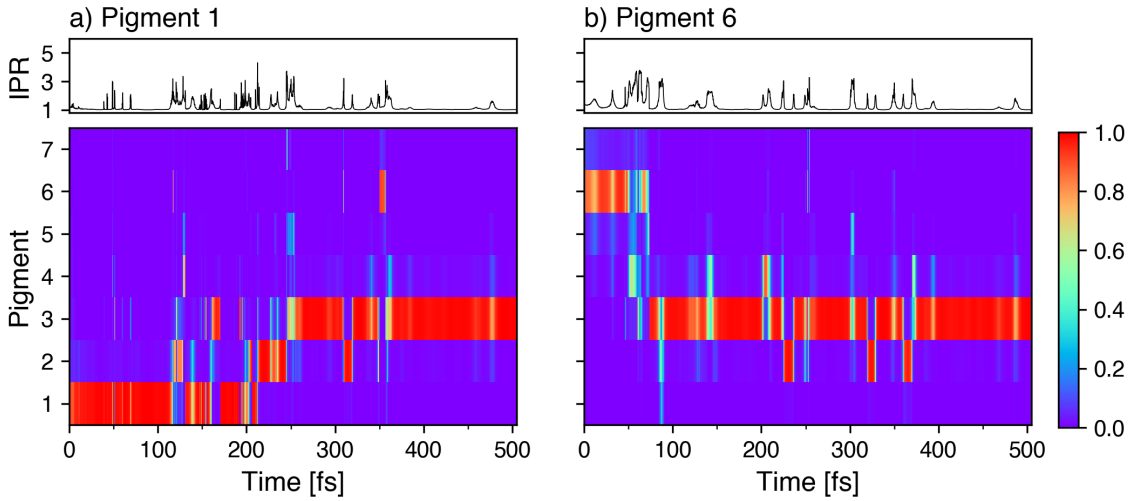


Fig. 8.9.: One NAMD simulation at 77 K of *set 3* with $\lambda = 35$ meV, scaled but not screened couplings and the exciton initially localized on pigment 1 (a) or pigment 6 (b). The top row shows the inverse participation ratio (IPR) and the bottom row shows the populations of the pigments for the first 500 fs of the simulation.

In the simulation of an exciton that originates at pigment 1 (left panel), the excitation remains at pigment 1 for the initial 100 fs. The sharp spikes in the IPR around 50 fs indicate attempted hopping events. Over the next 100 fs, the first three pigments are populated from time to time, and the exciton most often localizes on the respective pigments. The final transfer to pigment 3 occurs at 250 fs. Subsequently, the IPR remains close to one. Initiating the simulation with pigment 6 changes the result qualitatively. The right panel shows a stable delocalization lasting for the first 100 fs, during which the exciton is distributed among three to four pigments. Subsequently, it localizes on pigment 3 and remains there, with some attempted transfer events to neighboring sites.

Delocalized transition states are more likely when the exciton is initially populated on pigment 6. The discrepancy in behavior between incoherent hopping and transiently stable delocalization could potentially explain the fast transfer from pigment 6 to pigment 3. This fact is of importance for the subsequent chapter, in which the transfer mechanism within the two subsystems of the LH2 complex will be analyzed more deeply.

8.4. Conclusion and Outlook

In this chapter, the potential of simulating absorption spectra and energy transfer in the photosynthetic FMO complex was investigated. As the experiments are typically conducted at cryogenic temperatures, simulations were analyzed at both room temperature and 77 K. It has been demonstrated that reproducing cryogenic experiments with a single trajectory is unfeasible due to trapping in local minima. However, a cold temperature trajectory was identified that best aligned with the properties at room temperature. Comparisons to experimental absorption spectra revealed the quality of the underlying theoretical method.

Non-adiabatic molecular dynamics simulations enabled a comparison of exciton transfer at 300 and 77 K. The varying transfer speeds were attributed to delocalized excitations.

The delocalized collective excitations of the seven pigments in the FMO complex were calculated in the form of excitons. Excitons are described with pigment-wise excitation energies (site energies) and inter-pigment couplings. The size of the couplings had minimal effect on the absorption spectrum due to the dominant line broadening resulting from intense fluctuations of site energies. These fluctuations were quantified with spectral densities, providing a reliable estimate of the critical parameter of reorganization energy. After downscaling the raw spectral densities, the experimental spectra were reproduced qualitatively. The corresponding reorganization energy was found to be two times lower than the one calculated with DFT/B3LYP.

To simulate exciton dynamics, swarms of 100 non-adiabatic molecular dynamics simulations with a wide range of input parameters were performed at cryogenic and room temperature. Analyzing simulations at 300 K revealed a set of most appropriate input parameters, i.e. a reorganization energy of 35 meV and scaled couplings. Simulations at 77 K with these input parameters refuted the intuitive hypothesis that cold temperatures lead to slower transfer. Instead, the long-lasting delocalization that results from reduced energetic fluctuations enhances excitonic transfer.

This work demonstrates the potential of atomistic simulations of energy transfer in reproducing ensemble experimental findings. The effect of molecular properties, such as site energies, couplings, and reorganization energies, on ensemble properties could be discriminated. These findings contribute to a deeper understanding of energy transfer in photosynthetic pigments and pave the way for elucidating the transfer mechanism in a comparable protein complex in the next chapter.

9. Elucidating Exciton Transport Mechanism in the Light-Harvesting Complex 2

Reprinted in parts with permission from
David S. Hoffmann, Philipp M. Dohmen, Monja Sokolov, Ulrich Kleinekathöfer and Marcus Elstner: Exciton Transfer Simulations in a Light-Harvesting 2 Complex Reveal the Transient Delocalization Mechanism
***J. Phys. Chem. B.* 2025, 129 (13), 3345-3365**

DOI: 10.1021/acs.jpcb.5c00320

Author Contributions:
The results presented in this chapter were generated in collaboration. The LH2 system was provided by the group of Prof. Ulrich Kleinekathöfer. Philipp Dohmen trained the neural networks and created example trajectories. The method was first applied to the FMO system, as described in the previous chapter. Building on this experience, I systematically carried out MD and NAMD simulations and subsequently performed all analyses.

9.1. Introduction

In purple bacteria, the photosynthetic process is anoxygenic, starting with the absorption of light by light-harvesting 2 (LH2) complexes. Subsequently, the excitation energy is transferred to reaction centers located within LH1 complexes. The structure of the light-harvesting complexes in purple bacteria has been the subject of extensive study over many years[115,50]. For example, the structure of the LH2 complex of *Rhodoblastus acidophilus* (*Rbl. acidophilus*), previously known as *Rhodopseudomonas acidophila*, was solved in 1995[184,51]. Other examples include LH2 rings from other organisms such as *Rhodospirillum molischianum*[145], *Marichromatium purpuratum*[88], and *Rhodopseudomonas palustris*[227].

All of these LH2 complexes exhibit a ring-like structure with N-fold symmetry. The N-fold symmetry is a consequence of the repetition of a fundamental heterodimer subunit comprising two transmembrane polypeptides, designated as α and β, arranged in a circular configuration. In nature, different ring sizes have been observed, with rings comprising seven, eight, and nine $\alpha - \beta$ dimers[227]. The rationale behind this symmetric configuration

has also been a subject of investigation[49]. In all these LH2 complexes, each $\alpha - \beta$ subunit contains three Bacteriochlorophyll a (BChl a) molecules non-covalently bound to it, resulting in two rings of pigment molecules, the B850 and B800 rings, which are named after their Q_y absorption band (in nm). Moreover, the N Bchl a molecules in the B800 ring exhibit greater distances between adjacent chromophores than the $2N$ tightly coupled pigments in the B850 ring.

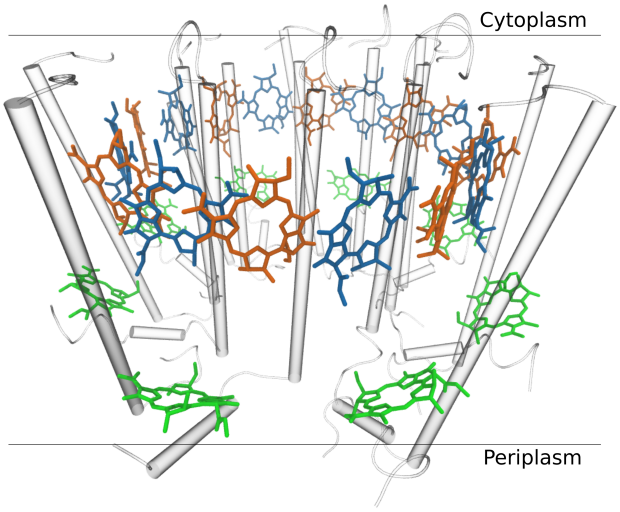


Fig. 9.1.: Side view of the LH2 complex with the cytoplasmic side of the membrane on top and the periplasmic side on the bottom. Polypeptide helices are shown as transparent cylinders, with the eight inner helices denoted as α and the eight outer helices as β. BChl a molecules of the B800 ring, coordinated by β helices, are shown in green. The B850 α pigments are shown in blue, and B850 β pigments are shown in orange.

Experimentally, LH2 complexes have been studied with a variety of techniques, including linear absorption and circular dichroism[93], transient absorption[225], three-pulse photon echo measurements[134], and femtosecond pump-probe experiments[38]. Eventually, these early experiments have been reconciled[277] to study key properties such as couplings, temperature dependence of spectra, and transition rates. Moreover, single-molecule experiments elucidated the intricacies of the dynamics within the LH2 rings[207,168,251]. Even single-photon absorption and emission have been studied for LH2 rings[164]. Recently, polarization-controlled two-dimensional electronic spectroscopy has been conducted at cryogenic temperatures with simultaneous optimal spectral and temporal resolution to obtain information on dynamic and static disorder in the LH2 complexes[138].

As in the previous chapter on the FMO complex, a substantial body of model calculations and molecular simulations has been conducted on LH2 complexes. These investigations encompass a range of approaches, from calculations on Frenkel Hamiltonians with parameters primarily derived from experimental data[38,234,140,100] to molecular modeling of individual complexes.[54,65,208,175] Moreover, studies have been conducted on membrane patches comprising multiple LH1 and LH2 complexes[41,314,295] up to an entire chromatophore with 63 LH2 complexes[265]. In addition, various spectroscopic properties have been calculated early on.[166,30,207,100]

Many of these calculations, particularly those pertaining to spectroscopic properties, were based on Redfield theory with a perturbative approach to the system-bath coupling. Although the theory of Förster resonance energy transfer (FRET) has been successfully employed to describe the slow, incoherent transfer within the B800 ring, it significantly underestimates the energy transfer rate of the B800-B850 transition[277]. Given that the spectral overlap of the 0-0 transitions is insufficient to account for the rapid transfer observed in the B800-B850 transition, it has been proposed that the higher-lying states of the 850 band, the so-called 850* states, may play a role in this transition[255,203,154]. The multichromophoric Förster resonance energy transfer (MC-FRET) method was developed with the objective of extending the standard Förster method to include coherence within complexes[126,315].

While the incoherent hopping mechanism is typified by excitons localized at a single site[130,224], which can also be described by the Marcus rate equation[274], a second transport mechanism that cannot be described with Förster theory has been discussed already for a long time. This mechanism is characterized by the delocalization of excitations over several sites, as indicated by an exciton band structure in the absorption spectrum. In this model, the exciton dynamics is described by a phonon-induced transition between exciton levels, which differs from the incoherent hopping model. The two models can be viewed as two qualitatively distinct limiting cases[277], which can be elucidated by the use of complementary methods. In the hopping picture, Förster rates are used to propagate the exciton population using a master equation approach, while the delocalized exciton is described by a Frenkel Hamiltonian coupled to a thermal environment.

The large couplings in the B850 ring lead to the observation of significant Davydov splittings, indicative of molecular excitonic states. This raises the question if quantum effects impose a distinct and more efficient transport mechanism compared to a random walk based on Förster rates. Electronic coherence in the sense of a superposition of eigenstates has been ruled out to be the reason for the exciton transfer efficiency[35,79].

To this end, comparisons between Förster and Redfield theories have been made, focusing on different aspects[122,206,261,2,204,241]. Among the high-level methods, the Hierarchy Equation of Motion (HEOM) approach[278,276] and its stochastic variant, the Hierarchy of Pure States Equations (HOPS), are examples of theories capable of covering both limiting cases and have been applied to the dynamics in LH2 rings within a Frenkel Hamiltonian picture[295]. Noteworthy is also the trajectory-based Numerical Integration of the Schrödinger Equation (NISE) approach[128,129,110], alternatively known as Ehrenfest scheme without back-reaction[5], which has recently been applied to chlorosomes with thousands of pigments[78,77]. In addition, this approach has been applied to an LH2 complex and at the same time compared to the surface hopping approach[296]. In some instances, the dynamics between LH2 complexes has also been the subject of investigation[295,314]. Most of the theoretical studies mentioned are aimed at relating molecular properties such as couplings and excitation energies to observables like the absorption spectrum or transfer rates. However, the detailed mechanism of the exciton transport remains unclear.

The transfer of excitons has traditionally been considered to occur within two distinct regimes, distinguished by the relative values of excitonic couplings and reorganization

energies. In these regimes, the excitonic couplings V, are either much stronger or much weaker than the reorganization energy λ, i.e., $V >> \lambda$ and $V << \lambda$. In the first coherent band-like transport regime, an electronic structure model is proposed that is analogous to that one observed in certain metals or organic semiconductor (OSC) crystals. In contrast, the latter regime is associated with an incoherent hopping mechanism, which is analogous to that one seen in many biomolecules involving electron transport and certain solid state materials. However, it has been acknowledged that the case of LH systems is particularly complex and exhibits an intermediate situation, with $V \approx \lambda$, where couplings and reorganization energy have comparable magnitudes.

As previously noted by Scholes and coworkers, the large coupling between neighboring molecules in LH systems can also lead to an efficient exciton transport regime, which has been reported for several organic semiconductor architectures[254]. Interestingly, the intermediate transport regime has been discussed in the field of OSCs, and an effective transport theory called “Transient Localization Theory” has been developed. It is capable of describing and understanding efficient charge transport in highly mobile molecular semiconductors[85,48]. Although initially proposed as a transport model for charge transfer, it has been demonstrated to also apply to exciton transfer in OSCs. Explicit atomistic simulations of charge[105,307,238,68,97,95] and exciton transfer[96,268,94] have confirmed that transfer efficiency (mobility, diffusion) is strongly correlated with transient delocalization. These multiscale non-adiabatic approaches have recently been applied to the Fenna-Matthews-Olson complex (FMO), indicating an exciton transfer regime comparable to that observed in OSCs. While the energy landscape in FMO is not iso-energetic, the energy differences are sufficiently small compared to k_BT to allow for such a transport mechanism, because the λ/V ratio is roughly two in this system[272].

The non-adiabatic molecular dynamics (NAMD) simulations extend beyond the Born-Oppenheimer adiabatic picture by simultaneously propagating the coupled electronic and nuclear degrees of freedom using surface hopping (SH) techniques. Exciton dynamics are driven by a time-dependent Frenkel Hamiltonian that does not rely on any averaging or a timescale separation within the system. The unbiased Frenkel Hamiltonian is significant because the Förster theory’s fundamental assumption may be violated if the dynamics of the nuclei and electronic degrees of freedom occur on the same timescale. Pullerits and Sundström have already proposed this idea for the LH2 system[226], arguing that the actual transport mechanism may be a combination of the aforementioned limiting cases.

Based on the knowledge of the FMO system presented in the previous chapter, this work simulates exciton transfer in the LH2 system. Site energies and couplings were used to calculate an absorption spectrum that matched the experimental spectrum. The adiabatic states were analyzed in terms of populations, delocalization and fluctuations. Swarms of NAMD simulations were performed at room temperature and compared with experimental observations. Analyzing individual NAMD simulations revealed the "Transient Localization Mechanism" in the LH2 system. Finally, the obtained values are compared against those of different OSCs.

9.2. Computational Details

9.2.1. System Setup and MD Simulations

The studied LH2 complex from *Rs. molischianum* (PDB: 1LGH) was obtained from Ref. [25]. It contains 24 BChl a molecules embedded in a membrane protein complex consisting of 16 helical polypeptides. For the protein and membrane molecules, the CHARMM27 force field was used, water was modeled with the TIP3P model and the BChl a parameters were taken from Ref. [65]. Using Gromacs 2021.5[1,211,19], a 200 ns long MD simulation was performed on the equilibrated structure from Ref. [25]. The temperature of 300 K was controlled by the Nosé-Hoover thermostat[201,111] with a time step of 1 fs and the Parrinello-Rahman barostat[216,202] was used to control the pressure of 1.013 bar. Long range electrostatic calculations were performed using the Particle Mesh Ewald method.

9.2.2. Excited States Properties

Site Energies and Couplings: BChl a molecules were treated as described in section 8.2.2: The phytyl tail has been excluded from the QM region, excitation energies and transition charges were calculated using the semi-empirical TD-LC-DFTB2, and the pairwise couplings were determined as Coulomb interactions between the transition charges. Excitonic coupling values have been scaled so that the transition dipole moments match the experimental value of $\approx 6.1D$[142,3]. No additional screening effect as suggested by Scholes et al.[253] was taken into account, as described in section 8.3.4.

Reorganization energy: The critical parameter λ can be obtained from relaxations from the Franck-Condon region to the minimum geometry (see also chapter 4). B3LYP based reorganization energies in gas phase have been calculated before.[272] The reorganization energy can also be obtained from the integral of the spectral density or from the width of the histograms of the excitation energies. In the harmonic approximation, all three formulations of reorganization energies are equivalent (section 4.4).[21].

Here, λ was estimated based on the histogram of site energies calculated along the ground-state MD simulation. This was done by dividing the variance of the absorption energies by a factor of $2k_BT$ at 300 K. To obtain the absolute reorganization energy for absorption and emission, this value was doubled. One geometry per 40 ps was used to calculate the site energies, resulting in 1 000 single point calculations per BChl a molecule. To estimate the internal (vacuum) relaxation λ_{vac}, the site energies were calculated without electrostatic embedding, i.e., using only the geometry of the respective pigment without considering the external MM charges. The total reorganization energy λ_{tot} was calculated including electrostatic embedding, which includes the ESP of the surrounding atoms in the TD-LC-DFTB2 calculations. In addition, the outer-sphere reorganization energy $\lambda_{outer-sphere}$ was estimated from $\sigma(E^{tot} - E^{vac})^2$, i.e., the variance of the difference of the site energies with and without electrostatic embedding as done in Ref. [61].

Absorption spectra: Spectral densities from a 60 ps QM/MM simulation and absorption spectra with the cumulant expansion method were generated as described in the previous chapter.

9.2.3. NAMD Simulations

The partial charges of the BChl a molecules are updated with the difference between ground and excited state charges, weighted with the excitonic occupation of the respective BChl a pigment (section 8.2.3).

NAMD simulations were performed with the in-house charge and exciton transfer code as described in section 8.2.3. From the 200 ns MD trajectory, 200 equally spaced snapshots from the first 40 ns were used as starting geometries for the NAMD simulations. In all NAMD simulations, the total number of atoms in the MM zone was identical and only the pigments of interest were additionally treated as the QM zone. Therefore, a NAMD simulation of the B850 ring alone, as discussed in the Results section, is defined as a simulation in which both rings are present in the MD simulation, but only the B850 molecules form the diabatic basis of the Frenkel Hamiltonian. Therefore, the exciton is forced to stay on the B850 ring, and only the partial charges of the B850 molecules are updated during these simulations.

The simulation time was set to approximately ten times the experimentally determined transfer rate of interest, resulting in 5 ps, 1 ps and 10 ps, for the B800 ring alone, the B850 ring alone, and the entire LH2 complex, respectively.

NAMD simulations were initialized with 100 % occupation on one of the pigments (B800-1 or B850-1). An adiabatic state, as one eigenfunction of the Frenkel Hamiltonian at the respective moment in time, was not considered as initial condition. Such a “true” adiabatic state, e.g., excited by a laser pulse in a spectroscopic experiment, is difficult to be defined. First, the energetic order of the adiabatic states and the diabatic occupation within an adiabatic state changes within tens of femtoseconds. Therefore, even with a femtosecond laser pulse, it is not clear which adiabatic state is excited. Secondly, the uncertainty principle applied to femtosecond laser pulses leads to a broadening in the range of a few up to tens of meV. As demonstrated in Figure 9.4, the adiabatic states differ by only a few meV. Therefore, it is unclear whether a single adiabatic state can be excited at room temperature.

Delocalization measurements: There are numerous ways to quantify the delocalization length, including the occupations of the pigments or the autocorrelation function of the wave function[38]. Here, the inverse participation ratio (IPR) and an alternative quantity suggested by Jang, Dempster and Silbey 2001[125] (JDS) was used. These quantities are calculated as follows:

$$\mathrm{IPR} = \frac{1}{\sum_n |b_{in}|^4}$$

and

$$\mathrm{JDS} = \sum_n \min\{1, 2N_c|b_{in}|^2\}|$$

with b_{in} being the n-th diabatic coefficients of the i-th adiabatic state and N_c being the total number of pigment molecules. The number of chromophores N_c has values of 8, 16 and 24 for B800, B850 and LH2, respectively.

9.2.4. Machine Learning

An approach similar to that described in the previous section was used. For the NN for Coulomb couplings, no environmental effects in form of ESPs have been considered, because the transition charges are hardly effected by the environment (FigureD.7). The resulting NN models have R^2 values of 0.987 and 0.999 and mean absolute errors of 4.1 meV and 0.3 meV for the site energies and couplings, respectively.

9.3. Results and Discussion

9.3.1. Frenkel Elements from MD

9.3.1.1. Site Energies

The LH2 system from the bacterium *Rsp. molischianum* has an 8-fold symmetry with one B800 and two B850 (α, β) molecules per monomer. All BChl a pigments are chemically identical in their relevant parts, so the only difference is their specific conformation and electrostatic environment within the protein. The latter factor influences the tuning of the site energies. In a first step, we estimate the spectra in the uncoupled case, i.e., we compute the excitation energy distribution for the individual chromophores. Given that the system is 8-fold symmetric with three pigments in each monomer, we expect three distinct distributions when averaging over the monomers. Figure 9.2a presents the averaged site energies of B800, B850α and B850β pigments throughout the 40 ns long MD simulation. The differences in site energies illustrate the influence of the varying electrostatic environments that characterize these sites. Figure D.4b demonstrates that the site energies of the B800 pigments are not entirely identical even after 40 ns. This indicates that the Frenkel Hamiltonian of the B800 ring cannot be considered to consist of equal site energies at all relevant time scales. The distributions for the B850 pigments (Figure D.4c,d) show only minimal differences, fulfilling the expectations for a symmetrical system.

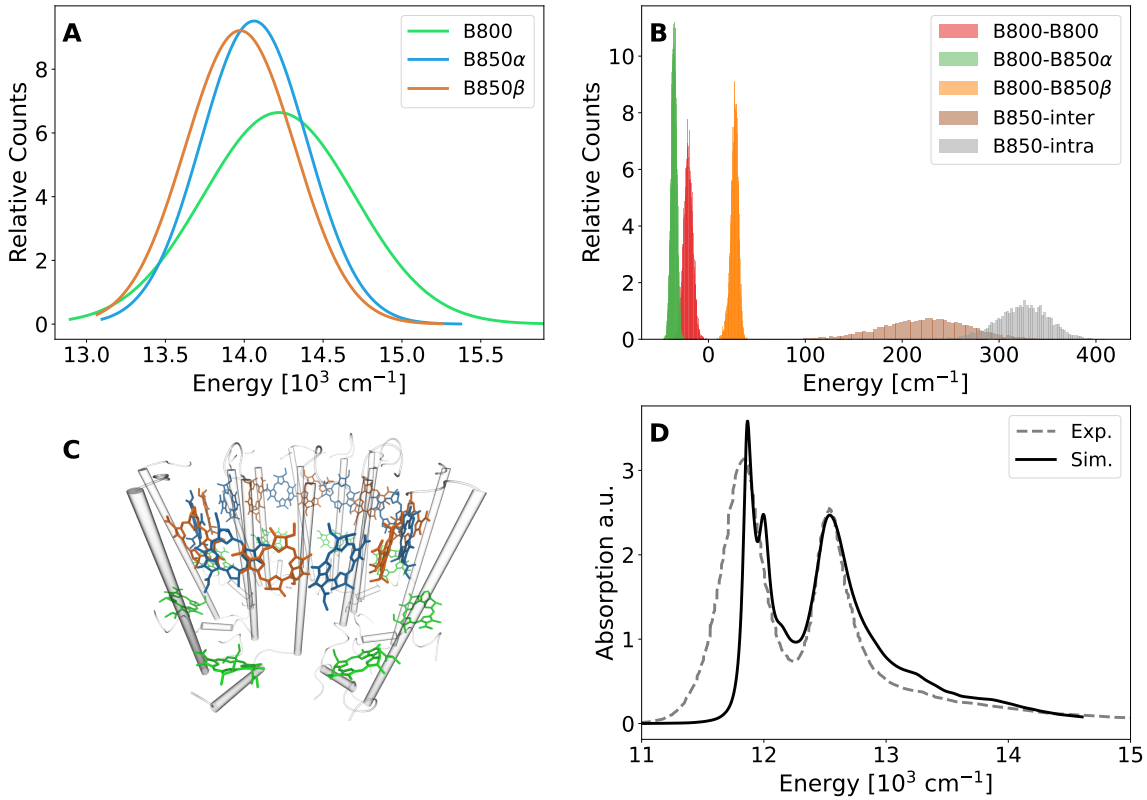


Fig. 9.2.: Distributions of site energies (a) and couplings (b) along a 40 ns MD simulation at 300 K. The curves show accumulated values of eight individual pigments and pigment pairs. The LH2 complex (c) presents three pigment types B800, B850α and B850β in green, blue and orange, respectively. The calculated spectrum (d) obtained from the time averaged hamiltonian is compared to the measured spectrum[313].

The B800 molecules are blue shifted in comparison to B850α and B850β by 158 cm^{-1} and 246 cm^{-1}, respectively. This aligns well with previous work[93,296,59,228] as detailed in the appendix (section D.2). It is commonly accepted that this blue shift can be attributed to the B850 molecules being embedded in a hydrophobic protein scaffold, whereas the B800 pigments have a hydrophilic and partially aqueous environment[145] (see also Figure D.5).

The differences between B850α and B850β are minor, which are likely overruled by the dynamic disorder, as demonstrated in Figure D.6 for 500 fs windows. At longer time scales, this disorder may be termed (quasi-)static disorder, which manifests as inhomogeneous line broadening. However, the term “static” lacks a clear definition, and it is therefore necessary to refer to a specific process[188].

Extending the trajectories to 200 ns, we found a significant amount of static disorder introduced by single water molecules entering the binding sites of some B850 pigments (see Figure D.9). When the electrostatic environment is disregarded, nearly identical distributions are obtained for all pigments. Accordingly, on time scales much longer than the exciton transfer, structural changes of the environment may occur that could influence the dynamics and spectroscopic properties of LH2 ensembles in aqueous solvation. An analogous finding has recently been reported for the CP24 complex of plants[246]. This

effect merits further investigation, as the stability of water molecules in hydrophobic environments, parameterized for the bulk phase, is a delicate matter. TIP3P water is undoubtedly overpolarized, and this may lead to artificial binding or an overestimation of stability due to the absence of polarization effects, as has recently been detailed.[213,160]

9.3.1.2. Reorganization Energies

The static disorder effect of water molecules in the B850 ring is not present for the first 40 ns of the MD simulation, i.e., in this time scale we didn't find a significant variation of average site energies. This allows the analysis of the dynamic disorder (homogeneous line broadening) expressed in the fluctuations of the site energies. The dynamic disorder plays a central role, because it is related to the reorganization energy λ, as described in section 9.2.1, via the variance of the site energy fluctuations divided by $k_B T$[21]. While the harmonic approximation is not applicable to pigment-protein complexes[157] and delocalized systems[181], a qualitative understanding of the effects can be gained by comparing the site energy fluctuations of individual chromophores. To isolate the dynamic disorder, we calculated the reorganization energy of all BChl a pigments individually and averaged the eight values for each pigment type, i.e., B800, B850α, and B850β. The results are shown in Table 9.1.

Tab. 9.1.: Excitation reorganization energy λ in cm^{-1} (in meV in the brackets) obtained from the 40 ns MD simulation at 300 K.

	B800	B850α	B850β
λ_{vac}	516 (64.0)	500 (62.4)	524 (65.0)
λ_{tot}	1074 (133.2)	580 (72.0)	546 (67.6)
$\lambda_{outer-sphere}$	378 (46.8)	58 (7.4)	46 (5.8)

The fluctuations of the geometries without considering the electrostatic effect, represented by the parameter λ_{vac}, are comparable among all pigments and lie around 510 cm^{-1} (63 meV). Energy fluctuations solely based on the geometries reproduce the gas phase reorganization energy, also referred to as the inner-sphere reorganization energy.

Although based on a force-field parametrization, the value of 510 cm^{-1} falls within the window of reasonable estimates for the inner-sphere reorganization energy. In our previous work on FMO[272], we used the B3LYP calculated value of 280 cm^{-1} (35 meV) in gas phase as the lower limit and the experimentally determined value of 520 cm^{-1} (65 meV)[229] in triethylamine (TEA) as the upper limit. Please note, that the NAMD simulations do not depend on the force field estimate, but make use of λ_{vac} as a parameter computed from B3LYP within the implicit relaxation method as detailed above. The total (inner-sphere and outer-sphere) reorganization energy λ_{tot}, which includes both geometry fluctuations and altered electrostatic environments, demonstrates a different pattern. The B800 pigments have a value of 1074 cm^{-1} (133 meV), while the B850 molecules have a total reorganization energy of approximately 564 cm^{-1} (70 meV). Therefore, outer-sphere reorganization energy,

which is captured in our simulations by the QM/MM term, is a sizable factor for the B800 chromophores, while it vanishes for the B850 due to shielding from solvent effects. This difference results from a different solvent exposure, as analyzed in more detail in section D.2.

9.3.1.3. Couplings

The second key component of the Frenkel Hamiltonian is the coupling between pigments. TD-LC-DFTB2 computed couplings have been shown to compare very well with those of higher-level quantum chemical methods for a variety of molecular interactions[25,250]. Previous work by our group[272,25] has demonstrated that TD-LC-DFTB2 is an accurate method for calculating couplings in light-harvesting complexes. A direct comparison with other methods found in the literature also demonstrates excellent agreement, as detailed in Table D.4.

We consider the two dominant couplings of B850 molecules (intra- and intermonomeric), together with the three strongest couplings for the B800 pigments (B800-B800, B800-B850α, B800-B850β) and disregard couplings between next neighbors. As illustrated in Figure 9.2b, the three distributions of couplings involving B800 molecules are sharper than those of the two B850 couplings. This is due to the fact that the B850 neighbors have only an Mg-Mg distance of 8 to 10 Å, whereas the B800 pigments have their next coupling partner within 19 to 22 Å[188].

Adding the B3LYP calculated inner-sphere reorganization energy and the MD based estimation of the outer-sphere contribution, the absolute reorganization energy for B850 chromophores is about 330 cm^{-1} (41 meV). From the distribution of couplings, it is evident that the B850 molecules fall into the $V \approx \lambda$ regime as discussed in the Introduction.

9.3.2. Eigenstates, Active Environments and Fluctuations

Combining the mean values of site energies and couplings together with the time averaged transition dipoles of all pigments and a calculated spectral density for one BChl a molecule, the experimental absorption spectrum in Figure 9.2d is reproduced with the full cumulant expansion method[58,170]. The split between the B800 and B850 peaks might be underestimated due to the neglect of charge transfer states in the B850 ring.[60] Because the calculated elements of the Frenkel Hamiltonian agree well with other computaional studies and the experimental spectrum is reproduced quite well, the position and width of the adiabatic states resulting from the diagonalization of the Frenkel Hamiltonian is adressed in this chapter. To gain insight into the non-trivial distribution and time dependence of the adiabatic states, it is necessary to perform some averaging. Two approaches may be considered for this: (i) averaging the Frenkel Hamiltonian along the MD trajectory and then diagonalizing the averaged Hamiltonian, or (ii) diagonalizing at each MD step and then averaging the resulting adiabatic states.

For the first approach, we created the averaged and symmetrized Hamiltonian by placing the three site energies, averaged per pigment type, i.e., B800, B850α and B850β, at the respective diagonal matrix positions. The five different couplings were assigned in the same manner to the off-diagonal entries, while all other couplings were disregarded. The second approach was not applied, due to the fact that averaging the n-th adiabatic state with respect to energy leads to non-physical results, since the states change their order along the trajectory. An adiabatic state with the lowest energy at t=0 is unlikely to have similar properties than the lowest adiabatic state after a few femto- or picoseconds. An averaging just following an energetic time-dependent ordering of these states leads to a very broad splitting, as demonstrated in Figure D.10. Thus, in the following, the Hamiltonian was diagonalized at every MD step resulting in time-dependent eigenstates and -energies. The results are depicted in Figure 9.3 for the two rings and the entire LH2 system. The averaged and symmetrized Hamiltonian leads to a narrow range of B800 states, situated at an energetic midpoint between the B850 states. All resulting states are highly analogous to those observed in a previous study, where the Hamiltonian was determined for a single symmetrized crystal structure[286].

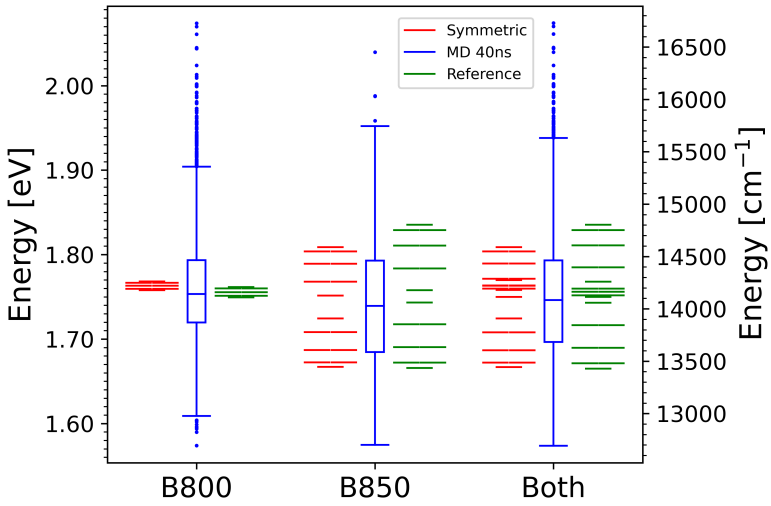


Fig. 9.3.: A total of 1 000 Frenkel Hamiltonians, spaced by 40 ps, were generated and diagonalized over the course of the 40 ns MD simulation. This yields 8 000, 16 000 and 24 000 adiabatic energies for B800, B850 and both rings, respectively. They are shown as box plots (MD 40 ns, blue), with the box extending from the first to the third quartile and the median value being represented as a horizontal bar. The whiskers extend from the box edges to 1.5 times the inter-quartile range. Outliers are indicated as blue dots. Additionally, a symmetric Hamiltonian (symmetric, red) was constructed from time-averaged values for site energies and couplings from Figure 9.2. The eigenvalues are compared to those obtained from a geometry-optimized crystal structure by Tretiak et al.[286], which was used by Kundu et al.[156,155] (Reference, green) with the site energies shifted by 4 400 cm^{-1} to match the mean values obtained from TD-LC-DFTB2.

To visualize the 24 data points of the LH2 complex (both rings) obtained per snapshot, we present the eigenvalues together with the squared eigenvectors (occupations) of one snapshot in Figure 9.4B, alongside the results obtained from the symmetrized Hamiltonian (Figure 9.4A).

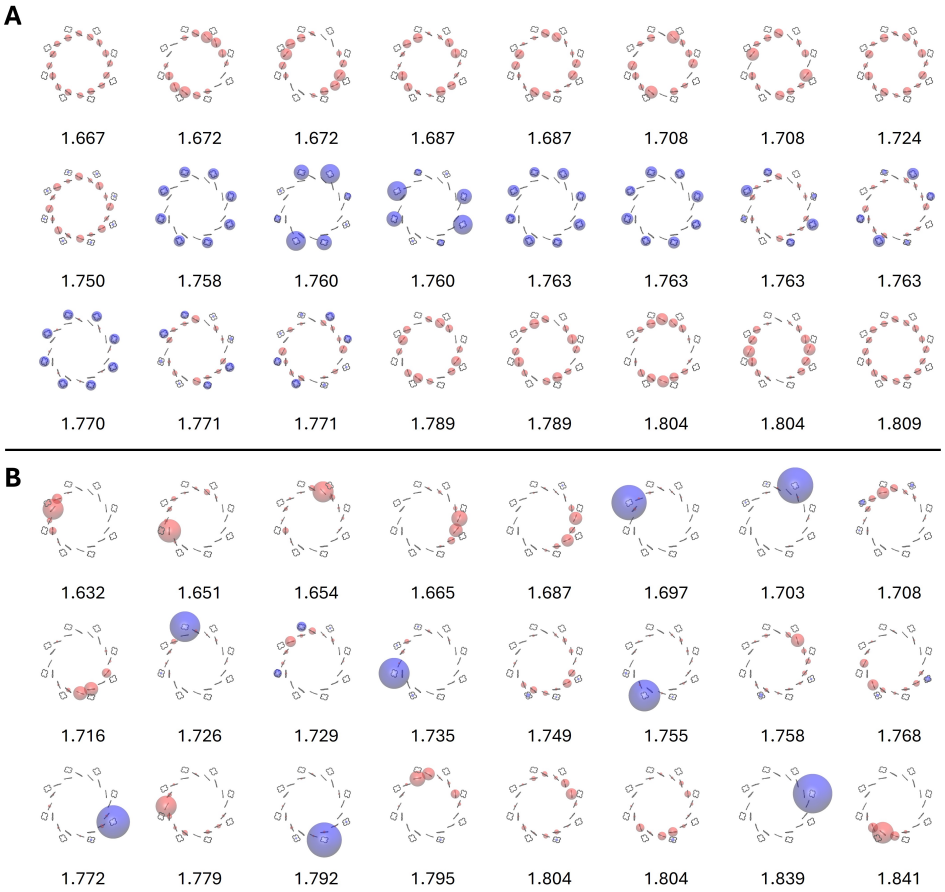


Fig. 9.4.: Occupations of the eigenstates of the LH2 complex are shown with their excitonic energy in eV. The symmetric Hamiltonian (A) is compared against the Hamiltonian from the first snapshot of the MD simulation (B). The porphyrin ring of all BChl a chromophores are shown as black lines. The magnesium atoms are represented with a radius that is proportional to the squared eigenvectors. Red magnesium atoms are from the B850 ring and blue ones are from the B800 ring.

The states from the symmetrized Hamiltonian exhibit complete delocalization over the ring system, aligning closely with the findings of Kundu et al.[155]. This is a highly idealized and artificial representation that contrasts with the expectation of complete localization in the B800 ring, where excitons are transferred by a Förster-type hopping mechanism. Delocalization to opposite pigments is very unlikely, given the negligible couplings between them. Regarding the B850 ring, while the couplings are significantly larger, it is also expected that some degree of localization will occur at a limited number of sites due to dynamic disorder. Section 9.3.4.1 contains a detailed analysis about delocalization. It is evident that disorder causes a break of symmetry, both in terms of the energy distribution and the eigenvectors[283,267].

This situation is illustrated using the first snapshot of the MD simulation in Figure 9.4B. The same visualization for the second snapshot is shown in the appendix (Figure D.11), demonstrating that the characteristics of each adiabatic state change significantly between one snapshot and the next.

The difference is unsurprising given the separation of snapshots by 60 ps. Consequently, an analysis was performed using a simulation with a time step of 0.1 fs. Figure 9.5 demonstrates the impact of fluctuations on the lowest adiabatic state, that was computed along the geometries of a NAMD simulation.

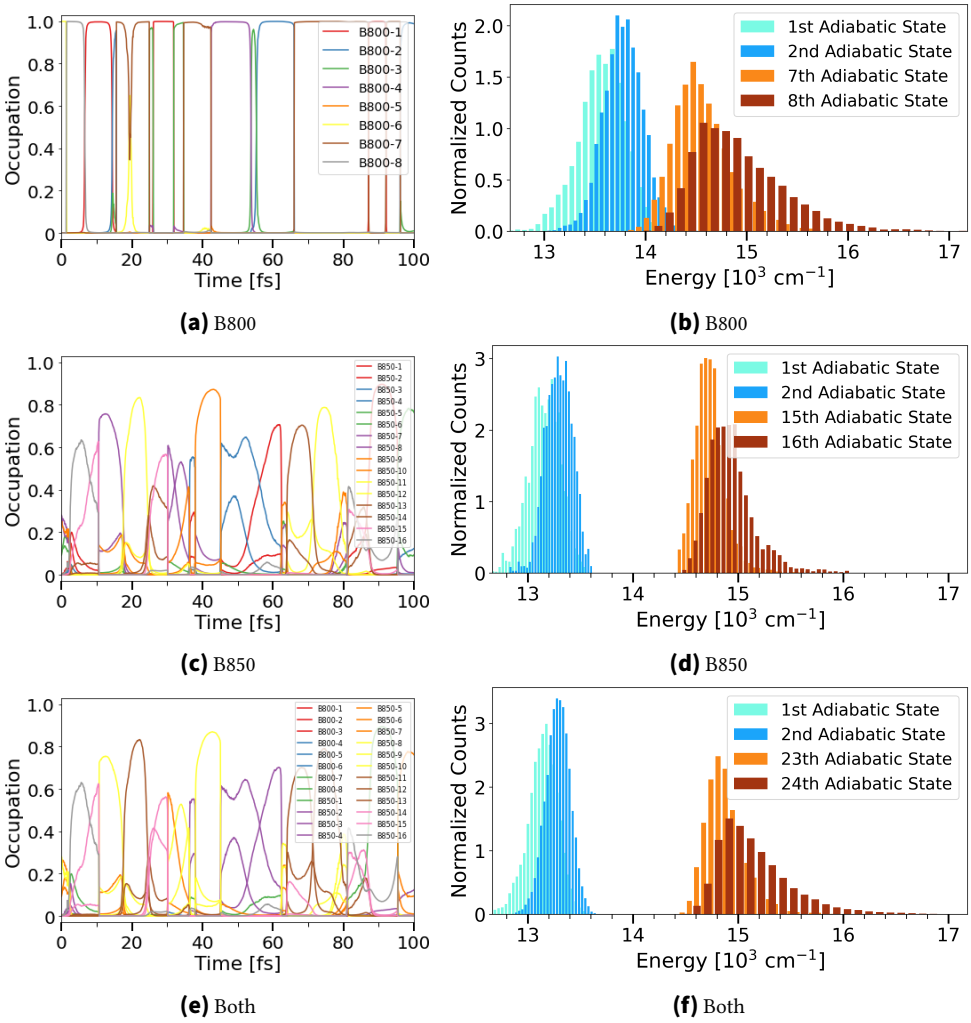


Fig. 9.5.: The diabatic occupations of the pigments of the lowest eigenstate during the first 100 fs of one NAMD simulation (a,c,e) are shown together with the distributions of eigenvalues of the first two and last two adiabatic states, i.e. the adiabatic surfaces with the n-th highest energy (b,d,f). The histograms were created throughout the full simulation time of 5 ps, 1 ps, 10 ps, for B800, B850 and LH2, respectively. All NAMD simulations started from the initial snapshot of the 40 ns long MD trajectory. In Figure D.12, the same analysis was performed for a NAMD simulation started from the 100th snapshot.

The occupations of the lowest adiabatic surface is plotted for the initial 100 fs, together with a histogram of the eigenvalues of the first two and last two adiabatic states throughout the whole simulation time. Please note that the histograms represent the n-th adiabatic state with respect to energy and do not track a specific state, such as one dominated by a particular pigment. Therefore, the mean value of each histogram represents one mean value when performing the wrong averaging process mentioned above (Figure D.10).

Figure 9.5 leads to three conclusions. At first, it is evident that the nature of the lowest adiabatic state changes rapidly within ≈20 fs. This is consistent with the fluctuations in site energies and couplings (Figure D.3), which occur on this same time scale, and can be attributed to the aromatic carbon-carbon vibrations. These small changes in the site energies result in unavoided crossings, which occur when two spatially separated and therefore non-interacting states intersect. The large 'jumps' in the wave function demonstrates a breakdown of the Born-Oppenheimer approximation and the Aufbau principle. While this is not a problem in principle, it may require an explicit time-dependent treatment of electronic and nuclear degrees of freedom, one has 'to overcome the static image' as Beratan and coworkers put it[18]. Especially when considering that the interactions with the environment will be different for a localized vs. a delocalized state. In view of this issue, NAMD simulations employ a correction scheme to address these trivial crossings through a state tracking algorithm[307,298]. The fluctuations are much faster than any other relevant process in this system. Consequently, both averaging processes fail to produce a meaningful result. A moving exciton does not experience an instantaneous energy spectrum from one MD snapshot and it is not possible to make statements about the adiabatic surface of the exciton without tracking it over many frames. On the other hand, an average derived from multiple nanoseconds of MD simulation is not a relevant representation, as well. Similar problems have been discussed in different contexts: One case is the tunneling-problem of a particle through a fluctuating barrier. Büttiger and Landauer discussed two limiting cases, slow vs. fast tunneling with respect to the timescales of the bridge fluctuations, which leads to a static barrier on the one hand, and a time-averaged barrier on the other[32].

Secondly, Figure 9.5 illustrates the localization of the lowest adiabatic state in the B800 ring with occupations approaching one and the presence of delocalized states in the B850 ring, which validates the trend observed in one snapshot in Figure9.4(B) over a period of 100 fs. Thirdly, the highest (24th) eigenstate in LH2 has a pronounced shoulder towards the blue light (f), which is absent in the highest (16th) eigenstate of the B850 ring (d). This is due to the fact that the highest adiabatic state in LH2 can be occupied either by B800 or by B850 molecules (Figure 9.4(B) and Figure D.11). This is also reflected in the greater number of high-energy outliers for the adiabatic states in the LH2 complex (Figure 9.3, Both) compared to the B850 ring alone (Figure 9.3, B850).

This preliminary analysis indicates that it is not possible to consider energetically separated, delocalized eigenstates where the dynamics of the excitons are driven by the coupling to the vibrational states of the system. These eigenstates are necessarily localized and change within 10-20 fs. This is a very short timescale compared to the pulse duration with which these systems are probed in experiments, or compared to the motion of the exciton. Therefore, neither the time-averaged Hamiltonian nor the averages of eigenstates along a simulation seem to be suitable basis for the dynamical treatment of the exciton transfer. This is reminiscent of problems discussed in the context of the application of Marcus theory to charge transfer in organic and biological materials: Marcus theory requires a clear separation of time scales of the dynamics of the charge carrier and the environmental response, and there are clear examples in the literature where this separation is not valid: e.g. Troisi has shown, that Marcus theory breaks down for charge transport in

high-mobility organic semiconductors[287], and Matyushov[180] has shown, that the protein response on longer time scales can interfere with charge transfer time scales, requiring a modified approach of the determination of reorganization energies.

The issue of time scale separation and the influence of exciton dynamics on the nuclear system, the back reaction, makes the use of non-adiabatic MD schemes, as applied by Martinez and coworkers in 2017 for this system[266], an optimal methodological choice. It should be noted that these methods are also based on approximations, such as the use of classical nuclei. In our application, the fragmentation of the electronic structure due to the application of a Frenkel Hamiltonian is an additional approximation. Furthermore, surface hopping techniques are not the most sophisticated methods for describing the dynamics of the electronic system. However, these methods appear to effectively address the core aspects of these intricate systems, as evidenced by the numerous applications in biological and organic systems, as detailed in our previous work[272]. Furthermore, in instances where a comparison to experimental data is available, such surface hopping based approaches have been shown to very closely resemble the experimental findings[238,96,94].

9.3.3. Exciton Transfer Rates

As outlined in section 9.2.3, we performed NAMD simulations for the B800 and B850 rings individually, as well as for the entire system, simulating a swarm of 200 trajectories starting from different initial conditions. Initially, the exciton was localized to one of the chromophores in B800 or B850, and the subsequent dynamic evolution was then observed. It should be noted that the implicit relaxation approximation used in this study has been demonstrated to overestimate the diffusion constants by a factor of two[94]. Consequently, the transition rates may be slightly overestimated.

Figure 9.6 shows the diabatic occupations for the active surface, with the data averaged over the swarm of 200 NAMD trajectories. In modelling anisotropy decay experiments, the decay of the initially excited pigment was fitted with an exponential function. In order to simulate 800 nm pump, 850 nm probe experiments, the increase in B850 occupations subsequent to an initial excitation of a B800 molecule was fitted.

Intraband B800 Dynamics: As a first step, the B800 ring was included solely in the QM zone of the simulations. Figure 9.6a shows the decay of the initially populated site and the subsequent increase in occupations of neighboring sites. It is clear that 5 ps is insufficient for achieving complete exciton equilibrium, as evidenced by the observed occupations after 5 ps. In an eight-fold symmetric system, one would anticipate that all chromophores exhibit an occupation of 12.5%. It can be observed that the B800-8 molecule (blue) carries the majority of the occupation after 5 ps, whereas B800-3 (green) reaches a plateau of approximately 20% occupation at 1.5 ps, but is poorly occupied after 5 ps. This effect is attributed to the site energies of the pigments which are higher for B800-3 and lower for B800-8 than for the other B800 molecules (Figure D.4b). The couplings between B800-1 and its left and right neighbors (B800-2, B800-8) differ on average by only 2 cm^{-1}.

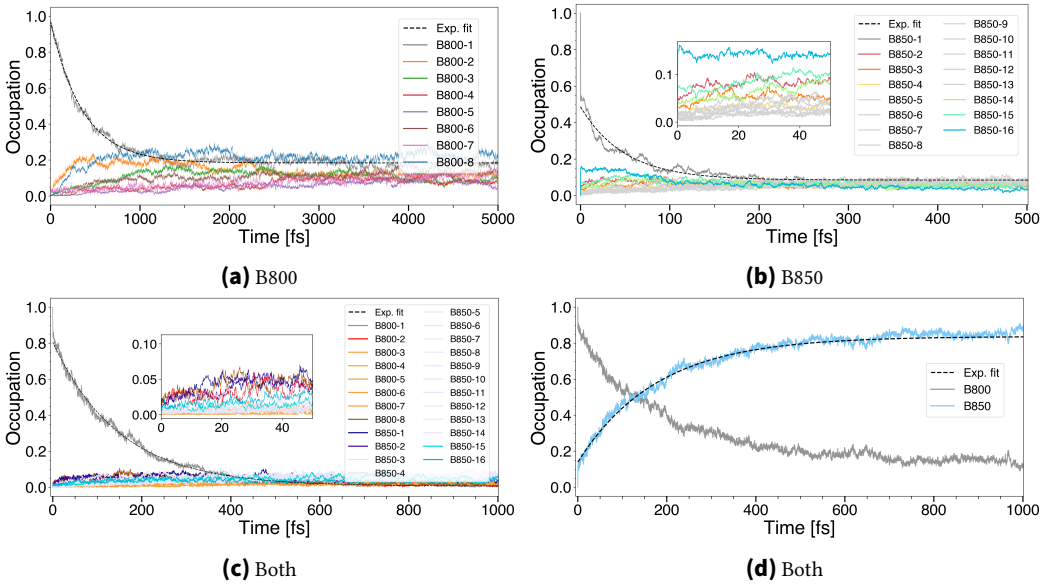


Fig. 9.6.: Averaged diabatic occupation over 200 single NAMD simulations at 300 K with $\lambda = 35meV$. The evolution of occupations are shown for 5 ps, 500 fs, and 1 ps for B800, B850, and both rings, respectively. Exponential fits are shown in dotted lines. The B800→B850 transfer in both rings is illustrated for all pigments (c) and for all pigments within the B800 or B850 ring (d).

As illustrated in Figure 9.6a, the exciton displays equal speed in both directions (B800-2, yellow and B800-8, blue).

The fitted time constant of 359 fs for the B800-B800 transfer is in good agreement with experimental estimates, which range from 0.4 to 1 ps[277,120,285], as well as other theoretical estimates of 410 fs[286], even without correction for the IR approximation. A higher reorganization energy of 65 meV yielded a time constant of 541 fs (see FigureD.14). The Marcus equation $k_{et} = \sqrt{\frac{\pi}{\lambda k_B T}} \frac{V^2}{\hbar} \exp(-\frac{\lambda}{4k_B T})$ with the B800-B800 couplings and λ = 35 meV, yields a time constant of 2.4 ps for hopping between neighboring sites. Given that the exciton of a B800 chromophore can be transferred to two neighboring molecules, we have determined a time constant of 1.2 ps for the deexcitation of a B800 chromophore. This value is slightly larger than the result obtained using the NAMD method, but remains within the range of experimental estimates.

Intraband B850 Dynamics: The initial observation in the B850 ring is the immediate relaxation of the artificially localized exciton. In contrast to the B800 ring, the occupation of the B850-1 molecule declines from 100% to 55% during the first 0.5 fs.

The immediate equilibration of the exciton into an adiabatic state is equivalent to the initial conditions of an adiabatic surface. In order to emulate anisotropy decay, it would be necessary to utilize the adiabatic surface with the highest population of the pigment of interest (B850-1), which is, in any case, the surface in which the system relaxes within the first 0.5 fs. In contrast to the B800 ring, there is no adiabatic surface with one pigment being fully occupied (see Figure 9.4B). As this initial equilibrium is not a form of EET, we have fitted the exponential decay to the time period subsequent to this equilibrium. Figure 9.6b,

inset, illustrates the path of the exciton initially localized on the B850-1 pigment. The molecules have been numbered in a manner that identifies B850-1 and B850-16 as the left and right neighbour, respectively. Therefore, B800-1, B850-1 and B850-16 are considered as one monomer. The intra-monomer couplings in B850 are observed to be 1.5 times larger than the inter-monomer couplings (see Table D.4). It can thus be postulated that the adiabatic surface that is dominated by B850-1 is more likely to be shared with B850-16 (blue) than with B850-2 (purple). During the 500 fs interval, the occupation of B850-16 decays and reaches a minimal value at approximately 480 fs. Although the slightly blue-shifted site energy of B850-16 (Figure D.4c) appears to be insignificant, its impact can already be observed in the form of low occupations after 400 fs.

In experimental studies, the B850-B850 transitions have been observed to occur at the fastest rates around 100 fs[188,138]. The results of the exponential fit from our simulations indicate a time constant of 49 fs and 67 fs for the reorganization energies of 35 meV and 65 meV, respectively. This rate will be examined in greater detail below through a comparison with analogous excitation transfer rates observed in organic semiconductors. Espinosa and Kananenka[239] employed the HEOM approach to simulate exciton transfer in synthetic porphyrin nanorings, identifying faster transition rates of approximately 10 fs between neighboring chromophores. This finding is consistent with their excitonic couplings of 570 cm^{-1}, which are considerably larger than those in the B850 ring.

B800-B850 Transitions: In order to examine this transition, the simulations were initiated with a single B800 site populated and both rings were included in the QM zone. As evidenced by the simulation of the B850 ring (Figure 9.6b), an initial equilibration of the exciton is also observed to occur within the first 0.5 fs (Figure 9.6c). The coupling between the initially occupied B800-1 molecule and the B850-1 and B850-16 molecules is greater than that observed within the B800 ring, but less than that within the B850 ring (Table D.4). Consequently, the exciton initially equilibrates into an adiabatic state that is occupied by the B800-1 pigment to 80%. This falls between the 95% and 55% observed in simulations of solely the B800 and B850 rings, respectively.

In order to represent the results of pump-probe experiments, the site occupations were summed up within the B800 and B850 rings. Subsequently, the rise of occupations in the B850 ring was fitted (Figure 9.6d). The resulting time constant of 177 fs is significantly faster than the experimental values, which range between 500 fs and 1.5 ps[277,120,138]. It should be noted, however, that (i) as previously stated, the IR approximation overestimates the transition rates, (ii) the energetic split between the B800 and B850 band is underestimated and (iii) the arrival of the occupation is measured in all B850 states, whereas pump-probe experiments monitoring the rise of the B850 occupation can only access those states with the dominant oscillator strengths. These are typically low-lying B850 states. For exciton wave functions with complete symmetry, these are exclusively the states with $k = \pm 1$ (Table D.5). Disorder and the inclusion of charge transfer states have been demonstrated to result in a slight broadening of the distribution of oscillator strengths. However, the head-to-tail arrangement of the B850 pigments suggests lower oscillator strengths in energetically higher states. Consequently, transitions from B800 to the higher-lying dark B850 states are not directly detectable in emission or pump-probe spectra[285].

The time constant of 177 fs is faster than the value of 359 fs, which was obtained for the intra B800 transfer. Interestingly, the couplings for B800-B800 and B800-B850 are of comparable magnitude. However, the outer-sphere reorganization energy for the B850 chromophores is less than that of the B800 sites, and the transition is slightly downhill. Given the exciton transfer parameters, it is reasonable to expect a slightly higher rate for this transfer, thereby suggesting that the discrepancy with experimental measurements may indeed be due to the method of detection. This is in accordance with the results of a recent study conducted by Makri and coworkers[155], which reported a rapid initial relaxation to higher-lying states and a subsequent relaxation to the (symmetric) ground or first ($k = \pm 1$) excited state on a time scale comparable to that observed in the experimental data. The Marcus equation yields a value of 1.4 ps for the transfer between a B800 and a B850 chromophore. The occupations in Figure 9.6d converge towards a 80/20 ratio between the B850 and B800 rings.

Equilibrium: It has previously been proposed, that quantum effects must be involved in order to describe the equilibration of energy between the two ring systems[155]. As illustrated in Table 9.2, the NAMD simulations indicate that the B850 ring is occupied by 82% after 1 ps.

Tab. 9.2.: Calculated ratio of populations obtained from Boltzmann distributions compared to the occupations obtained from NAMD simulations. As energy differences ΔE, we considered 735 cm^{-1} from the experimental spectrum (850 nm, 11 765 cm^{-1} vs. 800 nm, 12 500 cm^{-1}) and 403 cm^{-1} from the difference of the lowest adiabatic surfaces from B800 and B850 (1.685 eV, 13 590 cm^{-1} vs. 1.635 eV, 13 187 cm^{-1}) along the MD simulation. For the NAMD propagations, we used the converged values of Figure 9.6(d)

	Spectrum	MD	NAMD (35 meV)
$\frac{P_{B850}}{P_{B800}}$	$\frac{97}{3}$	$\frac{87}{13}$	$\frac{82}{18}$

The Boltzmann distribution $\frac{P_1}{P_2} = \exp(-\frac{\Delta E}{k_B T})$ indicates that the energy difference between the B850 and B800 rings yields an occupation of 97% for the B850 ring. However the result from the energy difference of the adiabatic states obtained from our MD simulation, is smaller, leading to an occupation of 87%. The adiabatic states from MD lead to a slightly smaller gap than the 735 cm^{-1} obtained from the spectrum. This can be attributed to the neglect of charge transfer states in the simulation, which have recently been shown to red-shift the B850 band by more than 100 cm^{-1}[160]. The discrepancy between the NAMD result and the MD simulation is likely due to convergence.

9.3.4. Single Trajectories

In section 9.3.3, we have computed a large set of individual trajectories. In order to compare with experimental data, we have to resort to ensemble properties by averaging, as for spectra, transition rates, etc. In SH-type approaches, individual trajectories are followed, and the sampling must cover both, the spreading of the wave function and the sampling of the phase space of the classical nuclei.

SH methods are widely used in computational photochemistry, and it is instructive to analyze single trajectories as well, although they are sometimes considered a computational construct, since only averaging leads to physical observables. However, it is a computational necessity to the extent that the environment plays a central role, for example in QM/MM simulations. In Ehrenfest dynamics or density propagation schemes, the environmental response refers to an ensemble rather than a single exciton, thereby leading to a phenomenon known as mean-field description, which can lead to poor description of the environmental effects, as manifested, e.g., in a massive underestimation of the reorganization energy[153]. The use of single trajectories also allows a simple determination of the exciton delocalization, which in Ehrenfest or density matrix propagation schemes refers to an ensemble.

9.3.4.1. Delocalization

The averaged delocalization length is a misleading value, because it suggests a steady delocalized state. For the B800 ring, its value is generally accepted to be close to one[127]. For the B850 ring, however, estimates from experiments in the late 1990s range from two to four pigments[134,190,38] up to the entire ring[161]. Damjanovic et al. calculated a delocalization length of 6.4 without the presence of an exciton and 5.4 with an exciton[66]. The 'small exciton' interpretation takes into account the disorder in site energies and assumes delocalization over approximately four chromophores, as recently pointed out by Zigmantas and coworkers[283]. Calculations by Scholes et al. show that in the LH2 complex a delocalization over the entire ring is transiently possible, with the steady-state delocalization length being only two sites, though[267].

We calculated the delocalization lengths as described in section 9.2.3 based on the diabatic occupations of the active surface. Depending on the reorganization energy and the method used, we obtain delocalization lengths between 1.0 and 1.3 for the B800 ring, between 2.2 and 5.3 for the B850 ring and between 2.0 and 5.6 for the whole LH2 complex containing both rings (Table D.6). We also computed delocalization lengths obtained from the squared eigenvectors of the lowest eigenvalue of the Frenkel Hamiltonians computed every 0.1 fs along classical MD simulation. They lie between our two limiting cases for reorganization energies of 35 and 65 meV. Higher lying states may have a smaller or larger delocalization length (see Figure D.15).

In Figure 9.7, we compare 100 fs of NAMD simulations with 100 fs of MD simulation.

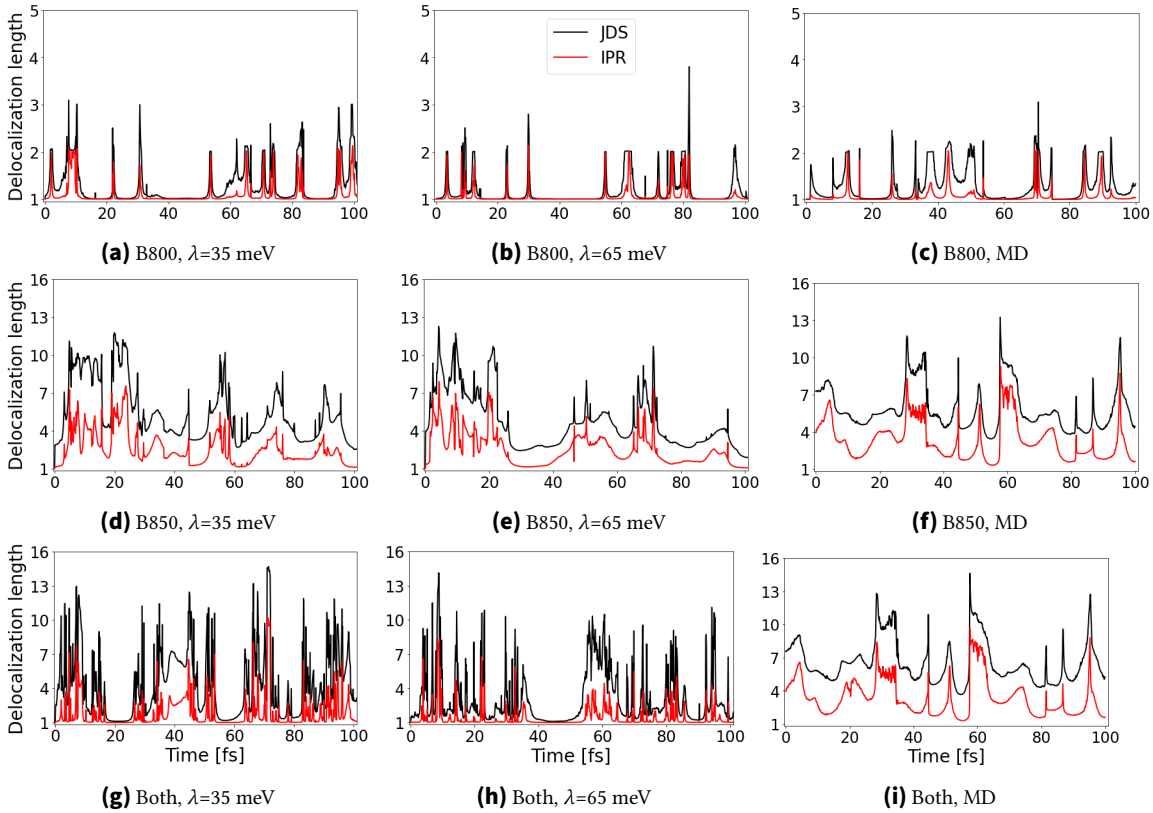


Fig. 9.7.: Delocalization lengths over 100 fs. The Occupations of the active surfaces of NAMD simulations with reorganization energies of 35 meV (a,d,g) and 65 meV (b,e,h) are compared with the occupations of the lowest adiabatic surface along an MD simulation (c,f,i). All simulations had identical initial conditions and a time step of 0.1 fs. The maximum delocalization length is equivalent to the total number of pigments, which is 8, 16, 24 for B800, B850 and Both rings, respectively.

For the B800 ring, the exciton is temporarily delocalized during hops between surfaces, and the peaks in NAMD based delocalization lengths are comparable to the ones obtained from the MD simulation (Figure 9.7a-c). The 'presence' of the exciton seems to impact the electronic structure of the system slightly. On the other hand, in the B850 ring and especially in the whole LH2 complex containing both rings (g,h), the peaks from NAMD simulations are sharper than the ones from the MD simulation (i). For given couplings and site energies, it is the reorganization energy that determines whether an exciton is localized or delocalized. In general, one would expect any value of the reorganization energy to lead to localization, because the site energy is reduced and therefore the symmetry is distorted. However, there is a concurrent effect of an excitation into higher lying states during the NAMD simulation ('transient delocalization'), which will be discussed in more detail below.

9.3.4.2. Insights into the Exciton Dynamics

The analysis of individual trajectories offers further insight into the exciton dynamics within and between the rings, as illustrated in Figure 9.8. They show examples of the time-

dependent diabatic occupations from the active surface, together with the time dependence of the delocalization, which displays a characteristic pattern.

Dynamics inside the B800 ring: Figure 9.8a+b illustrates the exciton dynamics for two different NAMD simulations of the B800 ring over the initial 500 fs. In all simulations, a clear hopping pattern is observed, with the exciton predominantly localized at a single site. The high localization within the B800 ring is a direct consequence of the weak couplings and the high outer-sphere reorganization energy exhibited by B800 pigments. During a single hop, the exciton undergoes a transition state in which it is delocalized over two sites. In these instances, the IPR assumes a value of two and there is no further delocalization. Therefore, the average values of IPR> 1 discussed above do not indicate delocalization. Rather, they result from averaging over hopping events. In the transition state, both site energies are equal and the orbital is delocalized. In very few cases more than two site energies match, and occupations of more than two sites (e.g. in Figure 9.8a at 80 fs) are considered a rare event.

B850 dynamics: The dynamics observed in the B850 ring are characterized by an exciton that is delocalized across 2-5 sites, depending on the measure employed (IPR vs. JDS). Notably, the delocalization exhibits significant fluctuations, indicating that the exciton does not have a constant extension during the dynamics. Localization on a single site is a very rare event, and there seems to be an extension of about 2-3 sites as a baseline exhibiting large fluctuations of up to 9 sites. Please note, that this behavior is typical for the so-called transient delocalization mechanism, originally proposed for charge transfer, and recently described also for excitons in organic semiconductors with unusually high exciton diffusion lengths[96,94]. As detailed by Blumberger and coworkers[96], this transport mechanism is characterized by frequent transitions into higher-lying states, which are more delocalized. As illustrated by the B800 dynamics, a conventional transition is characterized by a transition state where the exciton is delocalized across one additional site. Consequently, fluctuations in IPR by +1 are to be expected for this mechanism. Blumberger et al. define 'transient delocalization' as events wherein the IPR increases by more than one. The large IPR values account for a significant amount of the simulation time, and allow for large jumps of the exciton through abrupt delocalization and subsequent localization at a site distant from the initial site, as can be seen in many instances in Figure 9.8c+d. For example, in Figure 9.8c, at approximately 190 fs, the exciton starts to localize on the B850-1 molecule and is then rapidly displaced over pigments 16, 1, 2, 3, 4, 5, and 6 arriving in a localized state at B850-5 at the time step of 200 fs. The transfer is not restricted to nearest neighbor hops, as in the B800 case, nor is it characterized by a drift of an exciton along the ring with IPR fluctuations of +1. Rather, it is characterized by large fluctuations and sudden movements, which have a significant impact on the mobility. The fluctuation of the IPR is induced by an excitation to a higher adiabatic eigenstate, which is more delocalized, as shown in detail by Blumberger and coworkers[96]. This mechanism significantly enhances the diffusion length. For the exciton diffusion in B850, the simulation details indicate, that the transient delocalization mechanism is also at work in light-harvesting systems.

B800 to B850 transition: Starting the trajectories from an occupied B800 site, but also including the B850 ring in the simulations, shows an interesting pattern. Figure 9.8e+f

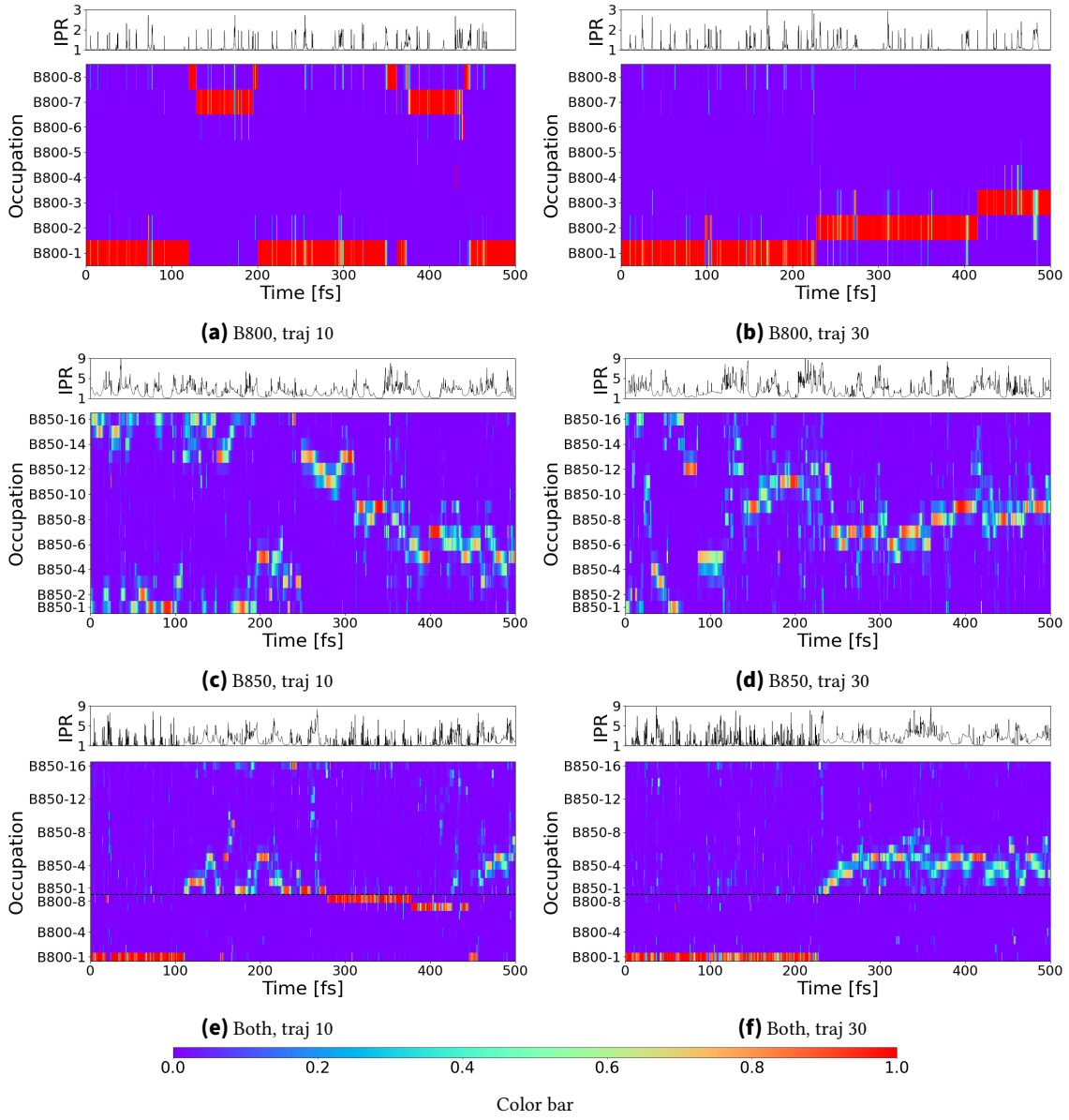


Fig. 9.8.: Two exemplary trajectories were taken from the swarm of 200 NAMD trajectories for each system. The diabatic occupations are color coded and plotted against time. The inverse participation ratio is shown on top. The exciton was initially localized on pigment B800-1 for simulations of the B800 ring and the whole LH2 system, and on pigment B850-1 for simulations of the B850 ring. In case of both rings, the dotted line represents the border between the B800 ring and the B850 ring.

show trajectories where the exciton spends 100 fs and more in the B800 sites before entering the B850 ring. Interestingly, even during this time frequent occupations of the B850 ring can be seen. These events occur as sharp and short-lived spikes, indicating that high energy states are accessed and the exciton quickly falls back to the B800 ring. Figure 9.8f also shows that the exciton finally transitions to the B850 ring at about 230 fs. Note that the IPR pattern now changes and shows a constant delocalization over at least 2-3 sites. In Figure 9.8e the exciton returns to the B800 ring at about 280 fs, and the pattern

is reversed, again showing very sharp spikes of short delocalization events into the B850 ring. At about 450 fs the exciton transitions back into the B850 ring and is again delocalized over 2-3 sites. This shows that thermal fluctuations can lead to mixed B800-B850 states that extend into the B850 region, but are only transiently occupied.

The aforementioned properties of the B800 to B850 transfer serve to highlight the influence of fluctuations in site energies. From the transient delocalization events, it is evident that the calculated transition rate of 177 fs is significantly faster than that of other calculations and experiments. The occupations of all B850 pigments were averaged over the trajectories. Consequently, all the transient delocalization events were incorporated, resulting in an apparent increase in the occupation of the B850 ring. However, as these states are not stable within the B850 ring, it is unclear whether they can participate in the absorption or emission processes observed in experiments at defined wavelengths. It is therefore necessary to consider the transition dipoles of all pigments together with the coefficients obtained from the active surface. In Table D.5, the linear combination of transition dipole moments based on the symmetric Hamiltonian (Figure 9.3) is presented. The next step in this system is to resolve the time-dependent spectrum via coefficients and transition dipole moments at each time step.

In Figure D.13 we also show trajectories simulated for 100 ps, which give an impression of the frequency of transitions between B800 and B850 rings. Furthermore, on this time scale the exciton will diffuse around the rings several times, showing the high mobility in both rings. Since the time scale for the transition between two LH2 complexes is in the order of 30 ps, this highlights a potential starting point for simulating more steps of the whole photosynthetic apparatus.

9.3.5. Comparison to Synthetic Materials

Our NAMD implementations employ an empirical decoherence correction and utilizes classical nuclear dynamics. Consequently, they disregard nuclear quantum effects and long-lived electronic coherences from the beginning. The latter was found to be not responsible for the astonishing effienciency of the exciton transport[35] in biological systems. Interestingly, the high efficiency of charge and exciton transport in organic semi-semicondcutors has also been explained using similar semi-classical NAMD simulations.

As described in the previous chapter, we find a transition time of about 50 fs within the B850 ring. Since this is in the order of ultrafast chromophore vibrations, it is interesting to compare it with other systems where such fast transfers have been reported. In the field of organic semiconductors (OSCs), charge transfer is characterized by mobilities μ and exciton transport is characterized by the diffusion constant D or the diffusion length $L = \sqrt{D\tau}$, where τ is the exciton lifetime. Of particular interest for applications in organic photovoltaics are materials with large diffusion lengths, and recently new materials like perylene diimide (PDI) derivates have been reported with diffusion constants exceeding the one of simple semiconductors like anthracene (ANT) by a magnitude[9,80].

Due to technical reason, the ring structure of B850 does not allow the extraction of the diffusion constant, so we mapped the system onto a linear chain, a procedure also used in our work on organic semiconductors[94]. 200 NAMD simulations with λ = 35 meV were performed with one pigment eliminated from the Frenkel Hamiltonian. The resulting ring with one missing element was considered as a linear chain with pigment-by-pigment distances equal to the COM-COM distances of the pigments. At one end of the chain, the propagation was started with 100% occupation and the mean squared displacement (MSD) was calculated for each frame. The slope of the MSD gives the diffusion constant. The obtained value was divided by two to correct for the known overestimation of the surface hopping approach based on implicit relaxation and serves as an estimation of the magnitude of the excitonic diffusion.

To be able to compare to organic materials, we computed the activation energy, i.e., the barrier to overcome, for the transfer of an exciton from one BChl a to another one. The following equation can be derived assuming a two-state approximation[96,94]

$$\Delta A = \frac{\lambda}{4} - < |V_{ij}|^2 >^{1/2} + \frac{1}{\lambda} < |V_{ij}|^2 > \quad .$$

It shows that the barrier vanishes if the couplings become larger than half of the reorganization energy, which indicates very efficient, delocalized exciton transfer. Giannini et al.[96] demonstrate the linear dependence of the logarithm of diffusion constants and activation energy.

Figure 9.9 compares diffusion constants with the activation energies for exciton transfer in several organic semiconductors. Our results are in line with the previously observed relation between activation energy and diffusion constant. Remarkably, the tuning of couplings and reorganization energy in the biological system lead to a smaller activation energy than the ones found in any organic semiconductor. This explains the outstanding efficiency of the LH system. In comparison to the most efficient OSCs, the BChl a system outperforms them with a diffusion constant of 0.25 cm^2/s, which is 50 times greater than the 0.005 cm^2/s observed for anthracene[9]. Anthracene exhibits a Marcus-type hopping transfer as in B800, while in PDI the above described transient localization mechanism is at work. Even the B800 hopping case features a very small barrier due to the extremely low reorganization energy, as already stated in our work on FMO[272].

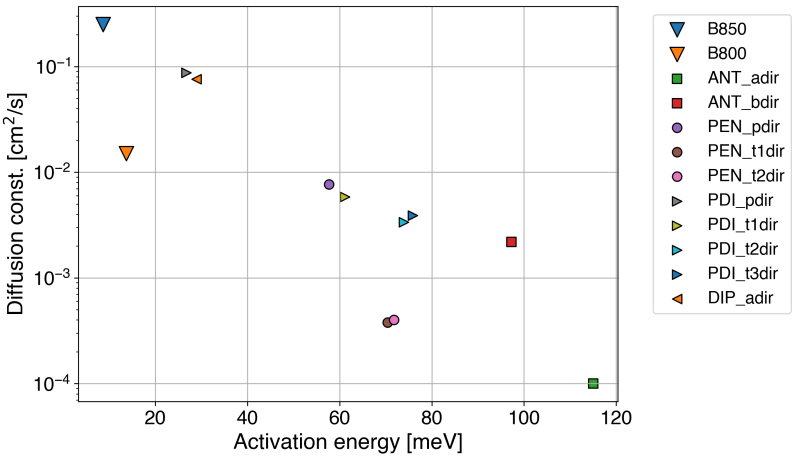


Fig. 9.9.: Diffusion constants for the B800 and B850 "linear chains" compared to values for some organic semiconductors including anthracene (ANT), pentacene (PEN), perylene diimide (PDI) and diindenoperylene (DIP) calculated in different directions of the crystal[94].

9.4. Conclusion and Outlook

In this study, we examined the exciton transfer in LH2 from *Rs. molischianum* using surface hopping-based non-adiabatic molecular dynamics simulations. The applied multiscale model is based on a fragmenation of the system and the use of neural networks to accelerate the calculations, which is the key for the simulation of such large complexes on nanosecond time scales. The electronic structure is described using a Frenkel Hamiltonian, reproducing the experimental absorption spectrum quite well. The time dependence of the Hamiltonian demonstrates that energetic disorder significantly reduces the symmetry of the exciton, leading to localization. Using trajectory based SH methods, we are able to seamlessly cover different transport regimes. On the one hand, the method is capable to reproduce the rates for hopping in the Förster limit ($V << \lambda$), on the other hand, it is able to correctly describe the ultrafast delocalized coherent transfer regime in the Redfield limit ($V >> \lambda$). Chenu and Scholes[44] proposed a third regime with $V \approx \lambda$ to be relevant for exciton transfer in the B850 ring of LH2.

In the context of organic semiconductors, a new charge transport theory has been proposed in order to rationalize the efficient transport in the $V \approx \lambda$ regime, known as "Transient Localization Theory".[48] It has recently been demonstrated to be also applicable to exciton transfer[96,94,196]. Its main feature is the fluctuation of delocalization, whereby the exciton does not "drift" through the material with constant delocalization. Instead, sharp spikes in delocalization are essential for explaining the high diffusion constants observed in these materials. Our simulations accurately reproduce the incoherent hopping transitions within the B800 ring, as well as the dynamics within the B850 ring, where the averaged exciton delocalization of 4±2 sites is described in good agreement with experimental estimates[190,225,188]. The time-resolved picture of the exciton dynamics provides mechanistic insights into the B850 ring, indicating a localization on approximately two to three sites, accompanied by large spikes with delocalization of up to 10 or more sites. The pronounced

spikes suggest that the transient delocalization transport mechanism, as described in OSCs, is also at work in the B850 subsystem of LH2.

The diffusion constant of exciton transfer within the B850 ring is comparable to that of the most efficient organic materials to date, which is related to the unusual properties of BChl a chromophores. The inner-sphere reorganization energy is surprisingly small, while the minimal change in dipole moment upon excitation results in a negligible outer-sphere contribution. It even vanishes for the B850 chromophores due to shielding from solvent exposure. As a result, the reorganization energies and couplings are within the same range, $V \approx \lambda$, indicating that the third transport regime between hopping and band-like transport is at work in B850. This regime was previously described independently in the field of organic electronics[85,48,96] as well as for LH2 systems[267] and it provides an explanation for the efficient exciton transport observed in light-harvesting systems. The physicochemical properties of the chromophores are responsible, thus negating the necessity to invoke non-trivial quantum effects resulting from the protein environment.

10. Conclusion and Outlook

This work demonstrated how machine learning models integrated into computational chemistry protocols can uncover molecular mechanisms that explain biological and spectroscopic experiments.

In Chapter 6, the molecular recognition of nematode-specific pheromones in the fungus *Arthrobotrys flagrans* was investigated. The hypothesis that two short sequence motifs, the R- and S-motif (RFAF and SLIL/SLIY), are key contributors to ligand binding was tested. AlphaFold2-based receptor models revealed that the fungal and the nematode receptor share several structural similarities despite lacking sequence identity. Docking analyses, supported by experimental validation, identified the binding pockets. The results showed that the R-motif is only sequentially conserved. While being essential for ligand binding in the fungal receptor, it was not involved in the nematode receptors. By contrast, the S-motif was found to be sequentially and structurally conserved. Its intracellular location suggests a role in signaling. Based on the structural analysis, experiments on chimeric receptors with or without R- and S-motifs were explained. The results demonstrate the utility and limits of molecular docking: Predicted binding poses can help to interpret phenomenological outcomes on a molecular level, but they need experimental validation due to approximations in scoring functions. Ongoing work on the related receptor, GprB, is expected to shed more light on the molecular origins of inter-species sensing between fungi and nematodes.

In Chapter 7, the spectroscopic properties of the pigment bacteriochlorophyll a (BChl a) were analyzed. Configurational sampling was performed either with force fields or with the semi-empirical DFTB3/3ob-f method. Excitation energies calculated with TD-LC-DFTB2 were used to obtain spectral densities from energy gap autocorrelation functions. The absorption spectra were calculated either as oscillator-strength–weighted histograms (ensemble) or with the full cumulant expansion (FCE) method. The analysis showed that, besides the well-known role of bond length alternation, the curvature of the BChl a ring influences the excitation energy, whereas the well reported effect of the acetyl dihedral angle could not be reproduced. Electrostatic interactions were found to influence both the average excitation energies and their fluctuations, with the resulting reorganization energy λ arising in equal parts from inner-sphere geometrical and outer-sphere electrostatic fluctuations. Comparison of the ensemble and FCE method showed that FCE not only reproduced the correct line shape of the absorption spectrum, but also yielded the solvatochromic shift more accurately than the ensemble approach. Whereas the inability of TD-LC-DFTB2 to capture the dihedral effect is considered a methodological limitation, the out-of-plane motion of the magnesium atom in DFTB3 trajectories may be an artifact

or a true part of the phase space. Both aspects call for a clarification through higher-level theories. The developed protocol for absorption spectrum calculations can be readily applied to future studies of solvatochromic shifts in other molecules. For the present case of BChl a, the insights gained were essential for the subsequent chapters, where identical pigments are studied within proteins that tune their excitation energies.

In chapter 8, exciton transfer in the Fenna–Matthews–Olson (FMO) complex was investigated with non-adiabatic molecular dynamics (NAMD) simulations using a trajectory based surface hopping approach. Geometries were sampled with a force field, while neural networks yielded excitation energies and couplings to construct a Frenkel Hamiltonian on-the-fly. This Hamiltonian was used to propagate the pigment-wise excitonic populations using the time-dependent Schrödinger equation. This bottom-up approach allowed the exciton dynamics to be controlled solely by geometric and electrostatic fluctuations of nuclei, without empirical bias or pigment-specific parameters. The feedback of the exciton on nuclear motion was included in the form of altered partial charges of the pigments, proportional to their excitonic populations. First, the impact of two methodological approximations were analyzed: the downscaling of the couplings and the implicit relaxation parameter, i.e. the inner-sphere reorganization energy that is subtracted from a pigment's site energy in the Hamiltonian if occupied by the exciton. The analysis at 300 K revealed that a combination of downscaling the couplings to gas phase transition dipoles and setting the inner-sphere reorganization energy to the gas phase calculated value of 35 meV reproduced experimental transfer times. Absorption spectra generated at 77 K and 300 K further suggested that a reorganization energy of 15 meV is needed to achieve agreement with experiments. NAMD simulations at 77 K additionally revealed that exciton transfer speed is highly sensitive to the degree of delocalization. A limitation of the study lies in the difficulty of reproducing cryogenic experiments with MD simulations due to trapping in local minima. Nevertheless, the benchmarked protocol, including optimized input parameters, proved transferable and was directly applied in the subsequent chapter.

Chapter 9 extends the NAMD simulations of exciton transfer to the light-harvesting complex 2 (LH2). This complex contains 24 BChl a chromophores: eight in the loosely packed B800 ring and 16 in the densely packed B850 ring. To study exciton dynamics in both rings, the surface hopping method was applied. It has the advantage of covering transport regimes from the localized limit with excitonic couplings V being much smaller than the reorganization energy λ up to the band-like transfer regime in the limit of $V \gg \lambda$. Although the calculated time-averaged Hamiltonian of the system produced an absorption spectrum that agreed with experimental results, it was found that the adiabatic states resulting from the collective excitation of the 24 pigments, must be considered time-dependently. One property of adiabatic states is the delocalization of the exciton among pigments. The experimentally determined value of four pigments[188] is misleading because it suggests a steady delocalized state. The NAMD simulations reproduced this value on average, but they also showed that fully localized excitons in the B800 ring transfer to their neighbors in discrete hops. In contrast, the B850 ring exhibits delocalization of approximately two pigments, accompanied by short delocalization events that can extend beyond half of the ring. Such a pattern aligns with the "transient localization theory",[48] which was originally formulated for the intermediate $V \approx \lambda$ regime in organic

semiconductors.[96,94,196] This highly efficient transfer regime is possible only because of the small inner-sphere reorganization energy of BChl a and the unpolar protein environment, which shields the chromophore from solvent exposure, leading to a negligible outer-sphere contribution.

Although the method can discriminate between qualitatively different transport regimes, studying the fine details of more complex photosynthetic systems, like those in plants, requires an extension of the method. The modular NAMD method can be expanded in various ways. For instance, it could incorporate non-adiabatic coupling vectors, charge transfer states, or ground state QM sampling. Additionally, hydrogen bonds to BChl a can be used as an additional descriptor in the NNs to obtain more accurate site energies.

The ML-accelerated modeling of exciton dynamics has generated a deeper understanding of the molecular origins of efficient energy transport in photosynthetic proteins. Future applications can be seen for example in the field of rational design of organisms, as was recently pioneered by the engineering of plants with increased photosynthetic productivity.[300]

Appendix

A. Protein Structure Prediction and Molecular Docking

Tab. A.1.: Docking scores (kcal/mol) of the best binding pose for various receptors and ligands: Ascarosides (A#n), glucose (Gluc) and sucrose (Sucr). n.d.: not determined.

Receptor	A#1	A#2	A#3	A#5	A#18	Gluc	Sucr
GprB	-6.25	-6.11	-7.04	-6.09	-7.02	-5.73	-5.97
GprC	-7.26	-6.90	-8.31	-6.62	-8.29	-5.63	-6.85
GprD	-5.08	-5.69	-6.30	-5.00	-5.85	-4.86	-5.64
SRBC64	-6.99	-6.61	-7.61	-6.50	-7.51	-6.00	-6.57
SRBC66	-6.36	-6.34	-6.18	-6.05	-6.94	-5.36	-5.77
Daf37	-6.06	-5.40	-6.25	-5.24	-6.76	-4.70	-6.32
Daf38	-7.03	-6.40	-7.64	-5.54	n.d.	n.d.	n.d.
SRBC64–GprC	-7.06	-6.93	-7.48	-5.66	n.d.	n.d.	n.d.
Daf37–GprC	-7.06	-6.79	-7.01	-6.18	n.d.	n.d.	n.d.
$SRBC64^{N2.60A}$	-6.90	-6.68	-7.45	-6.10	n.d.	n.d.	n.d.
$GprC^{N2.53A}$	-6.97	-7.09	-7.91	-6.29	n.d.	n.d.	n.d.
$GprC^{N2.57A}$	-7.13	-6.96	-7.57	-6.12	n.d.	n.d.	n.d.
$GprC^{R6.45A}$	-6.65	-6.50	-6.69	-5.92	n.d.	n.d.	n.d.

Tab. A.2.: Sequences of the three proteins GprC, SRBC64 and DAF37.

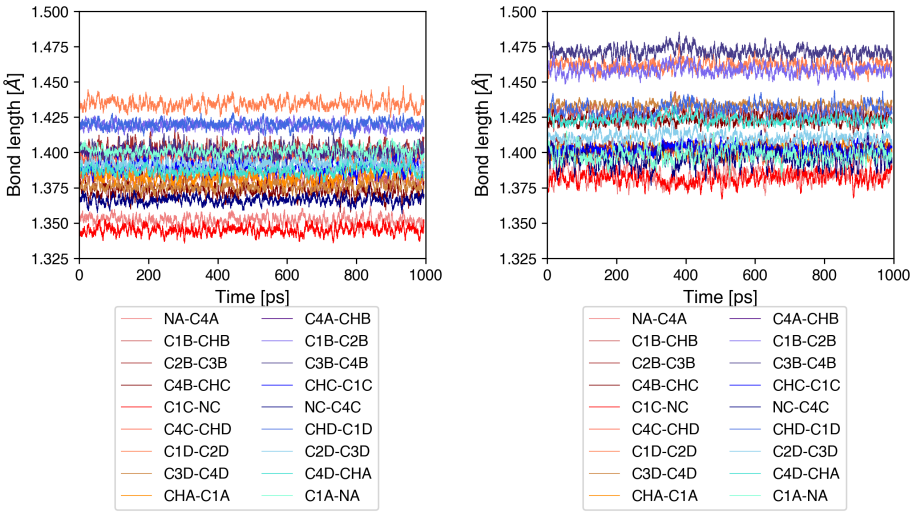
Receptor	Sequence
GprC	MAFTTLSFTIAVLNLIGSLSSFFGSGFIVVTYLLLPIKRHFRHSLILNLAIA DFINSSNNSTSGLWRLLKRREIPDSPGCVTNGFLGQLSVQATDTSILAI AIVTVWSLTRKTTICETLPRTTTALICATTWILPITTSFIALGMNRYGP VSGNWCWIKAEPSYFRYVMTHGWRFAFILSEIVMYTYLHFYIRKRFG AFLDASRCSSHTGRSTVRDVGTEFRNDSMAIRLDSVAIRNDSLAPAEE PQRSDSQMPFSITDEGDIRPIGARGQDLESGEAFVRTEVWISETYDDK TLPAHAVTTTISSGNNSLEQPNTDLRPETANTIESRRPKFSNPFAHRRA DEDDLADVDVNDENERKIRQQNASRSRRVRRILLLNAYPAMYILLWI PGIINRLIEASGGKSSVMQIMQASTQFVGLANAITYGWNERVWRQLK DYIADRWGSDRKRSVEELAVQKVELENRLRVGLKGSKERLREIRDGL RSPNKKDDDEVPLDNRGQLMYINGTWQWSQTASINSQDGDADDRE ISRDPGPRRE
SRBC64	MPEIVIILNNFLSVVFSQTTCILNFYLLFSIFYSKRIAWKSELSLIYTRFAF DMFYTFFVPHNKIYYIIRQIFPGFVVKNLSFYLIWPTIPLGCIRATLVFLIT LDRVIALFFPISYHNHRFKISLSAIIISVSCFGASDHLILFGYCRYTVDVPL ECDNFNCVIKTCFYSYWVSRDQIQYILTGTFSVILLFRLVIWNNFVASQ RNQNLSRATRIALLESIVIICFSIIPSLIFSTFPSLNFVSVGPWTIVLKHAG FMIEALIICKFLLKRKSKNGIGRNSTVLTWSQ
DAF37	MSAITITCNSLSILFPIITCSINCYLIYSIFYSKRITWKSEFSLIYTRFAIDIVY TFFVPHAKIYYVLRQISDIFVMKNLTFYLVWPTIPLGAIRATLVLLITLD RVVASFFPIFYHNHRRRIPILAVISCVITFGLSDHIVLFEYCKYTVDVPLA CDNFNCVINQCFFNYWSLRDQVQYFIIGVLMDVIGNITDLSPTVSGIP DECGLEPHDFLEVKFFLISVVGTLIGLFGLFGNATTALILTRPSMRNPN NLFLTALAVFDSCLLITAFFIYAMEYIIEYTAAFDLYVAWLTYLRFAFA LSHISQTGSVYITVAVTIERYLAVCHPKSSKNMCGPGGAAWTILGVTT FAVVFNCTKFFELQVTVNPSCPDGSNWQSYILLPSAMASNPIYQQVYS LWVTNFVMVFFPFLTLLLFNAIIAYTIRQSLEKYDFHNQKSVVAALSA SVNLPRNIAGISSRNELKEKSREATLVLVIIVFIFLGCNFWGFVLTLLERI MGQETLMVEHHIFYTFSREAINFLAIINSSINFVIYLLFGKDFRKELVVV YGCGIRGISLRLPVQDKFVIWRHWKRTKSRISMNTTNRTRHKISLPQT LVEHANLERLEETRFLAHHEDGVQTQVSPIHALRNGSTPKIDTLQDL TSNGRPCKTSIIDDNGTVVCTVTEFP

Tab. A.3.: Coordinating residues of three receptors labeled with the Ballesteros–Weinstein numbering. Orthosteric site reflects the typical binding pocket of GPCR proteins. Bold numbers indicate 5 positions that show highest interaction frequencies of about 40% in >200 GPCR class A crystal structures from ref[40]. The most conserved residue in TM7 could not be clearly defined for GprC and SRBC64, therefore, the most likely positions are shown.

Helix	Orthosteric site	GprC	SRBC64	DAF37
TM1	–	–	–	–
TM2	2.60	N2.53	Y2.53	L2.53
	2.63	N2.57	V2.57	Y2.61
		S2.60	N2.60	
		R2.63	K2.61	
TM3	3.28	N3.28	I3.29	S3.32
	3.29	Q3.33	T3.32	H3.33
	3.32	V3.36	I3.33	Q3.36
	3.33		C3.37	
	3.36		R3.39	
	3.37			
TM4	–	–	–	–
TM5	5.39	T5.39	W5.39	T5.42
	5.40	H5.40	R5.42	N5.43
	5.43	R5.44	D5.43	M5.46
	5.44		Q5.46	V5.47
	5.47		Y5.47	
TM6	6.44	W6.48	I6.43	N6.48
	6.48	R6.55	S6.47	L6.54
	6.51		S6.51	T6.55
	6.52			
	6.55			
	6.58			
TM7	7.31	Q7.39	I7.32	R7.35
	7.34	T7.42	K7.35	I7.38
	7.38	Q7.43	H7.36	N7.39
	7.41			A7.42
	7.42			N7.45
Loop4–5	**45.52**	W45.52	F45.48	–
		W45.54		

B. Spectroscopic Properties of Bacteriochlorophyll a

B.1. Vacuo Simulations



(a) Force Field **(b)** DFTB3

Fig. B.1.: Bond lengths of the 18 atoms in the aromatic bacteriochlorin ring. Single bonds are shown in red and brown colors, double bonds are shown in blue and cyan colors. The two single bonds NA-C4A and NC-C1C are shorter than most double bonds due to the nitrogen atom involved.

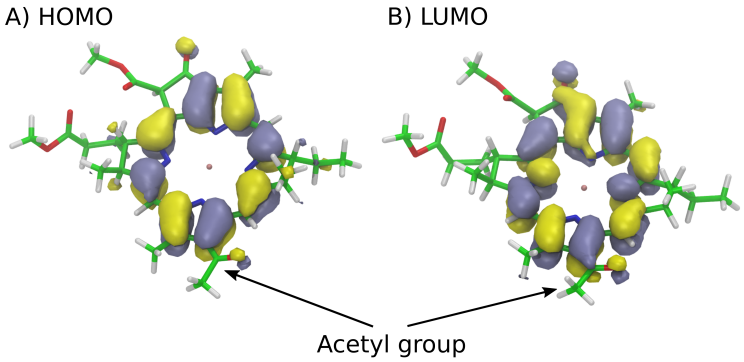


Fig. B.2.: Highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of the first snapshot of the 1 ns long QM/MM simulation of BChl a in gas phase.

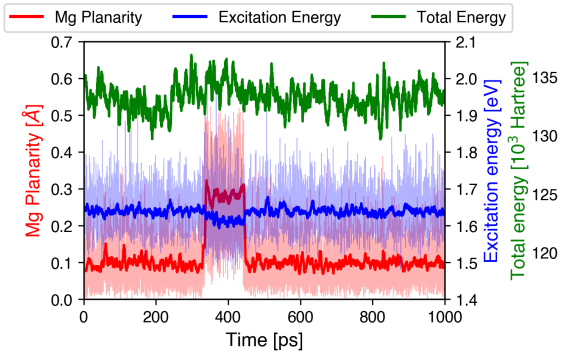


Fig. B.3.: Planarity (red), excitation energies (blue) and the total energy of the ground state (green) along the 1 ns long MD simulation performed with DFTB3. Running averages were taken over 50 ps windows are shown in bold color.

B.2. Solvent Simulations

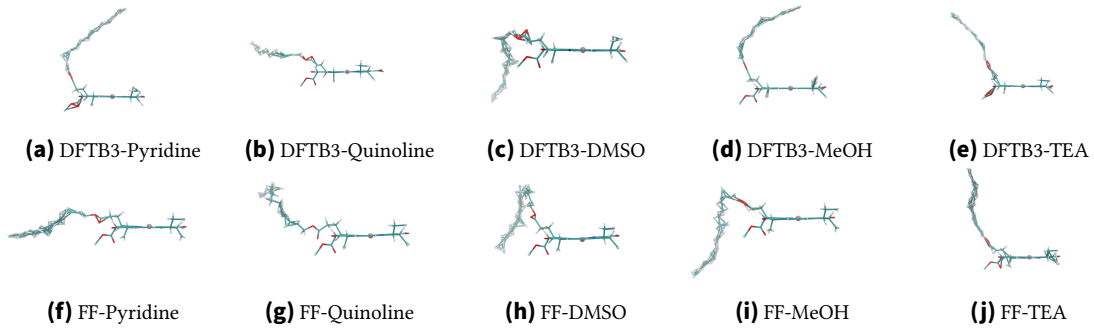


Fig. B.4.: Average Structures of the 1 ns long MD simulation in different solvents. The BChl a molecule was treated with DFTB3 (upper row) or with the force field (FF, lower row).

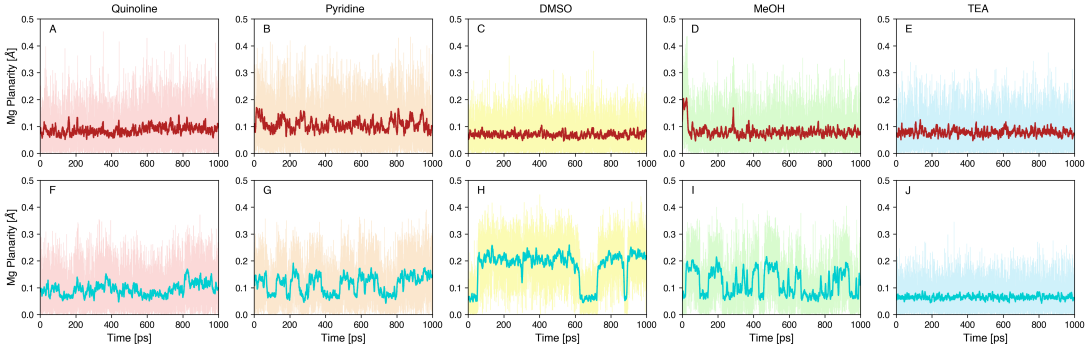


Fig. B.5.: Mg-Planarity during the 1 ns long MD simulation of all solvents with DFTB3 (upper row) or force fields (bottom row). Data is shown in transparent colors, running averages over 50 ps windows are shown in red and turquoise for DFTB3, and force fields, respectively.

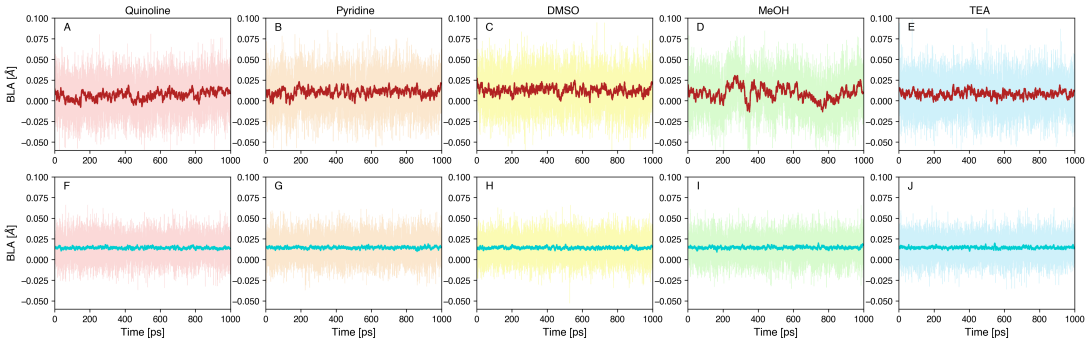


Fig. B.6.: Bond length alternations during the 1 ns long MD simulation of all solvents with DFTB3 (upper row) or force fields (bottom row). Data is shown in transparent colors, running averages over 50 ps windows are shown in red and turquoise for DFTB3, and force fields, respectively.

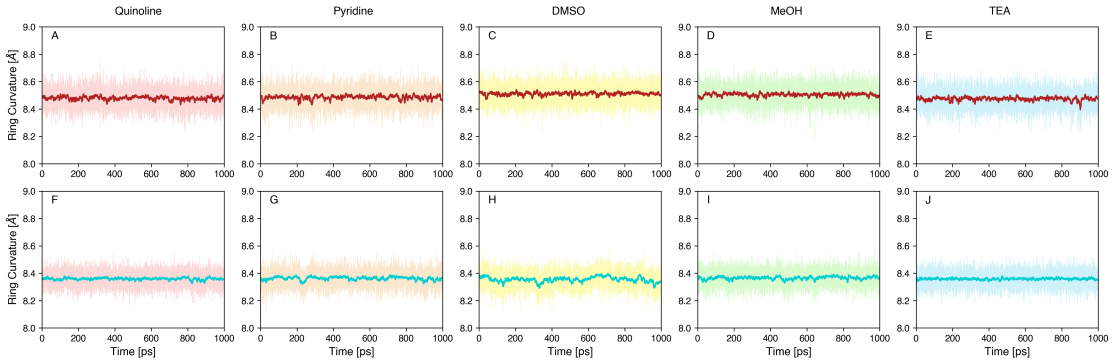


Fig. B.7.: Ring curvature during the 1 ns long MD simulation of all solvents with DFTB3 (upper row) or force fields (bottom row). Data is shown in transparent colors, running averages over 50 ps windows are shown in red and turquoise for DFTB3, and force fields, respectively.

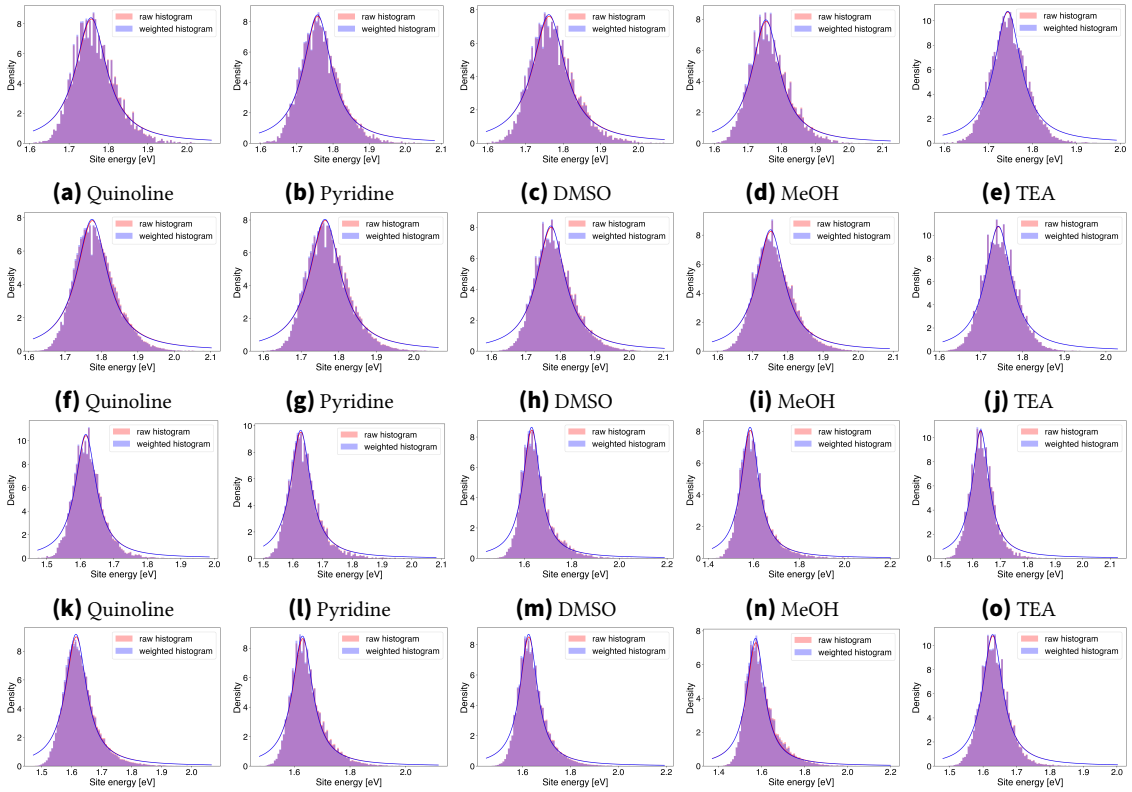


Fig. B.8.: Distributions of 10 000 and 60 000 site energies obtained from the 1 ns long (row 1, 3) or 60 ps long (row 2, 4) simulations, respectively. The effect of the weighting by oscillator strength was analyzed. The results from MM simulation (upper two rows) and the QM/MM simulation (bottom two rows) is shown. Six solvents were analyzed: quinoline (red), pyridine (orange), DMSO (yellow), methanol (green), and triethylamine (turquoise).

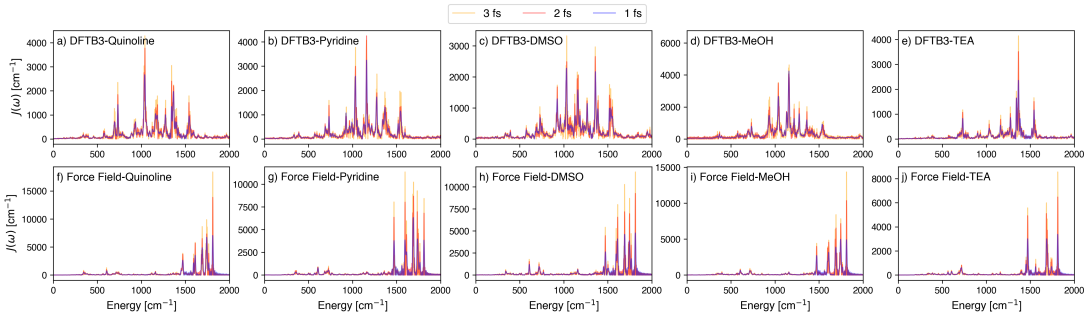


Fig. B.9.: Spectral densities with the influence of the environment calculated with a time step of 1 (blue), 2 (red), and 3 (yellow) femtoseconds along 60 ps long simulations with DFTB3 (upper row) or with the force field (lower row)

B.3. Absorption Spectra

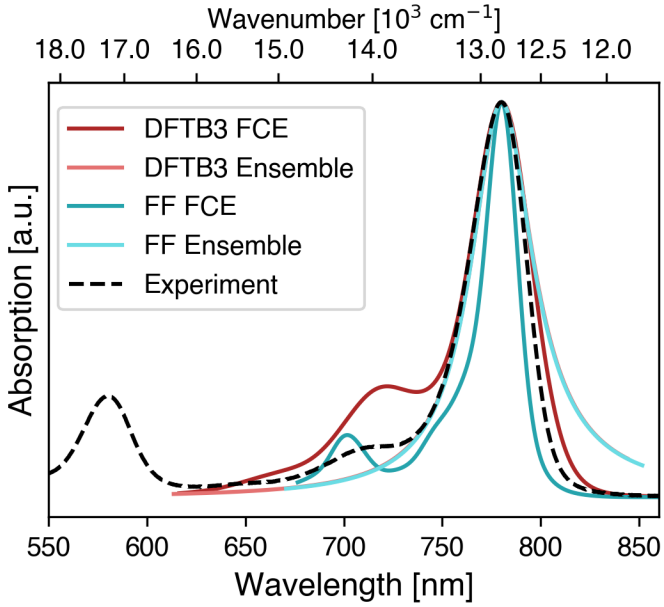
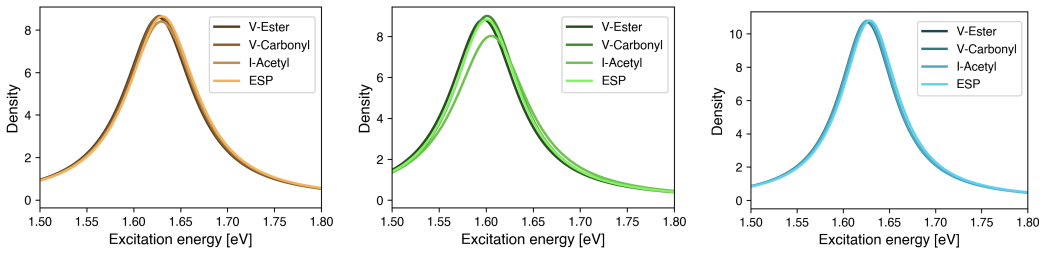


Fig. B.10.: Absorption spectra in nm obtained from solvent simulations in pyridine either with DFTB3 (red) or force field (blue). The Full-Cumulant expansion method (dark colors) was compared against the ensemble method with weighted histograms (bright colors). Positions were shifted and intensities were scaled to match the experimental absorption spectrum of BChl a in pyridine, which was taken from Ref. [197]. The Q_x band at 580 nm (2.13 eV) is indicated for comparison with the width of the Q_y peak.

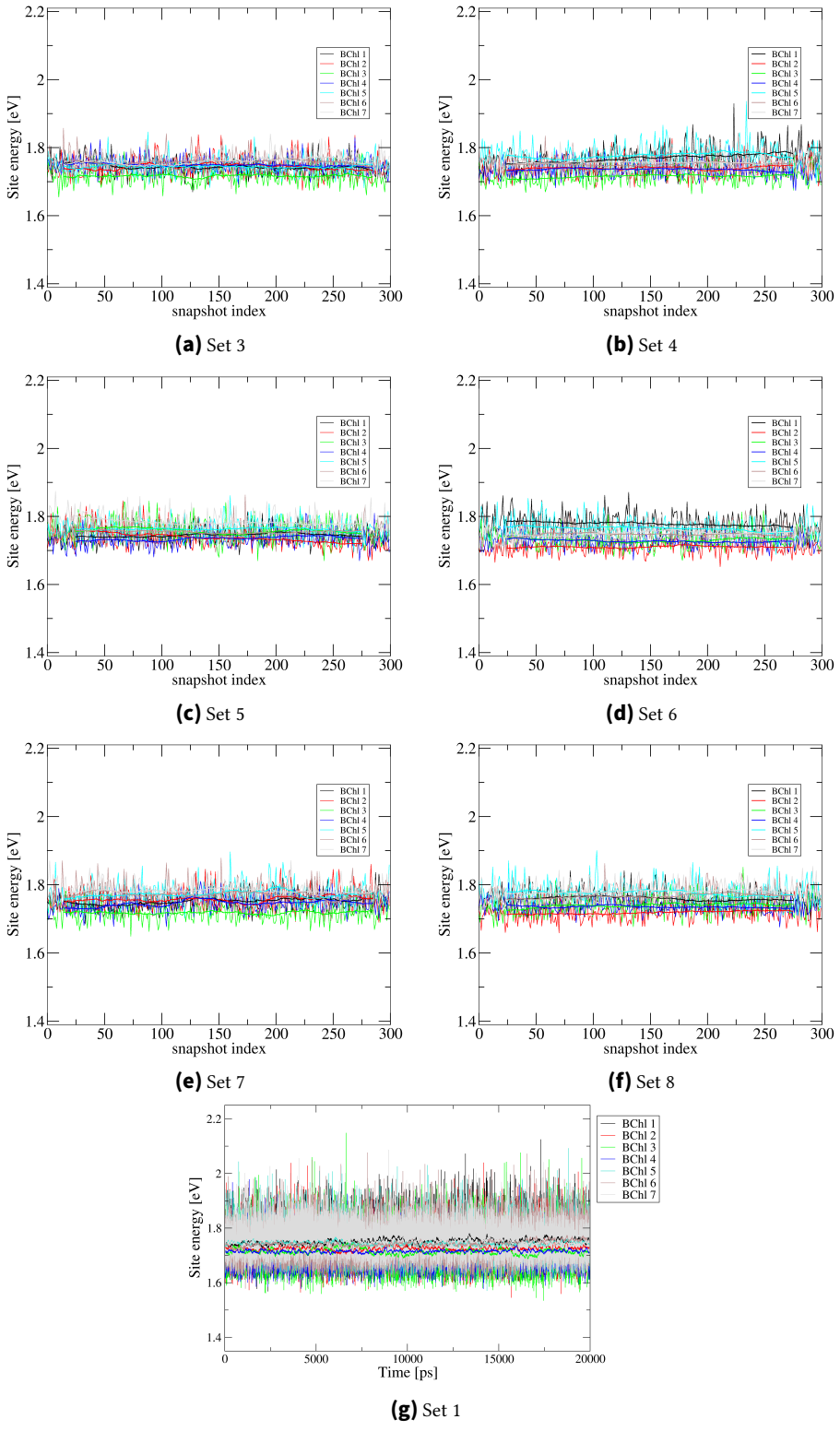
B.4. QM solvent interactions



(a) Pyridine **(b)** MeOH **(c)** TEA

Fig. B.11.: Charge transfer between hydrogen bond donating solvents and BChl a. Histogram of excitation energies calculated with the closest solvent molecule to three oxygen atoms in BChl a (Figure 7.1) included in the QM-zone. The calculations without additional solvent in the QM-zone are denotes as ESP.

C. Exciton Transfer in the Fenna–Matthews–Olson Complex at Cryogenic Temperatures



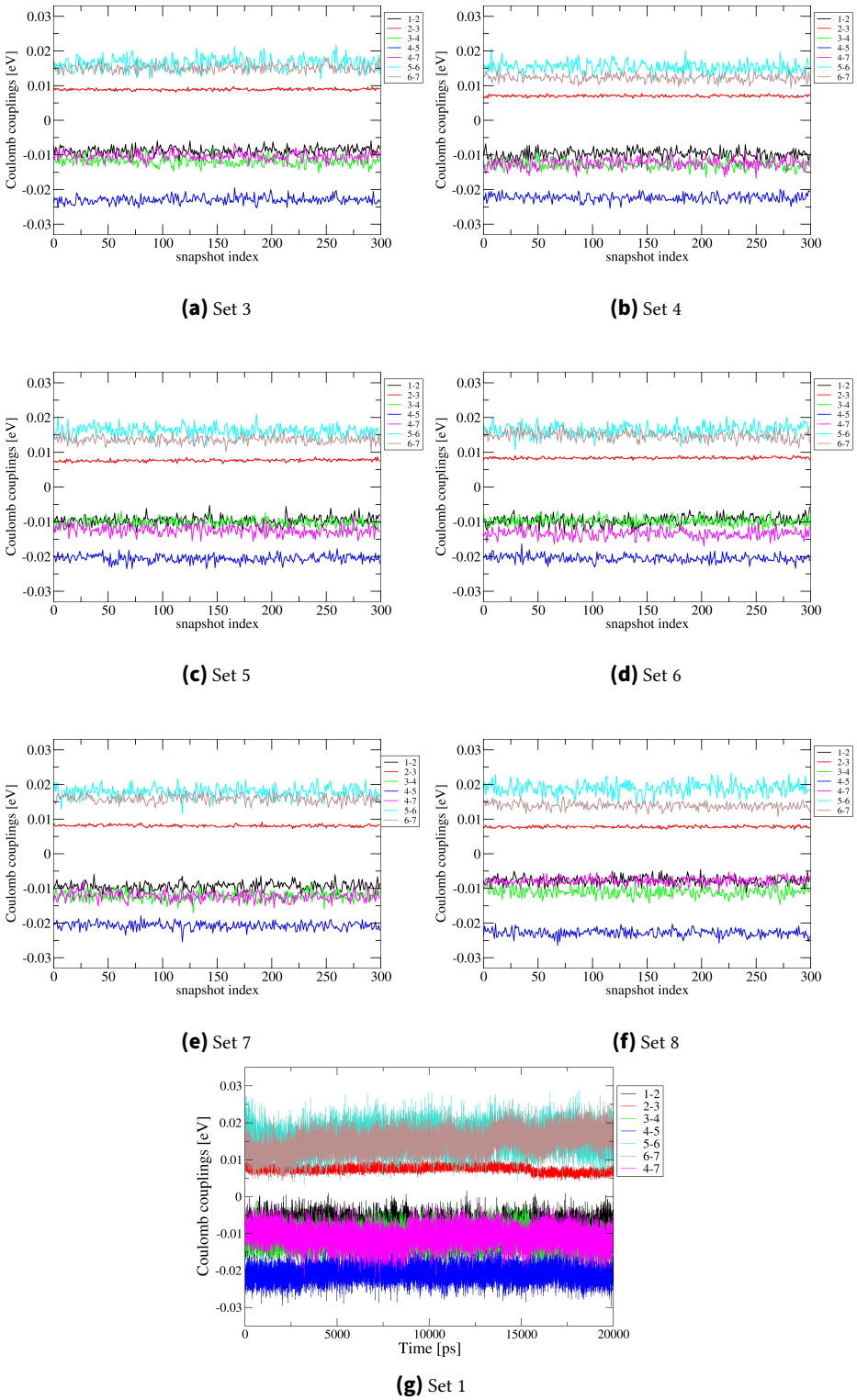
(a) Set 3 **(b)** Set 4

(c) Set 5 **(d)** Set 6

(e) Set 7 **(f)** Set 8

(g) Set 1

Fig. C.1.: Site energies along 100 ns of MD simulations at 77 K for set 3 to set 8 (a-f). The running averages over 50 data points are also shown. The snapshots are separated by 330 ps. For comparison, the site energies along 20 ns of MD simulations at 300 K (set 1) is also shown (g).



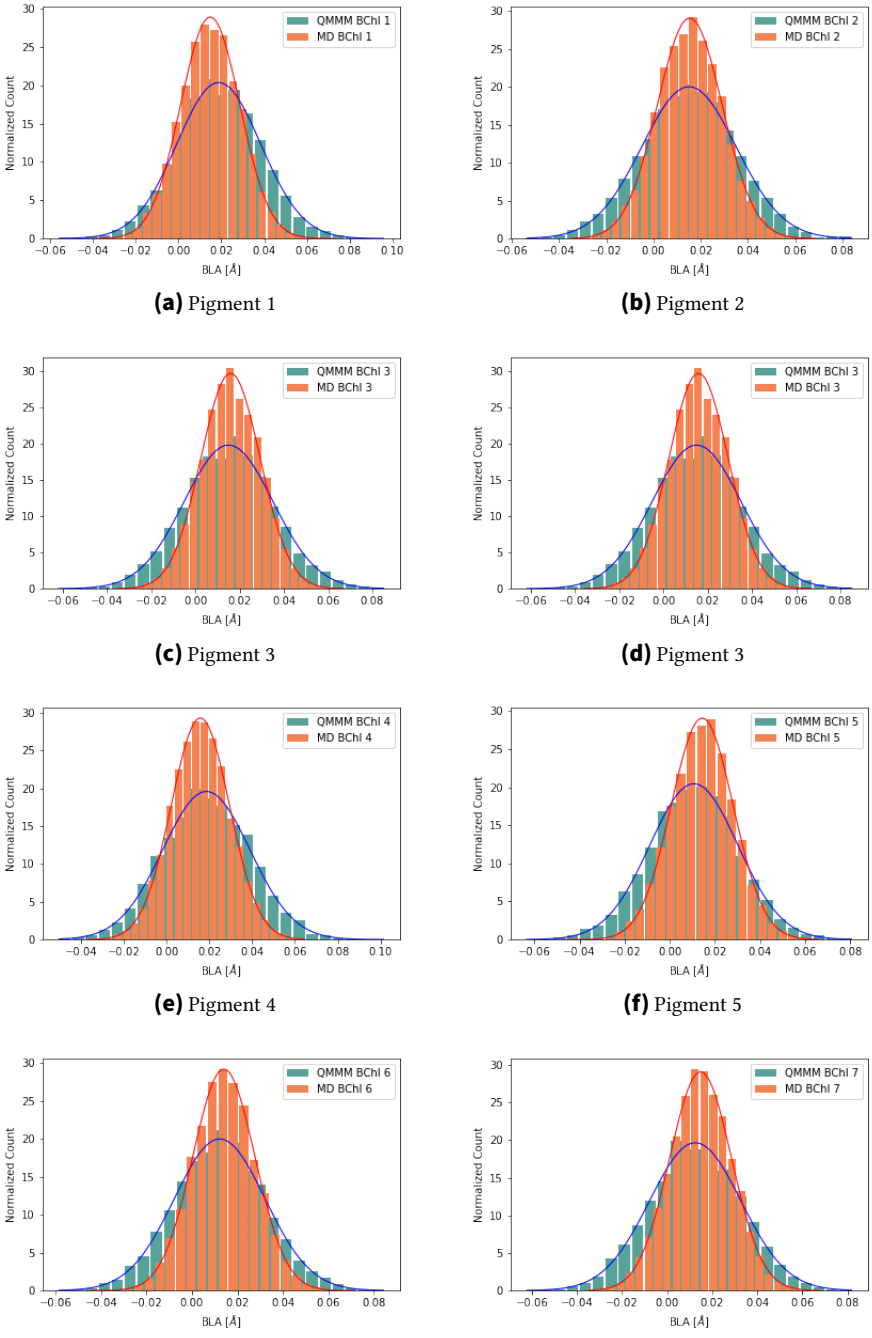
(a) Set 3 **(b)** Set 4

(c) Set 5 **(d)** Set 6

(e) Set 7 **(f)** Set 8

(g) Set 1

Fig. C.2.: Unscaled Coulomb couplings along 100 ns of MD simulations at 77 K for set 3 to set 8 (a-f). The running averages over 50 data points are also shown. The snapshots are separated by 330 ps. For comparison, the site energies along 20 ns of MD simulations at 300 K (set 1) is also shown (g).



(a) Pigment 1 **(b)** Pigment 2

(c) Pigment 3 **(d)** Pigment 3

(e) Pigment 4 **(f)** Pigment 5

(g) Pigment 6 **(h)** Pigment 7

Fig. C.3.: Bond length alternations of individual pigments. The force field geometries extracted from the 20 ns long MD simulation at 300 K (set 1) are shown in red. For each pigment, a separate 60 ps long QM/MM simulation was performed. The bond length alternations calculated on the respective geometries are shown in green.

D. Elucidating Exciton Transport Mechanism in the Light-Harvesting Complex 2

D.1. Machine Learning

The architecture of the neural network (NN) was chosen to be the same as in Ref. [272]: the inverse distances (Coulomb matrix without charges) of a Bchl a molecule together with the electrostatic potential (ESP) of the whole simulation box acting on each atom of the Bchl a molecule were used as descriptors for the site energy NN. For the couplings, the upper right square matrix of the inverse distances between two Bchl a molecules was used, so that only the intermolecular distances and no ESPs were included. The single-valued targets for the two NNs were site energies and couplings, respectively. For both models, 100 000 data points were used for training and 20 000 for testing. In a hyperparameter optimization using the hyperband algorithm, we explored a search space of 20 to 1000 neurons, 2 to 6 hidden layers and learning rates of 1e-3, 5e-4, 1e-4, 5e-5, 1e-5, 5e-6 and 1e-6. It was determined that small models perform as well as complex ones. Consequently, the final models were selected to comprise 30 neurons per layer accross two layers. The models were trained using the Adam optimizer with a learning rate of 10^{-4} and an early stop with patience 300. The activation function was leaky softplus. Because of the large amount of available training data, the current NNs have a more straightforward architecture in comparison to other NNs for chlorophyll molecules, as outlined in ref. [47, 169]. In these instances, the Coulomb matrix was also employed as a descriptor, but the regression to the target value (site energy) is more complex than our comparatively simple feed-forward NNs. As demonstrated in Figure D.1, it is clear that much more than 5 000 training data points are required. Such a large number of training data could not be easily created with DFT-based or even costlier methods. The semi-empirical TD-LC-DFTB2 was benchmarked on Bchl a molecules[25], so we can consider the training data set to be of high quantity and quality. For LH2, we had to recreate the training set for the Bchl a molecules, because the NN, which was trained on FMO molecules, didn't perform well on the LH2 pigments. Therefore, from the 200 ns long MD simulation at 300 K, we extracted snapshots every 40 ps, resulting in 5 000 frames. Because there are 24 Bchl a molecules in the system, we obtained 120 000 data points for site energies. For the couplings, we considered five different types of couplings, resulting in 600 000 data points for coulomb couplings. The learning curves of the two networks for site energies and couplings are illustrated in Figure D.1 and show convergence.

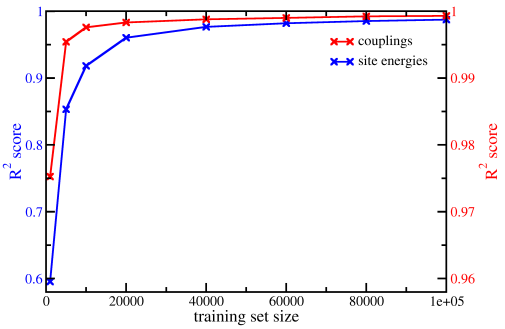


Fig. D.1.: Learning curves of models predicting site energies (blue) and couplings (red). The values of the R^2 score as calculated on the test set is plotted against the training set size.

For the couplings, the neural network has to learn the transition densities and their $1/R$ relation. Because the effect of the geometries on the transition charges is marginal, coulomb couplings are easier to predict than the site energies. Therefore, we didn't use all 600 000 data points, but reduced them to the amount that was used for site energies. Consequently, for both networks, we used 100 000 data points for training and 20 000 for testing. The final accuracy is shown in Table D.1 and is highlighted in form of scatter plots in Figure D.2. The time series in Figure D.3 confirms that the quality metrics (mean absolute error and R^2 values) do not only indicate good prediction on average, but show that the models are able to correctly predict the relevant fluctuations.

Tab. D.1.: Mean absolute error (MAE, in meV) and coefficient of determination (R^2) for the two models predicting site energies or couplings.

	site energies	couplings
MAE	4.109	0.305
R^2	0.987	0.999

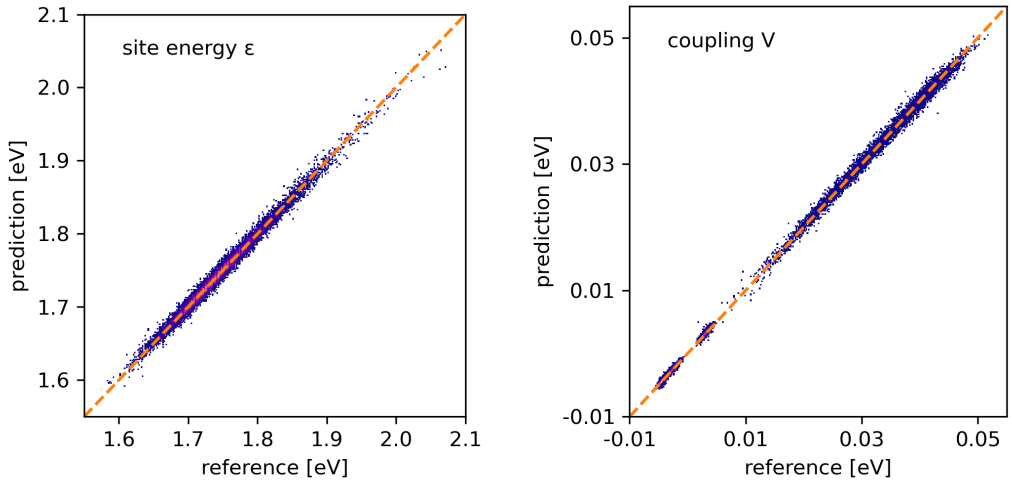


Fig. D.2.: Scatter plots of TD-LC-DFTB2 targets vs. model predictions on the test set: (left) site energies and (right) coulomb couplings. One purple scatter represents one data point and bright colors indicate high data density.

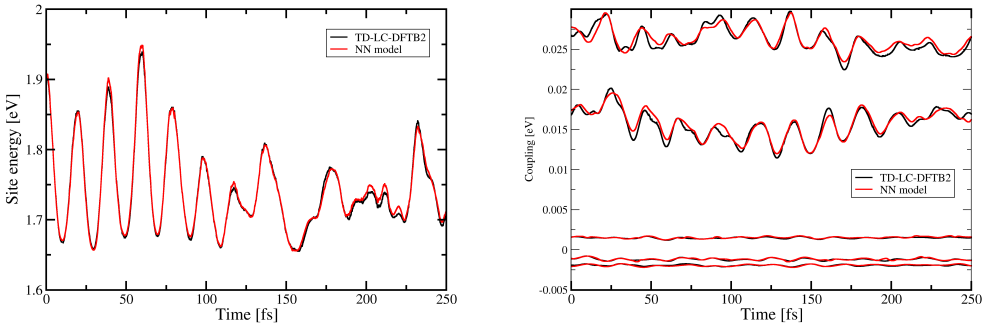
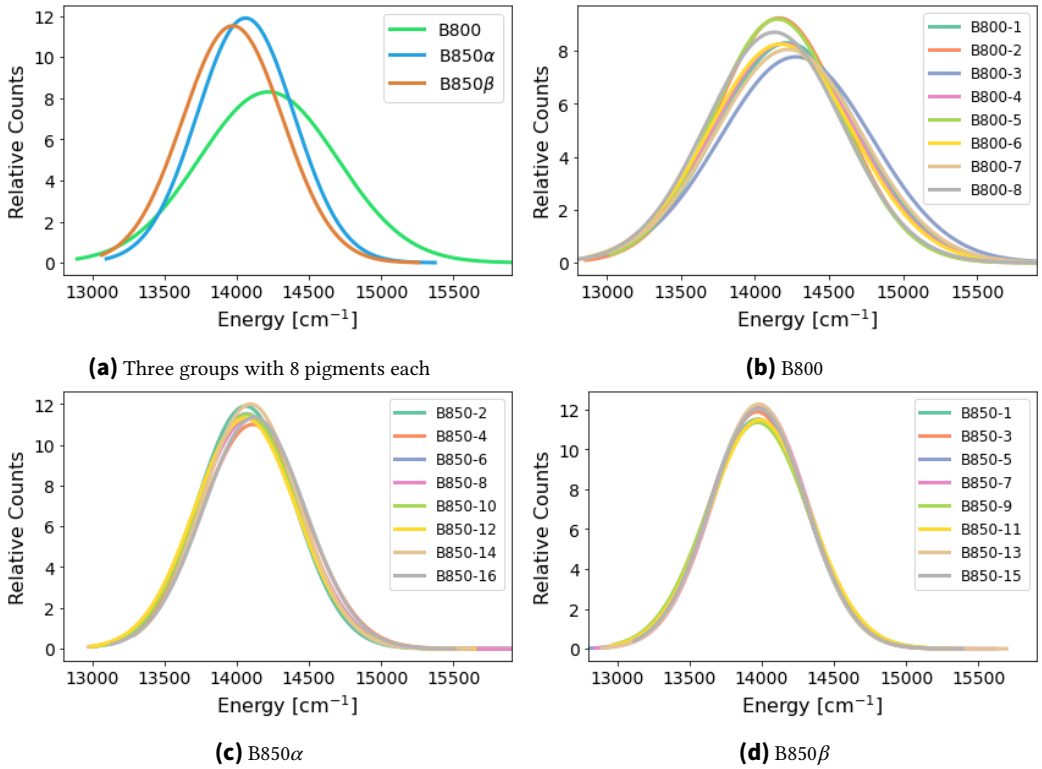


Fig. D.3.: Time series of the elements of the Frenkel-Hamiltonian computed along the MD simulation with a resolution of 1 fs. The values obtained from the reference method TD-LC-DFTB2 (black) is compared to the predictions of the respective neural network (red) for site energies (a) and forces (b).

D.2. Properties of BChl a Molecules

Site Energies and Reorganization Energy

The B800 molecules (14 220 cm^{-1}) are blue shifted in comparison to B850α (14 062 cm^{-1}) and B850β (13 974 cm^{-1}) by 158 cm^{-1} and 246 cm^{-1}, respectively. The average blue shift of 202 cm^{-1} between the B800 and B850 sites is consistent with previous findings of blue-shifted B800 pigments (336 cm^{-1}[93], 360 cm^{-1}[296]) and even in good quantitative agreement with recent work from the Mennucci group (200 cm^{-1}[59], 241 cm^{-1}[228]). The observed blue shift can be attributed to the B850 molecules being embedded in a



(a) Three groups with 8 pigments each

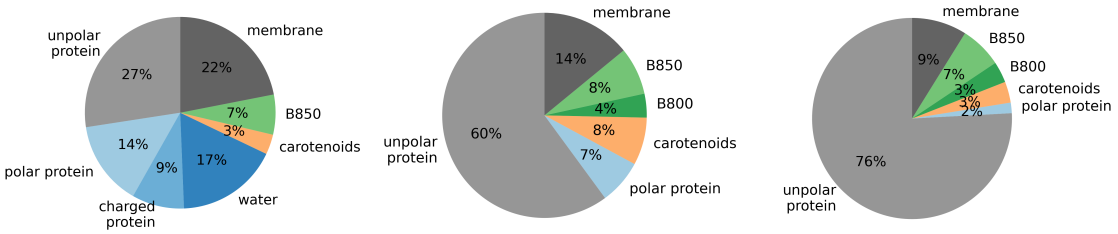
(b) B800

(c) B850α

(d) B850β

Fig. D.4.: Distribution of site energies along a 40 ns MD at 300 K for the three pigments in the system (a) and individual molecules belonging to the groups B800 (b), B850α (c) and B850β (d) calculated with LC-TD-DFTB in electrostatic embedding.

hydrophobic protein scaffold, whereas the B800 pigments have a hydrophilic and partially aqueous environment[145] as illustrated in Fig. D.5.



(a) B800

(b) B850α

(c) B850β

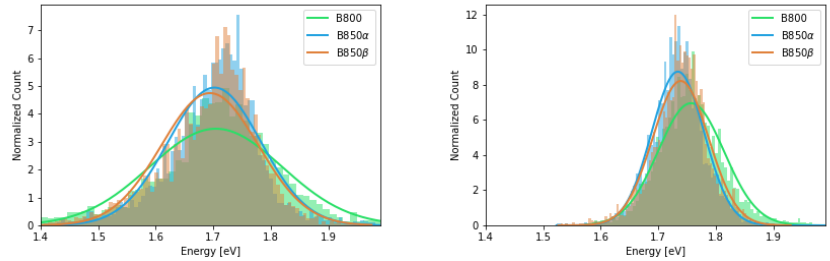
Fig. D.5.: Pie Diagrams of molecular origins from all atoms within 3 Åof the BChl molecule, averaged over all eight monomers and the 1 000 frames of the 40 ns long MD simulation. The environments of B800 (a), B850α (b) and B850β (c) are shown.

Tab. D.2.: Site energies of the pigments in eV obtained from TD-LC-DFTB2 compared to literature values. The symmetric LH2 ring consists of eight monomers, each having one B800 and two B850 (α and β) BChl a molecules. For each molecule, 1000 site energies have been calculated along the 40 ns long MD simulation. The mean value of Gaussian fits to 8000 site energies is reported. Tretiak et al.[286] used INDO/S/CIS on one crystal geometry, Vegte et al.[296] used the charge density coupling method with TD-DFT and fitted parameters. Fujimoto et al. and Sisto et al.[87, 266] used TD-DFT whereas Shibl et al. and Scholes et al.[264, 252] reported values fitted to stationary spectra. Vac refers to site energies calculated in the complex without electrostatic embedding. Acidoph refers to *Rps. acidophila*, other values refer to *Rs. molischianum*

System	ϵ_{B800}	ϵ_{B850a}	ϵ_{B850b}
TD-LC-DFTB2 vac	1.731	1.742	1.736
TD-LC-DFTB2 env	1.759	1.748	1.734
Tretiak 2000 vac	1.208	1.208	1.198
Tretiak 2000 $\epsilon = 9$	1.558	1.588	1.608
van der Vegte, Jansen 2015 (acidoph)	1.534	1.494	1.484
Fujimoto, Yanai 2021 (acidoph) vac	1.628	1.608	1.608
Fujimoto, Yanai 2021 (acidoph)	1.609	1.588	1.588
Sisto, Martinez 2017 (acidoph) vac	1.565	1.565	1.565
Scholes, Flemming 1999 (acidoph.)	1.562	1.497	1.497
Shibl, Kühn 2017 (acidoph.)	1.549	1.573	1.497

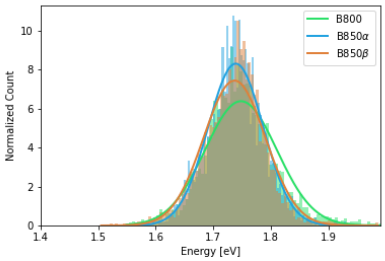
The B850α molecules are blue-shifted by 88 cm^{-1} compared to the B850β pigments. This finding is quantitatively consistent with van der Vegte et al. (80 cm^{-1})[296], and qualitatively aligned with experimental work (300 cm^{-1}[93], 530 cm^{-1}[255]). The energetic difference between these two identical molecules is challenging to evaluate using QM/MM simulations. They may be influenced by sampling, or by the specifics of the QM/MM approach. For instance, Tretiak et al. [286] reported a red-shift of 166 cm^{-1} using a geometry-optimized crystal structure and INDO/S/CIS as the quantum method. Ramos et al. [228] observed a blue-shift of 75 cm^{-1} using TD-DFT. However, they also found a red-shift of 29 cm^{-1} when they explicitly included certain hydrogen bonds in the high-light form of LH2. In many approaches, the energy differences were omitted, and identical energies were assigned to the B850α and B850β molecules[87, 266, 252].

The differences between B850α and B850β are minor, which are likely overruled by the dynamic disorder, as demonstrated in Figure D.6 for 500 fs windows. At longer time scales, this disorder may be termed (quasi-)static disorder, which manifests as inhomogeneous line broadening. However, the term “static” lacks a clear definition, and it is therefore necessary to refer to a specific process [188]. For example, different conformational structures in an ensemble of identical proteins, defects in ordered crystals, or varying coordination patterns of the magnesium atom over µs time scales [246] can all be classified as static disorder.

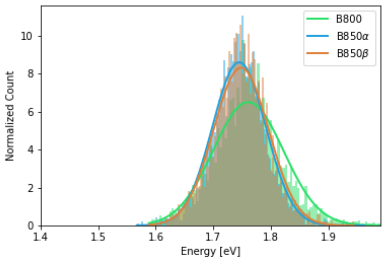


(a) 8 ps - 8.5 ps

(b) 1 ns - 1.0005 ns



(c) 10 ns - 10.0005 ns



(d) 20 ns - 20.0005 ns

Fig. D.6.: Visualization of Dynamic Disorder. Site energies calculated every 1 fs along MD Simulations for 500 fs, starting at different snapshots from the 40 ns long MD simulation, namely at 8 ps, 1 ns, 10 ns, 20 ns.

Tab. D.3.: Reorganization energy λ in meV obtained from the 40 ns MD simulation, as described in the computational details (section 9.2).

B800	B800-1	B800-2	B800-3	B800-4	B800-5	B800-6	B800-7	B800-8	mean
λ_{vac}	31.8	30.0	33.4	32.0	29.6	33.4	32.3	33.3	32.0
λ_{tot}	68.7	55.5	78.3	69.2	55.9	69.5	73.0	62.5	66.6
λ_{outer}	25.3	17.5	29.9	24.5	19.8	22.4	27.3	20.5	23.4

B850α	B850-2	B850-4	B850-6	B850-8	B850-10	B850-12	B850-14	B850-16	mean
λ_{vac}	30.2	32.9	32.4	30.0	32.2	34.0	30.0	28.2	31.2
λ_{tot}	33.5	39.2	35.7	37.3	35.9	37.0	32.9	36.8	36.0
λ_{outer}	2.9	3.9	2.8	4.6	3.2	2.7	2.8	6.1	3.6

B850β	B850-1	B850-3	B850-5	B850-7	B850-9	B850-11	B850-13	B850-15	mean
λ_{vac}	33.6	32.0	31.0	31.6	35.6	34.4	31.0	30.8	32.5
λ_{tot}	35.7	33.5	32.4	32.0	36.7	35.9	31.4	32.6	33.8
λ_{outer}	3.1	3.0	2.9	3.1	2.7	3.2	2.6	3.1	2.9

Couplings

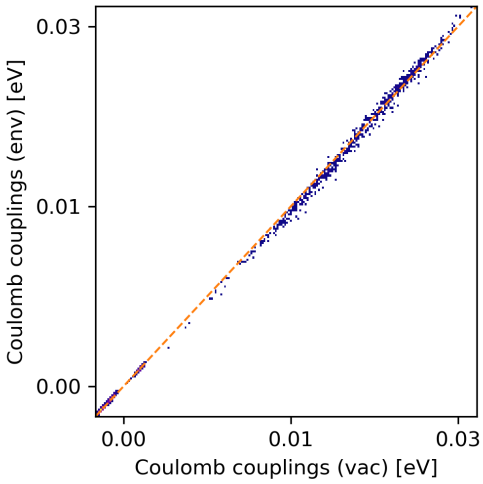
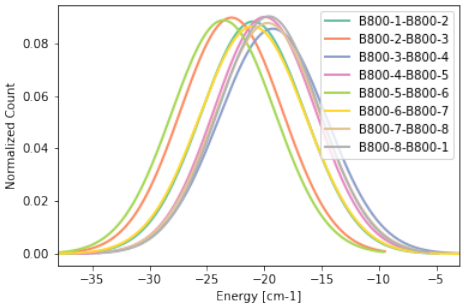


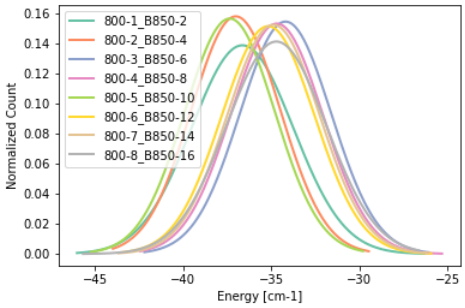
Fig. D.7.: Scatter plot of couplings calculated from coulomb interaction of transition charges obtained with TD-LC-DFTB2 in electrostatic embedding (env) or in gas phase (vac).

Tab. D.4.: Coulomb couplings in cm^{-1} of the pigments obtained from transition charges from TD-LC-DFTB2 compared to literature values. For each molecule, 1000 transition charges have been calculated along the 40 ns long MD simulation and nearest neighbor couplings are reported as mean values of the Gaussian fits to 8000 data points per coupling type. Tretiak et al.[286] used INDO/S/CIS on one crystal geometry, Sisto et al.[266] used TD-DFT, Shibl et al.[264] and Scholes et al.[252] used transition densities from CIS. Vac refers to site energies calculated in the complex without electrostatic embedding. Acidoph refers to *Rps. acidophila*, other values refer to *Rs. molischianum*.

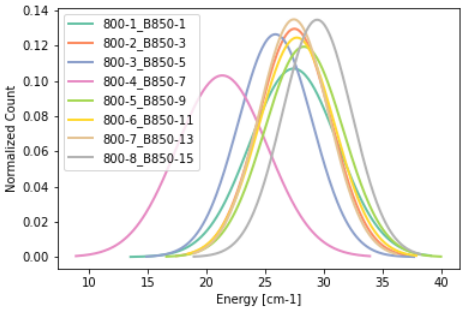
System	V_{B800}	$V_{B800-850\alpha}$	$V_{B800-850\beta}$	$V_{B850-intra}$	$V_{B850-inter}$
TD-LC-DFTB2	-20.8	-35.3	26.2	326	223
Tretiak 2000 env	-25	-53	38	363	320
Tretiak 2000 vac	-19	-36	25	258	210
Sisto 2017	-39	-46	46	440	125
Scholes 1999 (acidoph.)	-27	-13	23	320	255
Shibl 2017 (acidoph.)	-25	-12	27	307	237



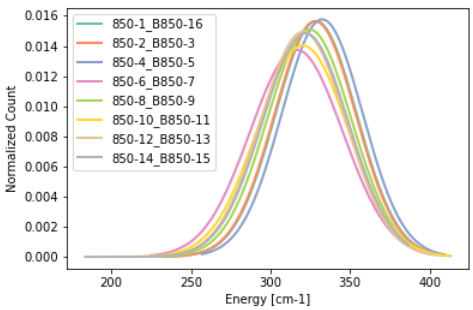
(a) B800s



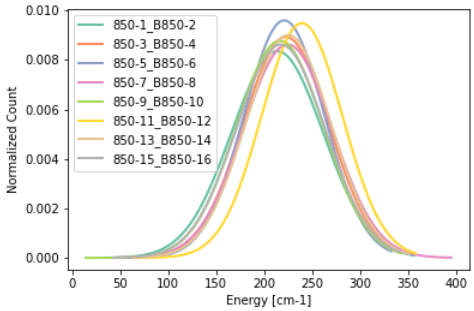
(b) B800-B850α



(c) B800-B850β



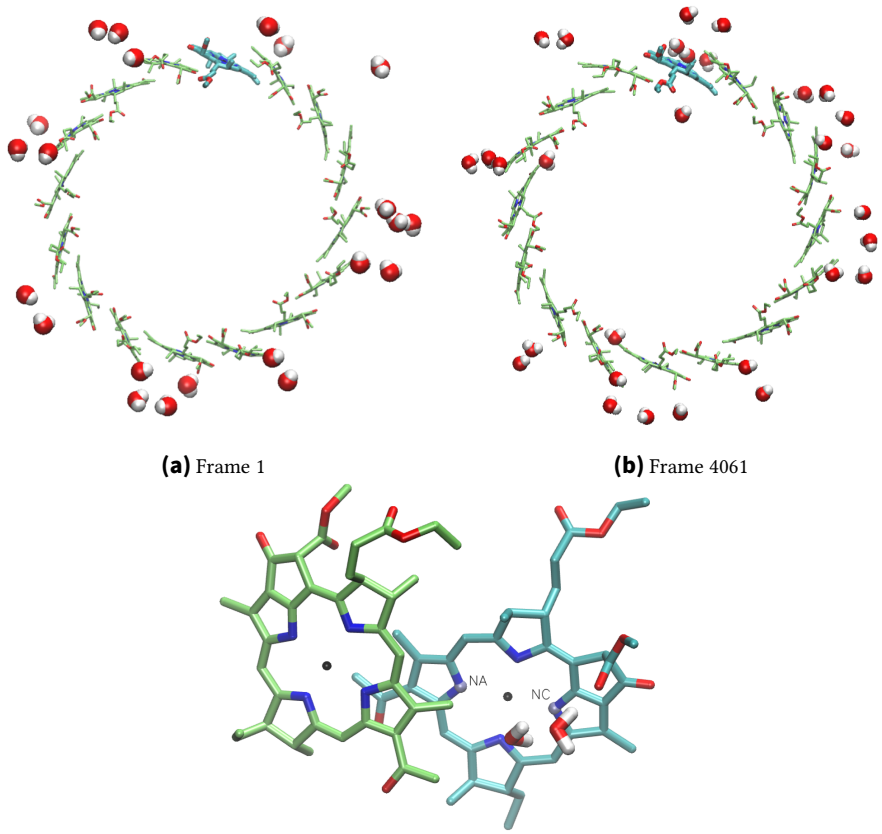
(d) B850s intra dimer



(e) B850s inter dimer

Fig. D.8.: Distribution of coulomb couplings along 40 ns MD at 300 K.

Static Disorder



(a) Frame 1

(b) Frame 4061

(c) Frame 4061, zoom into B850-16 (cyan) and B850-15 (green) with two water molecules in front

Fig. D.9.: Water around 6 Angstroms of B850 pigments at frame 1 (a) and frame 4061 (b). Pigment B850-16 is shown in cyan on top of the ring. A zoom into B850-16 explains deviations in site energies of the two pigments B850-15 and B850-16 (c).

D.3. Properties of Excitonic States

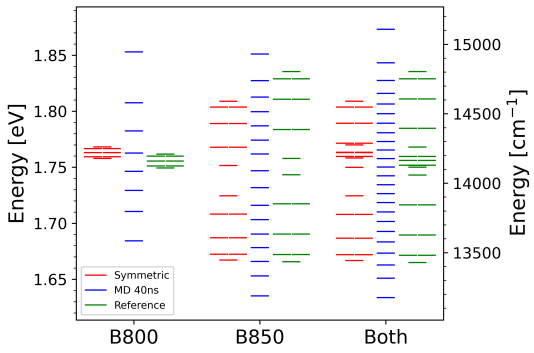


Fig. D.10.: A total of 1 000 Frenkel Hamiltonians, spaced by 40 ps, were generated and diagonalized over the course of the 40 ns MD simulation. This yields 8 000, 16 000 and 24 000 adiabatic energies for B800, B850 and both rings, respectively. The *n*-th adiabatic states with respect to energy were averaged and their mean values were plotted as horizontal bars (blue). Additionally, a symmetric Hamiltonian was constructed from time-averaged values from Table D.4 and Table D.2 (Symmetric, red). The eigenvalues are compared to those obtained from a geometry-optimized crystal structure by Tretiak et al.[286], which was used by Kundu et al.[156, 155] (Reference, green).

Tab. D.5.: Excitonic transition dipoles of the B850 ring obtained from the symmetric Hamiltonian.

State	E (cm^{-1})	μ (D)
k = 0	13446	1
k = ±1	13488	21
k = ±2	13606	1
k = ±3	13777	1
k = 8	13909	1
k = 9	14127	1
k = ±5	14259	1
k = ±6	14429	1
k = ±7	14548	1
k = 16	14590	4

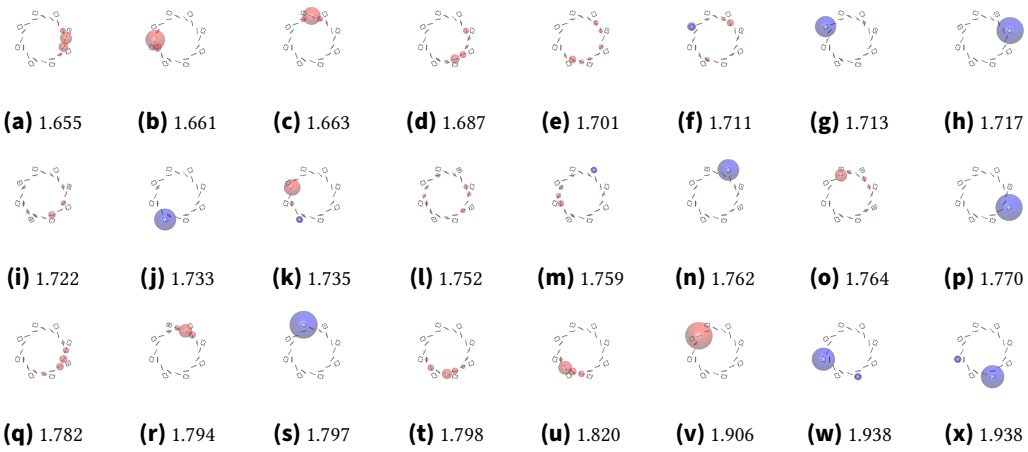


Fig. D.11.: Frame 2 of the MD simulation: Eigenstates of the LH2 complex are shown with their excitonic energy in eV.

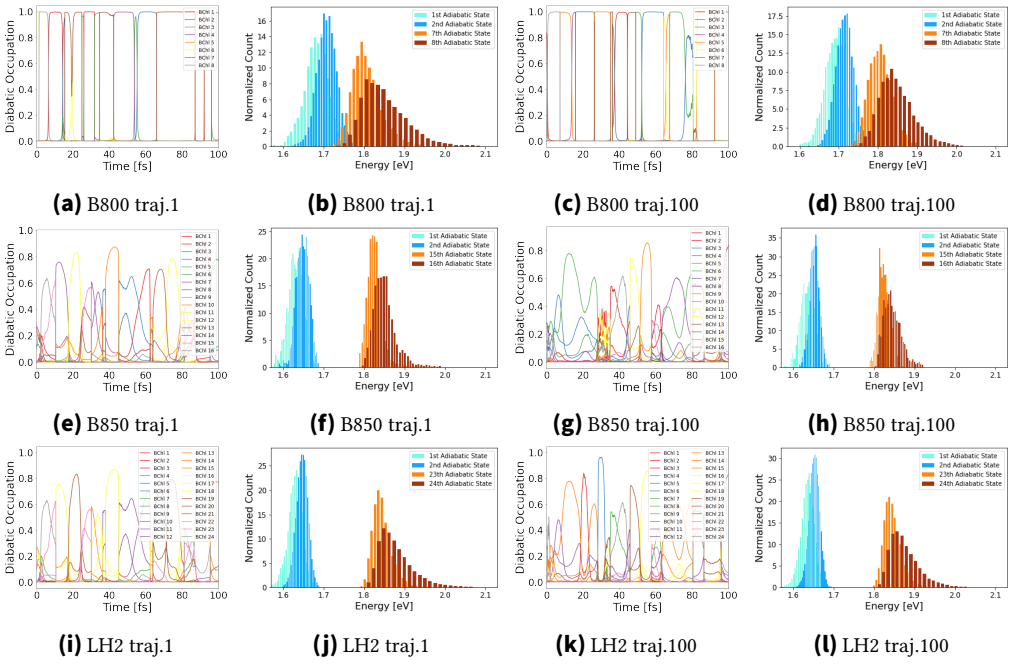
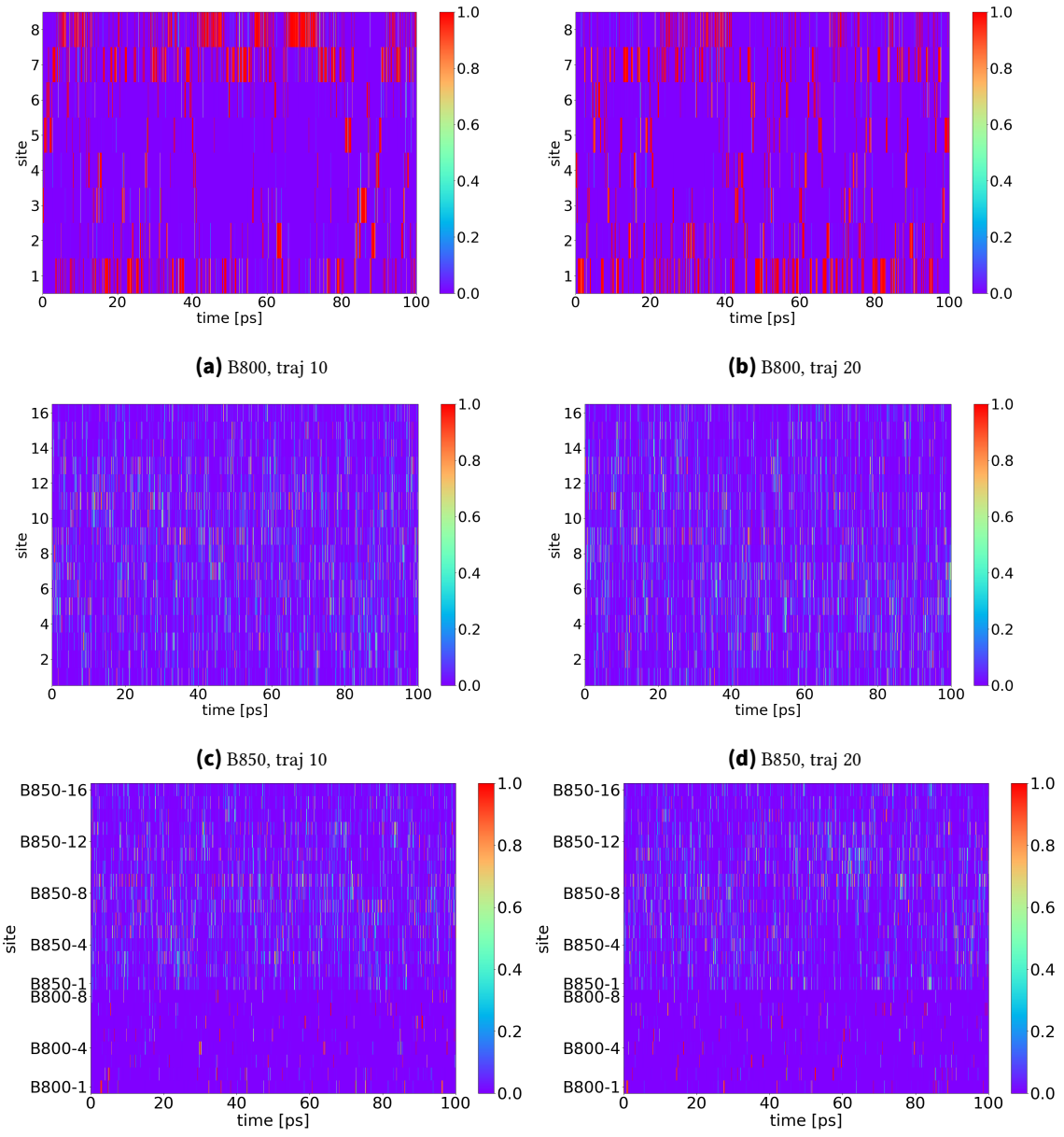


Fig. D.12.: Diabatic occupations of the pigments of the lowest eigenstate during the first 100 fs of one NAMD simulation together with the distribution of eigenvalues of the first two and last two adiabatic states, i.e. the adiabatic surfaces with the nth highest energy along the full simulation time of 5 ps, 1 ps, 10 ps, for B800, B850, LH2, respectively. The resulting squared eigenvectors of the lowest eigenvalue are shown in (a,c,e,g,i,k) and the histogram of the lowest two and highest two eigenvalues are shown in (b,d,f,h,j,l) for NAMD simulation 1 and 100.

NAMD simulations

Because of the computational efficiency of the method, the movement of the exciton was simulated for 100 ps.



(a) B800, traj 10

(b) B800, traj 20

(c) B850, traj 10

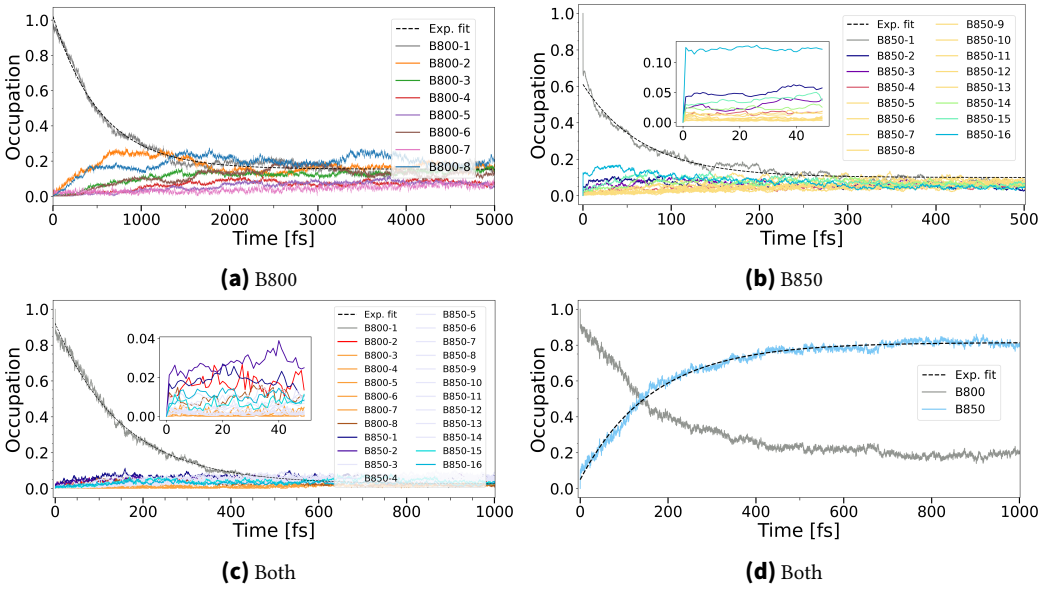
(d) B850, traj 20

(e) Both rings, traj 10

(f) Both rings, traj 20

Fig. D.13.: 100 ps long NAMD simulations of the exciton.

As it was done in the FMO protein, simulations with the implicit reorganization set to 65 meV have also been performed.



(a) B800 **(b)** B850 **(c)** Both **(d)** Both

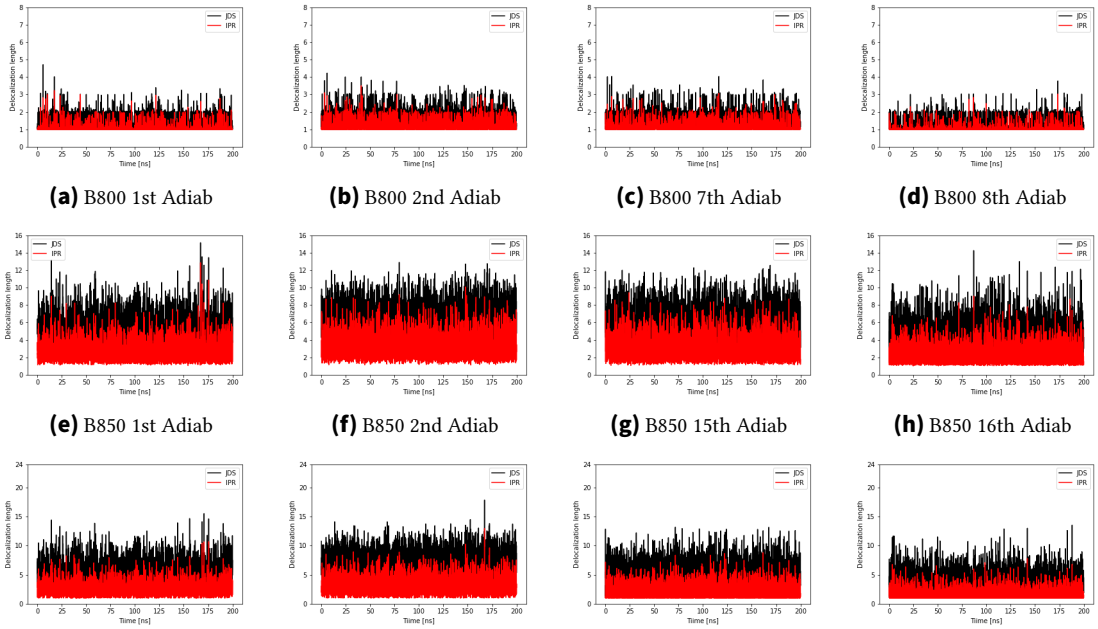
Fig. D.14.: Averaged diabatic occupation over 200 single NAMD simulations at 300 K with $\lambda = 65meV$. The evolution of occupations are shown for 5 ps, 500 fs, and 1 ps for B800, B850, and both rings, respectively. Exponential fits are shown in dotted lines. The B800→B850 transfer in both rings is illustrated for all pigments (c) and for all pigments within the B800 or B850 ring (d).

Delocalization

To compare the delocalization length of the different adiabatic states with each other, the IPR and JDS of the first two and last two adiabatic states are plotted for the 200 ns long MD simulation.

Tab. D.6.: Average delocalization lengths for the 200 NAMD simulations compared to the ones from the lowest adiabatic state in 40 ns long MD simulation. Systems B800, B850, LH2 consist of 8, 16 and 24 pigments, respectively. The ratio of occupied pigments $\frac{L}{n_{sites}}$ is also shown.

	NAMD $\lambda = 35meV$			NAMD $\lambda = 65meV$			40 ns MD		
	B800	B850	LH2	B800	B850	LH2	B800	B850	LH2
L_{IPR}	1.08	2.63	2.45	1.04	2.17	2.00	1.07	2.61	2.54
L_{JDS}	1.25	5.27	5.59	1.13	4.43	4.50	1.23	5.03	5.44
$\frac{L_{IPR}}{n_{sites}}$	14%	16%	10%	13%	14%	8%	13%	16%	11%
$\frac{L_{JDS}}{n_{sites}}$	16%	33%	23%	14%	28%	19%	14%	31%	23%



(a) B800 1st Adiab **(b)** B800 2nd Adiab **(c)** B800 7th Adiab **(d)** B800 8th Adiab

(e) B850 1st Adiab **(f)** B850 2nd Adiab **(g)** B850 15th Adiab **(h)** B850 16th Adiab

(i) LH2 1st Adiab **(j)** LH2 2nd Adiab **(k)** LH2 23th Adiab **(l)** LH2 24th Adiab

Fig. D.15.: Delocalization length for the 200 ns long MD simulation. For each of the 5000 frames, the Hamiltonians built up by the pigments of B800 (a,b,c,d), B850 (e,f,g,h) and the whole LH2 system (i,j,k,l) were diagonalized. The eigenvectors of the highest two and lowest two eigenstates were considered to calculate delocalization (first two and last two columns, respectively) as described in the computational details in the main text.

List of Figures

List of Tables

Bibliography

[1] M. J. Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1 (2015), pp. 19–25.

[2] Darius Abramavicius and Shaul Mukamel. “Energy-Transfer and Charge-Separation Pathways in the Reaction Center of Photosystem II Revealed by Coherent Two-Dimensional Optical Spectroscopy”. In: *J. Chem. Phys.* 133.18 (2010). ISSN: 00219606. DOI: 10.1063/1.3493580.

[3] J. Adolphs and T. Renger. “How Proteins Trigger Excitation Energy Transfer in the FMO Complex of Green Sulfur Bacteria”. In: *Biophys. J.* 91 (2006), pp. 2778–2797.

[4] Julia Adolphs and Thomas Renger. “How Proteins Trigger Excitation Energy Transfer in the FMO Complex of Green Sulfur Bacteria.” In: *Biophys. J.* 91.8 (2006), pp. 2778–2797. DOI: 10.1529/biophysj.105.079483.

[5] Mortaza Aghtar et al. “Juxtaposing Density Matrix and Classical Path-Based Wave Packet Dynamics”. In: *J. Chem. Phys.* 136.21 (2012), p. 214101. DOI: 10.1063/1.4723669.

[6] Hernan Alonso, Andrey A Bliznyuk, and Jill E Gready. “Combining docking and molecular dynamic simulations in drug design”. In: *Medicinal research reviews* 26.5 (2006), pp. 531–568.

[7] Andre Anda, Thorsten Hansen, and Luca De Vico. “Multireference Excitation Energies for Bacteriochlorophylls a within Light Harvesting System 2”. In: *J. Chem. Theory Comput.* 12 (2016). PMID: 26796483, pp. 1305–1313. DOI: 10.1021/acs.jctc.5b01104. eprint: http://dx.doi.org/10.1021/acs.jctc.5b01104.

[8] André Anda, Thorsten Hansen, and Luca De Vico. “Q_y and Q_x Absorption Bands for Bacteriochlorophyll a Molecules from LH2 and LH3”. In: *J. Phys. Chem. A* 123.25 (2019), pp. 5283–5292. ISSN: 1089-5639. DOI: 10.1021/acs.jpca.9b02877.

[9] Juan Aragó and Alessandro Troisi. “Regimes of exciton transport in molecular crystals in the presence of dynamic disorder”. In: *Advanced Functional Materials* 26.14 (2016), pp. 2316–2325.

[10] Hilke Bahmann and Martin Kaupp. “Efficient self-consistent implementation of local hybrid functionals”. In: *Journal of chemical theory and computation* 11.4 (2015), pp. 1540–1548.

[11] Juan A Ballesteros and Harel Weinstein. "[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors". In: *Methods in neurosciences*. Vol. 25. Elsevier, 1995, pp. 366–428.

[12] Albert P Bartók, Risi Kondor, and Gábor Csányi. "On representing chemical environments". In: *Physical Review B—Condensed Matter and Materials Physics* 87.18 (2013), p. 184115.

[13] Ilyes Batatia et al. "A foundation model for atomistic materials chemistry". In: *arXiv preprint arXiv:2401.00096* (2023).

[14] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". In: *J. Chem. Phys.* 98 (1993), p. 5648.

[15] Michael Becker, V Nagarajan, and William W Parson. "Properties of the excited singlet states of bacteriochlorophyll a and bacteriopheophytin a in polar solvents". In: *Journal of the American Chemical Society* 113.18 (1991), pp. 6840–6848.

[16] Lyudmila V Begunovich et al. "Absorption spectra of the purple nonsulfur bacteria light-harvesting complex: a DFT study of the B800 part". In: *Journal of Photochemistry and Photobiology A: Chemistry* 450 (2024), p. 115454.

[17] Jörg Behler and Michele Parrinello. "Generalized neural-network representation of high-dimensional potential-energy surfaces". In: *Physical review letters* 98.14 (2007), p. 146401.

[18] David N. Beratan et al. "Charge Transfer in Dynamical Biosystems, or The Treachery of (Static) Images". In: *Acc. Chem. Res.* 48.2 (2015), pp. 474–481. ISSN: 0001-4842. DOI: 10.1021/ar500271d.

[19] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation". In: *Comp. Phys. Comm.* 91 (1995), pp. 43–56.

[20] Peter H Berens and Kent R Wilson. "Molecular dynamics and spectra. I. Diatomic rotation and vibration". In: *The Journal of Chemical Physics* 74.9 (1981), pp. 4872–4882.

[21] Jochen Blumberger. "Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions". In: *Chem. Rev.* 115 (2015), pp. 11191–11238. DOI: 10.1021/acs.chemrev.5b00298.

[22] Juliana Bois and Thomas Körzdörfer. "How bond length alternation and thermal disorder affect the optical excitation energies of π-conjugated chains: a combined density functional theory and molecular dynamics study". In: *Journal of Chemical Theory and Computation* 12.4 (2016), pp. 1872–1882.

[23] B. M. Bold et al. "Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes". In: *Phys. Chem. Chem. Phys.* 22 (2020), pp. 10500–10518.

[24] B. M. Bold et al. “Correction: Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes”. In: *Phys. Chem. Chem. Phys.* 25 (2023), pp. 22535–22537.

[25] Beatrix Mirinda Bold et al. “Benchmark and Performance of Long-Range Corrected Time-Dependent Density Functional Tight Binding (LC-TD-DFTB) on Rhodopsins and Light-Harvesting Complexes”. In: *Phys. Chem. Chem. Phys.* 22 (2020), pp. 10500–10518. DOI: 10.1039/C9CP05753F.

[26] Max Born and W Heisenberg. “Zur quantentheorie der molekeln”. In: *Original Scientific Papers Wissenschaftliche Originalarbeiten* (1985), pp. 216–246.

[27] T. Brixner et al. “Two-dimensional spectroscopy of electronic couplings in photosynthesis”. In: *Nature* 434 (2005), pp. 625–628.

[28] Bernard R Brooks et al. “CHARMM: a program for macromolecular energy, minimization, and dynamics calculations”. In: *Journal of computational chemistry* 4.2 (1983), pp. 187–217.

[29] Bernard R Brooks et al. “CHARMM: the biomolecular simulation program”. In: *Journal of computational chemistry* 30.10 (2009), pp. 1545–1614.

[30] B. Brüggemann and V. May. “Exciton Exciton Annihilation Dynamics in Chromophore Complexes II. Intensity Dependent Transient Absorption of the LH2 Antenna System”. In: *J. Chem. Phys.* 120 (2004), p. 2325.

[31] Rebecca A Butcher. “Small-molecule pheromones and hormones controlling nematode development”. In: *Nature chemical biology* 13.6 (2017), pp. 577–586.

[32] M Büttiker and R Landauer. “Traversal time for tunneling”. In: *Physical Review Letters* 49.23 (1982), p. 1739.

[33] Carl Caleman et al. “Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant”. In: *Journal of chemical theory and computation* 8.1 (2012), pp. 61–74.

[34] Patricia M Callahan and Therese M Cotton. “Assignment of bacteriochlorophyll a ligation state from absorption and resonance Raman spectra”. In: *Journal of the American Chemical Society* 109.23 (1987), pp. 7001–7007.

[35] Jianshu Cao et al. “Quantum biology revisited”. In: *Science Advances* 6.14 (2020), eaaz4888. DOI: 10.1126/sciadv.aaz4888.

[36] Mark E Casida. “Time-dependent density functional response theory for molecules”. In: *Recent Advances In Density Functional Methods: (Part I)*. World Scientific, 1995, pp. 155–192.

[37] M. Ceccarelli, P. Procacci, and M. Marchi. “An Ab Initio Force Field for the Cofactors of Bacterial Photosynthesis”. In: *J. Comput. Chem.* 24 (2003), pp. 129–132. DOI: 10.1002/jcc.10198.

[38] Mirianas Chachisvilis et al. “Excitons in Photosynthetic Purple Bacteria: Wavelike Motion or Incoherent Hopping?” In: *J. Phys. Chem. B* 101.37 (1997), pp. 7275–7283. ISSN: 1520-6106. DOI: 10.1021/jp963360a.

[39] Marten L. Chaillet et al. “Static Disorder in Excitation Energies of the Fenna-Matthews-Olson Protein: Structure-Based Theory Meets Experiment”. In: *J. Phys. Chem. Lett.* (2020), pp. 10306–10314. DOI: 10.1021/acs.jpclett.0c03123.

[40] HC Stephen Chan et al. “New binding sites, new opportunities for GPCR drug discovery”. In: *Trends in biochemical sciences* 44.4 (2019), pp. 312–330.

[41] Danielle E. Chandler et al. “Light Harvesting by Lamellar Chromatophores in Rhodospirillum Photometricum”. In: *Biophys. J* 106 (2014), pp. 2503–2510. DOI: 10.1016/j.bpj.2014.04.030.

[42] Suryanarayanan Chandrasekaran et al. “Influence of force fields and quantum chemistry approach on spectral densities of BChl a in solution and in FMO proteins”. In: *The Journal of Physical Chemistry B* 119.31 (2015), pp. 9995–10004.

[43] J.-H. Chen et al. “Architecture of the photosynthetic complex from a green sulfur bacterium”. In: *Science* 370 (2020), pp. 931–938.

[44] Aurélia Chenu and Gregory D Scholes. “Coherence in energy transfer and photosynthesis”. In: *Annual review of physical chemistry* 66.1 (2015), pp. 69–96.

[45] E. Cignoni et al. “The Atomistic Modeling of Light-Harvesting Complexes from the Physical Models to the Computational Protocol”. In: *J. Chem. Phys.* 156 (2022), p. 120901.

[46] Edoardo Cignoni, Lorenzo Cupellini, and Benedetta Mennucci. “A Fast Method for Electronic Couplings in Embedded Multichromophoric Systems”. In: *J. Phys.: Condens. Matter* 34.30 (2022), p. 304004. DOI: 10.1088/1361-648x/ac6f3c.

[47] Edoardo Cignoni, Lorenzo Cupellini, and Benedetta Mennucci. “Machine Learning Exciton Hamiltonians in Light-Harvesting Complexes”. In: *J. Chem. Theory Comput.* 19 (2023), pp. 965–977. DOI: 10.1021/acs.jctc.2c01044.

[48] S. Ciuchi, S. Fratini, and D. Mayou. “Transient localization in crystalline organic semiconductors”. In: *Phys. Rev. B* 83 (2011), p. 081202.

[49] Liam Cleary et al. “Optimal Fold Symmetry of LH2 Rings on a Photosynthetic Membrane”. In: *Proc. Natl. Acad. Sci. USA* 110 (2013), pp. 8537–8542. DOI: 10.1073/pnas.1218270110.

[50] R. J. Cogdell, A. Gall, and J. Köhler. “The Architecture and Function of the Light-Harvesting Apparatus of Purple Bacteria: From Single Molecules to in Vivo Membranes”. In: *Q. Rev. Biophys.* 39 (2006), pp. 227–324. DOI: 10.1017/s0033583506004434.

[51] R. J. Cogdell et al. “The Structural Basis of Light-Harvesting in Purple Bacteria”. In: *FEBS Lett.* 5555 (2003), pp. 35–39. DOI: 10.1016/s0014-5793(03)01102-5.

[52] UniProt Consortium. “UniProt: a worldwide hub of protein knowledge”. In: *Nucleic acids research* 47.D1 (2019), pp. D506–D515.

[53] Irene Conti et al. “Ultrafast spectroscopy of photoactive molecular systems from first principles: Where we stand today and where we are going”. In: *Journal of the American Chemical Society* 142.38 (2020), pp. 16117–16139.

[54] Marshall G. Cory et al. “Electronic Excitations in Aggregates of Bacteriochlorophylls”. In: *J. Phys. Chem. B* 102.39 (1998), pp. 7640–7650. ISSN: 1520-6106. DOI: 10.1021/jp980966l.

[55] Christopher J Cramer. *Essentials of computational chemistry: theories and models*. John Wiley & Sons, 2013.

[56] Roberta Croce et al. *Light harvesting in photosynthesis*. CRC press, 2018.

[57] Lorenzo Cupellini and Filippo Lipparini. *FCE Program to Compute Optical Spectra with the Full Cumulant Expansion*. OpenAIRE, DOI: 10.5281/ZENODO.3900199. 2020. DOI: 10.5281/ZENODO.3900199.

[58] Lorenzo Cupellini, Filippo Lipparini, and Jianshu Cao. “Absorption And Circular Dichroism Spectra Of Molecular Aggregates With The Full Cumulant Expansion”. In: *J. Phys. Chem. B* 124 (2020), pp. 8610–8617. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.0c05180.

[59] Lorenzo Cupellini et al. “An Ab Initio Description of the Excitonic Properties of LH2 and Their Temperature Dependence”. In: *J. Phys. Chem. B* 120 (2016), pp. 11348–11359. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.6b06585.

[60] Lorenzo Cupellini et al. “Coupling to Charge Transfer States Is the Key to Modulate the Optical Bands for Efficient Light-harvesting in Purple Bacteria.” In: *J. Phys. Chem. Lett.* 9 (2018), pp. 6892–6899. DOI: 10.1021/acs.jpclett.8b03233.

[61] Lorenzo Cupellini et al. “Quantum chemical elucidation of a sevenfold symmetric bacterial antenna complex”. In: *Photosynthesis Research* 156.1 (2023), pp. 75–87.

[62] Basile FE Curchod and Todd J Martínez. “Ab initio nonadiabatic quantum molecular dynamics”. In: *Chemical reviews* 118.7 (2018), pp. 3305–3336.

[63] C. Curutchet and B. Mennucci. “Quantum Chemical Studies of Light Harvesting”. In: *Chem. Rev.* 117 (2017), pp. 294–343.

[64] M Dahlbom et al. “Exciton delocalization in the B850 light-harvesting complex: comparison of different measures”. In: *The Journal of Physical Chemistry B* 105.23 (2001), pp. 5515–5524.

[65] A. Damjanović et al. “Excitons in a Photosynthetic Light-Harvesting System: A Combined Molecular Dynamics, Quantum Chemistry and Polaron Model Study”. In: *Phys. Rev. E* 65 (2002), p. 031919. DOI: 10.1103/PhysRevE.65.031919.

[66] A. Damjanović et al. “Excitons in a photosynthetic light-harvesting system: A combined molecular dynamics, quantum chemistry, and polaron model study”. In: *Phys. Rev. E* 65 (2002), p. 031919.

[67] Luca De Vico et al. “Macrocycle ring deformation as the secondary design principle for light-harvesting complexes”. In: *Proceedings of the National Academy of Sciences* 115.39 (2018), E9051–E9057.

[68] Philipp M Dohmen et al. “Modeling Charge Transport in Organic Semiconductors Using Neural Network Based Hamiltonians and Forces”. In: *Journal of Chemical Theory and Computation* 19.13 (2023), pp. 3825–3838.

[69] J. Dostál, J. Pšenčík, and D. Zigmantas. “In situ mapping of the energy flow through the entire photosynthetic apparatus”. In: *Nat. Chem.* 8 (2016), pp. 705–710.

[70] H.-G. Duan et al. “Nature does not rely on long-lived electronic quantum coherence for photosynthetic energy transfer”. In: *Proc. Natl. Acad. Sci.* 114.32 (2017), pp. 8493–8498.

[71] H.-G. Duan et al. “Quantum coherent energy transport in the Fenna-Matthews-Olson complex at low temperature”. In: *Proc. Natl. Sci. U.S.A.* 119.49 (2022), p. 2212630119.

[72] Hong-Guang Duan et al. “Quantum Coherent Energy Transport in the Fenna-Matthews-Olson Complex at Low Temperature”. In: *Proc. Natl. Acad. Sci.* 119.49 (2022). DOI: 10.1073/pnas.2212630119.

[73] Jerome Eberhardt et al. “AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings”. In: *Journal of chemical information and modeling* 61.8 (2021), pp. 3891–3898.

[74] Marcus Elstner. “The SCC-DFTB method and its application to biological systems”. In: *Theoretical Chemistry Accounts* 116 (2006), pp. 316–325.

[75] Marcus Elstner and Gotthard Seifert. “Density functional tight binding”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2011 (2014), p. 20120483.

[76] Marcus Elstner et al. “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”. In: *Physical Review B* 58.11 (1998), p. 7260.

[77] Vesna Erić et al. “Observation of Dark States in Two-Dimensional Electronic Spectra of Chlorosomes”. In: *J. Phys. Chem. B* 128 (2024), pp. 3575–3584. DOI: 10.1021/acs.jpcb.4c00067.

[78] Vesna Erić et al. “Ultrafast Anisotropy Decay Reveals Structure and Energy Transfer in Supramolecular Aggregates”. In: *J. Phys. Chem. B* 127 (2023), pp. 7487–7496. DOI: 10.1021/acs.jpcb.3c04719.

[79] Francesca Fassioli et al. “Photosynthetic Light Harvesting: Excitons and Coherence”. In: *J. Roy. Soc. Interface.* 11.92 (2014), p. 20130901. DOI: 10.1098/rsif.2013.0901.

[80] Kevin M Felter et al. “Interplay between charge carrier mobility, exciton diffusion, crystal packing, and charge separation in perylene diimide-based heterojunctions”. In: *ACS Applied Energy Materials* 2.11 (2019), pp. 8010–8021.

[81] R. E. Fenna and B. W. Matthews. “Chlorophyll Arrangement in a Bacteriochlorophyll Protein from *Chlorobium Limicola*”. In: *Nature* 258.5536 (1975), p. 573. DOI: 10.1038/258573a0.

[82] Reinhard Fischer and Natalia Requena. “Small-secreted proteins as virulence factors in nematode-trapping fungi”. In: *Trends in Microbiology* 30.7 (2022), pp. 615–617.

[83] Xavier Fradera, Ronald MA Knegtel, and Jordi Mestres. "Similarity-driven flexible ligand docking". In: *Proteins: Structure, Function, and Bioinformatics* 40.4 (2000), pp. 623–636.

[84] E. M. Franken et al. "A Permanent Hole Burning Study of the FMO Antenna Complex of the Green Sulfur Bacterium *Prosthecochloris aestuarii*". In: *Biochemistry* 37 (1998), pp. 5046–5051.

[85] Simone Fratini et al. "A map of high-mobility molecular semiconductors". In: *Nature materials* 16.10 (2017), pp. 998–1002.

[86] M. J. Frisch et al. *Gaussian 09, Revision A.02*. 2009.

[87] Kazuhiro J Fujimoto, Takumi Minoda, and Takeshi Yanai. "Spectral tuning mechanism of photosynthetic light-harvesting complex II revealed by ab initio dimer exciton model". In: *The Journal of Physical Chemistry B* 125.37 (2021), pp. 10459–10470.

[88] Alastair T. Gardiner et al. "The 2.4 Å Cryo-Em Structure of a Heptameric Light-Harvesting 2 Complex Reveals Two Carotenoid Energy Transfer Pathways". In: *Sci. Adv.* 7 (2021). DOI: 10.1126/sciadv.abe4650.

[89] Johann Gasteiger and Mario Marsili. "Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges". In: *Tetrahedron* 36.22 (1980), pp. 3219–3228.

[90] M. Gaus, A. Goez, and M. Elstner. "Parametrization and Benchmark of DFTB3 for Organic Molecules". In: *J. Chem. Theory Comput.* 9 (2013), pp. 338–354.

[91] Michael Gaus, Qiang Cui, and Marcus Elstner. "DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)". In: *J. Chem. Theory Comput.* 7.4 (2011), pp. 931–948. ISSN: 1549-9618. DOI: 10.1021/ct100684s.

[92] Michael Gaus, Albrecht Goez, and Marcus Elstner. "Parametrization and Benchmark of DFTB3 for Organic Molecules". In: *J. Chem. Theory Comput.* 9.1 (2013), pp. 338–354. DOI: 10.1021/ct300849w.

[93] S. Georgakopoulou et al. "Absorption and CD Spectroscopy and Modeling of Various LH2 Complexes from Purple Bacteria". In: *Biophys. J.* 82 (2002), pp. 2184–2197. DOI: 10.1016/S0006-3495(02)75565-3.

[94] Farhad Ghalami et al. "Nonadiabatic Simulation of Exciton Dynamics in Organic Semiconductors Using Neural Network-Based Frenkel Hamiltonian and Gradients". In: *Journal of Chemical Theory and Computation* (2024).

[95] S. Giannini and J. Blumberger. "Charge Transport in Organic Semiconductors: The Perspective from Nonadiabatic Molecular Dynamics". In: *Acc. Chem. Res.* 55 (2022), pp. 819–830.

[96] S. Giannini et al. "Exciton transport in molecular organic semiconductors boosted by transient quantum delocalization". In: *Nat. Commun.* 13 (2022), p. 2755.

[97] Samuele Giannini et al. “Quantum localization and delocalization of charge carriers in organic semiconducting crystals”. In: *Nature communications* 10.1 (2019), p. 3843.

[98] Colm G. Gillis and Garth A. Jones. “A Theoretical Investigation into the Effects of Temperature on Spatiotemporal Dynamics of EET in the FMO Complex”. In: *J. Phys. Chem. B* 119.11 (2015), pp. 4165–4174. ISSN: 1520-6106. DOI: 10.1021/jp509103e.

[99] Saulius Gražulis et al. “Crystallography Open Database–an open-access collection of crystal structures”. In: *Applied Crystallography* 42.4 (2009), pp. 726–729.

[100] Rienk van Grondelle and Vladimir I. Novoderezhkin. “Energy Transfer in Photosynthesis: Experimental Insights and Quantitative Models”. In: *Phys. Chem. Chem. Phys.* 8.7 (2006), pp. 793–807. ISSN: 1463-9076. DOI: 10.1039/B514032C.

[101] Thomas A Halgren and Wolfgang Damm. “Polarizable force fields”. In: *Current opinion in structural biology* 11.2 (2001), pp. 236–242.

[102] Inbal Halperin et al. “Principles of docking: An overview of search algorithms and a guide to scoring functions”. In: *Proteins: Structure, Function, and Bioinformatics* 47.4 (2002), pp. 409–443.

[103] Janna Hastings et al. “ChEBI in 2016: Improved services and an expanding collection of metabolites”. In: *Nucleic acids research* 44.D1 (2016), pp. D1214–D1219.

[104] D. Hayes and G. S. Engel. “Extracting the Excitonic Hamiltonian of the Fenna-Matthews-Olson Complex Using Three-dimensional Third-Order Electronic Spectroscopy”. In: *Biophys. J.* 100 (2011), pp. 2043–2052. DOI: 10.1016/j.bpj.2010.12.3747.

[105] Alexander Heck et al. “Multi-Scale Approach to Non-Adiabatic Charge Transport in High-Mobility Organic Semiconductors”. In: *J. Chem. Theory Comput.* 11.11 (2015), pp. 5068–5082. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.5b00719.

[106] J. S. Higgins et al. “Photosynthesis tunes quantum-mechanical mixing of electronic and vibrational states to steer exciton energy transfer”. In: *Proc. Natl. Sci. U.S.A.* 118.11 (2021), p. 2018240118.

[107] S. Höfener. “Coupled-cluster frozen-density embedding using resolution of the identity methods”. In: *J. Comput. Chem.* 35.23 (2014), pp. 1716–1724.

[108] Pierre Hohenberg and Walter Kohn. “Inhomogeneous electron gas”. In: *Physical review* 136.3B (1964), B864.

[109] J Michael Hollas. *Modern spectroscopy*. John Wiley & Sons, 2004.

[110] Yannick Holtkamp et al. “Machine-Learned Correction to Ensemble-Averaged Wave Packet Dynamics”. In: *J. Chem. Phys.* 159 (2023), p. 094107. DOI: 10.1063/5.0166694.

[111] W. G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Phys. Rev. A* 31.3 (1985), pp. 1695–1697.

[112] Viktor Hornak et al. “Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters”. In: *Proteins: Struct., Funct., Bioinf.* 65.3 (2006), pp. 712–725.

[113] B. Hourahine et al. “DFTB+, a Software Package for Efficient Approximate Density Functional Theory Based Atomistic Simulations”. In: *J. Chem. Phys.* 152.12 (2020), p. 124101. ISSN: 0021-9606. DOI: 10.1063/1.5143190.

[114] Yen-Ping Hsueh et al. “Nematode-trapping fungi eavesdrop on nematode pheromones”. In: *Current Biology* 23.1 (2013), pp. 83–86.

[115] X. Hu et al. “Photosynthetic Apparatus of Purple Bacteria”. In: *Q. Rev. Biophys.* 35 (2002), pp. 1–62. DOI: 10.1017/s0033583501003754.

[116] Jinrong Huang et al. “Ammonia and nematode ascaroside are synergistic in trap formation in Arthrobotrys oligospora”. In: *Pathogens* 12.9 (2023), p. 1114.

[117] W. F. Humphrey, A. Dalke, and K. Schulten. “VMD – Visual Molecular Dynamics”. In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38. DOI: 10.1016/0263-7855(96)00018-5.

[118] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD: visual molecular dynamics”. In: *Journal of molecular graphics* 14.1 (1996), pp. 33–38.

[119] Medhat Ibrahim et al. “Analysis of the structure and vibrational spectra of glucose and fructose”. In: *Ecletica quimica* 31 (2006), pp. 15–21.

[120] Janne A. Ihalainen et al. “Energy Transfer in LH2 of *R*hodospirillum Molischianum, Studied by Subpicosecond Spectroscopy and Configuration Interaction Exciton Calculations”. In: *J. Phys. Chem. B* 105.40 (2001), pp. 9849–9856. ISSN: 1520-6106. DOI: 10.1021/jp010921b.

[121] Vignir Isberg et al. “GPCRDB: an information system for G protein-coupled receptors”. In: *Nucleic acids research* 42.D1 (2014), pp. D422–D425.

[122] A. Ishizaki et al. “Quantum Coherence and Its Interplay with Protein Environments in Photosynthetic Electronic Energy Transfer”. In: *Phys. Chem. Chem. Phys.* 12 (2010), pp. 7319–7337. DOI: 10.1039/C003389H.

[123] Akihito Ishizaki and Graham R Fleming. “Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature”. In: *Proceedings of the National Academy of Sciences* 106.41 (2009), pp. 17255–17260.

[124] Ajay N Jain. “Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities”. In: *Journal of computer-aided molecular design* 10 (1996), pp. 427–440.

[125] S. Jang, S. E. Dempster, and R. J. Silbey. “Characterization of the Static Disorder in the B850 Band of LH2”. In: *J. Phys. Chem. B* 105 (2001), pp. 6655–6665. DOI: 10.1021/jp010169e.

[126] Seogjoo Jang, Marshall D. Newton, and Robert J. Silbey. “Multichromophoric Förster Resonance Energy Transfer from B800 to B850 in the Light Harvesting Complex 2: Evidence for Subtle Energetic Optimization by Purple Bacteria”. In: *J. Phys. Chem. B* 111 (2007), pp. 6807–6814. DOI: 10.1021/jp070111l.

[127] Seogjoo J. Jang and Benedetta Mennucci. “Delocalized Excitons in Natural Light-Harvesting Complexes”. In: *Rev. Mod. Phys.* 90 (3 2018), p. 035003. DOI: 10.1103/RevModPhys.90.035003.

[128] T. L. C. Jansen and J. Knoester. "Nonadiabatic Effects in the Two-Dimensional Infrared Spectra of Peptides: Application to Alanine Dipeptide". In: *J. Phys. Chem. B* 110 (2006), pp. 22910–22916. DOI: 10.1021/jp064795t.

[129] Thomas L. C. Jansen. "Simple Quantum Dynamics with Thermalization". In: *J. Phys. Chem. A* 122.1 (2018), pp. 172–183. ISSN: 1089-5639. DOI: 10.1021/acs.jpca.7b10380.

[130] John M Jean et al. "Excitation transport and trapping on spectrally disordered lattices". In: *Biophysical journal* 56.6 (1989), pp. 1203–1215.

[131] F. Jensen. *Introduction to Computational Chemistry*. 2nd. West Sussex: Wiley, 2007.

[132] Frank Jensen. *Introduction to computational chemistry*. John wiley & sons, 2017.

[133] X. Jia et al. "Hybrid QM/MM study of FMO complex with polarized protein-specific charge". In: *Sci. Rep.* 5 (2015), p. 17096.

[134] R. Jimenez et al. "Three-Pulse Photon Echo Measurements on LH1 and LH2 Complexes of *Rhodobacter sphaeroides*: A Nonlinear Spectroscopic Probe of Energy Transfer". In: *J. Phys. Chem. B* 101 (1997), pp. 7350–7359.

[135] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *nature* 596.7873 (2021), pp. 583–589.

[136] S. Jurinovich, C. Curutchet, and B. Mennucci. "The Fenna-Matthews-Olson Protein Revisited: A Fully Polarizable (TD)DFT/MM Description". In: *ChemPhysChem* 15 (2014), pp. 3194–3204.

[137] Ephraim Katchalski-Katzir et al. "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques." In: *Proceedings of the National Academy of Sciences* 89.6 (1992), pp. 2195–2199.

[138] Erika Keil et al. "Light Harvesting in Purple Bacteria Does Not Rely on Resonance Fine-Tuning in Peripheral Antenna Complexes". In: *Photosynth. Res.* (2024). DOI: 10.1007/s11120-024-01107-4.

[139] John A Keith et al. "Combining machine learning and computational chemistry for predictive insights into chemical systems". In: *Chemical reviews* 121.16 (2021), pp. 9816–9872.

[140] U. Kleinekathöfer et al. "Memory Effects in the Fluorescence Depolarization Dynamics Studied Within the B850 Ring of Purple Bacteria". In: *J. Phys. Chem. B* 107 (2003), pp. 14094–14102. DOI: 10.1021/jp035801d.

[141] R. S. Knox and B. Q. Spring. "Dipole Strengths in the Chlorophylls". In: *Photochem. Photobiol.* 77.5 (2003), pp. 497–501.

[142] Robert S Knox and Bryan Q Spring. "Dipole Strengths in the Chlorophylls". In: *Photochem. Photobiol.* 77.5 (2003), pp. 497–501. DOI: 10.1562/0031-8655(2003)0770497DSITC2.0.CO2.

[143] Wolfram Koch and Max C Holthausen. *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[144] Julia Koehler Leman et al. “Sequence-structure-function relationships in the microbial protein universe”. In: *Nature communications* 14.1 (2023), p. 2351.

[145] J. Koepke et al. “The Crystal Structure of the Light Harvesting Complex II (B800-850) from *Rhodospirillum molischianum*”. In: *Structure* 4 (1996), pp. 581–597. DOI: 10.1016/s0969-2126(96)00063-9.

[146] Walter Kohn and Lu Jeu Sham. “Self-consistent equations including exchange and correlation effects”. In: *Physical review* 140.4A (1965), A1133.

[147] Dávid Péter Kovács et al. “Mace-off: Short-range transferable machine learning force fields for organic molecules”. In: *Journal of the American Chemical Society* 147.21 (2025), pp. 17598–17611.

[148] J. J. Kranz and M. Elstner. “Simulation of Singlet Exciton Diffusion in Bulk Organic Materials”. In: *J. Chem. Theory Comput.* 12.9 (2016), pp. 4209–4221.

[149] J. J. Kranz et al. “Time-dependent extension of the long-range corrected density functional based tight-binding method”. In: *J. Chem. Theory Comput.* 13.4 (2017), pp. 1737–1747.

[150] Daniela Kruschel and Bojan Zagrovic. “Conformational averaging in structural biology: issues, challenges and computational solutions”. In: *Molecular Biosystems* 5.12 (2009), pp. 1606–1616.

[151] Andriy Kryshtafovych et al. “Critical assessment of methods of protein structure prediction (CASP)—Round XIV”. In: *Proteins: Structure, Function, and Bioinformatics* 89.12 (2021), pp. 1607–1617.

[152] T. Kubař and M. Elstner. “Coarse-Grained Time-Dependent Density Functional Simulation of Charge Transfer in Complex Systems: Application to Hole Transfer in DNA”. In: *J. Phys. Chem. B* 114 (2010), pp. 11221–11240.

[153] Tomáš Kubař and Marcus Elstner. “Efficient Algorithms for the Simulation of Non-Adiabatic Electron Transfer in Complex Molecular Systems: Application to DNA”. In: *Phys. Chem. Chem. Phys.* 15.16 (2013), pp. 5794–5813. DOI: 10.1039/c3cp44619k.

[154] Chawntell Kulkarni et al. “Theoretical study of the influence of the photosynthetic membrane on B800-B850 energy transfer within the peripheral light-harvesting complex LH2”. In: *arXiv preprint arXiv:2407.12591* (2024).

[155] Sohang Kundu, Reshmi Dani, and Nancy Makri. “B800-to-B850 Relaxation of Excitation Energy in Bacterial Light Harvesting: All-State, All-Mode Path Integral Simulations”. In: *J. Chem. Phys.* 157.1 (2022), p. 015101. DOI: 10.1063/5.0093828.

[156] Sohang Kundu and Nancy Makri. “Real-Time Path Integral Simulation of Exciton-Vibration Dynamics in Light-Harvesting Bacteriochlorophyll Aggregates”. In: *J. Phys. Chem. Lett.* 11 (2020), pp. 8783–8789. DOI: 10.1021/acs.jpclett.0c02760.

[157] David N LeBard and Dmitry V Matyushov. “Glassy protein dynamics and gigantic solvent reorganization energy of plastocyanin”. In: *The Journal of Physical Chemistry B* 112.16 (2008), pp. 5218–5227.

[158] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation energy formula into a functional of the electron density". In: *Phys. Rev. B* 37 (1988), p. 785.

[159] Sumin Lee et al. "Evaluating GPCR modeling and docking strategies in the era of deep learning-based protein structure prediction". In: *Computational and Structural Biotechnology Journal* 21 (2023), pp. 158–167.

[160] Igor V. Leontyev and Alexei A. Stuchebrukhov. "Accounting for Electronic Polarization in Non-polarizable Force Fields". In: *Phys. Chem. Chem. Phys.* 13.7 (2011), pp. 2613–2626. DOI: 10.1039/c0cp01971b.

[161] D Leupold et al. "Size enhancement of transition dipoles to one-and two-exciton bands in a photosynthetic antenna". In: *Physical review letters* 77.22 (1996), p. 4675.

[162] Yi-Fen Li et al. "Crystal structure of the bacteriochlorophyll a protein from Chlorobium tepidum". In: *Journal of molecular biology* 271.3 (1997), pp. 456–471.

[163] Pengfei Li and Kenneth M Merz Jr. "Taking into account the ion-induced dipole interaction in the nonbonded model of ions". In: *Journal of chemical theory and computation* 10.1 (2014), pp. 289–297.

[164] Quanwei Li et al. "Single-Photon Absorption and Emission from a Natural Photosynthetic Complex". In: *Nature* 619.7969 (2023), pp. 300–304. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06121-5.

[165] J. Linnanto and J. Korppi-Tommola. "Quantum Chemical Simulation of Excited States of Chlorophylls, Bacteriochlorophylls and Their Complexes". In: *Phys. Chem. Chem. Phys.* 8 (2006), pp. 663–667. DOI: 10.1039/b513086g.

[166] J. Linnanto, J. E. I. Korppi-Tommola, and V.M. Helenius. "Electronic States, Absorption Spectrum and Circular Dichroism Spectrum of the Photosynthetic Bacterial LH2 Antenna of *Rhodopseudomonas acidophila* As Predicted by Exciton Theory and Semiempirical Calculations". In: *J. Phys. Chem. B* 103 (1999), pp. 8739–8750. DOI: 10.1021/jp9848344.

[167] Nanna Holmgaard List et al. "Toward Reliable Prediction of the Energy Ladder in Multichromophoric Systems: A Benchmark Study on the FMO Light-Harvesting Complex". In: *J. Chem. Theory Comput.* 9 (2013), pp. 4928–4938. ISSN: 1549-9618. DOI: 10.1021/ct400560m.

[168] Alexander Löhner, Richard Cogdell, and Jürgen Köhler. "Contribution of Low-Temperature Single-Molecule Techniques to Structural Issues of Pigment–Protein Complexes from Photosynthetic Purple Bacteria". In: *J. Roy. Soc. . Interface* 15.138 (2018), p. 20170680. ISSN: 1742-5662. DOI: 10.1098/rsif.2017.0680.

[169] Dongyu Lyu et al. "Excitation Energy Transfer between Porphyrin Dyes on a Clay Surface: A Study Employing Multifidelity Machine Learning". In: *Advanced Theory and Simulations* (2024), e00271.

[170] Jian Ma and Jianshu Cao. "Förster Resonance Energy Transfer, Absorption and Emission Spectra in Multichromophoric Systems. I. Full Cumulant Expansions and System-Bath Entanglement". In: *J. Chem. Phys.* 142.9 (2015), p. 094106. DOI: 10.1063/1.4908599.

[171] M. E. Madjet, A. Abdurahman, and T. Renger. "Intermolecular Coulomb Couplings from Ab Initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers". In: *J. Phys. Chem. B* 110 (2006), pp. 17268–17281.

[172] S. Maity et al. "DFTB/MM Molecular Dynamics Simulations of the FMO Light-Harvesting Complex". In: *J. Phys. Chem. Lett.* 11 (2020), pp. 8660–8667.

[173] Sayan Maity et al. "DFTB/MM Molecular Dynamics Simulations of the FMO Light-Harvesting Complex". In: *J. Phys. Chem. Lett.* 11 (2020), pp. 8660–8667. DOI: 10.1021/acs.jpclett.0c02526.

[174] Sayan Maity et al. "Electric field susceptibility of chlorophyll c leads to unexpected excitation dynamics in the major light-harvesting complex of diatoms". In: *The Journal of Physical Chemistry Letters* 15.9 (2024), pp. 2499–2510.

[175] Maria Ilaria Mallus et al. "Environmental Effects on the Dynamics in the Light-Harvesting Complexes LH2 and LH3 Based on Molecular Simulations". In: *Chem. Phys.* 515 (2018), pp. 141–151. DOI: 10.1016/j.chemphys.2018.08.013.

[176] WG Mäntele et al. "Infrared spectroelectrochemistry of bacteriochlorophylls and bacteriopheophytins: Implications for the binding of the pigments in the reaction center from photosynthetic bacteria". In: *Proceedings of the National Academy of Sciences* 85.22 (1988), pp. 8468–8472.

[177] Seth R Marder et al. "Relation between bond-length alternation and second electronic hyperpolarizability of conjugated organic molecules". In: *Science* 261.5118 (1993), pp. 186–189.

[178] Valerio Mariani et al. "lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests". In: *Bioinformatics* 29.21 (2013), pp. 2722–2728.

[179] D. V. Matyushov. "Reorganization energy of electron transfer". In: *Phys. Chem. Chem. Phys.* 25 (2023), pp. 7589–7610.

[180] Dmitry V Matyushov. "Protein electron transfer: Dynamics and statistics". In: *The Journal of Chemical Physics* 139.2 (2013).

[181] Dmitry V Matyushov and Gregory A Voth. "Reorganization parameters of electronic transitions in electronically delocalized systems. 2. Optical spectra". In: *The Journal of Physical Chemistry A* 104.27 (2000), pp. 6485–6494.

[182] Volkhard May and Oliver Kühn. *Charge and energy transfer dynamics in molecular systems*. John Wiley & Sons, 2023.

[183] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. "Dynamics of folded proteins". In: *nature* 267.5612 (1977), pp. 585–590.

[184] G. McDermott et al. “Crystal Structure of an Integral Membrane Light-Harvesting Complex from Photosynthetic Bacteria”. In: *Nature* 374 (1995), pp. 517–521. DOI: 10.1038/374517a0.

[185] K. McLuskey et al. “The Crystallographic Structure of the B800-820 LH3 Light-Harvesting Complex from the Purple Bacteria *Rhodopseudomonas Acidophila* Strain 7050”. In: *Biochemistry* 40.30 (2001), pp. 8783–8789. ISSN: 0006-2960. DOI: 10.1021/bi010309a.

[186] Naveen Michaud-Agrawal et al. “MDAnalysis: a toolkit for the analysis of molecular dynamics simulations”. In: *Journal of computational chemistry* 32.10 (2011), pp. 2319–2327.

[187] Milot Mirdita et al. “ColabFold: making protein folding accessible to all”. In: *Nature methods* 19.6 (2022), pp. 679–682.

[188] T. Mirkovic et al. “Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms”. In: *Chem. Rev.* 117.2 (2017), pp. 249–293. DOI: 10.1021/acs.chemrev.6b00002.

[189] J. M. Moix, Y. Zhao, and J. Cao. “Equilibrium-Reduced Density Matrix Formulation: Influence of Noise, Disorder, and Temperature on Localization in Excitonic Systems”. In: *Phys. Rev. B* 85 (11 2012), p. 115412. DOI: 10.1103/PhysRevB.85.115412.

[190] René Monshouwer et al. “Superradiance and exciton delocalization in bacterial photosynthetic light-harvesting systems”. In: *The Journal of Physical Chemistry B* 101.37 (1997), pp. 7241–7248.

[191] Shaul Mukamel. “Fluorescence and absorption of large anharmonic molecules-spectroscopy without eigenstates”. In: *The Journal of Physical Chemistry* 89.7 (1985), pp. 1077–1087.

[192] Shaul Mukamel. *Principles of nonlinear optical spectroscopy*. Vol. 6. Oxford university press New York, 1995.

[193] P. Nalbach, D. Braun, and M. Thorwart. “Exciton transfer dynamics and quantumness of energy transfer in the Fenna-Matthews-Olson complex”. In: *Phys. Rev. E* 84 (2011), p. 041926.

[194] F. Neese. “Software update: the ORCA program system, version 5.0”. In: *WIRES Comput. Molec. Sci.* 12.1 (2022), e1606. DOI: 10.1002/wcms.1606.

[195] Frank Neese. “The ORCA program system”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2.1 (2012), pp. 73–78.

[196] Tahereh Nematiaram et al. “Practical computation of the charge mobility in molecular semiconductors using transient localization theory”. In: *The Journal of Physical Chemistry C* 123.12 (2019), pp. 6989–6997.

[197] Dariusz M Niedzwiedzki and Robert E Blankenship. “Singlet and triplet excited state properties of natural chlorophylls and bacteriochlorophylls”. In: *Photosynthesis research* 106.3 (2010), pp. 227–238.

[198] Thomas A Niehaus et al. “Tight-binding approach to time-dependent density-functional response theory”. In: *Physical Review B* 63.8 (2001), p. 085108.

[199] Michael A Nielsen. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, USA, 2015.

[200] Raquel Norel et al. “Electrostatic contributions to protein–protein interactions: fast energetic filters for docking and their physical basis”. In: *Protein Science* 10.11 (2001), pp. 2147–2161.

[201] S. Nosé. “A molecular dynamics method for simulations in the canonical ensemble”. In: *Mol. Phys.* 52.2 (1984), pp. 255–268.

[202] S. Nosé and M. L. Klein. “Constant pressure molecular dynamics for molecular systems”. In: *Mol. Phys.* 50.5 (1983), pp. 1055–1076.

[203] V. Novoderezhkin, M. Wendling, and R. van Grondelle. “Intra- and Interband Transfers in the B800-B850 Antenna of *R*hodospirillum molischianum: Redfield Theory Modeling of Polarized Pump-probe Kinetics”. In: *J. Phys. Chem. B* 107 (2003), pp. 11534–11548. DOI: 10.1021/jp035432l.

[204] V. I. Novoderezhkin. “Combined Förster-Redfield Theory for Modeling Energy Transfer in Plant Photosynthetic Antenna Complexes”. In: *Biochemistry* 6.4 (2012), pp. 314–319. ISSN: 1990-7494. DOI: 10.1134/s199074781204006x.

[205] V. I. Novoderezhkin et al. “Energy-Transfer Dynamics in the LHCII Complex of Higher Plants: Modified Redfield Approach”. In: *J. Phys. Chem. B* 108 (2004), p. 10363. ISSN: 1520-6106. DOI: 10.1021/jp0496001.

[206] Vladimir Novoderezhkin and Rienk van Grondelle. “Spectra and Dynamics in the B800 Antenna: Comparing Hierarchical Equations, Redfield and Förster Theories”. In: *J. Phys. Chem. B* 117.38 (2013), pp. 11076–11090. ISSN: 1520-6106. DOI: 10.1021/jp400957t.

[207] Vladimir I. Novoderezhkin, Danielis Rutkauskas, and Rienk van Grondelle. “Dynamics of the Emission Spectrum of a Single LH2 Complex: Interplay of Slow and Fast Nuclear Motions”. In: *Biophys. J.* 90.8 (2006), pp. 2890–2902. ISSN: 0006-3495. DOI: 10.1529/biophysj.105.072652.

[208] C. Olbrich and U. Kleinekathöfer. “Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II”. In: *J. Phys. Chem. B* 114 (2010), pp. 12427–12437. DOI: 10.1021/jp106542v.

[209] Carsten Olbrich and Ulrich Kleinekathöfer. “Time-dependent atomistic view on the electronic relaxation in light-harvesting system II”. In: *The Journal of Physical Chemistry B* 114.38 (2010), pp. 12427–12437.

[210] Carsten Olbrich et al. “From atomistic modeling to excitation transfer and two-dimensional spectra of the FMO light-harvesting complex”. In: *The Journal of Physical Chemistry B* 115.26 (2011), pp. 8609–8621.

[211] Szilárd Páll et al. “Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS”. In: *Solving Software Challenges for Exascale*. Ed. by Stefano Markidis and Erwin Laure. Cham: Springer International Publishing, 2015, pp. 3–27. ISBN: 978-3-319-15976-8.

[212] Xingjie Pan and Tanja Kortemme. “Recent advances in de novo protein design: Principles, methods, and applications”. In: *Journal of Biological Chemistry* 296 (2021).

[213] Ziwei Pang et al. “Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model”. In: *Physical Chemistry Chemical Physics* 24.4 (2022), pp. 2441–2453.

[214] M. Z. Papiz et al. “The Structure and Thermal Motion of the B800-850 LH2 Complex from Rps. Acidophila at 2.0 Å Resolution and 100 K”. In: *J. Mol. Biol.* 326 (2003), p. 1523. DOI: 10.1016/S0022-2836(03)00024-X.

[215] Jun Young Park et al. “Ascaroside pheromones: chemical biology and pleiotropic neuronal functions”. In: *International journal of molecular sciences* 20.16 (2019), p. 3898.

[216] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190.

[217] David A Pearlman and Paul S Charifson. “Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system”. In: *Journal of medicinal chemistry* 44.21 (2001), pp. 3417–3423.

[218] Robert M Pearlstein. “Theory of the optical spectra of the bacteriochlorophyll a antenna protein trimer from Prosthecochloris aestuarii”. In: *Photosynthesis research* 31.3 (1992), pp. 213–226.

[219] John P Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized gradient approximation made simple”. In: *Physical review letters* 77.18 (1996), p. 3865.

[220] M. B. Plenio, J. Almeida, and S. F. Huelga. “Origin of long-lived oscillations in 2D-spectra of a quantum vibronic model: Electronic versus vibrational coherence”. In: *J. Chem. Phys.* 139 (2013), p. 235102.

[221] Per-Arno Plötz, Thomas Niehaus, and Oliver Kühn. “A new efficient method for calculation of Frenkel exciton parameters in molecular aggregates”. In: *The Journal of Chemical Physics* 140.17 (2014).

[222] Dirk Porezag et al. “Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon”. In: *Physical Review B* 51.19 (1995), p. 12947.

[223] “Protein Data Bank: the single global archive for 3D macromolecular structure data”. In: *Nucleic acids research* 47.D1 (2019), pp. D520–D528.

[224] Tonu Pullerits and Arvi Freiberg. “Kinetic model of primary energy transfer and trapping in photosynthetic membranes”. In: *Biophysical journal* 63.4 (1992), pp. 879–896.

[225] Tönu Pullerits, Mirianas Chachisvilis, and Villy Sundström. "Exciton Delocalization Length in the B850 Antenna of Rhodobacter Sphaeroides". In: *J. Phys. Chem.* 100.25 (1996), pp. 10787–10792. DOI: 10.1021/jp953639b. eprint: http://pubs.acs.org/doi/pdf/10.1021/jp953639b.

[226] Tõnu Pullerits and Villy Sundström. "Photosynthetic light-harvesting pigment-protein complexes: toward understanding how and why". In: *Accounts of Chemical Research* 29.8 (1996), pp. 381–389.

[227] Pu Qian et al. "Cryo-EM structures of light-harvesting 2 complexes from *Rhodopseudomonas palustris* reveal the molecular origin of absorption tuning". In: *Proc. Natl. Acad. Sci.* 119.43 (2022). DOI: 10.1073/pnas.2210109119.

[228] Felipe Cardoso Ramos et al. "The molecular mechanisms of light adaption in light-harvesting complexes of purple bacteria revealed by a multiscale modeling". In: *Chemical Science* 10.42 (2019), pp. 9650–9662.

[229] Margus Rätsep et al. "Demonstration and Interpretation of Significant Asymmetry in the Low-resolution and High-resolution Qy Fluorescence and Absorption Spectra of Bacteriochlorophyll A". In: *J. Chem. Phys.* 134.2, 024506 (2011), p. 024506. DOI: 10.1063/1.3518685.

[230] Pradeep Anand Ravindranath et al. "AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility". In: *PLoS computational biology* 11.12 (2015), e1004586.

[231] T. Reinot et al. "Exciton Lifetime Distributions and Population Dynamics in the FMO Protein Complex from *Prosthecochloris aestuarii*". In: *ACS Omega* 6 (2021), pp. 5990–6008.

[232] Patrick Reiser et al. "Graph neural networks for materials science and chemistry". In: *Communications Materials* 3.1 (2022), p. 93.

[233] Indrek Renge and Koit Mauring. "Spectral shift mechanisms of chlorophylls in liquids and proteins". In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 102 (2013), pp. 301–313. ISSN: 1386-1425. DOI: https://doi.org/10.1016/j.saa.2012.10.034. URL: https://www.sciencedirect.com/science/article/pii/S1386142512010396.

[234] T. Renger, V. May, and O. Kühn. "Ultrafast Excitation Energy Transfer Dynamics in Photosynthetic Pigment-Protein Complexes". In: *Phys. Rep.* 343 (2001), pp. 137–254. DOI: 10.1016/S0370-1573(00)00078-8.

[235] Thomas Renger and Frank Müh. "Understanding Photosynthetic Light-Harvesting: A Bottom Up Theoretical Approach". In: *Phys. Chem. Chem. Phys.* 15 (2013), pp. 3348–3371. DOI: 10.1039/C3CP43439G.

[236] Eva Rivera et al. "Influence of Site-Dependent Pigment-Protein Interactions on Excitation Energy Transfer in Photosynthetic Light Harvesting". In: *J. Phys. Chem. B* 117.18 (2013), pp. 5510–5521. ISSN: 1520-6106. DOI: 10.1021/jp4011586.

[237] Luis E Herrera Rodríguez and Alexei A Kananenka. "Systematic study of the role of dissipative environment in regulating entanglement and exciton delocalization in the Fenna-Matthews-Olson complex". In: *Physical Review E* 111.1 (2025), p. 014143.

[238] Sara Roosta et al. "Efficient surface hopping approach for modeling charge transport in organic semiconductors". In: *Journal of Chemical Theory and Computation* 18.3 (2022), pp. 1264–1274.

[239] Kennet J Rueda Espinosa and Alexei A Kananenka. "Photoexcited Energy Relaxation in Porphyrin Nanorings". In: *The Journal of Physical Chemistry C* (2024).

[240] Peder Thusgaard Ruhoff and Mark A Ratner. "Algorithms for computing Franck–Condon overlap integrals". In: *International Journal of Quantum Chemistry* 77.1 (2000), pp. 383–392.

[241] Johan E. Runeson, Thomas P. Fay, and David E. Manolopoulos. "Exciton Dynamics from the Mapping Approach to Surface Hopping: Comparison with Förster and Redfield Theories". In: *Phys. Chem. Chem. Phys.* 26.6 (2024), pp. 4929–4938. ISSN: 1463-9084. DOI: 10.1039/d3cp05926j.

[242] M Saberi et al. "Energy transfer and quantum correlation dynamics in FMO light-harvesting complex". In: *Molecular Physics* 114.14 (2016), pp. 2123–2131.

[243] Keisuke Saito, Takumi Suzuki, and Hiroshi Ishikita. "Absorption-energy calculations of chlorophyll a and b with an explicit solvent model". In: *Journal of Photochemistry and Photobiology A: Chemistry* 358 (2018), pp. 422–431.

[244] Shinji Saito, Masahiro Higashi, and Graham R. Fleming. "Site-Dependent Fluctuations Optimize Electronic Energy Transfer in the Fenna-Matthews-Olson Protein". In: *J. Phys. Chem. B* (2019). ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.9b07456.

[245] Fabrizio Santoro et al. "Effective method for the computation of optical spectra of large molecules at finite temperature including the Duschinsky and Herzberg–Teller effect: The Qx band of porphyrin as a case study". In: *The Journal of chemical physics* 128.22 (2008).

[246] Pooja Sarngadharan, Yannick Holtkamp, and Ulrich Kleinekathöfer. "Protein Effects on the Excitation Energies and Exciton Dynamics of the CP24 Antenna Complex". In: *The Journal of Physical Chemistry B* (2024).

[247] Kenneth Sauer, John R Lindsay Smith, and Alfred J Schultz. "The Dimerization of Chlorophyll a, Chlorophyll b, and Bacteriochlorophyll in Solution1". In: *Journal of the American Chemical Society* 88.12 (1966), pp. 2681–2688.

[248] Sergei Savikhin and Walter S Struve. "Ultrafast energy transfer in FMO trimers from the green bacterium Chlorobium tepidum". In: *Biochemistry* 33.37 (1994), pp. 11200–11208.

[249] N. Schieschke et al. "Geometry dependence of excitonic couplings and the consequences for configuration-space sampling". In: *J. Comput. Chem.* 42 (2021), pp. 1402–1418.

[250] Nils Schieschke et al. “Geometry dependence of excitonic couplings and the consequences for configuration-space sampling”. In: *Journal of Computational Chemistry* 42.20 (2021), pp. 1402–1418.

[251] Gabriela S. Schlau-Cohen et al. “Single-Molecule Spectroscopy Reveals Photosynthetic Lh2 Complexes Switch between Emissive States”. In: *Proc. Natl. Acad. Sci. USA* 110 (2013), pp. 10899–10903. DOI: 10.1073/pnas.1310222110.

[252] G. D. Scholes et al. “Ab Initio Molecular Orbital Calculations of Electronic Couplings in the LH2 Bacterial Light-Harvesting Complex of Rps. Acidophila”. In: *J. Phys. Chem. B* 103 (1999), pp. 2543–2553.

[253] G. D. Scholes et al. “How Solvent Controls Electronic Energy Transfer and Light Harvesting”. In: *J. Phys. Chem. B* 111.25 (2007), pp. 6978–6982.

[254] G. D. Scholes et al. “Lessons from Nature About Solar Light Harvesting”. In: *Nat. Chem.* 3 (2011), pp. 763–764. DOI: 10.1038/nchem.1145.

[255] Gregory D Scholes and Graham R Fleming. “On the mechanism of light harvesting in photosynthetic purple bacteria: B800 to B850 energy transfer”. In: *The Journal of Physical Chemistry B* 104.8 (2000), pp. 1854–1868.

[256] Gregory D. Scholes et al. “How Solvent Controls Electronic Energy Transfer and Light Harvesting”. In: *J. Phys. Chem. B* 111.25 (2007), pp. 6978–6982. DOI: 10.1021/jp072540p.

[257] Jonathan D Schultz et al. “Coherence in Chemistry: Foundations and Frontiers”. In: *Chemical Reviews* (2024).

[258] The Royal Swedish Academy of Sciences. *The Nobel Prize in Chemistry 2024*. https://www.nobelprize.org/prizes/chemistry/2024/summary/. Accessed: 2025-09-11. 2024.

[259] Francesco Segatta et al. “A Quantum Chemical Interpretation of Two-Dimensional Electronic Spectroscopy of Light-Harvesting Complexes”. In: *J. Am. Chem. Soc.* 139 (2017), pp. 7558–7567. ISSN: 0002-7863. DOI: 10.1021/jacs.7b02130.

[260] Gotthard Seifert, Dirk Porezag, and Th Frauenheim. “Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme”. In: *International journal of quantum chemistry* 58.2 (1996), pp. 185–192.

[261] Melih Şener et al. “Förster Energy Transfer Theory As Reflected in the Structures of Photosynthetic Light-Harvesting Systems”. In: *ChemPhysChem* 12.3 (2011), pp. 518–531. ISSN: 1439-4235. DOI: 10.1002/cphc.201000944.

[262] Hans Martin Senn and Walter Thiel. “QM/MM methods for biomolecular systems”. In: *Angewandte Chemie International Edition* 48.7 (2009), pp. 1198–1229.

[263] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. “Activation functions in neural networks”. In: *Towards Data Sci* 6.12 (2017), pp. 310–316.

[264] Mohamed F Shibl et al. “Multilayer-MCTDH approach to the energy transfer dynamics in the LH2 antenna complex”. In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 50.18 (2017), p. 184001.

[265] Abhishek Singharoy et al. “Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism”. In: *Cell* 179 (2019), 1098–1111.e23. DOI: 10.1016/j.cell.2019.10.021.

[266] Aaron Sisto et al. “Atomistic Non-Adiabatic Dynamics of the LH2 Complex with a GPU-Accelerated Ab Initio Exciton Model”. In: *Phys. Chem. Chem. Phys.* 19 (2017), pp. 14924–14936. DOI: 10.1039/C7CP00492C.

[267] Cathal Smyth, Daniel G. Oblinsky, and Gregory D. Scholes. “B800-B850 Coherence Correlates with Energy Transfer Rates in the Lh2 Complex of Photosynthetic Purple Bacteria”. In: *Phys. Chem. Chem. Phys.* 17.46 (2015), pp. 30805–30816. ISSN: 1463-9076. DOI: 10.1039/C5CP00295H.

[268] Alexander J Sneyd et al. “Efficient energy transport in an organic semiconductor mediated by transient exciton delocalization”. In: *Science Advances* 7.32 (2021), eabh4232.

[269] Vladimir Sobolev et al. “Molecular docking using surface complementarity”. In: *Proteins: Structure, Function, and Bioinformatics* 25.1 (1996), pp. 120–129.

[270] M. Sokolov et al. “Analytical Time-Dependent Long-Range Corrected Density Functional Tight Binding (TD-LC-DFTB) Gradients in DFTB+: Implementation and Benchmark for Excited-State Geometries and Transition Energies”. In: *J. Chem. Theory Comput.* 17.4 (2021), pp. 2266–2282.

[271] Monja Sokolov. “Computation of optical properties of chromophores in different environments using QM/MM methods”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2022. 214 pp. DOI: 10.5445/IR/1000144349.

[272] Monja Sokolov et al. “Non-adiabatic molecular dynamics simulations provide new insights into the exciton transfer in the Fenna–Matthews–Olson complex”. In: *Physical Chemistry Chemical Physics* (2024).

[273] Fernand Spiegelman et al. “Density-functional tight-binding: basic concepts and applications to molecules and clusters”. In: *Advances in physics: X* 5.1 (2020), p. 1710252.

[274] Vera Stehr et al. “Singlet exciton diffusion in organic crystals based on marcus transfer rates”. In: *Journal of Chemical Theory and Computation* 10.3 (2014), pp. 1242–1255.

[275] P. J. Stephens et al. “Ab initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. In: *J. Phys. Chem.* 98 (1994), p. 11623.

[276] Johan Strümpfer and Klaus Schulten. “Open Quantum Dynamics Calculations with the Hierarchy Equations of Motion on Parallel Computers”. In: *J. Chem. Theory Comput.* 8.8 (2012), pp. 2808–2816. ISSN: 1549-9618. DOI: 10.1021/ct3003833.

[277] Villy Sundström, Tönunu Pullerits, and Rienk van Grondelle. “Photosynthetic Light-Harvesting: Reconciling Dynamics and Structure of Purple Bacterial LH2 Reveals Function of Photosynthetic Unit”. In: *J. Phys. Chem. B* 103.13 (1999), pp. 2327–2346. ISSN: 1520-6106. DOI: 10.1021/jp983722+.

[278] Yoshitaka Tanimura. “Numerically “Exact” Approach to Open Quantum Dynamics: The Hierarchical Equations of Motion (HEOM)”. In: *J. Chem. Phys.* 153 (2020), p. 020901. DOI: 10.1063/5.0011599.

[279] Jeffrey S Taylor and Roger M Burnett. “DARWIN: a program for docking flexible molecules”. In: *Proteins: Structure, Function, and Bioinformatics* 41.2 (2000), pp. 173–191.

[280] Janet M Thornton, Roman A Laskowski, and Neera Borkakoti. “AlphaFold heralds a data-driven revolution in biology and medicine”. In: *Nature Medicine* 27.10 (2021), pp. 1666–1669.

[281] Erling Thyrhaug et al. “Exciton Structure and Energy Transfer in the Fenna-Matthews-Olson Complex”. In: *J. Phys. Chem. Lett.* 7 (2016), pp. 1653–1660. DOI: 10.1021/acs.jpclett.6b00534.

[282] Erling Thyrhaug et al. “Identification And Characterization Of Diverse Coherences In The Fenna-Matthews-Olson Complex”. In: *Nat. Chem.* 10 (2018), pp. 780–786. DOI: 10.1038/s41557-018-0060-5.

[283] Erling Thyrhaug et al. “Intraband dynamics and exciton trapping in the LH2 complex of Rhodopseudomonas acidophila”. In: *The Journal of Chemical Physics* 154.4 (2021).

[284] Alexander B Tong et al. “Could AlphaFold revolutionize chemical therapeutics?” In: *Nature structural & molecular biology* 28.10 (2021), pp. 771–772.

[285] Ashley L. Tong et al. “Comparison of the Energy-transfer Rates in Structural and Spectral Variants of the B800-850 Complex from Purple Bacteria”. In: *J. Phys. Chem. B* 124.8 (2020), pp. 1460–1469. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.9b11899.

[286] S. Tretiak et al. “Exciton Hamiltonian for the Bacteriochlorophyll System in the LH2 Antenna Complex Of Purple Bacteria”. In: *J. Phys. Chem. B* 104 (2000), pp. 4519–4528.

[287] A. Troisi. “Charge transport in high mobility molecular semiconductors: classical models and new theories ”. In: *Chem. Soc. Rev.* 40 (2011), pp. 2347–2358.

[288] D. E. Tronrud, M. F. Schmid, and B. W. Matthews. “Structure and X-ray Amino Acid Sequence of a Bacteriochlorophyll a Protein from *Prosthecochloris aestuarii* Refined at 1.9 A Resolution”. In: *J. Mol. Biol.* 188 (1986), pp. 443–444.

[289] Oleg Trott and Arthur J Olson. “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”. In: *Journal of computational chemistry* 31.2 (2010), pp. 455–461.

[290] JohnáC Tully. “Mixed quantum–classical dynamics”. In: *Faraday Discussions* 110 (1998), pp. 407–419.

[291] CE Tyner, WD Drotning, and HG Drickamer. “Asymmetric line shapes of localized optical excitations in condensed systems”. In: *Journal of Applied Physics* 47.3 (1976), pp. 1044–1047.

[292] Stephanie Valleau, Alexander Eisfeld, and Alan Aspuru-Guzik. "On the Alternatives for Bath Correlators and Spectral Densities from Mixed Quantum-Classical Simulations". In: *J. Chem. Phys.* 137.22 (2012), pp. 224103–13. DOI: 10.1063/1.4769079.

[293] Stéphanie Valleau, Alexander Eisfeld, and Alán Aspuru-Guzik. "On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations". In: *The Journal of Chemical Physics* 137.22 (2012).

[294] Mihaly Varadi et al. "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models". In: *Nucleic acids research* 50.D1 (2022), pp. D439–D444.

[295] Leonel Varvelo et al. "Formally Exact Simulations of Mesoscale Exciton Diffusion in a Light-Harvesting 2 Antenna Nanoarray". In: *J. Phys. Chem. Lett.* 14.12 (2023), pp. 3077–3083. ISSN: 1948-7185. DOI: 10.1021/acs.jpclett.3c00086.

[296] Cornelis Petrus van der Vegte et al. "Atomistic Modeling of Two-Dimensional Electronic Spectra and Excited State Dynamics for a Light Harvesting 2 Complex". In: *J. Phys. Chem. B* 119 (2015), pp. 1302–1313. DOI: 10.1021/jp509247p.

[297] S. I. E. Vulto, A. M. Streltsov, and T. J. Aartsma. "Excited State Energy Relaxation in the FMO Complexes of the Green Bacterium *Prosthecochloris aestuarii* at Low Temperatures". In: *J. Phys. Chem. B* 101 (1997), pp. 4845–4850.

[298] Linjun Wang et al. "Surface hopping methods for nonadiabatic dynamics in extended systems". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 10.2 (2020), e1435.

[299] Xiaoqing Wang et al. "Open Quantum System Parameters for Light Harvesting Complexes from Molecular Dynamics". In: *Phys. Chem. Chem. Phys.* 17 (38 2015), pp. 25629–25641. DOI: 10.1039/C5CP03891J.

[300] Yu Wang et al. "Addition of longer wavelength absorbing chlorophylls into crops could increase their photosynthetic productivity by 26%". In: *Nature Communications* 16.1 (2025), p. 7933.

[301] Andrew Waterhouse et al. "SWISS-MODEL: homology modelling of protein structures and complexes". In: *Nucleic acids research* 46.W1 (2018), W296–W303.

[302] F. Weigend and R. Ahlrichs. "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design an assessment of accuracy". In: *Phys. Chem. Chem. Phys.* 7 (2005), pp. 3297–3305.

[303] F. Weigend et al. "RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency". In: *Chem. Phys. Lett.* 294 (1998), p. 143.

[304] Jianzhong Wen et al. "Native electrospray mass spectrometry reveals the nature and stoichiometry of pigments in the FMO photosynthetic antenna protein". In: *Biochemistry* 50.17 (2011), pp. 3502–3511.

[305] M. Wendling et al. "The quantitative relationship between structure and polarized spectroscopy in the FMO complex of *Prosthecochloris aestuarii*: refining experiments and simulations." In: *Photosynth. Res.* 71 (2002), pp. 99–123.

[306] Graham A Worth et al. “Using the MCTDH wavepacket propagation method to describe multimode non-adiabatic dynamics”. In: *International Reviews in Physical Chemistry* 27.3 (2008), pp. 569–606.

[307] Weiwei Xie et al. “Performance of mixed quantum-classical approaches on modeling the crossover from hopping to bandlike charge transport in organic semiconductors”. In: *Journal of Chemical Theory and Computation* 16.4 (2020), pp. 2071–2084.

[308] T. Yanai, D. P. Tew, and N. C. Handy. “A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”. In: *Chem. Phys. Lett.* 393 (2004), pp. 51–57.

[309] Dehua Yang et al. “G protein-coupled receptors: structure-and function-based drug discovery”. In: *Signal transduction and targeted therapy* 6.1 (2021), p. 7.

[310] Taimin Yang et al. “A comparison of structure determination of small organic molecules by 3D electron diffraction at cryogenic and room temperature”. In: *Symmetry* 13.11 (2021), p. 2131.

[311] Yang Yang et al. “Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction”. In: *The Journal of Physical Chemistry A* 111.42 (2007), pp. 10861–10873.

[312] Yang Yang et al. “Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction”. In: *The Journal of Physical Chemistry A* 111.42 (2007), pp. 10861–10873.

[313] Jian-Ping Zhang et al. “Mechanism of the carotenoid-to-bacteriochlorophyll energy transfer via the S1 state in the LH2 complexes from purple bacteria”. In: *The Journal of Physical Chemistry B* 104.15 (2000), pp. 3683–3691.

[314] Fulu Zheng et al. “Fully Quantum Modeling of Exciton Diffusion in Mesoscale Light Harvesting Systems”. In: *Materials* 14 (2021), p. 3291. DOI: 10.3390/ma14123291.

[315] Kai Zhong et al. “An Efficient Time-Domain Implementation of the Multichromophoric Förster Resonant Energy Transfer Method”. In: *J. Chem. Phys.* 158.6 (2023). ISSN: 1089-7690. DOI: 10.1063/5.0136652.

[316] Tim J Zuehlsdorff and Christine M Isborn. “Modeling absorption spectra of molecules in solution”. In: *International Journal of Quantum Chemistry* 119.1 (2019), e25719.

[317] Tim J Zuehlsdorff et al. “Optical spectra in the condensed phase: Capturing anharmonic and vibronic features using dynamic and static approaches”. In: *The Journal of Chemical Physics* 151.7 (2019).

[318] Tim J Zuehlsdorff et al. “Vibronic and environmental effects in simulations of optical spectroscopy”. In: *Annual Review of Physical Chemistry* 72.1 (2021), pp. 165–188.