Karlsruhe Institute of Technology

Computation of optical properties of chromophores in different environments using QM/MM methods

Zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

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

Dissertation

von

M. Sc. Monja Sokolov

aus Lahr/Schwarzwald

Erster Referent: Prof. Dr. Marcus Elstner
Zweiter Referent: PD Dr. Sebastian Höfener
Tag der mündlichen Prüfung: 08. Februar 2022

Danksagung

Zuerst möchte ich mich bei meinem Betreuer Prof. Dr. Marcus Elstner bedanken, für die interessanten Projekte, die vielen produktiven Meetings, die Hilfestellungen und Ermutigungen.
Thanks to all my colleagues for the great working atmosphere, the scientific discussions and the time together during coffee breaks, barbecues and other activities. Special thanks goes to Ziwei Pang for the perfect teamwork in the glucose sensor project. Thank you also for the organization of the hot pot sessions. Danke Philipp Dohmen, für den Austausch über die LH-Komplexe und Machine learning, aber auch besonders für das Zuhören, wenn mal wieder etwas nicht geklappt hat. Danke Beatrix Bold, für das stetige Beantworten meiner Fragen, die gute Zusammenarbeit in mehreren Projekten und die gemeinsamen Mittagspausen. Danke Mila Krämer, für die ausführliche Einweisung in Machine Learning, dein konstruktives Feedback und deine Ermutigungen. Danke Tomáš Kubař, für das Beantworten sämtlicher Fragen, die Unterstützung bei technischen Problemen und insbesondere deine Hilfe bei den QM/MM Simulationen.
Ich bedanke mich außerdem bei Thomas Niehaus für die Implementierung eines State-Trackers in DFTB+, bei Sebastian Höfener für seine Hilfe bei den schwingungsaufgelösten Spektren und bei Violetta Schneider für die vorbereiteten Strukturen des Glukosebindeproteins. Ich danke außerdem Ulrich Kleinekathöfer und Sayan Maity für die gute Zusammenarbeit im Projekt zu den Lichtsammelkomplexen.
I also say thank you to the members of the GRK2039 for the discussions on our GRK retreats and the nice time during workshops and other activities.
Ich bedanke mich bei Sabine Holthoff für die Hilfe bei diversen organisatorischen Problemen.
Ich danke auch meinen Studienkolleg*innen, besonders Änni Aslam, Anastasia Görtz und Kerstin Ohler. Ohne euch wäre die Studienzeit nicht halb so schön gewesen.
Ein besonderer Dank gilt außerdem Helena Drobjasko und Anna Heck. In unserer inzwischen schon jahrzehntelangen Freundschaft wart ihr immer verständnisvoll und Zeit mit euch bessert meine Laune jedes Mal erheblich.
Abschließend möchte ich mich bei meiner Familie und meinen Schwiegereltern bedanken. Ihr habt mich immer unterstützt, mich so oft verwöhnt und ich konnte mich jederzeit auf euch verlassen. Ich danke meinem Mann Sebastian Sokolov von ganzem Herzen, dass er mich so oft aufgemuntert und mir den Rücken freigehalten hat.

Zusammenfassung

Die theoretische Beschreibung der Wechselwirkung zwischen Molekülen und Licht kann herausfordernd sein, insbesondere dann, wenn es sich um flexible Farbstoffe in einer komplexen und dynamischen Umgebung handelt. Obgleich quantenmechanische (QM) Methoden den angeregten Zustand eines Moleküls beschreiben können, sind sie zu rechenaufwändig, um strukturelle Fluktuationen simulieren zu können. Darüber hinaus ist die mögliche Systemgröße, die beschrieben werden kann, durch die Rechenkosten begrenzt. Aus diesem Grund kommen für die Untersuchung von Farbstoffen in Proteinumgebung semiempirische und Multiskalenansätze ins Spiel.

Die semiempirische Time-Dependent Long-range Corrected Density Functional Tight Binding (TD-LC-DFTB2) Methode wurde als effiziente Alternative zu *ab initio* Methoden oder der Dichtefunktionaltheorie in Bezug auf Geometrien im angeregten Zustand und Anregungsenergien getestet. Sie wurde in QM/MM Simulationen angewandt, in denen sie einen angeregten Fluorophor beschrieb, dessen Umgebung von einem klassischen Kraftfeld beschrieben wurde. Diese neue Strategie für die Untersuchung von Fluoreszenz wurde sorgfältig anhand von Literaturergebnissen bewertet, indem die Ergebnisse sowohl mit experimentellen als auch mit theoretischen Studien, die auf anderen Ansätzen basieren, verglichen wurden. Es wurde herausgefunden, dass TD-LC-DFTB2 im Allgemeinen Geometrien und Anregungsenergien von ausreichender Qualität liefert, aber es wurden auch einige Schwächen entdeckt.

Außerdem wurde ein optischer Glukosesensor untersucht, der aus dem Glukosebindeprotein und einem angefügten Fluorophor besteht. Mit Hilfe von klassischen Molekulardynamiksimulationen (MD Simulationen) konnten Zusammenhänge zwischen der Anwesenheit von Glukose, den Proteinkonformationen und dem Aufenthaltsort des Farbstoffs gefunden werden. Daraus ergab sich ein starker Hinweis auf die Funktionsweise des Sensors.

Schließlich wurde der Energietransfer in einem Pigment-Protein-Komplex untersucht. Der Fenna-Matthews-Olson-Komplex von Photosynthese betreibenden grünen Schwefelbakterien beinhaltet mehrere Bakteriochlorophyll *a*–Pigmente in seinem Proteingerüst. Diese leiten die im Chlorosome gesammelte Anregungsenergie mit erstaunlicher Effizienz zum Reaktionszentrum weiter. Es wird Vorarbeit für eine Simulation der Exzitonenpropagation durch den Komplex gezeigt. Anregungsenergien und die Kopplungen zwischen den Pigmenten, das heißt die Elemente des exzitonischen Hamiltonoperators, wurden mit TD-LC-DFTB2 für Strukturen aus klassischen MD Simulationen berechnet. Dadurch wurde ein Eindruck zu deren Entwicklung über die Zeit und den Einfluss der Proteinumgebung gewonnen. Weiterhin wurden diese Daten genutzt, um neuronale Netze zu trainieren, die Anregungsenergien und Kopplungen noch schneller als TD-LC-DFTB2 vorhersagen können.

Abstract

The theoretical description of the interaction between molecules and light can be challenging, in particular if the chromophores are flexible and located within a complex and dynamical environment. While quantum mechanical (QM) methods are able to describe the excited state of a molecule, they are computationally too demanding to simulate structural fluctuations in this state. Moreover, the treated system size is limited due to the computational cost. For the study of chromophores in a protein environment thus semi-empirical methods and multiscale approaches come into play.

The semi-empirical Time-Dependent Long-range Corrected Density Functional Tight Binding (TD-LC-DFTB2) method was benchmarked with respect to excited state geometries and excitation energies as efficient substitute for *ab initio* methods or density functional theory. It was applied in QM/MM simulations describing a fluorophore in excited state embedded in an environment treated by a classical force field. This new strategy for the study of fluorescence was carefully evaluated in comparison to literature results, experimentally as well as theoretically based on different approaches. It was found, that TD-LC-DFTB2 in general provides geometries and excitation energies of sufficient accuracy but also some weaknesses were detected.

Besides, an optical glucose sensor consisting of the Glucose Binding protein and an attached fluorophore was studied. Relations were found between the presence of glucose, the protein conformation and the location of the dye by means of classical molecular dynamics (MD) simulations. This provided a strong hint on the working mechanism of this sensor.

Finally, the energy transfer in a pigment protein complex was investigated. The Fenna-Matthews-Olson complex of photosynthetic green sulfur bacteria embedds several bacteriochlorophyll *a* pigments in its protein scaffold. These pigments transfer the energy collected in the chlorosome to the reaction center with astonishing efficiency. Preliminary work is presented for the simulation of the exciton propagation through the complex. Excitation energies and couplings between the pigments, that is the elements of the excitonic Hamiltonian, were computed with TD-LC-DFTB2 using structures from classical MD simulations. This provided an idea on their development over time and the effect of the protein environment. Further, with these data neural networks were trained which are able to predict the excitation energies and couplings even faster than TD-LC-DFTB2.

Contents

1. Introduction **1**

2. Molecules and Light **3**

3. Methods **9**
3.1. Quantum Mechanics . . . 9
3.1.1. LCAO-MO-Method . . . 10
3.1.2. Hartree-Fock . . . 10
3.1.3. Density Functional Theory . . . 12
3.1.4. Density Functional Tight Binding . . . 19
3.2. Classical Mechanics . . . 23
3.3. Enhanced Sampling Techniques . . . 25
3.4. Multiscale Modelling . . . 26
3.5. Machine Learning . . . 28
3.6. Optimization Techniques . . . 31
3.6.1. Steepest Descent . . . 31
3.6.2. Conjugate Gradient . . . 31

4. Benchmark of TD-LC-DFTB2 Gradients **33**
4.1. Motivation . . . 33
4.2. Benchmark Set . . . 34
4.3. Geometries . . . 34
4.3.1. Computational Details . . . 34
4.3.2. Results and Discussion . . . 35
4.4. Transition Energies . . . 43
4.4.1. Computational Details . . . 43
4.5. Results and Discussion . . . 44
4.6. Vibrational Frequencies . . . 59
4.6.1. Computational Details . . . 59
4.6.2. Results and Discussion . . . 60
4.7. State Ordering . . . 64
4.8. Conclusion . . . 64

5. Validation and Application of Excited State QM/MM Simulations **65**
5.1. Introduction of the Fluorophores . . . 65
5.1.1. Flugi-2 . . . 65
5.1.2. Prodan . . . 66
5.1.3. 4-Aminophthalimide . . . 68
5.2. Solvatochromism of Flugi-2 . . . 69
5.2.1. Computational Details . . . 69
5.2.2. Results and Discussion . . . 70

5.2.3. Conclusion and Outlook . . . 78
5.3. Solvatochromism of Prodan . . . 79
5.3.1. Computational Details . . . 80
5.3.2. State Characterization . . . 81
5.3.3. State Following Approach . . . 84
5.3.4. Solvated Prodan . . . 87
5.3.5. Discussion and Outlook . . . 88
5.4. Solvatochromism of 4-Aminophthalimide . . . 89
5.4.1. Computational Details . . . 89
5.4.2. State Characterization . . . 90
5.4.3. Dynamics in Vacuum . . . 91
5.4.4. Solvated 4-AP . . . 92
5.4.5. Discussion and Outlook . . . 95
5.5. Concluding Remarks . . . 96

6. Understanding Optical Glucose Sensing: Glucose/Galactose Binding Protein Combined with Badan 97
6.1. The Biological System . . . 98
6.2. Experimentally Tested Glucose Sensors . . . 100
6.3. Motivation . . . 103
6.4. Computational Details . . . 105
6.4.1. Charge Models for Glucose . . . 105
6.4.2. Molecular Dynamics Simulations . . . 106
6.4.3. QM/MM Simulations . . . 107
6.5. Results and Discussion . . . 108
6.5.1. Glucose Charges . . . 108
6.5.2. Protein Conformations . . . 111
6.5.3. Badan Conformations and Interactions . . . 115
6.6. Conclusion and Outlook . . . 127

7. Exciton Transfer in the Fenna-Matthews-Olson Complex 129
7.1. Introduction . . . 129
7.2. Motivation . . . 134
7.3. Computational Details . . . 134
7.4. Results and Discussion . . . 137
7.4.1. Site Energies and Couplings Predicted by TD-LC-DFTB2 . . . 137
7.4.2. Site Energies and Couplings Predicted from a Neural Network . . . 144
7.5. Conclusion and Outlook . . . 148

8. Conclusion and Outlook 151

A. Appendix 155
A.1. TD-LC-DFTB2 Benchmark . . . 155
A.1.1. Tables of Values . . . 155
A.1.2. Analysis of States . . . 173
A.2. QM/MM Simulations: Results for the Ground State . . . 178
A.3. Optical Glucose Sensor: Partial Charges . . . 182
A.4. Exciton Transfer in the FMO Complex: Couplings . . . 184

Bibliography **185**

List of Abbreviations **205**

List of Figures **207**

List of Tables **211**

Contributions to publications **213**

1 Introduction

Interactions of molecules with light are omnipresent in our life and also a precondition for life on earth. Green plants and some bacteria are able to satisfy their energy demand by photosynthesis, a process using light to build energy-rich molecules from low-energy molecules. Many animals and we humans need light for vision. Furthermore, it strongly influences the circadian rhythm and has an impact on physical and mental health [1]. There are also many technical applications based on the interaction of molecules and light. Light can provide the energy to enable certain chemical reactions needed in industry or medicine. For instance, it can act as initiator of some polymer chain reactions, as used in dentistry [2]. Solar cells are an application, where light is used to generate electricity. The interaction of molecules with light can also be used for signalling. One example are luminescent probes applied in biology or medicine. Luminescent substances are also found in our daily life, in highlighters.

This thesis focusses on the very fundamental step of the theoretical description of chromophores in their complex environments and their interaction with light. This includes an extensive benchmark of a recent theoretical method and tests on its performance in simple systems. Then more complex systems are studied combining different methods. The structure of this thesis and a short introduction of the topics are presented in the following.

In chapter 2, the basic theoretical background to absorption and emission processes is summarized. Then the theoretical methods used in this study are introduced in chapter 3. It is shown that different methods are applicable for different system sizes and purposes and how they can be combined. This is particularly important for the study of environmentally sensitive dyes. The study of fluorescence properties is generally more difficult than the study of absorption since it requires knowledge about the dynamics in the excited state. This is a particular challenge if a picosecond to nanosecond timescale is of interest. Quantum mechanical methods are able to describe excited states but with a high computational cost. Semi-empirical methods are often a good compromise between accuracy and computational time.

In chapter 4, the accuracy of the fast semi-empirical Time-Dependent Long-range Corrected Density Functional Tight Binding (TD-LC-DFTB2) [3, 4] method is assessed in an extensive benchmark on excited state geometries and excitation energies in vacuum. In the next step, in chapter 5, the dynamics of three different dyes was studied in vacuum as well as their behaviour in solution. For the latter, combined quantum mechanical/molecular mechanical (QM/MM) simulations were performed with these dyes in excited state described by the TD-LC-DFTB2 method. The selected dyes are well studied environmentally sensitive fluorophores and thus allow the evaluation of this new approach of excited state QM/MM simulations with TD-LC-DFTB2. Such simulations can be a helpful tool in the study of the fluorescence properties of a dye in a complex environment, as for example in a protein.

In medicine and science optical sensors are widely used. The dependence of emission

properties, for example the wavelength (colour), intensity or lifetime, on the environment of a dye are used to characterize a system or to detect changes in it. Such kinds of sensors are also being developed for diabetes patients to painlessly measure their blood glucose level. One example for such a sensor is based on the glucose/galactose binding protein (GGBP), to which an environmentally sensitive dye (badan) was covalently linked [5]. Upon binding the sugar, the protein stabilizes in a different conformation. The fluorescence properties of the dye in turn depend on the protein conformation and possibly also directly on the presence of glucose. Hence, the sugar determines the observed fluorescence. To find a suitable dye and an appropriate linking position in the protein, several trial and error experiments are reported in the literature [6–9]. In chapter 6, the sensor system based on GGBP with badan is studied by all-atom molecular dynamics simulations to reveal the protein and dye conformations in presence and absence of glucose to promote the understanding of the working mechanism and to facilitate further development of such systems.

Solar cells are an energy source working without emission of CO_2 and thus represent an important pillar in the fight against climate change. However, regarding their efficiency there is a lot room for improvement. A biological system, the Fenna-Matthews-Olson (FMO) complex of green sulfur bacteria, transfers excitation energy from the light-harvesting chlorosomes to the reaction center with almost 100 % efficiency [10]. The reasons for this astonishing efficiency have been studied for decades but are still subject of debate. The FMO complex is studied in chapter 7. The protein scaffold embedds several bacteriochlorophyll a pigments which transfer the energy via delocalized excitons. The simulation of the exciton propagation through the complex could help in the understanding of the efficiency of the energy transfer. However, such simulations require a very fast computation of the excitation energies and couplings between the pigments, since they represent the elements determining the exciton energy transfer. For this purpose, neural networks were trained based on TD-LC-DFTB2 data to predict the excitation energies and couplings from the geometries of the pigments.

In the last chapter, the main findings are summarized and possible next steps are formulated.

2 | Molecules and Light

This chapter gives an introduction about the interactions of light with molecules. An overview of possible processes after irradiation of a molecule is given and some general characteristics of absorption and fluorescence spectra are explained. Further, important quantities characterizing the absorption and emission processes are introduced.

The nature of light is difficult to understand since it shows both, particle and wave character. It is described as electromagnetic radiation consisting of photons, which are quantum particles without mass but possessing an energy and momentum. The probability density of these photons is proportional to the intensity of the electromagnetic wave. [11]

The energy of the light is particularly important for the interaction with molecules. It is often expressed in accordance with the wave picture as wavelength λ, frequency ν or wavenumber $\tilde{\nu} = \frac{1}{\lambda}$. The photon energy E and ν are related by

$$E = h\nu \tag{2.1}$$

with h being the Planck constant. The energy of the light is also related to its colour. Light with low energy is in the red area and light with high energy in the blue to violet area of the spectrum. For orientation, some numbers in the visible range are given in Tab. 2.1. Adjacent in the spectrum are the infrared light (IR) with larger wavelengths and the ultraviolet (UV) area with lower wavelengths. [12]

E/eV	λ/nm	colour
1.77	700	red
2.14	580	yellow
2.34	530	green
2.64	470	blue
2.95	420	violet

Tab. 2.1.: The colour of light depends on its energy. Values are taken from [12].

When the energy of the light encountering a molecule matches the energy difference between two energy levels of the molecule, the light can be absorbed. By the absorption the molecule becomes excited from its ground state to a usually higher vibrational level of a higher electronically state. This process is very fast with times in the order of 10^{-15} s and thus retains the molecular geometry due to the slower movement of the nuclei in comparison to the electrons. The probability of the transition in a specific vibrational state is the higher, the larger the overlap between the wave functions of this vibrational state in the excited state and the vibrational state of the ground state (Franck-Condon principle). [12, 13]

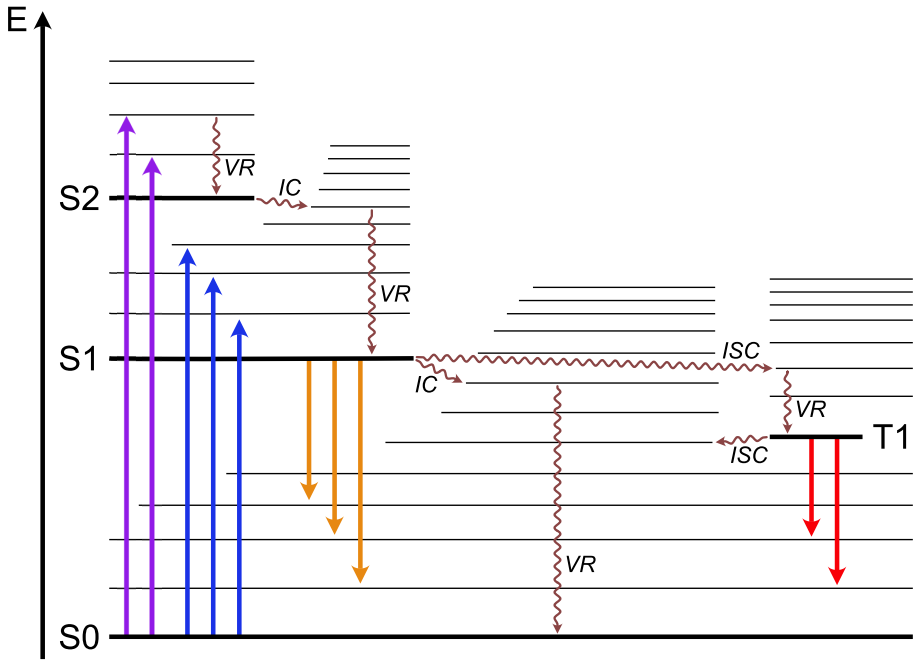


Fig. 2.1.: Jablonski diagram. **S0** denotes the electronical ground state, **S1** and **S2** the first and second electronically excited singlet states. **T1** is the first excited triplet state. Vibrational energy levels are shown above the electronically states, respectively. The purple and blue arrows symbolize absorption, the orange arrows fluorescence and the red arrows phosphorescence. The wiggly lines indicate different radiationless processes; **VR**: vibrational relaxation, **IC**: internal conversion, **ISC**: intersystem crossing.

The absorption of electromagnetic radiation by the molecule depends on the transition dipole moment μ. The transition dipole moment operator is defined as

$$\hat{\mu} = -e\sum_i \vec{r}_i + e\sum_I Z_I \vec{R}_I \tag{2.2}$$

with r_i and R_I as vectors from the center of charge to the respective electron and nuclei positions and the atomic number Z. Its expectation value is computed with the ground and excited state wave functions which are separated in an electronical and vibrational part, $\Psi^0 = \Psi_e^0\Psi_v^0$ and $\Psi^{\text{exc}} = \Psi_e^{\text{exc}}\Psi_v^{\text{exc}}$ (Born-Oppenheimer approximation). The electronical part depends on the electron positions and the vibrational part on the nuclei positions. The separation is possible due to the already mentioned slower movement of the nuclei. Considering that different electronic states (as the ground and excited state) are orthorgonal, the expectation value of the transition dipole moment finally reads:

$$\mu^{0\to exc} = -e\sum_i \int \Psi_e^{0*}\vec{r}_i\Psi_e^{exc}d\tau_e \int \Psi_v^{0*}\Psi_v^{exc}d\tau_v \tag{2.3}$$

$$= \mu_e^{0\to\text{exc}} S_v^{0\to\text{exc}} \tag{2.4}$$

S denotes the overlap integral and measures the similarity between the vibrational wave functions. $|S|^2$ is the Franck-Condon factor of the transition and $|\mu^{0\to\text{exc}}|^2$ is proportional

to the intensity of the transition. [12]

The absorption and following processes are visualized in Fig. 2.1 in a Jablonski diagram. Typically, at first fast vibrational relaxation (VR) occurs from the higher vibrational states to the vibrational ground state of the respective electronically states reached in the absorption, for example the first (S1) and second (S2) singlet excited states. Molecules which were excited into S2 or another higher electronically state show internal conversion (IC) from their vibrational ground state in a vibrational level of similar energy belonging to the lower electronically excited state. From there, the molecules relax further vibrationally to the vibrational ground state. If this is not S1, yet, again IC and VR take place. Usually, less than 10^{-12} s are needed to reach the vibrational ground state of S1. [11, 13]

There are now several ways how such an electronically excited molecule could return to its electronical ground state (S0). One possible relaxation process is the emission of light, which is generally called luminescence. The emission from a singlet excited state like S1 is named fluorescence. It occurs usually after $\approx 10^{-8}$ s. If instead the spin is converted then the first excited triplet state (T1) is reached, a process called intersystem crossing (ISC). From this state, the molecule can show phosphorescence. This is a luminous process which is spin-forbidden and thus slow. Alternatively, from either S1 or T1 the molecule can relax radiationless to the ground state via IC (S1) or ISC (T1) and VR. [13]

The finding, that molecules fluoresce from the first excited state and not from the state which is reached directly after absorption, is called *Kasha's rule*. It is a result of the smaller energy differences between the higher electronical states compared to the energy difference between S1 and the ground state and a more frequent crossing of the potential energy curves of the higher electronical states. Fluorescence from S1 explains why the emission spectrum is typically independent of the excitation energy. However, exceptions of this rule are possible. [11, 13]

The absorption and emission spectra of a dye often follow a so-called mirrow-image rule because similar transitions between vibrational levels are involved in the absorption and emission process. This can be traced back to a similar spacing of the vibrational levels in similar geometries in ground and excited state. If the absorption and emission spectra show less symmetry this may indicate that the molecule undergoes larger changes in the excited state, for example forms a complex or acts as acid or base. [13]

The emitted light is red-shifted compared to the absorbed light and one reason is also seen in the Jablonski diagram. After absorption, energy dissipates by vibrational relaxation and internal conversion to S1 and once more after the emission in a higher vibrational level of S0. Moreover, surrounding solvent molecules may rearrange according to the new electron distribution in S1 which leads to a further stabilization of this state. [13]

Not visible in the Jablonski diagram are the changes of the geometry of the molecule during the relaxation process. In Fig. 2.2, a scheme is shown with the potential energy surfaces of a molecule in ground and excited state, but for clearness without vibrational levels. The energy depends on the geometry of the molecule and the energy minima of the different states correspond to different geometries. The absorption process can be viewed as vertical excitation from the ground state minimum since the electronical rearrangement is much faster than the movement of the nuclei. However, until the molecule fluoresces, also the nuclei relaxed in the excited state and thus the fluorescence can be described as

a vertical deexcitation from the geometry at the excited state minimum. The difference between the energy immediately after the absorption and the energy of the minimum of the excited state is called the reorganization energy of the excited state. Similarly, the energy difference of the ground state minimum and the energy directly after the emission, i. e. the energy on the ground state potential energy curve at the position of the excited state minimum, is the ground state reorganization energy. The sum of both reorganization energies, which is the same as the difference between the absorption and emission energies, corresponds to the Stokes shift (without vibrational progression). The energy difference between the ground state minimum and the excited state minimum is called adiabatic excitation energy. The picture given in Fig. 2.2 is simplified in this respect as it does not include the zero-point vibrations. If they are added to the energies of the minima, respectively, the 0–0 transition energies can be computed as difference between the two corrected energy levels.

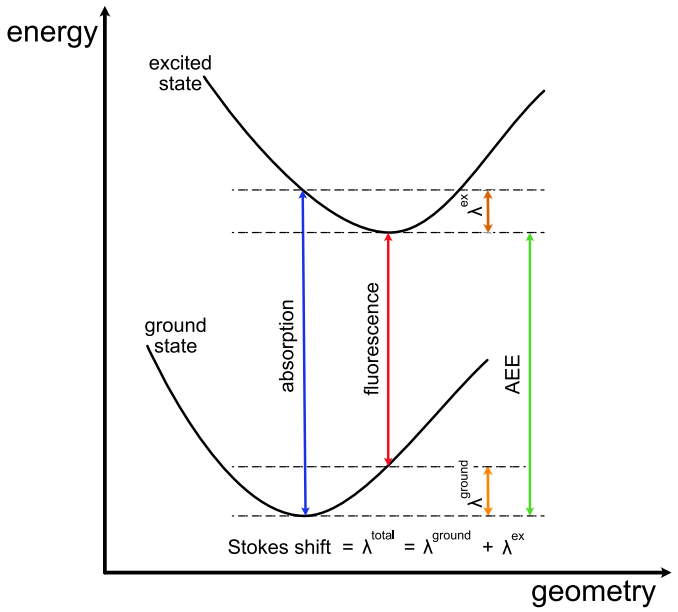


Fig. 2.2.: Scheme of potential energy surfaces in ground and excited state and important quantities. AEE: Adiabatic Excitation Energy; $\lambda^{\text{ground}}, \lambda^{\text{ex}}$: reorganization energy in ground and excited state; λ^{total}: total reorgnization energy which corresponds to one definition of the Stokes shift.

The various relaxation processes shown within the Jablonski diagram have different probabilities in different molecules. Phosphorescence is observed only if intersystem crossing is possible, as for instance in some molecules containing heavy atoms such as iodine. In fluorophores, photon emission (fluorescence) competes with radiationless relaxation. The fraction of absorbed photons which are finally emitted is the quantum yield (Q). The closer it is to unity, the more intense is the fluorescence. Expressed with the emissive k_e and non-radiative k_{nr} rate constants, it reads [13]:

$$Q = \frac{k_e}{k_e + k_{nr}} \tag{2.5}$$

Another important quantity to characterize a fluorophore is the fluorescence lifetime τ (Eq. 2.6). This means the time the molecule stays in excited state on average, which is typically in the order of few nanoseconds. The natural or intrinsic lifetime τ_n of a fluorophore describes the lifetime in case that there is no radiationless relaxation (Eq. 2.7). The larger the natural lifetime, the lower is the radiative rate constant. [13]

$$\tau = \frac{1}{k_e + k_{nr}} \tag{2.6}$$

$$\tau_n = \frac{1}{k_e} \tag{2.7}$$

The fluorescence of a fluorophore can be quenched due to different reasons. Contact of the excited fluorophore with a so-called quencher molecule prevents its fluorescence. This is called collisional or dynamic quenching and described by the Stern-Volmer equation:

$$\frac{F_0}{F} = 1 + K_q[q] = 1 + k_q\tau_0[q] \tag{2.8}$$

with F_0 being the fluorescence intensity without quencher, F the intensity in the presence of the quencher, K_q the Stern-Volmer quenching constant which combines the bimolecular quenching constant k_q and the lifetime without quencher τ_0, and $[q]$ the concentration of the quencher. If a plot of $\frac{F_0}{F}$ versus $[q]$ is linear, this indicates that there is only one kind of fluorophore and all of them are equally accessible to the quencher. A typical quencher molecule is molecular oxygen which quenches a wide variety of fluorophores. [13]

In static quenching, the fluorophore and the quencher build a complex in the ground state which is not fluorescent. It is described by the following equation:

$$\frac{F_0}{F} = 1 + K_s[q]. \tag{2.9}$$

K_s is the association constant of the complex and given by

$$K_s = \frac{[c]}{[f][q]} \tag{2.10}$$

with the concentration of the complex $[c]$, the concentration of the fluorophore $[f]$ and the concentration of the quencher $[q]$. Static and dynamic quenching can experimentally be distinguished for example by lifetime measurements. In the case of static quenching, only the lifetime of the unquenched fluorophore is measured, which remains the same ($\tau = \tau_0$). For dynamical quenching it holds $\tau = \tau_0 \cdot \frac{F}{F_0}$ instead. However, also combinations of static and dynamic quenching are possible. [13]

From the above explanations it becomes clear, that the fluorescence properties of a dye can be very sensitive to its surrounding. The important factors are finally summarized:

(i) The polarity of the environment can impact the emission wavelength by changing the energy of the excited state. For example, a dye with a large dipole moment in excited state is stabilized by a polar solvent. In fluorophores which possess close-lying $n{\rightarrow}\pi^*$ and $\pi{\rightarrow}\pi^*$ states, this effect can even change the order of these states [14,15]. Further, the polarity can influence the vibronic band intensities, as shown for pyrene [16]. (ii) Direct interactions as hydrogen bonds between dye and solvent can change the emission wave length but also the

intensity by offering an alternative relaxation pathway. Examples are 4-aminophthalimide or xanthene dyes like rose bengal and erythrosin B [14, 17]. The strength of a hydrogen bond between a dye and a solvent molecule depends on the electron density of the dye and thus may change after excitation [18]. $\pi - \pi$ stacking or even electron or proton transfer reactions are further examples for interactions influencing the fluorescence properties [13, 19]. (iii) The geometries sampled by the dye impact the fluorescence energies and oscillator strengths. In a dense environment, the geometrical space can be limited and lead to aggregation induced emission because a conical intersection is no more reached [20]. Of course, several of these factors can play a role at the same time which complicates the interpretation of the spectra. In some cases, there may be also different species of the dye which fluoresce. For instance, fluorophores with charge-transfer excited states typically show two emission bands, of which the first belongs to a locally excited state and the second origins from a twisted intramolecular charge-transfer state with full charge separation. The effect of the surrounding on these two states is usually different. Examples for fluorophores with an internal charge-transfer excited state are dimethylaminobenzonitrile and badan [21, 22].

An intersting case are systems with many interacting chromophores. One example studied in this work are photosynthetic pigment-protein complexes (PPCs) which transfer collected energy from sun light non-radiatively to the reaction centers with very high efficiencies. While the excitation energy is transferred via the pigments, the surrounding protein determines their orientation, geometrical freedom and electrostatic environment.

The theoretical description of energy transfer in a PPC depends on the distances between the pigments. If they are significantly larger than the pigment size, the theory of Förster can be applied. According to this model, the energy transfer efficiency depends on the amount of overlap between the emission spectrum of the donor pigment and the absorption spectrum of the acceptor pigment, the Coulomb coupling between these pigments and screening effects from the surrounding. However, if the pigments are closer but still too far from each other to exchange electrons, they can create delocalized exciton states. [23, 24]

In this process, the energy levels of the pigments split in new levels, the exciton energies. These are dependent on the coupling strength on one hand and on the excitation energies of the pigments on the other hand. The excitation energies of pigments in their individual local environment are called *site energies*. The excitonic Hamiltonian of the PPC in the site-basis contains the site energies on its diagonal and the couplings between the pigments as the off-diagonal elements. The eigenvalue problem to be solved can be expressed as

$$H_{site} \left|M\right\rangle = \mathcal{E}_m \left|M\right\rangle . \tag{2.11}$$

The eigenvalues $\mathcal{E}_m$ are the mth exciton energy, respectively. The eigenvectors are the exciton states ($\left|M\right\rangle$) and correspond to linear combinations of the local excited states of the m individual pigments $\left|m\right\rangle$:

$$\left|M\right\rangle = \sum_m c_m^{(M)} \left|m\right\rangle \tag{2.12}$$

The square of a coefficient, $|c_m^{(M)}|^2$, provides the probability of pigment m to be excited when the system is in state $\left|M\right\rangle$. [25, 26]

3 | Methods

In this section, the various methods applied in this study are introduced. It is started with quantum mechanics as the most accurate method used here to describe molecules in their ground and excited state. It focuses mainly on density functional theory and the semi-empirical density functional tight binding method. Afterwards, classical mechanics is described which allows sampling of structures of large molecules such as proteins. Enhanced sampling techniques follow, which help to sample difficult cases. The next part shows how quantum mechanics and classical mechanics can be combined to benefit from their respective advantages. Then a completely different technique, machine learning, is introduced. This method can be used to replace time-consuming calculations with faster models. Finally, a further subchapter explains optimization algorithms which are repeatedly applied in this work.

3.1. Quantum Mechanics

In quantum mechanics, the development of a system over time is described by the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(\vec{r},t)\right\rangle = \hat{H}\left|\Psi(\vec{r},t)\right\rangle. \tag{3.1}$$

i is the imaginary unit, $\hbar = \frac{h}{2\pi}$ the reduced Planck constant, $\hat{H}$ the Hamilton operator and $\Psi(\vec{r},t)$ a wave function depending on spatial coordinates and time.

Time-independent systems can be computed with the stationary Schrödinger equation:

$$\hat{H}\left|\Psi(\vec{r})\right\rangle = E\left|\Psi(\vec{r})\right\rangle. \tag{3.2}$$

It is an eigenvalue equation with the wave function $\Psi(\vec{r})$ as eigenfunction and the energy E as eigenvalue. The Hamilton operator contains several terms for the different energy contributions. For a molecule consisting of N_n nuclei with mass M_A and charge Z_A, and N_e electrons, it reads in atomic units:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N_e}\Delta_i - \frac{1}{2}\sum_{A=1}^{N_n}\frac{1}{M_A}\Delta_A - \sum_{i=1}^{N_e}\sum_{A=1}^{N_n}\frac{Z_A}{|\vec{r_i}-\vec{r_A}|} + \sum_{i<j}^{N_e}\frac{1}{|\vec{r_i}-\vec{r_j}|} + \sum_{A<B}^{N_n}\frac{Z_A Z_B}{|\vec{r_A}-\vec{r_B}|}. \tag{3.3}$$

The first two terms describe the kinetic energies of the electrons (nuclei), the third term describes the attraction between electrons and nuclei and the last two terms the repulsion of the electrons (nuclei) with each other. The second term is neglected in the Born-Oppenheimer approximation and the last term treated as a constant because the nuclei move significantly more slowly than the electrons.

The wave function itself has no physical meaning but its square represents the probability density of the electrons positions. Finding an appropriate wave function needs

some strategy. Generally, a meaningful function is guessed and further improved using the variational principle:

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

E_0 and Ψ_0 represent the true ground state energy and ground state wave function. The true energy is a lower bound to the energies of the trial wave functions. [27]

3.1.1. LCAO-MO-Method

A useful ansatz for a molecular wave function Φ is a linear combination of atomic orbitals (LCAO):

$$\Phi = \sum_{k=1}^{n} c_k \chi_k. \tag{3.5}$$

The atomic orbitals χ are known selected functions and the coefficients are determined in a calculation of the respective molecular system. The set of functions χ_k is called the basis set. The kind and number of included atomic orbitals can have an impact on the quality of the result.

Insertion of Eq. 3.5 in the stationary Schrödinger equation (Eq. 3.2) leads to

$$E = \frac{\langle \sum_{k=1}^{n} c_k \chi_k | \hat{H} | \sum_{l=1}^{n} c_l \chi_l \rangle}{\langle \sum_{k=1}^{n} c_k \chi_k | \sum_{l=1}^{n} c_l \chi_l \rangle} = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} c_k^* c_l H_{kl}}{\sum_{k=1}^{n} \sum_{l=1}^{n} c_k^* c_l S_{kl}}. \tag{3.6}$$

The introduction of the Hamiltonian matrix with the elements H_{kl} and the overlap matrix with the elements S_{kl} is a convenient way to write the integrals $\int \chi_k^* \hat{H} \chi_l d\vec{r}$ and $\int \chi_k^* \chi_l d\vec{r}$. The latter equals unity for $k = l$ if the basis functions are normalized, and it vanishes for $k \neq l$ if the basis functions are orthogonal. The minimum of the energy expression is found by solving the secular equations:

$$\sum_{l=1}^{n} (H_{kl} - E S_{kl}) c_l = 0 \quad (k = 1, 2, ..., n). \tag{3.7}$$

The matrix elements are computed and the energy values are obtained from the condition for a non-trivial solution of the secular equations,

$$det\,(H_{kl} - E S_{kl}) = 0. \tag{3.8}$$

The energy values can then be inserted in Eq. 3.7 for the determination of the coefficients. Every energy value corresponds to the energy of a molecular orbital constructed from the respective coefficients and basis functions. [28]

If an orthonormal basis set is used, the energy values and coefficients can be obtained in an easier way: diagonalization of the Hamiltonian matrix provides the lowest possible energies which are solutions of the eigenvalue equation. [29]

3.1.2. Hartree-Fock

The Hartree-Fock (HF) method represents the basis of many wave function based quantum chemistry methods. Here, the many-electron wave function is constructed from one-

electron wave functions to a Slater determinant:

$$\Psi(\vec{r_1}, \vec{r_2} \dots \vec{r_N})^{\mathrm{HF}} = \frac{1}{\sqrt{N!}} det \begin{pmatrix} \Phi_1(1) & \Phi_1(2) & ... & \Phi_1(N) \\ \Phi_2(1) & \Phi_2(2) & ... & \Phi_2(N) \\ ... & ... & ... & ... \\ \Phi_N(1) & \Phi_N(2) & ... & \Phi_N(N) \end{pmatrix}. \tag{3.9}$$

This ansatz is useful since it satisfies the antisymmetry condition which is required for fermions. The prefactor $\frac{1}{\sqrt{N!}}$ normalizes the Slater determinant.

The best possible Slater determinant is found using the variational principle (Eq. 3.4). The one-electron wave functions are varied but kept orthonormal. Using the Lagrangian multipliers ϵ_i, the final Hartree-Fock equations to solve are:

$$\hat{f}\Phi_i = \epsilon_i \Phi_i \quad (i = 1, ..., N_e). \tag{3.10}$$

The ϵ_i can be interpreted as orbital energies. The Fock operator $\hat{f}$ is an effective one-electron operator. It contains the kinetic energy, the electron-nuclei attraction and an averaged repulsion from the other $N_e - 1$ electrons in the system:

$$\hat{f}_i = -\frac{1}{2}\Delta_i - \sum_{A=1}^{N_n} \frac{Z_A}{|\vec{r_i} - \vec{r_A}|} + \sum_{j}^{N_e} (\hat{J}_j(i) - \hat{K}_j(i)). \tag{3.11}$$

The Coulomb operator $\hat{J}$ and the exchange operator $\hat{K}$ are defined as

$$\hat{J}_j(1) \left|\Phi_i(1)\right\rangle = \left\langle \Phi_j(2) \right| \frac{1}{|\vec{r}_1 - \vec{r}_2|} \left|\Phi_j(2)\right\rangle \left|\Phi_i(1)\right\rangle \tag{3.12}$$

$$\hat{K}_j(1) \left|\Phi_i(1)\right\rangle = \left\langle \Phi_j(2) \right| \frac{1}{|\vec{r}_1 - \vec{r}_2|} \left|\Phi_i(2)\right\rangle \left|\Phi_j(1)\right\rangle. \tag{3.13}$$

The effective repulsive potential between one electron and all others depends on the one-electron wave functions. Thus, the Hartree-Fock equations need to be solved in an iterative procedure. Starting from a set of trial wave functions, an improved set is obtained in every iteration, which is then used for the next cycle. As soon as the difference between the results of two iterations is smaller than a predetermined threshold, self-consistency is achieved. [27]

The LCAO ansatz can be applied to construct the functions Φ_i, which is described by the Roothaan-Hall formalism. The Roothaan-Hall equations expressed as matrix equation read:

$$FC = SC\epsilon \tag{3.14}$$

with the Fock matrix elements F, the matrix of the coefficients from the LCAO ansatz C and the overlap matrix S. The Fock matrix elements are computed as

$$\begin{aligned} F_{kl} = &\left\langle \chi_k(i) \middle| -\frac{1}{2}\Delta_i - \sum_{A=1}^{N_n} \frac{Z_A}{|\vec{r_i} - \vec{r_A}|} \middle| \chi_l(i) \right\rangle \\ &+ \sum_{a=1}^{n}\sum_{b=1}^{n} P_{ab} \left[\iint \chi_k^*(i)\chi_l(i) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_a^*(j)\chi_b(j) d\vec{r}_i d\vec{r}_j \right. \\ &\left. - \frac{1}{2} \iint \chi_k^*(i)\chi_b(i) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_a^*(j)\chi_l(j) d\vec{r}_i d\vec{r}_j \right]. \end{aligned} \tag{3.15}$$

In the case of a closed-shell system the density matrix is $P_{ab} = 2\sum_{m=1}^{N/2} c_{am}^* c_{bm}$. [28]

3.1.3. Density Functional Theory

The wave function of a molecular system depends on 4N variables (three spatial coordinates and one spin for each electron) which makes it very complicated for large systems. The basic idea of density funtional theory (DFT) is to use the electron density as quantity containing all information about the system but being dependent only on the three spatial dimensions. The first Hohenberg-Kohn theorem [30] proves that in the ground state the electron density uniquely determines the external potential (the attraction between electrons and nuclei) and hence also the Hamilton operator and all resulting properties of the system. The ground state energy is thus a functional of the ground state electron density:

$$E_0 = E_0[\rho_0(\vec{r})]. \tag{3.16}$$

The second Hohenberg-Kohn theorem [30] states that every trial ground state density provides an energy larger than the true ground state energy except it corresponds to the true ground state density (Eq. 3.17). This means that the ground state density can be found using the variational principle.

$$E_0 \leq E[\rho_{\text{trial}}(\vec{r})] \tag{3.17}$$

The energy of the system consists of the kinetic energy of the electrons T, the electron-electron interaction E_{ee} and the nuclei-electron attraction E_{ne}, which all need to be expressed as funtionals of the electron density:

$$E[\rho] = T[\rho] + E_{\text{ee}}[\rho] + E_{\text{ne}}[\rho]. \tag{3.18}$$

One tricky point is finding the relation between the electron density and the kinetic energy of the electrons. Kohn and Sham [31] proposed to calculate the kinetic energy of a reference system of non-interacting electrons (T_{ref}) in an effective potential by means of orbitals. The effective potential of the non-interacting system is set such that the electron density is the same as the one of the interacting system. The kinetic energy of the reference system is nonetheless only an approximation. The difference to the real kinetic energy of the interacting system needs to be considered in a correction term. The electron-electron interaction is also separated into two terms: the classical Coulomb repulsion J and a correction term for self-interaction, exchange and correlation effects. Summarizing all corrections in one term, E_{XC}, the energy expression reads [27]

$$\begin{aligned} E[\rho] &= T_{\text{ref}}[\rho] + J[\rho] + E_{\text{XC}}[\rho] + E_{\text{ne}}[\rho] && (3.19)\\ &= T_{\text{ref}}[\rho] + \frac{1}{2}\iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 + E_{\text{XC}}[\rho] + \int V_{\text{ne}}\rho(\vec{r}) d\vec{r} && (3.20) \end{aligned}$$

or in dependence of orbitals

$$\begin{aligned} E[\rho] = &-\frac{1}{2}\sum_i^N \langle\varphi_i|\,\Delta\,|\varphi_i\rangle + \frac{1}{2}\sum_i^N\sum_j^N \iint |\varphi_i(\vec{r}_1)|^2 \frac{1}{|\vec{r}_1 - \vec{r}_2|} |\varphi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \qquad (3.21)\\ &+ E_{\text{XC}}[\rho] - \sum_i^N \int \sum_A^{N_n} \frac{Z_A}{|\vec{r}_1 - \vec{r}_A|} |\varphi_i(\vec{r}_1)|^2 d\vec{r}_1. \end{aligned}$$

The sum of the squared orbitals corresponds to the electron density.

$$\sum_{i}^{N} |\varphi_i(\vec{r})|^2 = \rho_{\text{ref}}(\vec{r}) = \rho_0(\vec{r}) \tag{3.22}$$

Using the variational principle while keeping the orbitals orthonormal leads to the Kohn-Sham equations (Eq. 3.23). These one-particle equations provide the orbitals. They need to be solved iteratively since V_{eff} depends in turn on the orbitals via the density in the Coulomb term (Eq. 3.24).

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}(\vec{r})\right)\varphi_i = \epsilon_i \varphi_i \tag{3.23}$$

$$V_{\text{eff}}(\vec{r}) = \int \frac{\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_2 + V_{\text{XC}}(\vec{r}_1) - \sum_{A}^{N_k} \frac{Z_A}{|\vec{r}_1 - \vec{r}_A|} \tag{3.24}$$

The main challenge is to find an appropriate approximation for the exchange-correlation potential $V_{\text{XC}} = \frac{\delta E_{\text{XC}}[\rho]}{\delta \rho}$. If the exact exchange-correlation energy and the corresponding potential were known, the Schrödinger equation could be solved exactly with DFT. The different classes of exchange-correlation functionals are discussed in the following four subchapters. Before that, the contributions to include are specified and the concept of *hole functions* is introduced.

The exchange energy arises from the fact that two electrons with same spin cannot occupy the same orbital (Pauli principle), which is considered in wave function theory by the antisymmetry of the wave functions. In DFT, an explicit term for this energy is needed. The correlation energy is missing in HF, and is in fact defined as the difference between the exact non-relativistic energy and the HF energy. It can be further divided into *dynamical* and *non-dynamical* correlation. The dynamical correlation describes the avoidance of the electrons to come too close to each other, which in the HF scheme is only captured on average. Non-dynamical correlation means deficiencies of describing the ground state by only a single Slater determinant in HF. The correlation energy in DFT is not exactly the same as the HF correlation energy since here the nucleus-electron attraction and classical Coulomb repulsion are computed exactly and the density is by construction the real ground state density. This is not the case in HF where the correlation energy includes contributions from these interactions and also the kinetic energy. Usually, the exchange energy contribution is larger than the correlation energy contribution. [27]

While the electron density (Eq. 3.25) shows the probability to find one electron in a certain place (with arbitrary spin), the pair density $\rho_p(\vec{x}_1, \vec{x}_2)$ (Eq. 3.26) describes the probability to find two electrons at two certain places at the same time (with defined spin).

$$\rho(\vec{r}) = N \int \cdots \int |\Psi(\vec{x}_1, \vec{x}_2 \ldots \vec{x}_N)|^2 ds_1 d\vec{x}_2 \ldots d\vec{x}_N \tag{3.25}$$

$$\rho_p(\vec{x}_1, \vec{x}_2) = N(N-1) \int \cdots \int |\Psi(\vec{x}_1, \vec{x}_2 \ldots \vec{x}_N)|^2 d\vec{x}_3 \ldots d\vec{x}_N \tag{3.26}$$

The electron density integrates to the total number of electrons, tends to zero for large distances from the nuclei and shows maxima at nuclei position. These are cusps with a

discontinuity in the gradient. The pair density provides information about exchange and correlation effects and is used to define the *exchange-correlation hole*:

$$h_{\mathrm{xc}}(\vec{x}_1;\vec{x}_2) = \frac{\rho_p(\vec{x}_1,\vec{x}_2)}{\rho(\vec{x}_1)} - \rho(\vec{x}_2) \tag{3.27}$$

h_{xc} corresponds to the difference between the probability to find an electron at $\vec{x}_2$ when there is another electron at $\vec{x}_1$ and the probability to find that electron at $\vec{x}_2$ when the other electron is not present. The hole function carries its name because it can be imagined that there is a hole in the probability to find an electron in the proximity of another electron. It has some properties which are important for the development of the functionals. h_{xc} is formally divided in the Fermi hole considering the exchange and the Coulomb hole. The former is negative everywhere and integrates to -1, the latter is neutral in total but has positive areas far away from the reference electron and is negative at and around the reference electron.
E_{XC} contains not only the exchange and correlation contributions described by the hole function but also the correction to the kinetic energy. This can be included in the hole function by *adiabatic connection* providing h'_{xc}. Finally, the following expression is obtained:

$$E_{\mathrm{XC}} = \frac{1}{2}\iint \frac{\rho(\vec{r}_1)h'_{\mathrm{xc}}(\vec{r}_1;\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2. \tag{3.28}$$

Functionals can now be assessed based on the included hole functions which should possess the properties of the real hole function, although the functionals can be abstract mathematical expressions without physical motivation. [27]

Local Density Approximation

It is possible to derive a very accurate expression for the exchange-correlation functional of a uniform electron gas. In this model system the electron density is equal at every point in space. This is a reasonable approximation for a metal whose valence electrons are homogeneously distributed on the positive background charge of the cores but a rather poor model for a molecule. However, the exchange-correlation expression of the uniform electron gas is still useful for molecules as shown in the following.

Exchange-correlation functionals applying the *Local Density Approximation* (LDA) compute the exchange-correlation energy of a molecule as shown in Eq. 3.29.

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] = \int \rho(\vec{r})\varepsilon_{\mathrm{XC}}[\rho(\vec{r})]d\vec{r} \tag{3.29}$$

Therein, $\varepsilon_{\mathrm{XC}}[\rho(\vec{r})]$ corresponds to the exchange-correlation energy per electron in a uniform electron gas of density $\rho(\vec{r})$. This means, the exchange-correlation energy is calculated at every point in space as for the uniform electron gas with the density at this point in space, weighted by the probability to find an electron there and integrated over the complete space. Hence, the exchange-correlation energy is assumed to depend only on the *local* values of the electron density.

$\varepsilon_{\mathrm{XC}}[\rho(\vec{r})]$ is the sum of an expression for the exchange (*Slater exchange*) and a correlation contribution (e.g. from Ref. [32]). [27]

Generalized Gradient Approximation

More accurate results are obtained using the *Generalized Gradient Approximation* (GGA). Exchange-correlation functionals belonging to this class consider not only the local values of the density but also its gradient $\nabla\rho(\vec{r})$:

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho] = \int f[\rho(\vec{r}), \nabla\rho(\vec{r})]d\vec{r} \tag{3.30}$$

The form of the functional f varies between different functionals and can contain empirical parameters. A common property is that they meet the boundary conditions for the exchange-correlation hole mentioned above which are also satisfied by LDA functionals and crucial for physically meaningful results. Usually, f is splitted in an exchange and a correlation part and the expressions are determined independently. Important examples for GGA functionals are PBE [33], BP86 [34,35] and BLYP [34,36]. The correlation functional LYP is no more based on the uniform electron gas but on the helium atom. [27]

Hybrid functionals

Hybrid functionals contain a certain amount of HF exchange in their exchange term. The idea is to consider the exact exchange energy of a Slater determinant instead of approximate exchange functionals. However, the exact exchange hole is delocalized and the correlation holes of the DFT functionals are localized which results in a poor description of the total hole if both holes are simply added.

The problem is tackled as follows: The exchange-correlation energy can be obtained by integrating over the non-classical energy contributions which depend on a parameter λ switching on the electrostatic interactions between the electrons:

$$E_{\mathrm{XC}}^{\mathrm{hybrid}}[\rho] = \int_0^1 E_{\mathrm{ncl}}^{\lambda} d\lambda. \tag{3.31}$$

If $\lambda = 0$, the interaction is turned off and E_{ncl} contains only exchange energy. This energy can be described exactly by the exchange energy of a Slater determinant. For $\lambda = 1$ instead, E_{ncl} consists of the exchange and correlation energy and can be approximated by an LDA or GGA functional. For $0 < \lambda < 1$, the contributions of the exchange and correlation energy are mixed, while the exact partitioning depends on the functional.

A prominent example for a hybrid functional is B3LYP [36–38]. It contains three semi-empirical parameters which control the amount of exact exchange (a), exchange from Becke's B88 functional [34] (b) and correlation from the LYP functional [36] (c):

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a)E_{\mathrm{X}}^{\mathrm{LSD}} + aE_{\mathrm{XC}}^{\lambda=0} + bE_{\mathrm{X}}^{\mathrm{B}} + cE_{\mathrm{C}}^{\mathrm{LYP}} + (1-c)E_{\mathrm{C}}^{\mathrm{LSD}}. \tag{3.32}$$

The superscript LSD stands for *local spin density* approximation which is the unrestricted extension to LDA. The values of the coefficients were determined to $a = 0.20$, $b = 0.72$ and $c = 0.81$. [27]

Long-range corrected functionals

Approximate exchange-correlation functionals contain a certain amount of unphysical self-interaction, which becomes particularly visible at large distances. The Coulomb potential

then approaches N/r. The exchange-correlation potential V_{XC} needs to cancel the unphysical self-interaction included therein and thus, it should have a $-1/r$ asymptotic behaviour for increasing distances. However, this is not the case for the functionals introduced so far. [27, 39]

Long-range corrected functionals improve the description of long-range exchange interactions by incorporating the HF exchange integral in E_{XC}. In contrast to hybrid functionals, the amount of HF exchange is not equal for all distances but distance dependent. The electron-electron repulsion operator is divided into a short-range and a long-range part, mostly using the error function:

$$\frac{1}{|\vec{r}_1 - \vec{r}_2|} = \frac{1 - \mathrm{erf}(\mu|\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|} + \frac{\mathrm{erf}(\mu|\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|} \tag{3.33}$$

The parameter μ determines the partitioning. The first term is applied in the short-range DFT exchange functional E_X^{sr}. The second term accounts for the long-range interactions and is used in the HF exchange integral providing E_X^{lr}. The correlation energy E_C is computed by a DFT functional. Generally, E_{XC} is the sum of these contributions:

$$E_{XC}^{LC} = E_X^{sr} + E_X^{lr} + E_C. \tag{3.34}$$

An important LC-functional is for instance CAM-B3LYP [39]. CAM is the abbreviation for *Coulomb attenuating method*. Eq. (3.33) is extended by two additional parameters to

$$\frac{1}{|\vec{r}_1 - \vec{r}_2|} = \frac{1 - [\alpha + \beta \cdot \mathrm{erf}(\mu|\vec{r}_1 - \vec{r}_2|)]}{|\vec{r}_1 - \vec{r}_2|} + \frac{\alpha + \beta \cdot \mathrm{erf}(\mu|\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|}. \tag{3.35}$$

This makes it possible to vary the ratio of HF and DFT exchange in a distance dependent way. The parameters found to work best in the CAM-B3LYP functional have the following values: $\mu = 0.33$, $\alpha = 0.19$ and $\beta = 0.46$. The long-range correction particularly improves for polarisabilities of long chains and the description of Rydberg and charge-transfer excitations. [39–41]

Time-dependent Density Functional Theory

Time-dependent density functional theory (TD-DFT) is based on the Runge-Gross theorem and the van Leeuwen theorem. Runge and Gross [42] proved that a time-dependent potential $v(\vec{r}, t)$ acting on a fixed inital many-body state Ψ_0 corresponds to a uniquely determined time-dependent electron density $\rho(\vec{r}, t)$ and vice versa. The potential $v(\vec{r}, t)$ needs to be a scalar potential, expandable in a Taylor series around the initial time t_0 and is determined up to an additive time-dependent scalar function. The system of interest needs to be finite. The van Leeuwen theorem [43] states that there exists a many-body system with different interactions than the system of interest which, with an according different potential $v'(\vec{r}, t)$ and initial state Ψ_0', reproduces the same time-dependent density $\rho(\vec{r}, t)$ as the system of interest. Hence, this theorem clears the way for a time-dependent Kohn-Sham approach. The initial state Ψ_0' must provide the correct density and also its time derivative at t_0. [44]

Assuming a system in the ground state is exposed to an additional time-dependent external potential at time t_0, the total external potential is given as

$$v(\vec{r}, t) = v_0(\vec{r}) + v_1(\vec{r}, t)\theta(t - t_0). \tag{3.36}$$

v_0 is the potential of the nucleus-electron interactions and the step function θ switches on the additional external potential v_1 beginning at time t_0. At t_0, the wave function of the non-interacting reference system Φ_0 can be computed from the Kohn-Sham equations of stationary DFT as Slater determinant of the Kohn-Sham orbitals φ_i^0. After t_0, the Kohn-Sham orbitals need to be calculated from the time-dependent Kohn-Sham equations:

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}(\vec{r},t)\right)\varphi_i(\vec{r},t) = i\frac{\partial}{\partial t}\varphi_i(\vec{r},t) \tag{3.37}$$

with the effective potential

$$V_{\text{eff}}(\vec{r},t) = v(\vec{r},t) + \int \frac{\rho(\vec{r'},t)}{|\vec{r}-\vec{r'}|}d\vec{r'} + V_{\text{XC}}(\vec{r},t) \tag{3.38}$$

and the conditions:

$$\varphi_i(\vec{r},t_0) = \varphi_i^0(\vec{r}) \tag{3.39}$$

$$V_{\text{XC}}(\vec{r},t_0) = V_{\text{XC}}^0(\vec{r}) \tag{3.40}$$

$V_{\text{XC}}(\vec{r},t)$ is the time-dependent exchange-correlation potential which in the adiabatic approximation has the same form as the static exchange-correlation potential but uses the time-dependent density. Analogous to the static case, the density is computed from the Kohn-Sham orbitals:

$$\rho(\vec{r},t) = \sum_i^N |\varphi_i(\vec{r},t)|^2. \tag{3.41}$$

An important application of TD-DFT is the description of the dynamical process of an excitation. This is addressed in the following subchapter. [44]

Linear response and excitation energies

Linear response theory can be applied to describe the response of a system to a weak perturbation. The excitation of a molecule is such a case. When the molecular system is irradiated by light, the wave function is perturbed and as a consequence, the expectation values of quantum mechanical observables become time-dependent. The difference between a time-dependent observable and its ground state value before the perturbation is meant by its *response*. This response can be described by an expansion in powers of the external field acting on the system. The first term in this expansion is the linear response. [44]

In the frequency space, the linear response of the electron density to a perturbation $v_1(\vec{r},\omega)$ is given as

$$\rho_1(\vec{r},\omega) = \int \chi_{\rho\rho}(\vec{r},\vec{r'},\omega)v_1(\vec{r'},\omega)d\vec{r'} \tag{3.42}$$

with the density-density response function (Lehmann representation):

$$\chi_{\rho\rho}(\vec{r},\vec{r'},\omega) = \lim_{\eta\to 0^+}\sum_{n=1}^{\infty}\left(\frac{\langle\Psi_0|\,\hat{\rho}(\vec{r})\,|\Psi_n\rangle\langle\Psi_n|\,\hat{\rho}(\vec{r'})\,|\Psi_0\rangle}{\omega-\Omega_n+i\eta} - \frac{\langle\Psi_0|\,\hat{\rho}(\vec{r'})\,|\Psi_n\rangle\langle\Psi_n|\,\hat{\rho}(\vec{r})\,|\Psi_0\rangle}{\omega+\Omega_n+i\eta}\right) \tag{3.43}$$

Ω_n corresponds to the nth excitation energy, respectively. As seen in Eq. 3.43, $\chi_{\rho\rho}(\vec{r},\vec{r'},\omega)$ has poles for frequency values which equal the excitation energies. [44]

Excitation energies from TD-DFT and linear response can be computed by means of the Casida equation [44]:

$$\begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} \vec{X} \\ \vec{Y} \end{pmatrix} = \Omega \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{X} \\ \vec{Y} \end{pmatrix} \tag{3.44}$$

The excitation energies are denoted by Ω, A and B are matrices with the following elements [44]:

$$A_{ia\sigma,jb\sigma'}(\Omega) = \delta_{ij}\delta_{ab}\delta_{\sigma\sigma'}\omega_{jb\sigma'} + K_{ia\sigma,jb\sigma'}(\Omega) \tag{3.45}$$

$$B_{ia\sigma,jb\sigma'}(\Omega) = K_{ia\sigma,jb\sigma'}(\Omega) \tag{3.46}$$

The indices i,j are used for occupied orbitals, a,b for unoccupied (virtual) orbitals and σ,σ' for spin orbitals. K is the coupling matrix and consists of the following elements [44]:

$$K_{ia\sigma,jb\sigma'}(\Omega) = \iint \Phi_{ia\sigma}(\vec{r}) f_{\mathrm{H_{XC},\sigma\sigma'}}(\vec{r},\vec{r'},\Omega)\Phi^*_{ia\sigma'}(\vec{r'})d\vec{r'}d\vec{r} \tag{3.47}$$

$f_{\mathrm{H_{XC},\sigma\sigma'}}(\vec{r},\vec{r'},\Omega)$ is the combined Hartree-xc kernel. The Casida equation (Eq. 3.44) is an infinite-dimensional pseudo-eigenvalue equation. In the Tamm-Dancoff approximation the B matrix is neglected and the simpler equation

$$A\vec{X} = \Omega\vec{X} \tag{3.48}$$

is solved. [44]

A practical variant of the Casida equation is [44]:

$$C\vec{Z} = \Omega^2\vec{Z} \tag{3.49}$$

with

$$C = \sqrt{A-B}(A+B)\sqrt{A-B} \tag{3.50}$$

$$\vec{Z} = \sqrt{A-B}(\vec{X}-\vec{Y}). \tag{3.51}$$

The intensity of an excitation is related to its oscillator strength. This quantity describes the ratio of the absorption rate of a molecular system to the absorption rate of a classical oscillator with a frequency corresponding to the excitation energy. For the nth transition, it can be computed as

$$f_n = \frac{2\Omega_n}{3} \sum_{\alpha=\{x,y,z\}} |\left\langle \Psi_n \right| \hat{\mu}_\alpha \left| \Psi_0 \right\rangle|^2 \tag{3.52}$$

The oscillator strengths of all transitions in the system sum up to the total number of electrons. From the Casida equation, the oscillator strengths are obtained by

$$f_n = \frac{2}{3} \sum_{\alpha=\{x,y,z\}} |x_\alpha^T \sqrt{A-B}\vec{Z}_n)|^2. \tag{3.53}$$

[44, 45]

3.1.4. Density Functional Tight Binding

Density functional tight binding (DFTB) is a semi-empirical method based on DFT. It is about three orders of magnitude faster [3, 46] due to approximations explained in the following.

As tight binding method, DFTB relies on the assumption that core electrons are tightly bound by their nuclei and the valence electrons only are important to determine the electronic properties of a molecular system. In accordance, in the applied LCAO ansatz (Eq. 3.5) a minimal basis set is used which contains one function per valence orbital of an atom. These functions are determined from an atomic DFT calculation which includes an additional term to confine the wave function, respectively. This term contains a parameter, the compression radius r^{compr}. The exchange-correlation term is evaluated by means of the GGA functional PBE. [29, 46, 47]

The energy is determined in dependence of the electron density as done in DFT, but this density is differently computed. Starting from the DFT energy expression (Eq. 3.19), the density is replaced by the sum of a reference density ρ^0, and a fluctuation $\delta\rho$. The exchange-correlation energy E_{XC} is further expanded at the reference density, leading to the following expression for the total energy [48, 49]:

$$\begin{aligned} E[\rho^0+\delta\rho] = \sum_i n_i \left\langle \Psi_i \middle| -\frac{\Delta}{2} + V_{\mathrm{ne}} + \int \frac{\rho^{0\prime}}{|\vec{r}-\vec{r}^{\,\prime}|} d\vec{r}^{\,\prime} + V^{\mathrm{XC}}[\rho^0] \middle| \Psi_i \right\rangle \\ - \frac{1}{2}\iint \frac{\rho^0\rho^{0\prime}}{|\vec{r}-\vec{r}^{\,\prime}|} d\vec{r}d\vec{r}^{\,\prime} - \int V_{\mathrm{XC}}[\rho^0]\rho^0 d\vec{r} + E_{\mathrm{XC}}[\rho^0] + E_{\mathrm{nn}} \\ + \frac{1}{2}\iint \left(\frac{1}{|\vec{r}-\vec{r}^{\,\prime}|} + \left.\frac{\delta^2 E_{\mathrm{XC}}[\rho]}{\delta\rho\delta\rho'}\right|_{\rho^0,\rho^{0\prime}} \right) \delta\rho\delta\rho^{\prime} d\vec{r}d\vec{r}^{\,\prime} \\ + \frac{1}{6}\iiint \left.\frac{\delta^3 E_{\mathrm{XC}}[\rho]}{\delta\rho\delta\rho'\delta\rho''}\right|_{\rho^0,\rho^{0\prime},\rho^{0\prime\prime}} \delta\rho\delta\rho^{\prime}\delta\rho^{\prime\prime} d\vec{r}d\vec{r}^{\,\prime}d\vec{r}^{\,\prime\prime} \\ (+\text{higher orders}). \end{aligned} \tag{3.54}$$

The number of terms taken into account differs between the DFTB variants. Standard DFTB considers only the first two lines, DFTB2 truncates this Taylor series after 2nd order (third line in Eq. 3.54) and DFTB3 includes also the third order (fourth line in Eq. 3.54). The energy expression is further approximated, respectively, as follows.
Standard DFTB originates from solid-state physics and is best applicable to unpolar crystals consisting of only few different elements. The total energy of the system is written as a sum over the eigenvalues of non-self-consistent Kohn-Sham equations ϵ_i with occupation numbers n_i (for line one in Eq. 3.54) and short-range repulsive two-body potentials between atom pairs (for line two in Eq. 3.54), E_{rep} [49–51]:

$$E^0_{\mathrm{TB}} = E_{\mathrm{H}^0} + E_{\mathrm{Rep}} = \sum_{i=1}^{occ} n_i\epsilon_i + E_{\mathrm{Rep}} \tag{3.55}$$

Making use of the LCAO ansatz, the eigenvalues are determined from the secular equation

$$\sum_{\nu}^{M} c_{\nu i}(H^0_{\mu\nu} - \epsilon_i S_{\mu\nu}) = 0 \qquad \forall\mu, i. \tag{3.56}$$

with

$$H^0_{\mu\nu} = \langle \phi_\mu | \hat{H}^0 | \phi_\nu \rangle \qquad S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle \qquad \forall \mu \in \alpha, \nu \in \beta. \tag{3.57}$$

The orbitals ϕ stem from the minimal basis set described above with $r^{\text{compr}} = r^{\text{wf}}$. The Hamiltonian matrix is highly simplified: only two-center integrals are evaluated as

$$H^0_{\mu\nu} = \langle \phi^\alpha_\mu | \hat{T} + V^\alpha_0 + V^\beta_0 | \phi^\beta_\nu \rangle \qquad \alpha \neq \beta \tag{3.58}$$

with the atom indices α, β denoting the center of the functions and the spherical potentials V_0. The latter depend on the respective atomic densities which are determined from the basis functions with $r^{\text{compr}} = r^{\text{dens}}$. Additionally, the energies of the free neutral atoms are on the diagonal of the Hamiltonian matrix. They are obtained from DFT calculations without the confining term. The overlap integral S contains only two-center elements. For both, the Hamiltonian and the overlap matrix elements, precomputed values on a grid of distances are tabulated. The orbital orientation is accounted for by taking linear combinations of the tabulated results according to the Slater-Koster rules [52]. [48, 49, 51]

The repulsive energy E_{rep} in Eq. 3.55 can be computed from the difference of a reference SCF calculation with a higher level method and the tight-binding energies of a reference system (Eq. 3.59). [49, 51]

$$E_{\text{rep}}(R) = \left[E^{\text{SCF}}_{\text{ref}}(R) - \sum_{i=1}^{occ} n_i \epsilon_i(R) \right]_{\text{reference}} \tag{3.59}$$

E_{rep} can also be expressed as sum over short-ranged repulsive pair potentials [48, 50]:

$$E_{\text{rep}} = \frac{1}{2} \sum_{ab} V^{\text{rep}}_{ab}(\rho^0_a, \rho^0_b, R_{ab}) \tag{3.60}$$

These potentials are approximated by spline functions fitted to experimental or high-level theoretical reference data. Their range is a little longer than a typical covalent bond [47].

Heteronuclear molecules containing atoms of different electronegativity require the extensions DFTB2 and DFTB3. In DFTB2, an additional energy contribution E_γ is introduced (Eqs. 3.61 and 3.62), approximating the third line of Eq. 3.54. [48, 49]

$$E^{\text{DFTB2}} = E_{\text{H}^0} + E_\gamma + E_{\text{Rep}} \tag{3.61}$$

$$E_\gamma = \frac{1}{2} \sum_{\alpha\beta}^{N} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{3.62}$$

Δq_i is the net charge of atom i and $\gamma_{\alpha\beta}$ describes the electron-electron interaction. Thus, the electron density fluctuation is discretized to induced atomic charges. These are determined from a Mulliken population analysis. $\gamma_{\alpha\beta}$ approaches $\frac{1}{r_{\alpha\beta}}$ at large distances and $\gamma_{\alpha\alpha}$ is approximated by the Hubbard parameter U_α. This parameter can be computed with LDA or GGA DFT and is related to the chemical hardness of an element. A self-consistent procedure comes into play since the charges Δq depend on the orbitals Ψ_i. Due to this, DFTB2 is also called SCC-DFTB with SCC standing for *self-consistent charge*. [29, 46, 48, 49, 53]

DFTB2 describes charged molecules poorly because, first, U_α is independent of the charge state of the atom, and second, the atoms have fixed shapes. In DFTB3 this problem

is tackled by consideration of the derivative of $\gamma_{\alpha\beta}$ with respect to the charge. The energy expression reads [48]:

$$E^{\mathrm{DFTB3}} = E_{\mathrm{H}^0} + E_\gamma + E_\Gamma + E_{\mathrm{Rep}} \tag{3.63}$$

$$E_\Gamma = \frac{1}{3}\sum_{\alpha\beta} \Delta q_\alpha^2 \Delta q_\beta \Gamma_{\alpha\beta} \tag{3.64}$$

$$\Gamma_{\alpha\beta} = \left.\frac{\partial \gamma_{\alpha\beta}}{\partial q_\beta}\right|_{q_\beta^0}. \tag{3.65}$$

Long-range corrected Density Functional Tight Binding

DFTB suffers from the self-interaction error due to its DFT descent. This results for instance in an overdelocalization of the electrons and underestimated HOMO-LUMO gaps. Long-range corrected (LC) functionals are able to cure these issues to a large part. In LC-DFTB, the Coulomb operator is separated into a short-range and a long-range term using the Yukawa splitting (Eq. 3.66). The range-separation parameter ω determines the partioning between the short- and the long-range part. [3, 29, 54]

$$\frac{1}{|\vec{r}_1 - \vec{r}_2|} = \frac{\exp(-\omega \cdot |\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|} + \frac{1 - \exp(-\omega \cdot |\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|} \tag{3.66}$$

Eq. 3.67 shows the contributions to the exchange-correlation energy. The short-range DFT term is computed based on the functional of Baer, Neuhauser and Livshits (BNL) [55, 56], but with the correlation obtained from the PBE functional [33] and the exchange contribution obtained from LDA. The long-range part tends to the exact HF exchange for increasing distances. [3, 54]

$$E_{\mathrm{XC}} = E_{\mathrm{C}}^{\mathrm{PBE}} + E_{\mathrm{X}}^{\omega,\mathrm{LDA}} + E_{\mathrm{X}}^{\omega,\mathrm{HF}} \tag{3.67}$$

The expression for the total energy of LC-DFTB reads for a closed shell system [54]:

$$\begin{aligned} E^{\mathrm{LC-DFTB}} = &\sum_{\mu\nu} P_{\mu\nu} H^0_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu\alpha\beta} \Delta P_{\mu\nu} \Delta P_{\alpha\beta} \left[(\mu\nu|\alpha\beta) + f^{\omega,\mathrm{xc}}_{\mu\nu\alpha\beta}[\rho_0]\right] \\ &- \frac{1}{4}\sum_{\mu\nu\alpha\beta} \Delta P_{\mu\nu} \Delta P_{\alpha\beta} (\mu\alpha|\beta\nu)^{\mathrm{lr}} + E_{\mathrm{Rep}} \end{aligned} \tag{3.68}$$

with

$$\begin{aligned} H^0_{\mu\nu} = &h_{\mu\nu} + \sum_{\alpha\beta} \Delta P^0_{\alpha\beta} (\mu\nu|\alpha\beta) \\ &- \frac{1}{2}\sum_{\alpha\beta} P^0_{\alpha\beta} (\mu\alpha|\beta\nu)^{\mathrm{lr}} + v^{\omega,\mathrm{xc}}_{\mu\nu}[\rho^0] \end{aligned} \tag{3.69}$$

$h_{\mu\nu}$ is the integral accounting for the kinetic energy and the external potential. LC-DFTB uses a minimal valence basis set consisting of atom-centered functions like the traditional DFTB method but this time they are obtained with an LC functional. P stands for the density matrix containing the coefficients of the LCAO ansatz to describe the occupied molecular orbitals. $\Delta P = P - P^0$ is the difference between the final density matrix

at self-consistency and the density matrix of the reference density. $f^{\omega,\mathrm{xc}}$ represents the second derivative of the functional and $v^{\omega,\mathrm{xc}}$ the first derivative. $(\mu\nu|\alpha\beta)$ denotes the full-range Coulomb integral, $(\mu\nu|\alpha\beta)^{\mathrm{lr}}$ the long-range Coulomb integral. E_{Rep} is approximated by a sum of pair potentials which decay fast with increasing distance. The two-center approximation is applied to evaluate $H^0_{\mu\nu}$ as in usual DFTB. The four-center Coulomb integrals in Eq. 3.68 are reduced to the sum of four two-center integrals by means of the Mulliken approximation,

$$\phi_\mu(\vec{r})\phi_\nu(\vec{r}) \approx \frac{1}{2}S_{\mu\nu}(|\phi_\mu(\vec{r})|^2 + |\phi_\nu(\vec{r})|^2). \tag{3.70}$$

The resulting two-center integrals are approximated by a parametrized function $\gamma_{AB}(R_{AB})$, respectively. The included parameter τ_A, an atom specific decay constant, is determined from the LC-DFTB expression for the Hubbard parameter and the reference LC-DFT result for the Hubbard parameter. The final Hamiltonian used in the generalized Kohn-Sham equations reads [54]:

$$\begin{aligned} H_{\mu\nu} = H^0_{\mu\nu} &+ \frac{1}{4}\sum_{\alpha\beta}\Delta P_{\alpha\beta}S_{\mu\nu}S_{\alpha\beta}(\gamma^{\mathrm{fr}}_{\mu\alpha} + \gamma^{\mathrm{fr}}_{\mu\beta} + \gamma^{\mathrm{fr}}_{\nu\alpha} + \gamma^{\mathrm{fr}}_{\nu\beta}) \\ &- \frac{1}{8}\sum_{\alpha\beta}\Delta P_{\alpha\beta}S_{\mu\alpha}S_{\beta\nu}(\gamma^{\mathrm{lr}}_{\mu\beta} + \gamma^{\mathrm{lr}}_{\mu\nu} + \gamma^{\mathrm{lr}}_{\alpha\beta} + \gamma^{\mathrm{lr}}_{\alpha\nu}). \end{aligned} \tag{3.71}$$

LC-DFTB is computationally slightly more expensive than normal DFTB. [3, 54]

TD-LC-DFTB

For LC-DFTB as well as DFTB time-dependent variants were derived to compute excited states based on the Casida equation (Eq. 3.44) [3, 57–60]. The coupling matrix K in Eq. 3.45 reads in the case of TD-LC-DFTB [4]:

$$K_{ia\sigma,jb\sigma'} = \sum_{Al}\sum_{Bl'} q^{ia\sigma}_{Al}\Gamma^{\sigma\sigma'}_{Al,Bl'}q^{jb\sigma'}_{Bl'} - \delta_{\sigma\sigma'}q^{ij\sigma}_{Al}\gamma^{\mathrm{lr}}_{Al,Bl'}q^{ab\sigma'}_{Bl'}. \tag{3.72}$$

The indices A, B stand for two different atoms, l, l' are the angular momenta of the atom orbitals centered on these atoms. i, j label occupied MOs, a, b unoccupied MOs and p, q MOs in general. σ, σ' are the spin indices. $q^{pq\sigma}_{Al}$ denote the transition charges and are determined as follows [3, 57]:

$$q^{pq\sigma}_{Al} = \frac{1}{2}\sum_{\mu\in A,l}(c^\sigma_{\mu p}c^\sigma_{\mu q}S + c^\sigma_{\mu q}c^\sigma_{\mu p}S). \tag{3.73}$$

$\gamma^{\mathrm{lr}}_{Al,Bl'}$ is a function which approximates the following long-range Coulomb integral for $\mu = \{Alm\}, \nu = \{Bl'm'\}$ (m is the magnetic quantum number) [3]:

$$\gamma^{\mathrm{lr}}_{Al,Bl'} \approx \left(\mu\mu\left|\frac{1-\exp(-\omega\cdot|\vec{r}_1-\vec{r}_2|)}{|\vec{r}_1-\vec{r}_2|}\right|\nu\nu\right). \tag{3.74}$$

$\gamma^{\mathrm{fr}}_{Al,Bl'}$ approximates the same integral but with the full-range Coulomb operator plus $f^{\omega,\mathrm{xc}}$. Both γ functions depend on the distance of the atoms A and B and the Hubbard parameters $U_{Al}, U_{Bl'}$. $\Gamma^{\sigma\sigma'}_{Al,Bl'}$ is defined as

$$\Gamma^{\sigma\sigma'}_{Al,Bl'} = \gamma^{\mathrm{fr}}_{Al,Bl'} + \delta_\sigma\delta_{\sigma'}\delta_{AB}W_{Al,l'}. \tag{3.75}$$

$W_{Al,l'}$ are atomic constants. [3]

Eq. 3.44 is evaluated by solving the squared non-hermitian eigenvalue problem [4]:

$$(A-B)(A+B)(\vec{X}+\vec{Y}) = \Omega^2(\vec{X}+\vec{Y}) \tag{3.76}$$

$$(\vec{X}-\vec{Y}) = \Omega^{-1}(A+B)(\vec{X}+\vec{Y}). \tag{3.77}$$

Forces which correspond to the derivative of the energy with respect to the molecular coordinates are needed for geometry optimization or molecular dynamics simulation. The derivative of Ω with respect to a nuclear displacement s reads [4]:

$$\begin{aligned}\frac{d\Omega}{ds} = &\sum_{\mu\nu\sigma}\left(\frac{\partial H^0_{\mu\nu}}{\partial s}D_{\mu\nu\sigma} - \frac{\partial S_{\mu\nu}}{\partial s}W_{\mu\nu\sigma}\right) \\ &+ \sum_{\mu\nu\sigma\kappa\lambda\sigma'}\frac{\partial(\mu\nu|v_C + f^{\mathrm{XC},\omega}_{\sigma\sigma'}|\kappa\lambda)}{\partial s}\tilde{\Gamma}_{\mu\nu\sigma,\kappa\lambda\sigma'} \\ &+ \sum_{\mu\nu\sigma\kappa\lambda\sigma'}\frac{\partial(\mu\nu|v^{\mathrm{lr},\omega}_{\mathrm{C}}|\kappa\lambda)}{\partial s}\tilde{\Gamma}^{\mathrm{lr}}_{\mu\nu\sigma,\kappa\lambda\sigma'}\end{aligned} \tag{3.78}$$

with

$$\tilde{\Gamma}_{\mu\nu\sigma,\kappa\lambda\sigma'} = D_{\mu\nu\sigma}\Delta P_{\kappa\lambda\sigma'} + (\vec{X}+\vec{Y})_{\mu\nu\sigma}(\vec{X}+\vec{Y})_{\kappa\lambda\sigma'} \tag{3.79}$$

$$\begin{aligned}\tilde{\Gamma}^{lr}_{\mu\nu\sigma,\kappa\lambda\sigma'} = &-\frac{1}{2}\delta_{\sigma\sigma'}[2D_{\mu\lambda\sigma}\Delta P_{\nu\kappa\sigma} \\ &+ (\vec{X}+\vec{Y})_{\mu\lambda\sigma}(\vec{X}+\vec{Y})_{\kappa\nu\sigma} + (\vec{X}+\vec{Y})_{\kappa\mu\sigma}(\vec{X}+\vec{Y})_{\lambda\nu\sigma} \\ &+ (\vec{X}-\vec{Y})_{\kappa\mu\sigma}(\vec{X}-\vec{Y})_{\lambda\nu\sigma} - (\vec{X}-\vec{Y})_{\kappa\nu\sigma}(\vec{X}-\vec{Y})_{\mu\lambda\sigma}].\end{aligned} \tag{3.80}$$

D is the orbital-relaxed one-particle difference density matrix and composed of the unrelaxed difference density matrix T and the solution of the Z-vector equation, Z (Eq. 3.82).

$$D = T + Z \tag{3.81}$$

$$\sum_{jb\sigma'}(A+B)_{ia\sigma,jb\sigma'}Z_{jb\sigma'} = -Ria\sigma \tag{3.82}$$

R stands for a response expression. W in Eq. 3.78 denotes the energy-weighted difference density matrix. It contains the contributions arising from changes in the metric of the basis functions due to the geometrical changes. [4]

3.2. Classical Mechanics

In classical mechanics, molecular configurations are generated according to Newton's equations of motion. Newton's second law provides the connection between the force $\vec{F}$ acting

on a particle i with mass m and its acceleration $\vec{a}$, which is the second derivative of its position $\vec{r}$ with respect to the time t (Eq 3.83). $\vec{v}$ denotes the velocity.

$$\vec{F}_i = m_i \cdot \frac{d^2\vec{r}_i}{dt^2} = m_i \cdot \frac{d\vec{v}_i}{dt} = m_i \cdot \vec{a}_i. \tag{3.83}$$

The force can also be expressed as the negative of the derivative of the potential energy V at position $\vec{r}$:

$$\vec{F}_i = -\frac{\partial V(\vec{r}_1, \dots)}{\partial \vec{r}_i} = m_i \cdot \frac{d^2\vec{r}_i}{dt^2}. \tag{3.84}$$

The potential energy is approximated by a force field as described below. If initial coordinates and velocities are provided, the coordinates and velocities after a time step Δt can be computed from the Taylor expansion around the actual position vector:

$$\vec{r}_{t+\Delta t} = \vec{r}_t + \frac{d\vec{r}}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{r}}{dt^2}(\Delta t)^2 + \frac{1}{6}\frac{d^3\vec{r}}{dt^3}(\Delta t)^3 + ... \tag{3.85}$$

In this work, the leap-frog integrator is used to compute the new positions and velocities with time (Eqs. 3.86 and 3.87). It is correct to third order.

$$\vec{r}_{t+\Delta t} = \vec{r}_t + \vec{v}_{t+\frac{1}{2}\Delta t} \cdot \Delta t \tag{3.86}$$

$$\vec{v}_{t+\frac{1}{2}\Delta t} = \vec{v}_{t-\frac{1}{2}\Delta t} + \vec{a}_t \cdot \Delta t \tag{3.87}$$

The advantages of the leap-frog algorithm are its numerical accuracy, the time-reversibility and the direct occurence of the velocities. However, the velocities and the positions are not provided at the same points in time. [61, 62]

The potential energy arising from the interactions between the atoms in the simulated system is modeled by a so-called *force field*. It describes bonded and nonbonded interactions. The bonded interactions include covalent bonds, angles and dihedrals and the non-bonded interactions include the Van der Waals and Coulomb interactions. One example for a force field expression is given in Eq. 3.88. There, the energy of a bond or angle is modeled by a harmonic potential with a force constant k_b (k_θ) and an equilibrium value b_0 (θ_0). The energy contribution of a dihedral angle is described by a sum of several cosine functions with a multiplicity n, phase δ and amplitude k_ϕ. The nonbonded Van der Waals interactions are modeled for example by a Lennard-Jones 12-6 potential with the parameters σ and ε. σ corresponds to the distance where the potential is zero and ε is the depth of the potential well. $|\vec{r}_A - \vec{r}_B|$ is the distance between two atoms A and B and the sum runs over all pairs of atoms for which non-bonded interactions shall be calculated. The Van der Waals interactions include the Pauli repulsion (1st term) and the attractive dispersion interaction due to induced dipoles (2nd term). The electrostatic interaction between the atoms is considered by the Coulomb term with the atomic partial charges q and the vacuum permittivity ε_o. The parameters used in the force field expression can be determined from experiments or from high-level *ab initio* or density functional calculations. [63–66]

$$E_{\mathrm{MM}} = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} \sum_{n=1,2,3,4,6} k_{\phi n}[1+\cos(n\phi+\delta_n)] \tag{3.88}$$
$$+ \sum_{\substack{nonbonded\\pairs}} \left\{ 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{|\vec{r}_A - \vec{r}_B|} \right)^{12} - \left(\frac{\sigma_{AB}}{|\vec{r}_A - \vec{r}_B|} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_A q_B}{|\vec{r}_A - \vec{r}_B|} \right\}$$

3.3. Enhanced Sampling Techniques

Enhanced sampling techniques help to overcome high energy barriers which prevent the sampling of all relevant conformations in usual molecular dynamics simulations. Two important categories of these techniques are (i) replica exchange, like parallel tempering [67] or Hamiltonian replica exchange [68] and (ii) methods applying reaction coordinates (collective variables), as for instance thermodynamic integration [69], free energy perturbation [70], umbrella sampling [71], steered MD [72] or metadynamics [73]. Metadynamics simulations are explained in more detail, since they play an important role in chapter 6.

Metadynamics

In metadynamics, a history-dependent bias is added along pre-defined collective variables (CVs) which is intended to fill the free energy landscape of the system. Collective variables can be simple geometrical measures such as distances or angles but also complex functions. The essential point is that they represent coordinates on the free energy surface which feature the large barriers to overcome. In other words, they describe the connection between extrema of the conformations to be sampled. The bias consists of Gaussian functions which are deposited on the CVs every given time interval and then summed up. As soon as an energy minimum is filled, the barrier to another minimum vanishes and thus can be overcome (Fig. 3.1). When all minima are filled, the metadynamics simulation is converged and the free energy is obtained from the negative bias (Eq. 3.89).

Eq. 3.89 is the central equation showing the negative of the bias $V(\vec{s},t)$ constructed from a sum of Gaussians which tends at increasing time t to the free energy $F(\vec{s})$ up to a constant. $W(m\tau)$ is the height of the Gaussian, s_i denotes the ith collective variable, σ_i gives the width of the Gaussian and m times the time interval τ gives the time of the last Gaussian deposition. [73–76]

$$-V(\vec{s},t) = -\sum_{m\tau \le t} W(m\tau) \cdot \exp\left(-\sum_{i}^{d} \frac{(s_i(\vec{q}) - s_i(\vec{q}(m\tau)))^2}{2\sigma_i^2} \right) \rightarrow F(\vec{s}) + C \tag{3.89}$$

While in standard metadynamics simulations the height of the added Gaussian functions is constant, it is variable in well-tempered metadynamics (Eq. 3.90) [76–78]. In the latter the initial height W_0 is scaled depending on the value of the bias and thus also simulation time.

$$W(m\tau) = W_0 \cdot \exp\left(-\frac{V(\vec{s}(\vec{q}(m\tau)), m\tau)}{k_B \Delta T} \right) \tag{3.90}$$

The parameter ΔT, which has the dimension of temperature, can be used to tune the scaling. It can also be expressed in terms of the bias factor γ, which is defined as:

$$\gamma = \frac{T + \Delta T}{T} \tag{3.91}$$

with T being the temperature of the system. [77, 78]

In contrast to standard metadynamics, the well-tempered variant provides a smoothly filled free energy landscape and convergence is more easily achieved. However, it does not fill the landscape completely. The real free energy can be estimated as [77, 78]

$$\tilde{F}(\vec{s}, t) = -\frac{T + \Delta T}{\Delta T} V(\vec{s}, t)^{\text{well-tempered}}. \tag{3.92}$$

In sum, the main advantages of metadynamics are that it allows a faster sampling of conformations, promotes finding of unknown minima on the free energy surface and provides directly the relative free energies in the CV space. The disadvantage on the other hand is the need of CVs which are not always easy to define. They have to be chosen such that all free energy minima are accessible and distinguishable.

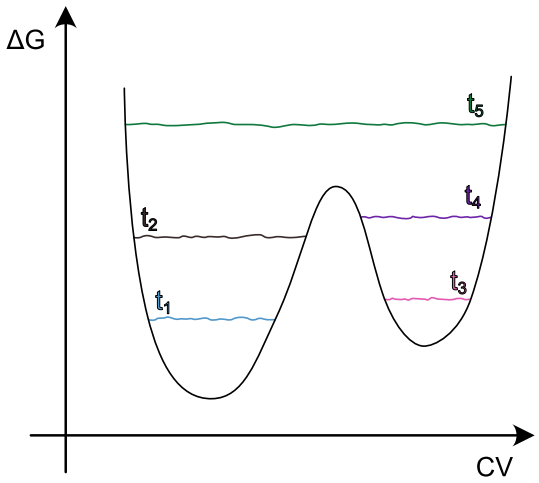


Fig. 3.1.: Working principle of metadynamics. The simulation starts in the left free energy minimum and with time, this minimum is filled by added Gaussian potentials such that the barrier to the right well can be overcome. At t_5 both minima are filled and the simulation converged.

3.4. Multiscale Modelling

Depending on the studied process, the system needs a more or less detailed description. For example, to study large conformational changes of a protein, it may be sufficient to use a coarse-grained model and to compute the interactions between pseudo-atoms representing a group of atoms with a classical force-field (e.g. [79]). If the interactions between for instance a protein and a ligand are of interest, the systems needs to be modelled in atomic

detail to detect specific hydrogen bonds (e.g. [80]). To describe the interaction of a molecule with light, quantum mechanics is needed. However, increasing the accuracy of the model also increases the computational cost. Hence, for the different levels of accuracy there are different limits regarding the system size and the possible simulation length.

An issue occurs when a process needs a high-level description but depends on long-range environmental effects. One example is the absorption of chromophores in a protein (e.g. [81]). In such cases, several different approaches need to be combined. The system is divided into subsystems according to the required accuracy and each subsystem is treated with a different method. Typically, the small high-level subsystem is surrounded by the larger subsystem treated with a less computationally demanding method. In doing so, the subsystems interact with each other and have an impact on each other. This can be implemented in different ways, as shown in the following.

In this work, quantum mechanics (QM) at a semiempirical level and classical mechanics (MM) are combined to describe the optical properties of chromophores. On the example of such QM/MM simulations, the different embedding variants of an additive QM/MM scheme are discussed. The total energy of the system is the sum of the energy of the QM part, the MM part and the energy due to the interaction between both parts:

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}. \tag{3.93}$$

The mechanical embedding is the simplest way to calculate $E_{\mathrm{QM/MM}}$. Here, the interactions between QM and MM region are described by the force field. For this, atomic charges need to be assigned to the atoms in the QM region. They can be obtained from a population analysis, so for example they are Mulliken charges. In this scheme, the MM environment has no impact on the electron density in the QM region which can be a big disadvantage.

In the electronic embedding scheme instead the atomic partial charges of the MM part are included in the QM calculation thus polarising the electron density in the QM part. The Hamiltonian for the electronic part of the QM/MM interaction includes thus the interaction of the MM point charges with the QM electrons and QM nuclei and reads in atomic units [64]:

$$H^{\mathrm{el}}_{\mathrm{QM/MM}} = -\sum_{i}^{N}\sum_{j}^{L}\frac{q_j}{|r_i - R_j|} + \sum_{\alpha}^{M}\sum_{j}^{L}\frac{q_j Q_\alpha}{|R_\alpha - R_j|}. \tag{3.94}$$

The q_j are L partial charges of the MM atoms at positions R_j, Q_α are the nuclear charges of the M QM atoms and r_i are the positions of the N electrons. This embedding scheme works well in many cases and is applied later in this work.

The polarised embedding scheme considers both, the polarisation of the QM part by the MM charges and the polarisation of the MM part from the QM charge distribution. This requires an iterative procedure which increases the computational cost significantly. [64,82]

Bonded interactions and Van-der-Waals interactions between the QM and MM regions are treated at the MM level. [64]

A critical point for such QM/MM simulations are boundaries of the different regions going through covalent bonds. The QM molecule(s) must be saturated to give meaningful results. In this work, the so-called link atom approach is used. This means, that a link atom is inserted on the bond connecting QM and MM region at a meaningful distance. This link atom is included as hydrogen in the QM calculation. Since this atom is close to

the MM region, the neighbouring MM charges could show a too large impact on this part of the QM region. Therefore, it often makes sense to set the charge of the closest MM atom to zero and divide it on other atoms in the proximity. In any case it is important to choose an unpolar bond like a C-C bond as QM/MM boundary. [64]

Mulliken Population analysis

In a population analysis, atomic gross charges are determined by assigning electrons from the overall electron density to the individual atoms. The Mulliken population analysis [83] is quite simple and can be easily derived for a diatomic molecule. Starting from an LCAO-MO ansatz with normalized AOs χ, $\Phi = c_r\chi_r + c_s\chi_s$, this equation is squared and multiplied on both sides by the number N of electrons in Φ, which is typically two. If the resulting expression is integrated over the whole space then the total number of electrons in the molecular orbital Φ is found to consist of three contributions:

$$N = Nc_r^2 + 2Nc_rc_sS_{rs} + Nc_s^2. \tag{3.95}$$

S_{rs} is the overlap integral $\int \chi_r\chi_s dV$. The first and the third term in Eq. 3.95 are the atomic gross populations of the atoms r and s, respectively, and the second term is the overlap population. This overlap population is equally distributed to both atoms.

In the general case of a polyatomic molecule, Eq. 3.95 is replaced by Eq. 3.96. The index i counts the molecular orbitals Φ_i which are a linear combination of r_k atomic orbitals. s_l denotes a different AO than r_k. k and l are the indices for the atoms.

$$N(i) = N(i)\sum_{r_k} c_{ir_k}^2 + 2N(i)\sum_{l>k} c_{ir_k}c_{is_l}S_{r_ks_l} \tag{3.96}$$

The final gross atomic population is the sum of the population of the atom and half of each overlap population with another atom. The net atomic charge (partial charge) is the difference between the number of electrons in the free, neutral atom in ground state ($N^0(k)$) and the computed population of the atom in the molecule (Eq. 3.97). [83]

$$\begin{aligned} Q_k =& N^0(k) - N(k) \\ =& N^0(k) - \sum_i \sum_r N(i)c_{ir_k}(c_{ir_k} + \sum_{l\neq k} c_{is_l}S_{r_ks_l}) \end{aligned} \tag{3.97}$$

3.5. Machine Learning

In simple words, doing Machine learning means to program a computer (*machine*) such that it improves on a task (*learns*) through data fed to it. Such a task could be for example to classify data or to make quantitative predictions based on regression. Machine learning is particularly helpful if input and output data show complex correlations or coding of a progam by hand would consume a lot of time for writing extensive lists of rules. Examples for machine learning application areas are image recognition, spam filtering or the prediction of molecular properties. Depending on the machine learning method, the trained machine is finally not only able for example to sort or predict data but the pattern between input and output becomes evident. [84]

The various machine learning techniques can be divided by different criteria. One is the supervision of the learning process. Supervised learning means that in the training process the input data is given together with the output data, called labels. Classification and regression belong to this category. Unsupervised learning in contrast means that the machine learns without labels. This is the case for instance by clustering or dimensional reduction. [84]

Another criterion to sort the different machine learning techniques is if incremental learning is possible or not. In the case of batch learning (offline-learning), the machine needs all available training data in the learning process and afterwards it can only predict. Thus, this variant is not useful if the system needs to learn new data continuously. In the case of incremental learning (online-learning), the machine is trained step by step by adding mini-batches of data and hence is able to adapt quickly to new data. Moreover, incremental learning needs less computing capacity than batch learning. [84]

Finally, it is also possible to divide machine learning techniques into instance-based and model-based learning. Instance based learning means that the machine gives an output for new data points based on their similarity to memorized input data points. In accordance to its name, model based learning uses a model, for example a linear function, and determines the parameters of the model by means of the input data. This can be done by minimizing a cost function. [84]

The success of machine learning is largely dependent on the amount and quality of the input data. In principle, the more input data is given to the machine the better it learns and thus the better are the predictions afterwards. Obtaining such a big amount of data can be challenging and time consuming. Furthermore, it is important that the data is representative. Predictions outside of the trained data range will likely be inaccurate. Besides, noisy training data makes it difficult for the machine to recognize an underlying pattern. Hence, the training data should be checked for outliers and noise before starting the learning process and cleaned if necessary. Another point to be aware of is what kind of input data is expected to be related to what kind of output. For example, it may be possible to obtain the total energy of a molecule as output if the molecular geometry is given as input. But the energy of the molecule will not be generally predictable from the number of atoms alone. [84]

Besides the input data, the success of machine learning depends on the algorithm used. In general, one needs to take care that the model does not adapt too closely to the input data so that predictions of unseen data become worse. This possible issue is called overfitting. It typically arises when the chosen model is too complex with respect to the amount of input data and the noise therein. One way to avoid overfitting is the regulation of the learning process by means of hyperparameters. These parameters are parameters of the learning algorithm in contrast to the parameters learned for the model and are set before the training. The training process can also be stopped at the moment when overfitting starts. For this, the predictions of the model are directly tested during the training with another data set not provided for training. At the beginning, the results on this validation data set improve with proceeding learning but at the moment overfitting starts, the results start to become worse. This is then the time to stop the training. Of course, there are some fluctuations in the loss function value, and so the training should only be stopped as soon as the loss function exhibits a clear increasing trend. Underfitting is the opposite of overfitting and occurs if a model is too simple. [84, 85]

Neural networks

Artificial neural networks (ANN) are the model of choice in this work. A simple architecture of an ANN is the single-layer perceptron. It consists of an input layer which only transmits the input data and one layer of linear threshold units (LTUs) which are a class of neurons. An LTU computes the weighted sum of its inputs and subsequently applies an activation function which determines the output. Thus, such a perceptron can work as a classifier. In the training process, the weigths of the input neurons used in the LTU calculation are learned. [84]

Larger ANNs (Fig. 3.2) contain additional layers between the input and output layer, so-called hidden layers. The outputs of the neurons in one layer serve as input for the next layer. Hence, the output of a neuron n_k in the ith layer of a neural network with Nj neurons per layer can be expressed as

$$n_j^{\mathrm{i,out}} = f\left(\sum_{j'}^{N_j} w_{jj'}^{i-1} n_{j'}^{i-1,\mathrm{out}}\right) \tag{3.98}$$

with j being the index of the neuron in layer i and j' the index of a neuron in layer $i-1$. The weights $w_{jj'}$ between neurons j and j' are the coefficients of the linear combination. The output of the j'th neuron from the previous layer, $n_{j'}^{i-1,\mathrm{out}}$, is itself the results of the activation function f for a linear combination of neuron outputs from layer $i-2$.

An ANN can be trained using the backpropagation algorithm. This means the output of the ANN for a given input is computed and compared to the expected output. Then it is analyzed, starting from the last hidden layer backwards, how much each neuron contributed to the wrong result. Finally, the weights used in the neurons are adapted. The backpropagation algorithm requires a differentiable activation function in contrast to the simple one-layer perceptron introduced before. [84]

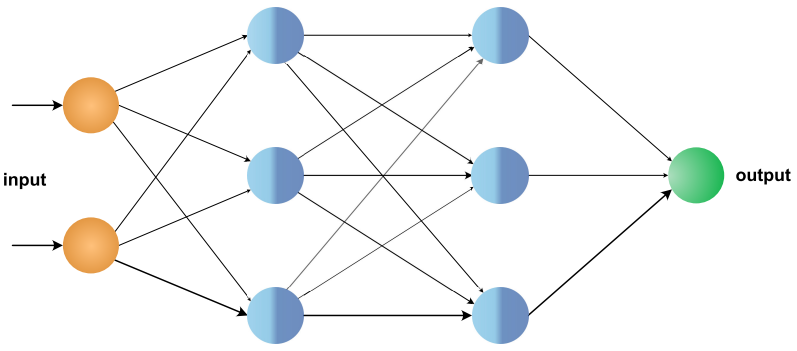


Fig. 3.2.: Architecture of a neural network: input layer (orange), hidden layers (blue), output layer (green). The neurons in the hidden layers make first a linear combination of the results from the previous layer (light-blue part) and then apply an activation function (dark-blue part). The weights used in this linear combinations can be imagined as the strength of the connection between two neurons of two neighboured layers and are symbolized by the different widths of the arrows connecting the neurons.

3.6. Optimization Techniques

In this subchapter two algorithms are introduced which can be used to minimize a function. They are suitable to find the energetically most favorable structure of a biomolecular system or the geometry with lowest energy of a single molecule.

3.6.1. Steepest Descent

The algorithm starts with the computation of the gradient vector $\vec{g}$. This vector reveals the steepest increase of the function. Thus, the strategy is to always make a step in the opposite direction $\vec{d}$ of the gradient:

$$\vec{d} = -\vec{g}. \tag{3.99}$$

The value of the function in this direction decreases first but at a certain point it will start to increase. If the step is made until the lowest possible point in that direction, then the next step in the opposite direction of the gradient at that point will be perpendicular to the previous step. However, in practice such a *line search* to find the optimal step size is usually not performed, since it is computationally demanding. Instead, a certain step size is chosen and dynamically adapted during the optimization: if a step ends on a lower point of the potential then the size of the next step is slightly increased, otherwise it is decreased. However, the algorithm may start to oscillate around the minimum. A threshold needs to be set for convergence or a maximum number of steps. A drawback of the steepest descent algorithm are the oscillations around the direct path to the minimum, even if a line search is performed. This makes this method unnecessary slow in the case of narrow valleys. On the other hand, the advantages of steepest descent are its simplicity, the low storage requirement and, if a line search is performed, the guarantee to come close to the minimum. [61]

3.6.2. Conjugate Gradient

The conjugate gradient algorithm starts with a steepest descent step but then does not follow the negative gradient any more but walks on a line which considers the negative gradient and the previous search direction:

$$\vec{d_i} = -\vec{g}_i + \beta_i \vec{d}_{i-1}. \tag{3.100}$$

This is done to reduce the number of steps needed to approach the minimum. The definition of β according to Polak-Ribiere is shown in Eq. 3.101.

$$\beta_i = \frac{\vec{g}_i^t(\vec{g}_i - \vec{g}_{i-1})}{\vec{g}_{i-1}^t \vec{g}_{i-1}} \tag{3.101}$$

The search directions are *conjugated* in the sense that the orthogonality condition,

$$d_i^t A d_j = 0 \tag{3.102}$$

holds for a matrix A containing the second derivatives of a quadratic surface. The conjugate gradient algorithm works well for (nearly) quadratic surfaces and then converges faster than the steepest descent algorithm. [61]

The steepest descent algorithm with line search and the conjugate gradient algorithm are compared in Fig. 3.3 showing the steps taken from the same starting point in a potential to the point of minimal energy. Without the line search, it would be more difficult to reach the minimum with steepest descent.

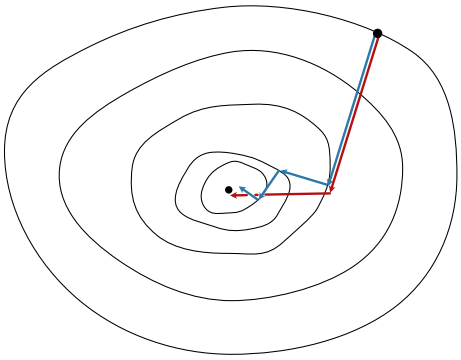


Fig. 3.3.: Different ways of the optimizers steepest descent (blue) and conjugate gradient (red) during an optimization in an approximately quadratic potential indicated by contour lines. The steps of steepest descent are perpendicular to each other as long as the step size is optimal. The conjugate gradient algorithm reaches the minimum faster.

4 Benchmark of TD-LC-DFTB2 Gradients

In parts reprinted with permission from
M. Sokolov, B. M. Bold, J. J. Kranz, S. Höfener, T. A. Niehaus and M. Elstner. 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. J. Chem. Theory Comput., 17(4):2266-2282, 2021.

DOI: 10.1021/acs.jctc.1c00095.

This chapter contains an extensive benchmark of TD-LC-DFTB2 gradients and transition energies. The motivation is explained first and then follow the computational details, the results and the discussion of the various quantities tested, respectively. The chapter closes with conclusions on the suitability of TD-LC-DFTB2 for fluorescence studies.

4.1. Motivation

The fluorescence of dye molecules in general is dependent on their dynamics and their specific environment. QM methods can be used to optimize the geometry which provides the most important structure in the case of a deep energy minimum. This can also be done under consideration of the environment, either implicitly or explicitly modelled. However, if the excited state potential energy surface of the dye molecule is relatively flat, then several conformations need to be taken into account to explain its fluorescence spectrum. These conformations can be obtained from sampling in the excited state, while the fluorophore can be embedded in implicit solvent or constitute the QM part in a QM/MM simulation. *Ab initio* methods and DFT are computationally too demanding for significantly long simulations and excitations cannot be described in the classical picture. However, the dynamics in excited state can be studied with semi-empirical methods. TD-LC-DFTB2 is a promising candidate as it is fast and incorporates a long-range correction for an improved description of charge-transfer states. Its ability to predict excitation energies is already assessed in refs [3,81,86]. In order to simulate the dynamics of a molecule, gradients of the energy (forces) are needed. For TD-LC-DFTB2, analytical forces in the excited state were recently derived and implemented for the GAMESS and DFTB+ programs [4,87,88].

Before these gradients are applied to simulate fluorophores, a detailed benchmark of their accuracy is needed. In this chapter, the performance of TD-LC-DFTB2 and in particular the gradients in the first excited state are benchmarked using DFTB+. The

first excited state is of special interest, since in most cases the fluorescence occurs from this state. Optimized geometries, different quantities depending on the geometries and potential energy surfaces (PES) are assessed. In particular, the geometries are analyzed with respect to bond length alternations, substituent orientation and resulting excitation energies. Further, adiabatic excitation energies and reorganization energies are evaluated and the PES are investigated via the vibrational frequencies. Finally, vibronic spectra are simulated.

It is aimed to evaluate the TD-LC-DFTB2 method first on a pure theoretical level instead of directly comparing to the experiment, since this allows a better understanding of the detailed strengths and weaknesses of the approach. DFT as the underlying QM method of LC-DFTB as well as high-level *ab initio* methods are eligible references. Furthermore, it is compared to DFTB/3OB which as very fast method may be viewed as an alternative to LC-DFTB2. DFTB3 is of third order in ground state while LC-DFTB2 is of second order. However, it does not contain the long-range correction but is based on a GGA functional.

4.2. Benchmark Set

The quality of a benchmark study depends on a meaningful reference as well as on the amount and diversity of the included molecules. The more than 100 molecules used for this benchmark are part of the benchmark sets from Ref. [89] and Ref. [90]. Ref. [89] provides accurate *ab initio* CC2 structures and DFT/B3LYP structures as reference. In this set several isomers are included, which makes it possible to study the ability of the method to reproduce trends between very similar molecules. In Ref. [90] the results of different DFT functionals including B3LYP are presented. This benchmark set shows a larger diversity of molecules compared to Ref. [89].

LC-DFTB2 repulsive parameters were only available for C, N, H, O, thus only molecules consisting of these atoms could be calculated[1]. From the benchmark set of Ref. [90] additionally small inorganic molecules were excluded since they do not belong to the application area of the tight binding methods. Often they cause problems in the calculations and they can be calculated instead in reasonable time with high-level *ab initio* methods or DFT. Moreover, only the first singlet excited states were taken into account.

4.3. Geometries

As first evaluation, the geometries optimized with TD-LC-DFTB2 are directly compared to the reference geometries. Differences in the bond length alternations in conjugated systems are studied quantitatively. Additionally, a qualitative comparison with respect to substituent orientation is given.

4.3.1. Computational Details

The molecules were optimized in the electronic ground state and the first excited state with (TD-)LC-DFTB2/OB2 in DFTB+ (development version based on 19.1). All three OB2 parameter sets for LC-DFTB2 from Ref. [92] were tested: $OB2^{0.3}$(base), $OB2^{0.3}$(shift) and $OB2^{0.3}$(split). The conjugate gradient algorithm was used and the default threshold for the

[1] In the meantime, a reparametrization including sulfur is available [91]

maximum force, 1×10^{-4} $\frac{Ha}{a_0}$. The SCC convergence criterium accounted to 1×10^{-8} electrons and the range-separation parameter to 0.3 a_0^{-1}. The thresholded screening method was applied with a threshold set to 1×10^{-16}. Subsequently, for each molecule a frequency analysis was carried out using the mode program of DFTB+ to check if the optimization converged into a minimum. In case of one or more imaginary frequencies the corresponding molecule was reoptimized starting from a different geometry. The excited state optimizations of hexatriene and styrene using the OB2 (base) parameters did not end in a minimum and thus are not considered in the further analysis. Dimethylaminobenzonitrile using the OB2 (split) and toluene using the OB2 (shift) or OB2 (split) parameters are also left out due to imaginary frequencies.

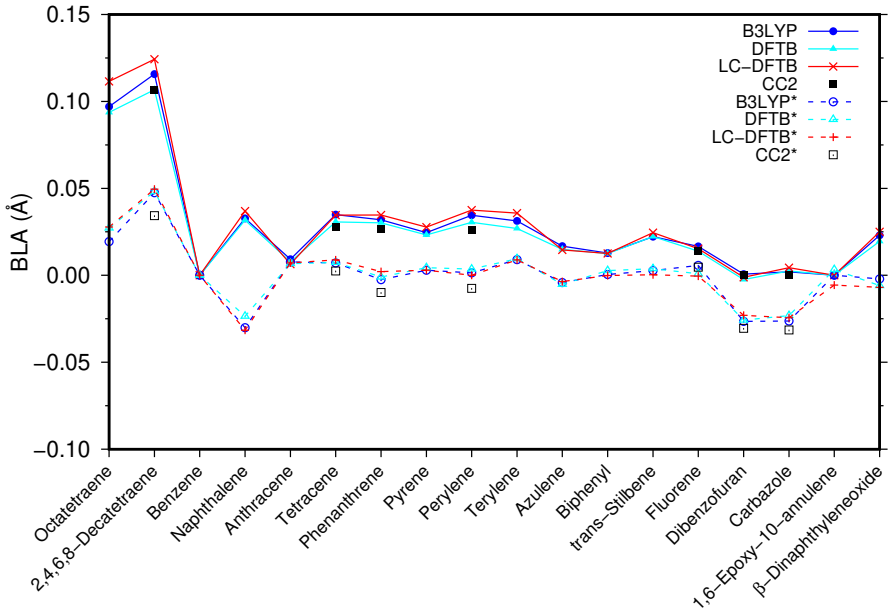
The DFTB calculations were performed under the same conditions. DFTB3 with 3OB parameters was applied for the ground state and the excited state, whereas the excited state is not of third order but computed based on the DFTB3 ground state density. The DFTB results for the tryptamine-anti-ph and tryptamine-anti-py isomers are left out since they converged to different isomers in the excited state. The optimization of formic acid did not converge in the excited state and is thus neglected.

TD-LC-DFT excited state optimizations were performed for a subset of molecules with flexible side chains. The CAM-B3LYP [39] and ωB97X [93] functionals were applied with the def2-TZVP basis set [94] in the ORCA program package (version 4.0.0.2) [95] using the Tamm-Dancoff approximation (TDA) [96]. The resolution of identity approximation (RI) [97] was used with the def2/J auxiliary basis set [98] and Coulomb integrals were treated with the RI-J approximation and the HF exchange with COSX numerical integration [99]. The optimizations were performed with an energy convergence criterium of 1×10^{-6} (keyword TightOpt), followed by a frequency analysis. The convergence tolerance of the excited state energies and for the norm of the residual vectors was set to 1×10^{-6}. Always only the first excited state was considered.

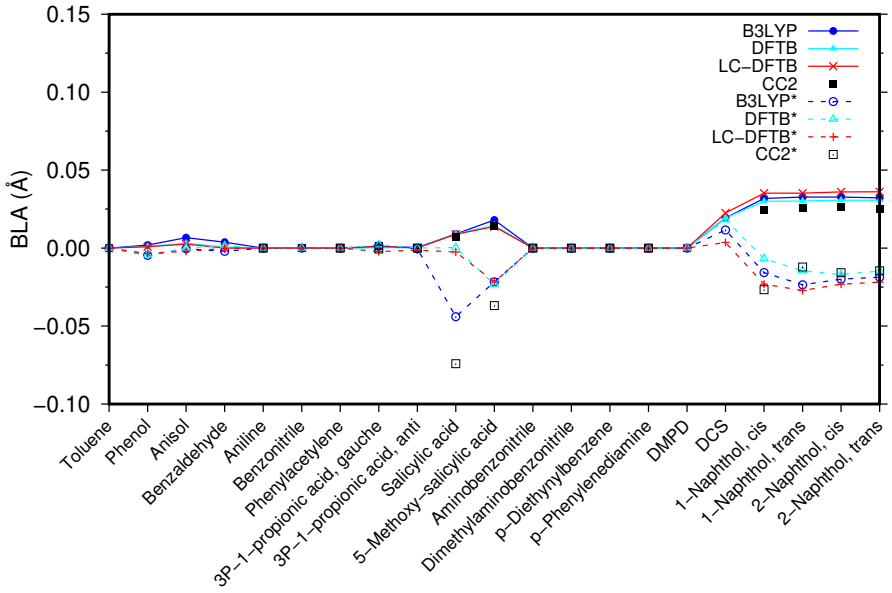
The bond length alternation (BLA) measures the similarity of the bond lengths in a conjugated system and is defined as the difference between the average length of the single bonds and the average length of the double bonds. In principle, it is an arbitrary choice which bonds are included in the calculation. Here generally all bonds belonging to the aromatic system of the respective molecules were included in the calculation. For tetraphenylporphine, porphycene, porphine and chlorin only the BLA of the diaza[18]annulene substructure was computed. Substituents of benzene rings were not included in the BLA calculation, except in methyl-4-hydroxycinnamate where the vinyl part is considered. The bonds were assigned to single and double bonds such that the BLA of the ground state geometry was positive. The BLA values were computed by means of VMD [100].

4.3.2. Results and Discussion

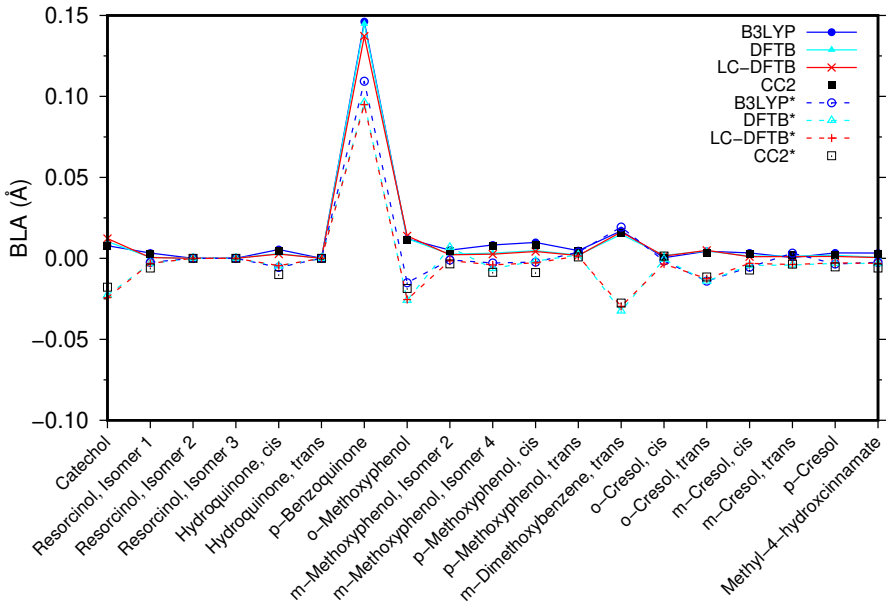
The results presented in the following were obtained with the OB2 (base) parameters. In general, the results obtained with the OB2 (shift) and OB2 (split) parameters are very similar.



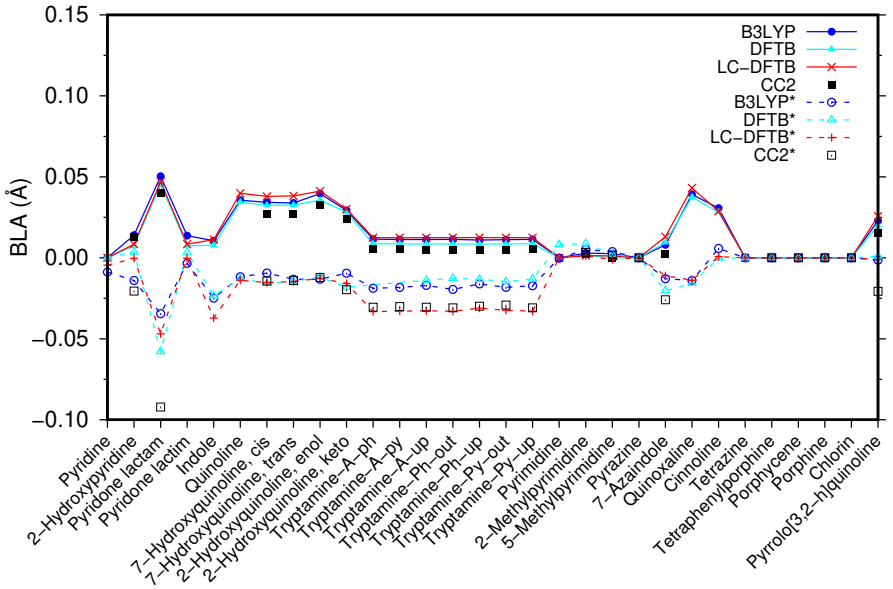
(a) Molecules without heteroatoms (until fluorene), dibenzofuran, carbazole, 1,6-epoxy-10-annulene and β-dinaphthyleneoxide.



(b) Monosubstituted benzenes, disubstituted benzenes and naphthols.



(c) Disubstituted benzenes.



(d) Ring systems containing one or more nitrogens.

Fig. 4.0.: BLA of ground (solid line) and excited state (dashed line) optimized structures. Molecules are divided into subgroups, see subplots (a) to (d). CC2 geometries were only available for the molecules from Ref. [89]. A negative BLA indicates a bond inversion in the excited state.

In Fig. 4.0, the BLA values in ground and excited state are presented for the B3LYP and CC2 reference structures, (TD-)LC-DFTB2/OB2 (base) and (TD-)DFTB/3OB. Since the BLA values depend on the included bond types (if for example a C-N bond is present the result cannot be compared to only C-C bonds), the molecules are not sorted according to their original benchmark set but according to their similarity. As mentioned above, CC2 reference values are only available for the molecules from the set of ref [89]. A negative BLA value indicates a bond inversion in the excited state. The closer the BLA is to zero, the more delocalized the system.

In general, the BLA obtained with (TD-)LC-DFTB2/OB2 (base) in ground and excited state agrees well with the B3LYP and CC2 reference values. Exceptions are (i) octatetraene and 2,4,6,8-decatetraene for which the BLA in ground state is slightly overestimated (octatetraene: 0.015 Å to B3LYP reference, decatetraene: 0.018 Å to CC2 reference); (ii) salicylic acid and (iii) 2-hydroxypyridine for which the BLAs in excited state are close to zero instead of significantly negative (deviations of TD-LC-DFTB2/OB2 (base) from CC2: 0.072 Å and 0.020 Å). For (iv) pyridone lactam (=2-pyridone) in the excited state the BLA of the CC2 reference structure is significantly more negative than the BLA of the other methods, too (deviation of TD-LC-DFTB2/OB2 (base) from CC2: 0.045 Å), but all methods predict a clear bond inversion. (v) The CC2 reference structure of pyrrolo[3,2-h]quinoline has a clear negative BLA in the excited state but the TD-LC-DFTB2/OB2 (base) value is close to zero. In this case, however, also the B3LYP value is almost zero.

The results of (TD-)DFTB/3OB are very similar to the (TD-)LC-DFTB2/OB2 (base) results. While the BLA values of octatetraene and decatetraene are better, it failes for salicylic acid and 2-hydroxypyridone, too, and provides similar BLA values for pyrrolo[3,2-h]quinoline and pyridone-lactam as B3LYP and TD-LC-DFTB2/OB2 (base). For the tryptamine isomers, TD-DFTB/3OB predicts values very close to B3LYP while TD-LC-DFTB2/OB2 (base) is very close to CC2.

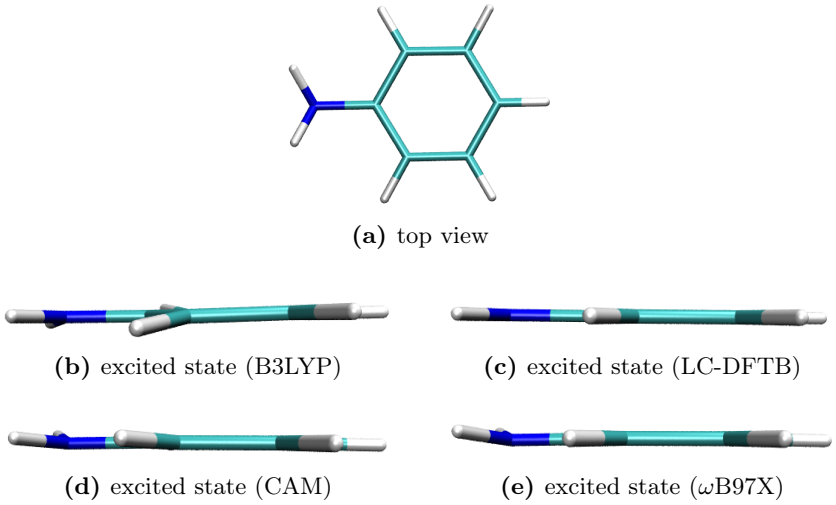
In sum, (TD-)LC-DFTB2/OB2 (base) provides reasonable BLA values for almost all molecules when taking CC2 and B3LYP results as reference and in this regard showed to be comparably accurate as (TD-)DFTB/3OB. This is promising since the quality of the BLA prediction can have an impact on the optical properties, as shown in refs. [101–103] for the carotenoids retinal and peridinin. The influence of the BLA of bacteriochlorophyll a on its excitation energies is discussed in Ref. [81].

After the analysis of the conjugated systems by means of the BLA comparisons, a look is taken on flexible side chains. A qualitative comparison of TD-LC-DFTB2/OB2 (base) optimized geometries and the respective reference geometries reveals frequently differences in side chain orientations. In Figs. 4.1-4.8 the excited state geometries of a couple of substituted aromatic compounds are compared. While the ground state LC-DFTB2/OB2 (base) geometries agree (by eye) well with the B3LYP reference geometries, in the excited state the substituent orientations are different in most cases. However, also the LC-DFT geometries frequently differ from B3LYP or from each other.

In detail, the optimized aniline geometry looks slightly different for all of the used methods (Fig. 4.1), but the aminobenzonitrile geometries of TD-LC-DFTB2/OB2 (base) and all DFT functionals agree well (Fig. 4.2). The predicted dimethylaminobenzonitrile and DMPD geometries again depend on the applied method (Figs. 4.3, 4.4). The TD-LC-DFTB2/OB2 (base) excited state geometries of para-phenylenediamine, phenol and

trans-stilbene differ from the B3LYP geometry but agree with the TD-LC-DFT geometries (Figs. 4.5, 4.6, 4.7). The pyridine geometry predicted by TD-LC-DFTB2/OB2 (base) instead deviates from the (LC-)DFT functionals (Fig. 4.8).

These few examples show that already the DFT functionals give different results in many cases. This may be traced back to the underlying functionals but also to the existence of several local minima on a flat potential energy surface. The latter is in particular expected for methyl groups where the hydrogens can rotate. The impact of differences in side chain orientations on optical properties is not obvious and must be checked individually.



(a) top view

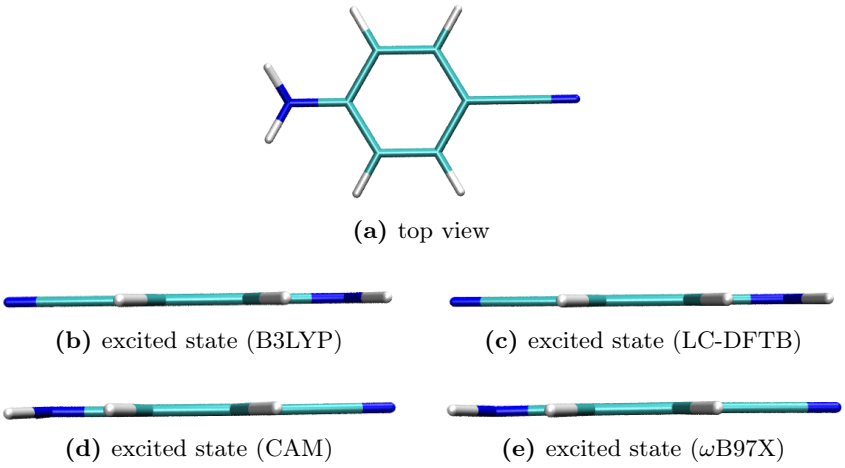
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.1.: Optimized geometries of aniline.



(a) top view

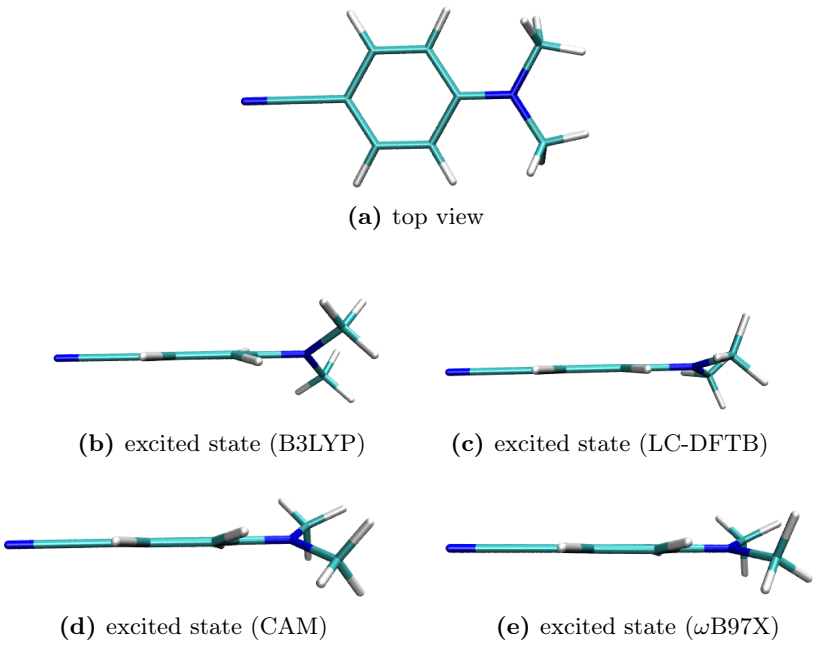
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.2.: Optimized geometries of aminobenzonitrile.



(a) top view

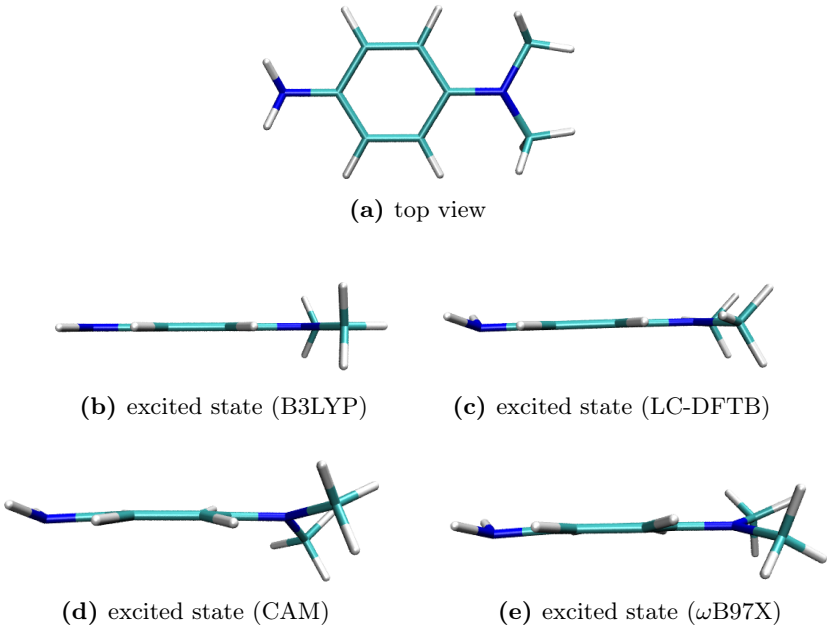
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.3.: Optimized geometries of dimethylaminobenzonitrile.



(a) top view

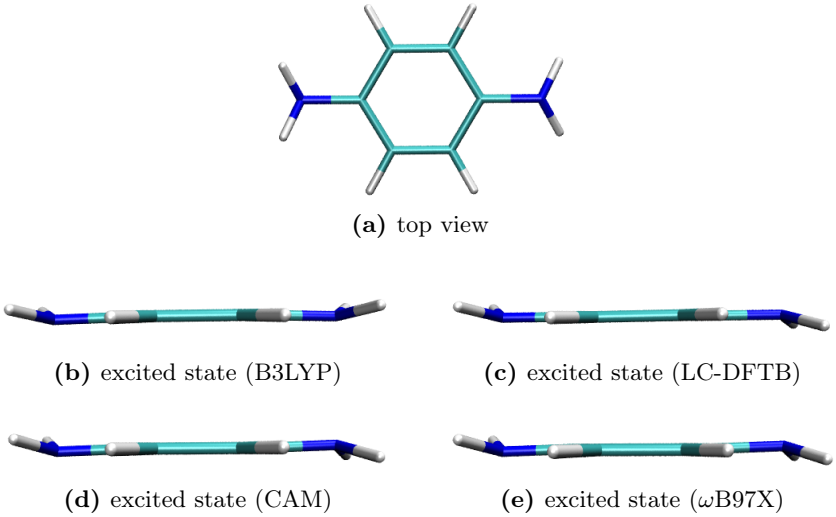
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.4.: Optimized geometries of DMPD.



(a) top view

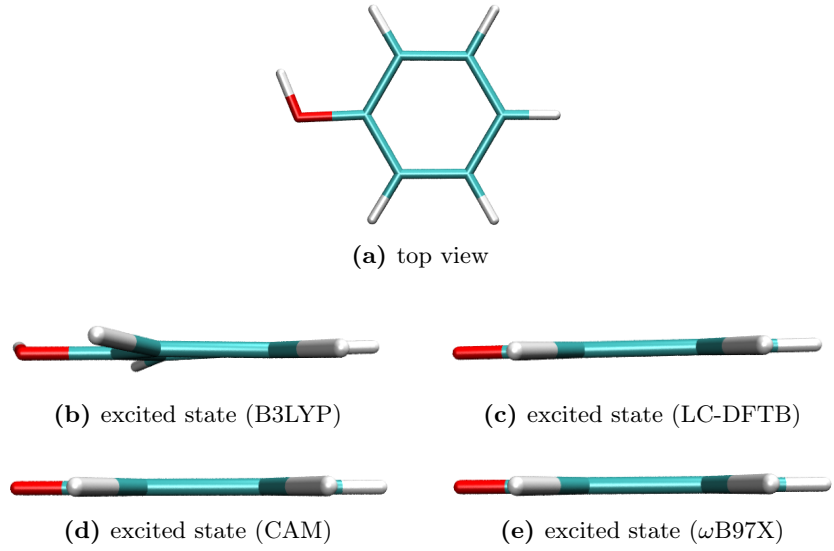
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.5.: Optimized geometries of para-phenylenediamine.



(a) top view

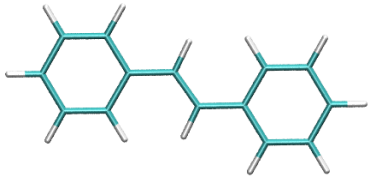
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.6.: Optimized geometries of phenol.



(a) top view

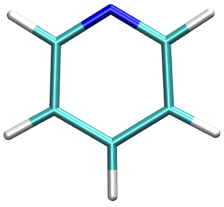
(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.7.: Optimized geometries of *trans*-stilbene.



(a) top view

(b) excited state (B3LYP)

(c) excited state (LC-DFTB)

(d) excited state (CAM)

(e) excited state (ωB97X)

Fig. 4.8.: Optimized geometries of pyridine.

4.4. Transition Energies

In the previous section, the similarities and differences of the TD-LC-DFTB2 and (LC-) DFT and CC2 geometries were analyzed. Now, adiabatic excitation energies, which depend on both, the accuracy of the geometries and the excitation energies, are investigated. The impact of geometrical differences with respect to excitation energies is revealed by the theoretical Stokes shifts presented afterwards. The Stokes shifts can also be viewed as total reorganization energies. The part of it corresponding to the reorganization energy in the excited state is evaluated afterwards.

4.4.1. Computational Details

Adiabatic excitation energies (AEEs) are defined as the difference between the total energies of two states at their respective equilibrium geometry (see Fig. 2.2). Here, the two states correspond to the ground state and the first excited state, respectively. Zero-point vibrational corrections were not taken into account and were accordingly subtracted from the reference data.

The energy of the first excited state was determined in two ways: to the total ground state energy at the equilibrium geometry of the first excited state (i) the first excitation energy obtained with the OB2 (base) parameters [92] was added or (ii) the first excitation energy obtained with a parameter set optimized for excitation energies [3, 104] was added. While the OB2 (base) parameters contain repulsive potentials, the latter parameter set is pure electronical. The variant (ii) which combines the parameter sets will be referred to as LC-DFTB2/comb.

Additionally, the AEEs were computed with the OB2 (shift) and OB2 (split) parameters. They contain repulsive parameters and in contrast to the base parameters it was particularly accounted for the balance of the sp^2 and sp^3 hybridizations of nitrogen in the paramterization by either a shift of the s-orbital energy or by a separate confining of the s and p orbitals [92]. These shift and split parameters were used for both optimization and excitation energy calculation to determine the AEEs.

The Stokes shift is defined as the energy difference between the maximum of the absorption and emission band in UV/vis spectra, which includes vibrational progression. In theoretical approaches, however, the difference of the vertical absorption and fluorescence energies of a dye, i.e. without vibrational progression, is often understood as Stokes shift, also in this work.

The structures optimized with TD-LC-DFTB2/OB2 (base) were used as well as the CC2 reference structures from Ref. [89] and DFT/B3LYP reference structures from Ref. [90] for comparison. The excitation energy calculations were performed with TD-LC-DFTB2/OB2 (base), TD-DFTB/3OB and TD-DFT employing the hybrid B3LYP [36–38] and the range-separated CAM-B3LYP [39] functionals together with the Karlsruhe basis set def2-TZVP, respectively. The DFT calculations were done with the ORCA program package (version 4.0.0.2) using the TightSCF keyword and applying the TDA and the RIJCOSX approximation. The target excited state was always the first excited state. Vertical excitation energies from B3LYP are already provided in the Supporting Information of Ref. [90]. However, these were recalculated to be consistent with our CAM-B3LYP calculations.

Additionally, Stokes shifts were computed with the OB2 (shift) and OB2 (split) parameter sets on the reference geometries and also with these parameters on the geometries

optimized with the respective same parameters.

The reorganization energy in the excited state is defined as the energy difference of the first excited state at the ground-state geometry and the relaxed excited state geometry. It can be computed as the difference between the vertical excitation energy on the ground state geometry and the adiabatic excitation energy.

In Ref. [90], vertical excitation energies were provided and used to compute the reference values. The molecules from Ref. [89] are left out here, because reference values are not available. In this work, reorganization energies were computed method consistently with TD-LC-DFTB2/OB2 (base) and TD-DFTB/3OB as difference between the respective vertical excitation energies and adiabatic excitation energies. Additionally, reorganization energies were obtained with CAM-B3LYP/def2-TZVP applying the TDA and RIJCOSX approximations and the def2/J auxiliary basis set using the ORCA program package (version 4.0.0.2) [95]. TightSCF criteria were applied. For the optimizations, an energy convergence criterium of 1×10^{-6} (keyword TightOpt) was used. Subsequently, a frequency analysis was performed. The following molecules did not converge to a minimum geometry with CAM-B3LYP: β-Dinaphthyleneoxide, biphenyl, DCS, terylene, toluene and trans-stilbene. For these molecules no reorganization energies are reported.

4.5. Results and Discussion

LC-DFTB2/OB2 (base) and LC-DFTB2/comb AEEs are compared to CC2 and B3LYP (with aug-cc-pVTZ correction) results of Ref. [89] and to B3LYP results of Ref. [90] in Figs. 4.9 and 4.10. These figures also include the AEEs obtained with DFTB/3OB. Qualitatively, it is observed that LC-DFTB2 is able to follow the trend of the reference values. However, it slightly overestimates the absolute values on average if only the OB2 parameters are used. DFTB/3OB instead generally underestimates the AEEs. These findings are in agreement with Ref. [105], where a TD-LC-DFTB2 benchmark is presented for TD-LC-DFTB2 implemented in GAMESS-US. A statistical evaluation is given in Tab. 4.1. It reveals that the DFTB/3OB and LC-DFTB2/OB2 (base) results are of similar quality. The maximal errors (MAX) of LC-DFTB2/OB2 (base) are smaller but the mean absolute error (MAE) are larger. The mean errors (ME) of LC-DFTB2/OB2 (base) and DFTB/3OB are almost equal for the set of Ref. [89], while the ME of LC-DFTB2/OB2 (base) is smaller for the set of Ref. [90]. The combination of the parameter sets improves the LC-DFTB2 results since the excitation energies provided by the second parameter set are lower than the excitation energies from the OB2 parameters. The LC-DFTB2/comb results are very close to the B3LYP results and outperform DFTB/3OB.

The reason for the overestimation of LC-DFTB2/OB2 (base) and the underestimation of DFTB/3OB could be the use of exact exchange in the functional of LC-DFTB2 in constrast to the GGA functional applied in the parametrization of DFTB. The electronical parameter set used to compute the excitation energies in the combined approach were derived with confinement radii $r_{d,wf}$ taken from the *mio-0-1* parameter set [49] but scaled by a factor of 0.8. The reduction of these confinement radii causes a decrease in the excitation energies and thus countersteers the overestimation of the AEEs in most cases. The results are promising but this does not necessarily mean that the geometries and excitation energies are perfect since it may be due to error cancellation.

Tab. 4.1.: Statistical evaluation of the errors for the AEEs in eV, using CC2 as reference for the set of Ref. [89] and B3LYP as reference for the set of Ref. [90]. LC-DFTB2/OB2 (base) is abbreviated by LC rep. and LC-DFTB2/comb is abbreviated by LC comb.

	Set [89],[a]				Set [89],[b]			Set [90],[b]		
	B3LYP	DFTB	LC rep.	LC comb.	DFTB	LC rep.	LC comb.	DFTB	LC rep.	LC comb.
MAX	-0.45	-1.30	0.84	-0.67	-1.36	0.77	-0.55	-1.39	1.36	0.80
ME	-0.04	-0.35	0.31	-0.08	-0.32	0.34	-0.04	-0.32	0.26	0.00
MAE	0.15	0.36	0.44	0.21	0.32	0.45	0.16	0.35	0.38	0.18

[a] Analysis with respect to CC2.
[b] Analysis with respect to B3LYP.

In addition, the AEEs obtained with the OB2 shift and split parameters are compared to the results of the base parameters and the reference values. Evidently, the AEEs of the three OB2 parameter sets are of comparable quality and very similar in most cases (Figs. 4.11 and 4.12). The largest differences between base and shift are found for tetrazine, cinnoline and DBH, which have in common that they contain N=N bonds. The shift parameter results for these molecules are more than 0.1 eV lower than the corresponding base parameter results. The base and split AEEs differ most in the case of para-phenylenediamine (0.1 eV).

It is not surprising, that the differences in the results from the various parameter sets are found for molecules with nitrogen, since this atom is differently treated in the parametrization process providing the shift and split parameters. Obviously, there is a strong effect on molecules with N=N bonds.

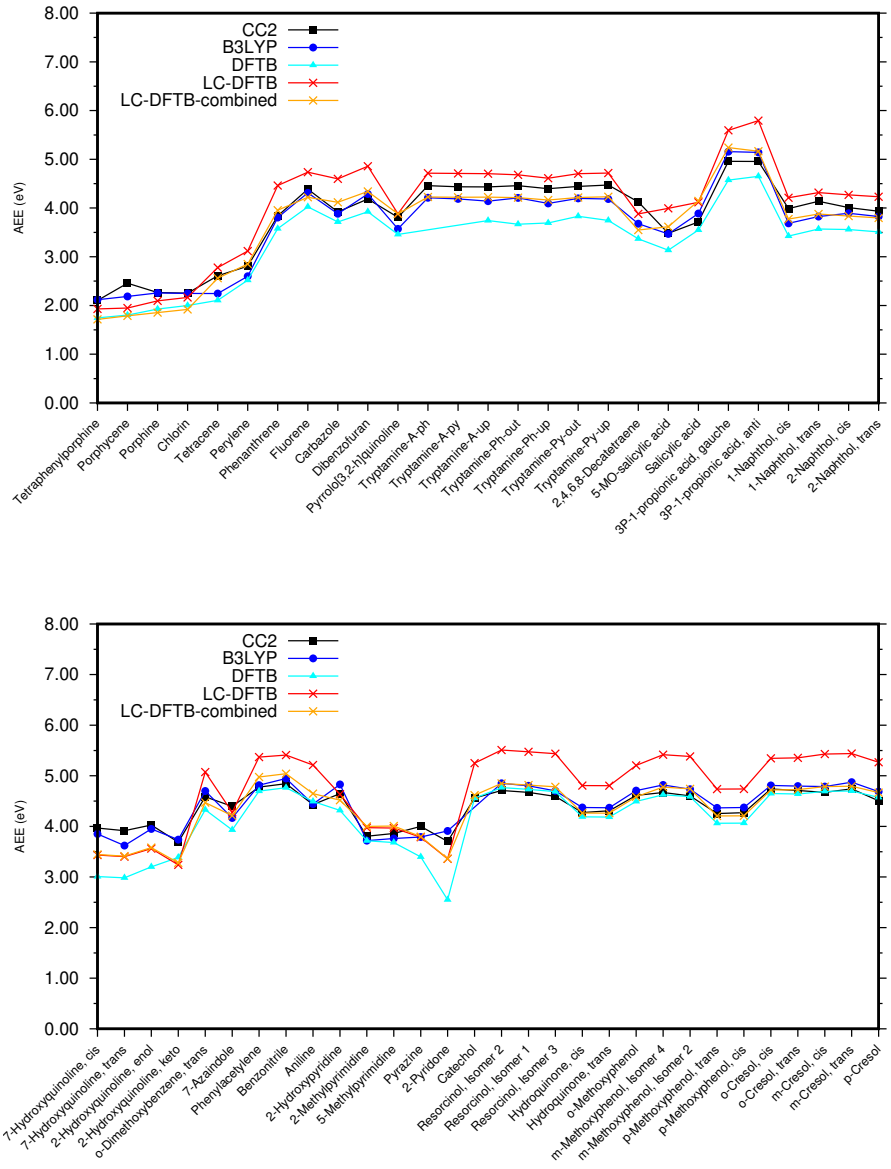


Fig. 4.9.: AEEs of the benchmark set from Ref. [89] adopting the same numbering of the isomers. The AEEs were computed with LC-DFTB2/OB2 (base), LC-DFTB/comb and DFTB/3OB. CC2 (reference) and B3LYP values were taken from Ref. [89].

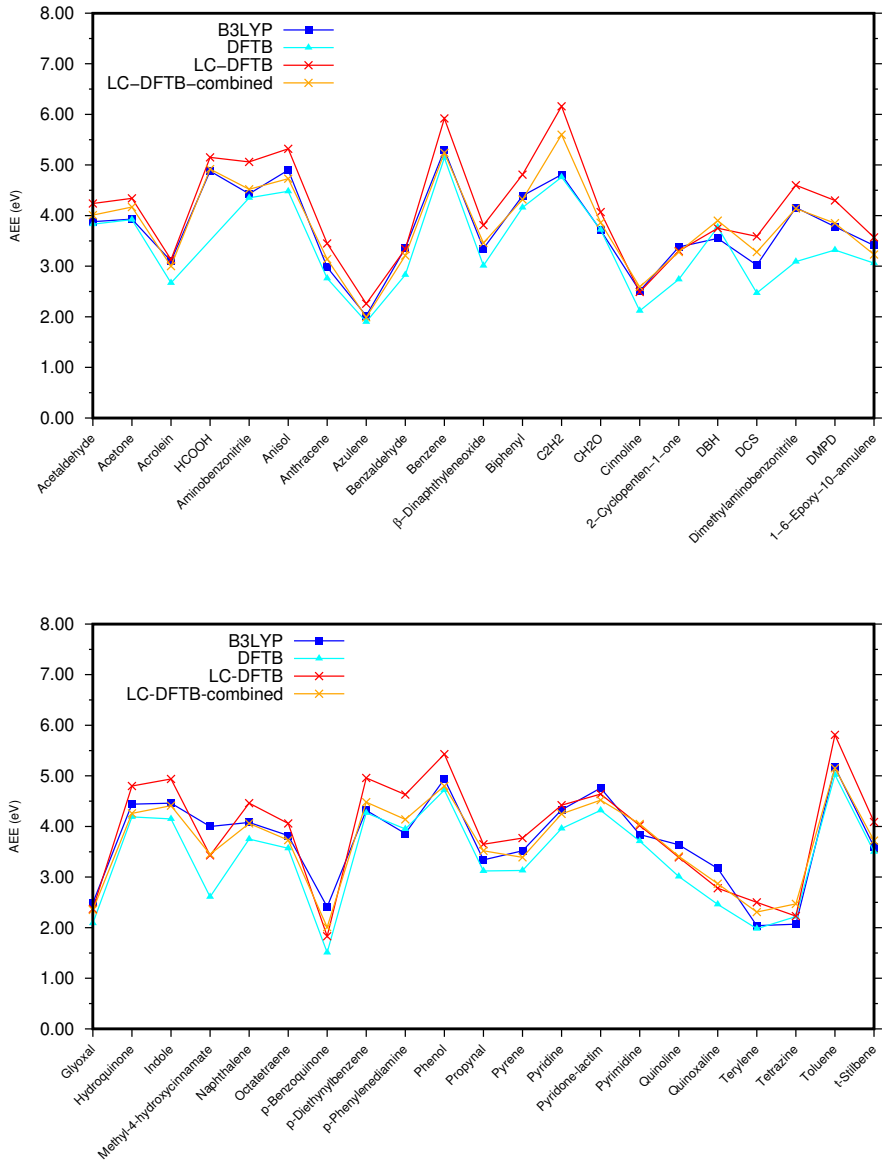


Fig. 4.10.: AEEs of the benchmark set from Ref. [90] computed with LC-DFTB2/OB2 (base), LC-DFTB/comb and DFTB/3OB. The B3LYP values (reference) were taken from Ref. [90].

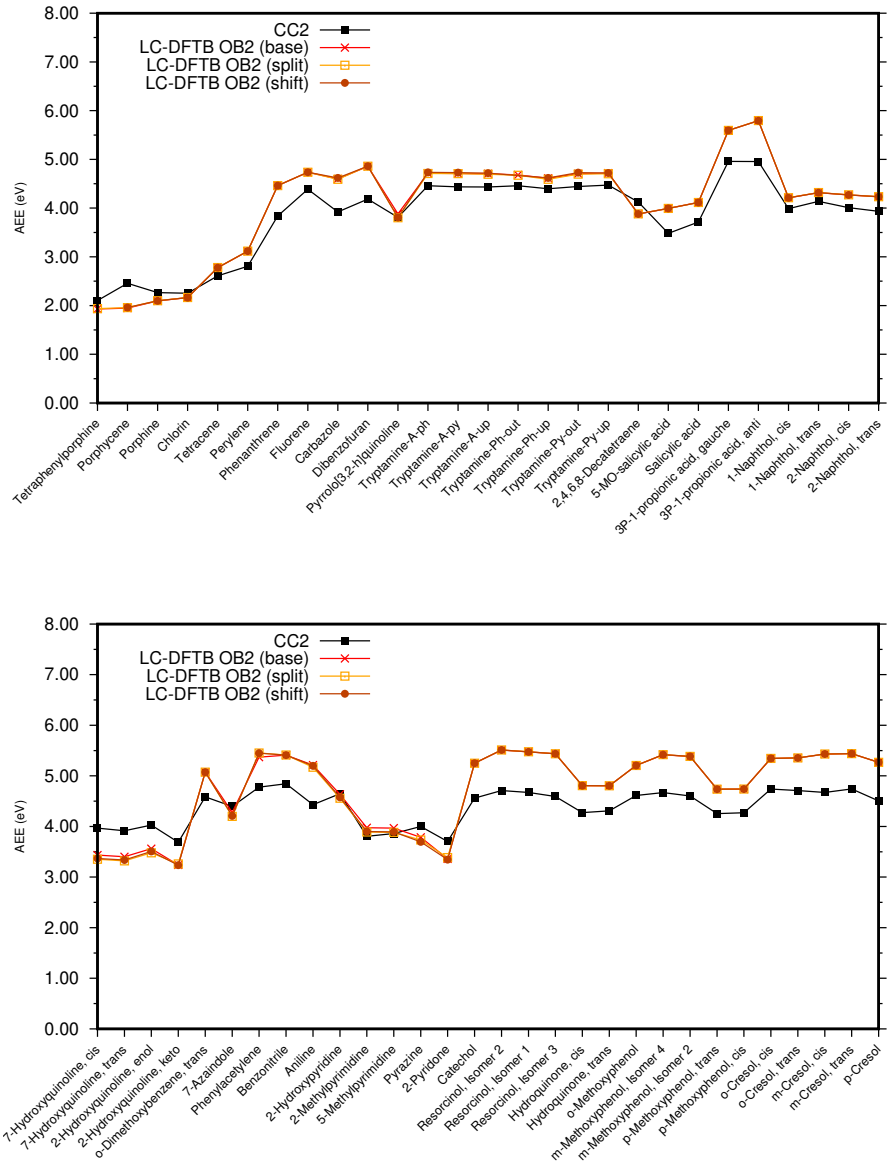


Fig. 4.11.: AEEs of the benchmark set from Ref. [89] computed with LC-DFTB2 and OB2 base, shift and split parameters. CC2 values (reference) were taken from Ref. [89].

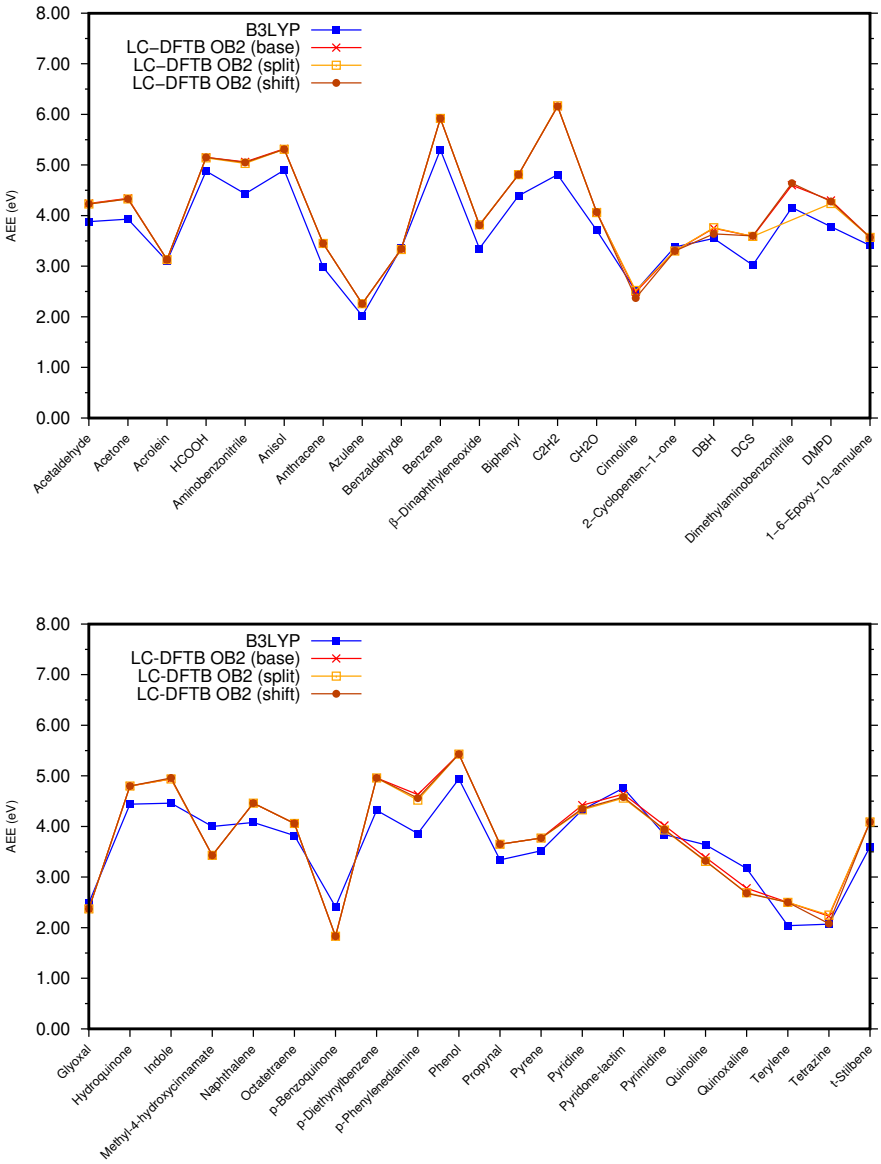


Fig. 4.12.: AEEs of the benchmark set from Ref. [90] computed with LC-DFTB2 and OB2 base, shift and split parameters. B3LYP values (reference) were taken from Ref. [90].

Now Stokes shifts are analyzed to assess the accuracy of the geometries. They are defined here as energy difference of the first excitation energy on the ground state geometry and the first excitation energy on the first excited state geometry. Thus, general systematic errors in the excitation energy calculations should have no impact on the results but geometrical differences determining the excitation energies become visible. Fig. 4.13 shows the Stokes shifts computed for LC-DFTB2/OB2 (base) (dashed lines) and CC2 (solid lines) geometries, excitation energies are evaluated with CAM-B3LYP, B3LYP, DFTB/3OB and LC-DFTB2/OB2 (base). In Fig. 4.14, Stokes shifts are presented which were computed for geometries optimized with B3LYP and LC-DFTB2/OB2 (base). In both figures most times the computed Stokes shifts agree sufficiently but some outliers are obtained.

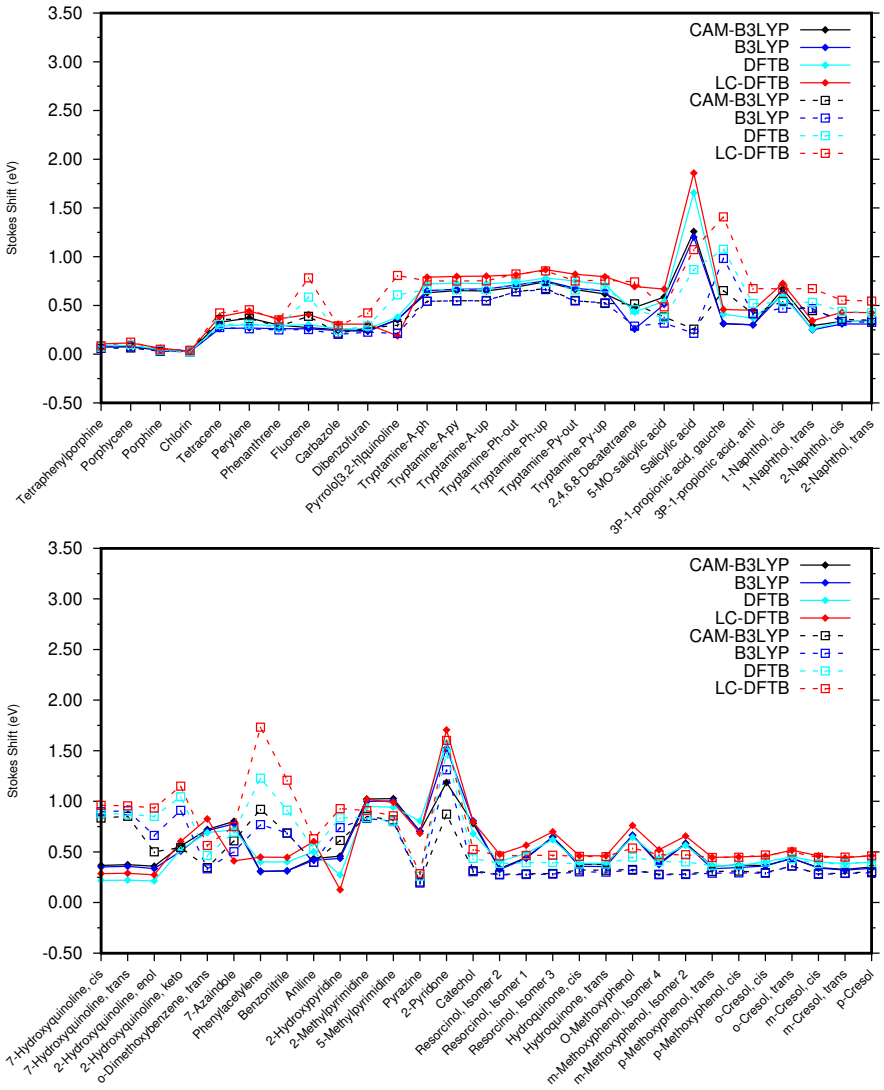


Fig. 4.13.: Stokes shift of first excited state for the molecules from Ref. [89]. The solid lines show the Stokes shifts obtained with CC2 optimized geometries and the dashed lines the Stokes shifts obtained with the LC-DFTB2/OB2 (base) optimized geometries. The different colors indicate the different methods used for the calculation of vertical excitation energies.

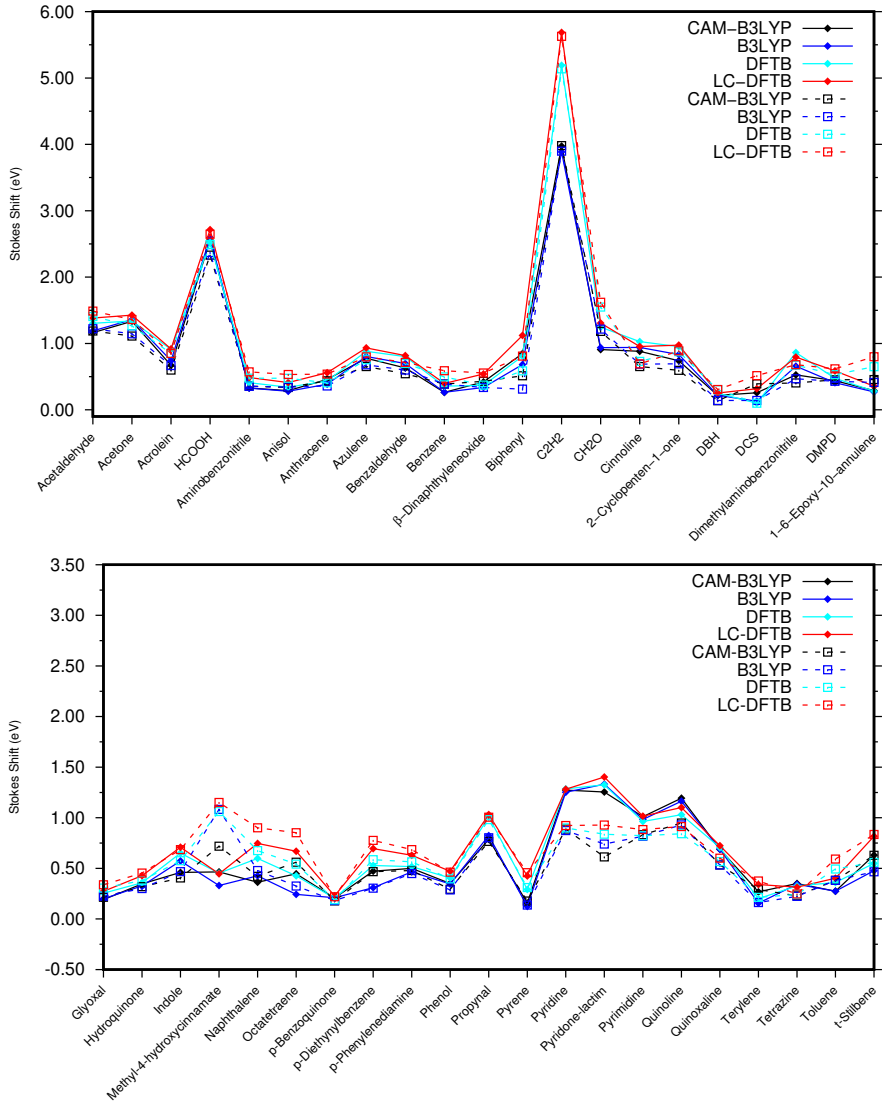


Fig. 4.14.: Stokes shift of first excited state for the molecules from Ref. [90]. The solid lines show the Stokes shifts obtained with B3LYP optimized geometries and the dashed lines the Stokes shifts obtained with the LC-DFTB2/OB2 (base) optimized geometries. The different colors indicate the different methods used for the calculation of vertical excitation energies.

A comparison of the LC-DFTB2/OB2 (base) Stokes shifts for the CC2 geometries and LC-DFTB2/OB2 (base) geometries reveals differences larger than 0.5 eV for pyrrolo[3,2-h]quinoline, salicylic acid, 3P-1 propionic acid (gauche), all of the four studied hydroxyquinolines, phenylacetylene, benzonitrile and 2-hydroxypyridine. Methyl-4-hydroxycinnamate from the set of Ref. [90], where the LC-DFTB2/OB2 (base) Stokes shifts for the LC-DFTB2/OB2 (base) and B3LYP geometries are compared, shows such a huge deviation, too.

The reasons for these large Stokes shift deviations are clearly visible on the geometries of salicylic acid, 3P-1 propionic acid (gauche), phenylacetylene and benzonitrile. Interest-

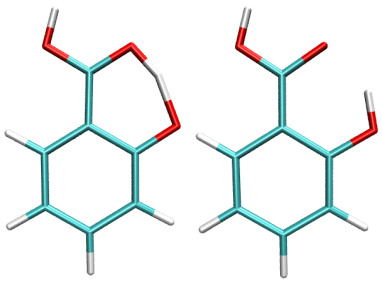
ingly, only quite small geometrical differences were found for pyrrolo[3,2-h]quinoline, the hydroxyquinolines, 2-hydroxypyridine and methyl-4-hydroxycinnamate.

CC2 predicts an intramolecular hydrogen bond for salicylic acid in the excited state which is not found by TD-LC-DFTB2/OB2 (base) (Fig. 4.15). This leads to the large differences for the predicted fluorescence energies. An additional DFT optimization with the long-range corrected functional CAM-B3LYP confirmed the presence of this hydrogen bond in excited state. The ground state geometries of CC2 and LC-DFTB2/OB2 (base) are instead very similar and without the hydrogen bond.

The optimized geometries of 3P-1-propionic acid (gauche) differ only in the relative orientation of the benzene ring and the propionic acid (Fig. 4.16). TD-LC-DFTB2/OB2 (base) provides a more than 1 eV higher fluorescence energy for the CC2 optimized excited state geometry than for the TD-LC-DFTB2/OB2 (base) excited state geometry whereas the absorption energies for the CC2 and LC-DFTB2/OB2 (base) geometries differ by less than 0.1 eV. The resulting CAM-B3LYP absorption energies are similar for both optimized ground state geometries as well (0.065 eV) but the fluorescence energy difference accounts to 0.405 eV.

For phenylacetylene CC2 predicts a planar excited state geometry while TD-LC-DFTB2/-OB2 (base) in contrast ends in a deformed benzene ring with an out-of-plane acetyl moiety (Fig. 4.17). A similar picture is obtained for benzonitrile. CC2 describes a planar geometry but TD-LC-DFTB2/OB2 (base) predicts hydrogens on the benzene ring and the nitrile moiety out of plane (Fig. 4.17).

In the case of methyl-4-hydroxycinnamate, different fluorescence energies are responsible for the deviation in the Stokes shift but only slight variations in the orientation of the chain and in the bond lengths of the benzene ring were observed. Fig. 4.18 shows an overlay of the B3LYP and TD-LC-DFTB2/OB2 (base) first excited state geometries.

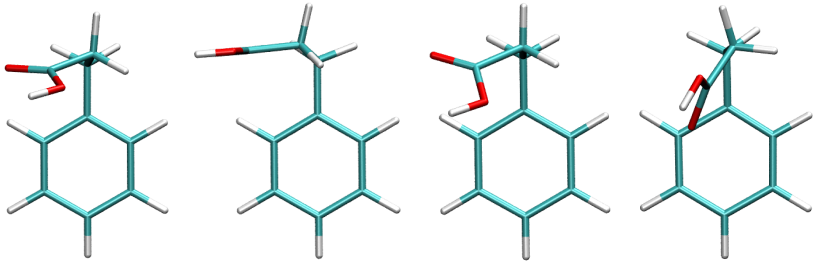


(a) excited state, CC2 **(b)** excited state, TD-LC-DFTB2

Fig. 4.15.: Structure of salicylic acid optimized with (a) CC2 [89] and (b) TD-LC-DFTB2/OB2 (base).

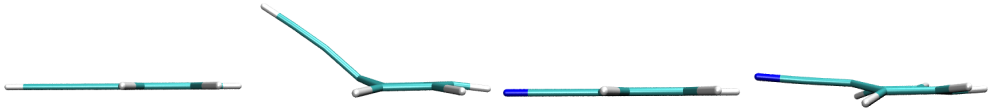
The deviations for pyrrolo[3,2-h]quinoline could be explained by a difference in the BLA. While CC2 predicts a clear bond length inversion in the excited state, the other methods including TD-LC-DFTB2/OB2 (base) predict a BLA of almost zero. The CC2 and TD-LC-DFTB2/OB2 (base) excited state geometries are overlain in Fig. 4.19 (e).

The TD-LC-DFTB2/OB2 (base) optimizations of the hydroxyquinolines resulted in geometries which are very similar to the CC2 geometries (Fig. 4.19, (a)-(d)) and also the bond length alternations do not differ significantly. Nonetheless, significant differences in the flu-



(a) ground state, CC2 **(b)** ground state, LC-DFTB2 **(c)** excited state, CC2 **(d)** excited state, TD-LC-DFTB2

Fig. 4.16.: Structure of 3P-1-propionic acid (gauche) in ground (a), (b) and excited state (c), (d) optimized with CC2 [89] and (TD-)LC-DFTB2/OB2 (base).



(a) excited state, CC2 **(b)** excited state, TD-LC-DFTB2 **(c)** excited state, CC2 **(d)** excited state, TD-LC-DFTB2

Fig. 4.17.: Structures of phenylacatylene (a), (b), benzonitrile (c), (d) in the first excited state optimized with CC2 [89] and TD-LC-DFTB2/OB2 (base).

orescence energies up to 0.75 eV are obtained when the LC-DFTB2/OB2 (base) excited state geometry is used instead of the CC2 one. The most striking difference observed in the excited state geometries are shorter bonds to the nitrogen in the ring system described by (TD-)LC-DFTB2/OB2 (base), which is also observed for pyrrolo[3,2-h]quinoline. Though, this cannot be the only reason for the different fluorescence energies since this is not the case in the keto form, which still shows a fluorescence energy difference of 0.51 eV. It needs to be mentioned, however, that the geometry of the keto isomer was identified as transition state in Ref. [89] and thus cannot be taken as reference for the correct excited state geometry here.

No clear structural deviations are observed for 2-hydroxypyridine between the CC2 and LC-DFTB2/OB2 (base) geometries, too, with the exception of the C-N bonds and slight differences at the hydroxy moiety (Fig. 4.19 (f)).

The Stokes shifts were recomputed using the shift and split parameter sets. In Figs. 4.20 and 4.21 the Stokes shifts obtained with CAM-B3LYP and the different parameter sets on the reference geometries are compared to the Stokes shifts obtained with the different parameter sets for optimization and excitation energy calculation. With these graphs two hints are gained. From a comparison of the excitation energies computed on the same geometries (from the reference methods, solid lines), the impact of the choice of the parameters (base, shift, split) on the excitation energies becomes visible. It is very small for all molecules: the largest difference to the base parameters is found for 7-azaindole (shift parameters: -0.05 eV, split parameters: -0.06 eV). The dashed lines from LC-DFTB2/OB2 show the resulting differences from optimization as well as excitation energy calculation

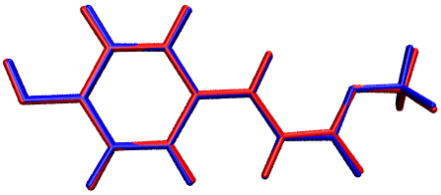
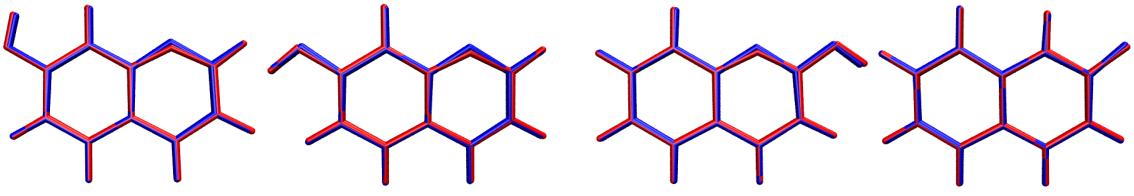


Fig. 4.18.: Overlay of structures of methyl-4-hydroxycinnamate in the first excited state optimized with B3LYP [90] and TD-LC-DFTB2/OB2 (base).

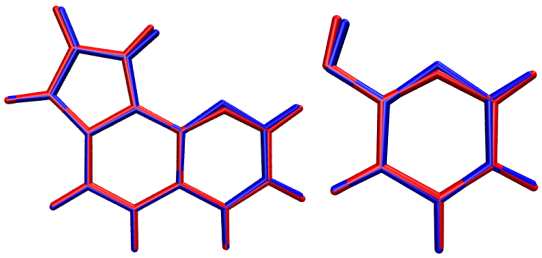


(a) 7-Hydroxyquinoline, cis

(b) 7-Hydroxyquinoline, trans

(c) 2-Hydroxyquinoline, enol

(d) 2-Hydroxyquinoline, keto



(e) Pyrroloquinoline

(f) 2-Hydroxypyridine

Fig. 4.19.: Overlay of excited state geometries. Blue: CC2 [89], red: TD-LC-DFTB2/OB2 (base).

with the three parameter sets. Generally, the resulting Stokes shifts are quite similar. In a few cases the shift and split parameters provide slightly improved Stokes shifts taking as reference the CAM-B3LYP Stokes shifts on the CC2 and B3LYP geometries, respectively: phenylacetylene, pyridine (only split), aniline, aminobenzonitrile and DMPD. The reason for these improvements are partly the different geometries and partly the different excitation energies obtained with the diverse parameter sets.

In sum, the Stokes shifts are a sensitive test for excited states geometries due to their extreme sensitivity to structural changes. In some cases, the structural changes are hardly visible although different fluorescence energies are obtained, for instance for the hydroxyquinolines. In other cases, TD-LC-DFTB2/OB2 shows clear geometrical deviations and some of them can be traced back to parametrization. One example is salicylic acid, where the proton transfer barrier is overestimated so that the shared proton is disfavored. Other geometrical deviations could be due to very shallow energy surfaces. In such cases it is not surprising that different methods lead to different geometries. One example is 3P-1-propionic acid (gauche) with its flexible propionic acid chain. In this case, conformational

sampling in a simulation could be meaningful to get an average over the possible geometries. This would be computationally feasible with TD-LC-DFTB2. Finally, phenylacetylene and benzonitrile revealed that TD-LC-DFTB2/OB2 in some cases has difficulties to describe the angles at triple bonds correctly. Hence, there is need to take care if similar molecules with triple bonds are calculated. However, not every such attempt fails as seen on the examples of aminobenzonitrile and dimethylaminobenzonitrile.

Excited-state reorganization energies are shown in Fig. 4.22. The B3LYP results from Ref. [90], and the CAM-B3LYP values computed for this work serve as references for the results from TD-DFTB/3OB and TD-LC-DFTB2/OB2 (base). CAM-B3LYP and TD-LC-DFTB2 contain a long-range correction in contrast to B3LYP and TD-DFTB. It is evident, that in general the DFT functionals B3LYP and CAM-B3LYP agree reasonably well, but there are notable exceptions (e.g. indole, naphthalene, p-Diethynylbenzene, pyridine). Both, TD-DFTB/3OB and TD-LC-DFTB2/OB2 (base) overestimate reorganization energies slightly in most cases. In Tab. 4.2, a statistical analysis with respect to CAM-B3LYP is presented. TD-LC-DFTB2/OB2 (base) is a bit closer to LC-DFT compared to TD-DFTB/3OB. In particular, the maximal error is reduced.

Tab. 4.2.: Statistical evaluation of the errors of the reorganization energies in eV, using CAM-B3LYP as reference.

	B3LYP	LC-DFTB	DFTB
MAX	0.16	1.41	1.73
ME	0.03	0.15	0.17
MAE	0.04	0.17	0.20

Tab. 4.3.: Reorganization energies in eV of polycyclic aromatic hydrocarbons calculated with different methods. λ^{ex} is the reorganization energy in the excited state and λ^{total} the sum of the reorganization energies in ground and excited state.

	$\lambda^{\mathrm{ground}}$			λ^{ex}			λ^{total}		
	CAM	LC-DFTB	DFTB	CAM	LC-DFTB	DFTB	CAM	LC-DFTB	DFTB
Anthracene	0.235	0.261	0.176	0.246	0.277	0.189	0.481	0.538	0.365
Tetracene	0.191	0.205	0.128	0.207	0.220	0.136	0.398	0.425	0.264
Pentacene	0.162	0.170	0.098	0.178	0.184	0.105	0.340	0.354	0.203

It is known that for larger molecules DFT-GGA underestimates reorganization energies significantly. For example, a SCS-CC2 reorganization energy λ^{total} of 0.533 eV has been reported for anthracene [106], which is underestimated by B3LYP and BLYP by about 0.1 eV and 0.2 eV, respectively. To test the LC-version of TD-DFTB, a closer look is taken on the reorganization energies of polycyclic aromatic hydrocarbons with CAM-B3LYP in comparison to TD-LC-DFTB2/OB2 (base) and TD-DFTB/3OB. The results are reported in Tab. 4.3. It is interesting that the tight-binding methods behave in analogy to DFT. TD-LC-DFTB2/OB2 (base) agrees well with LC-DFT in constrast to TD-DFTB/3OB which underestimates the reorganization energies. For anthracene, TD-LC-DFTB2/OB2 (base) is even closer to the *ab initio* value of 0.533 eV than CAM-B3LYP.

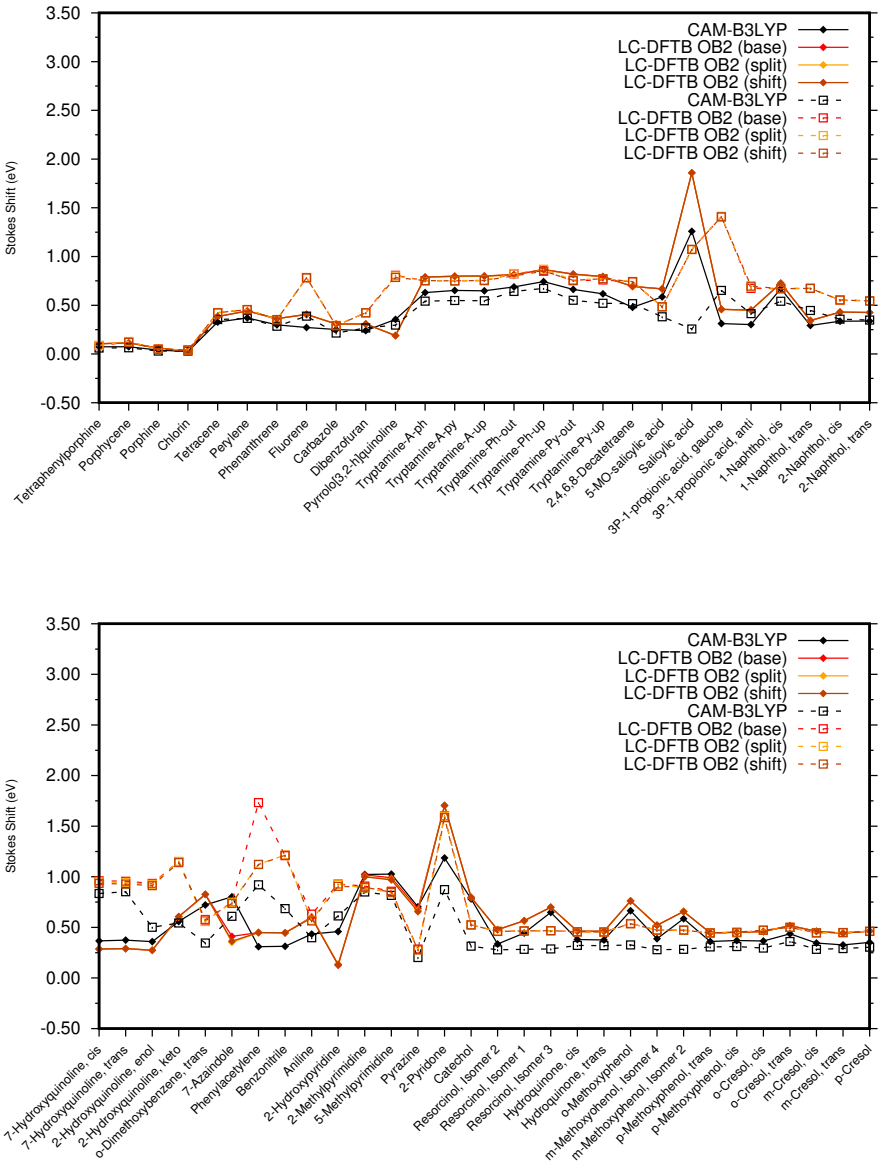


Fig. 4.20.: Stokes shifts of the benchmark set from Ref. [89] computed with LC-DFTB/OB2 base, shift and split parameters. The solid lines show the Stokes shifts obtained with CC2 optimized geometries from Ref. [89] and the dashed lines the Stokes shifts obtained with our LC-DFTB2/OB2 optimized geometries. The LC-DFTB2/OB2 geometries obtained with the respective parameters were used. The CAM-B3LYP excitation energies were computed on LC-DFTB2/OB2 (base) geometries.

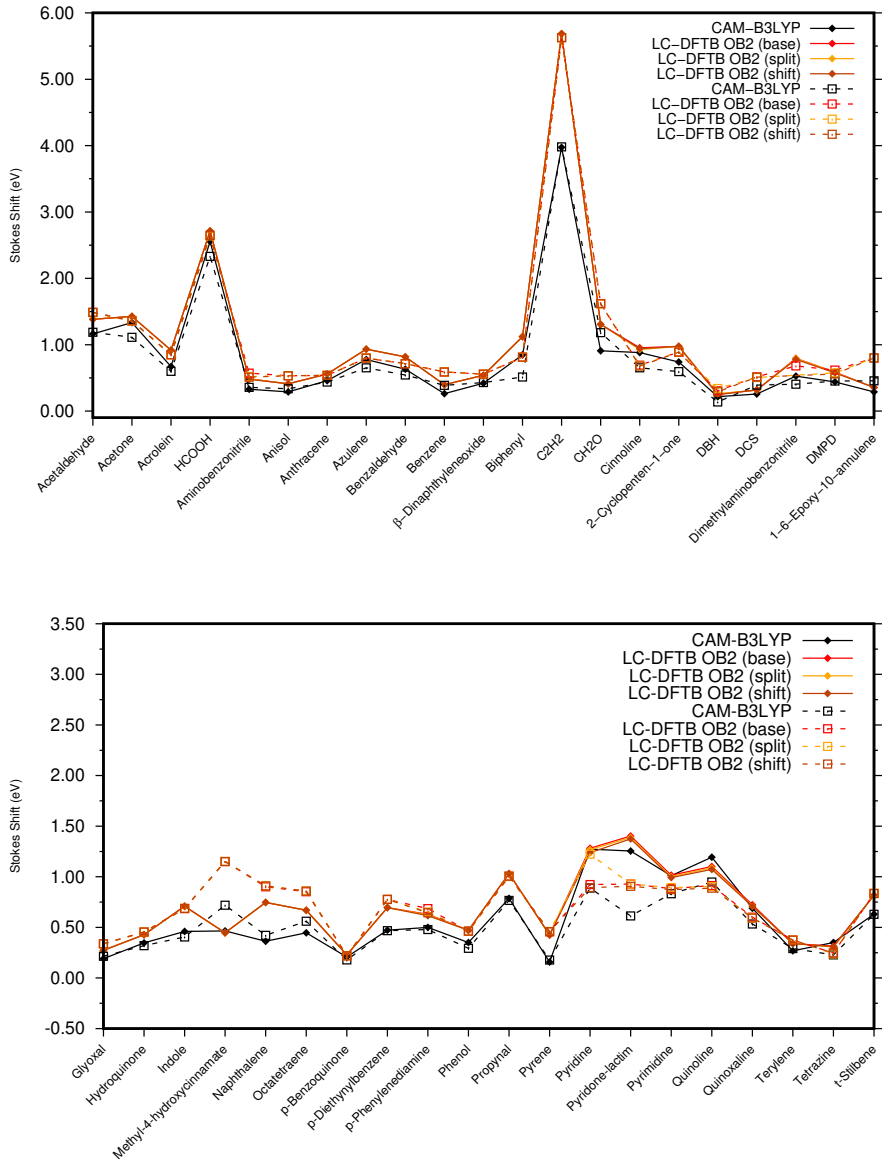


Fig. 4.21.: Stokes shifts of the benchmark set from Ref. [90] computed with LC-DFTB/OB2 base, shift and split parameters. The solid lines show the Stokes shifts obtained with B3LYP optimized geometries from Ref. [90] and the dashed lines the Stokes shifts obtained with our LC-DFTB2/OB2 optimized geometries. The LC-DFTB2/OB2 geometries obtained with the respective parameters were used. The CAM-B3LYP excitation energies were computed on LC-DFTB2/OB2 (base) geometries.

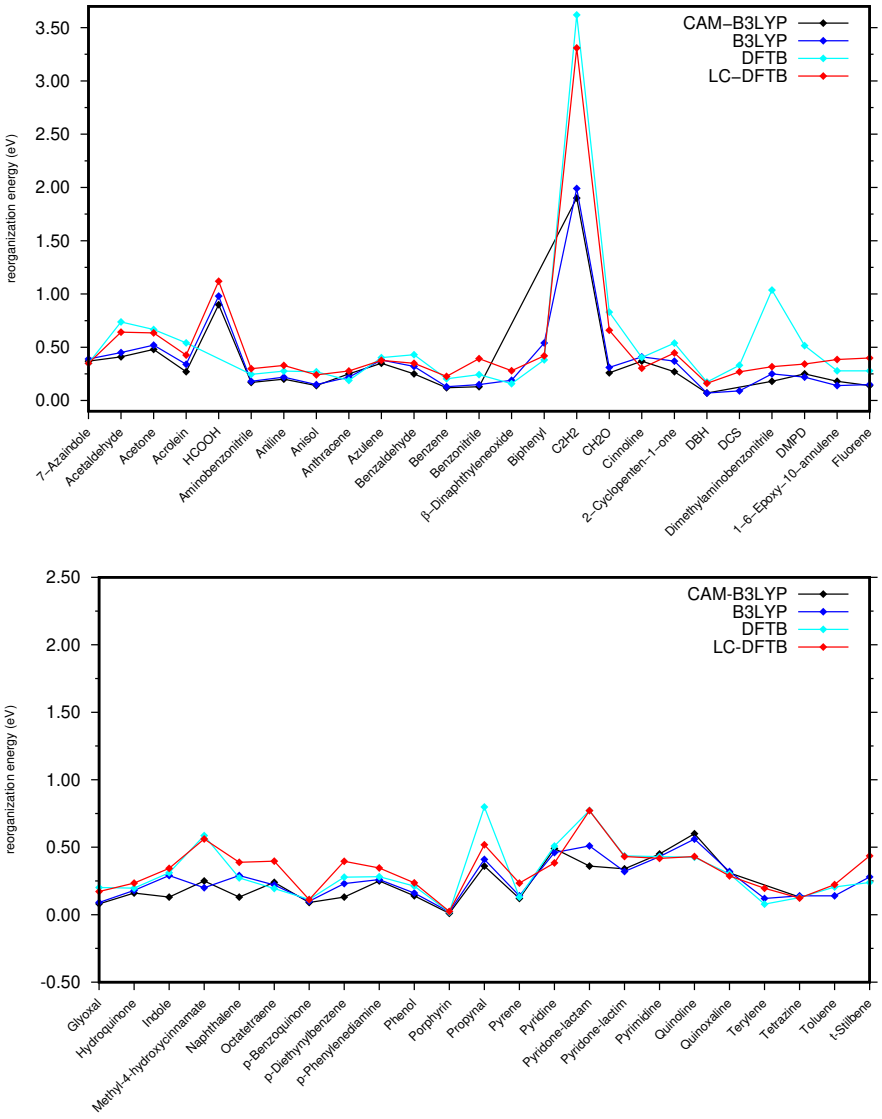


Fig. 4.22.: Excited state reorganization energies resulting from LC-DFTB and DFTB compared to the values from B3LYP presented in [90] and CAM-B3LYP.

4.6. Vibrational Frequencies

In the last part of this benchmark, vibrational frequencies are investigated. As eigenvalues of the Hessian matrix they serve as further indicator for the quality of the excited-state potential energy curvature. The analysis is done for two large but rigid dyes, anthanthrene and tetraazaperopyrene (TAPP), displayed in Fig. 4.23. Based on the optimized geometries and second derivatives of the potential energy surfaces of the ground and excited state, vibrational progression of UV/vis spectra of the two dyes were calculated and compared to DFT.

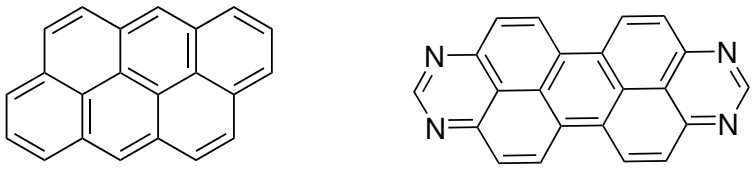
4.6.1. Computational Details

The dyes were optimized with (TD-)LC-DFTB2/OB2 (base) and DFTB/3OB-freq [47] in ground and excited state. The target state of anthanthrene was the first excited state and the target state of TAPP the third excited state. This state of TAPP has the largest oscillator strength and corresponds to the state for which experimental and theoretical vibrationally resolved spectra are available. The Hessians were computed semi-numerically with (TD-)LC-DFTB2/OB2 (base) and DFTB/3OB-freq from the analytical first derivatives. The script used for these calculations originates from the koala program and was adapted to DFTB+ by Sebastian Höfener [107–109]. It displaces every atom once in all directions (± x,y,z), reads the respective analytical gradient and determines the second derivative according to Eq. (4.1). The increment h was set to 0.005 bohr.

$$f''(x) \approx \frac{f'(x+h) - f'(x-h)}{2h} \tag{4.1}$$

The SCC tolerance was set very tight to 1×10^{-12} to get the energies at sufficient precision for the numerical differentiation. The eigenvectors and eigenvalues (frequencies) of the semi-numerical Hessians were determined adapting code from the KOALA program [107–109]. The 3OB-freq parameters used in the DFTB calculations are particularly parametrized for frequencies.

The dyes were optimized and Hessians were calculated at the B3LYP/def2-TZVP and CAM-B3LYP/def2-TZVP level using the ORCA program package, too. The target state of TAPP corresponded to the third excited state in the case of B3LYP and the first excited state in the case of CAM-B3LYP. The VeryTightSCF and VeryTightOpt keywords were applied and a DFT integration grid 7. The TDA and RIJCOSX approximations were used. The convergence tolerance of the excited state energies and for the norm of the residual vectors was set to 1×10^{-9} Ha. Analytical second derivatives were only available for the ground state.



(a) Anthanthrene **(b)** TAPP

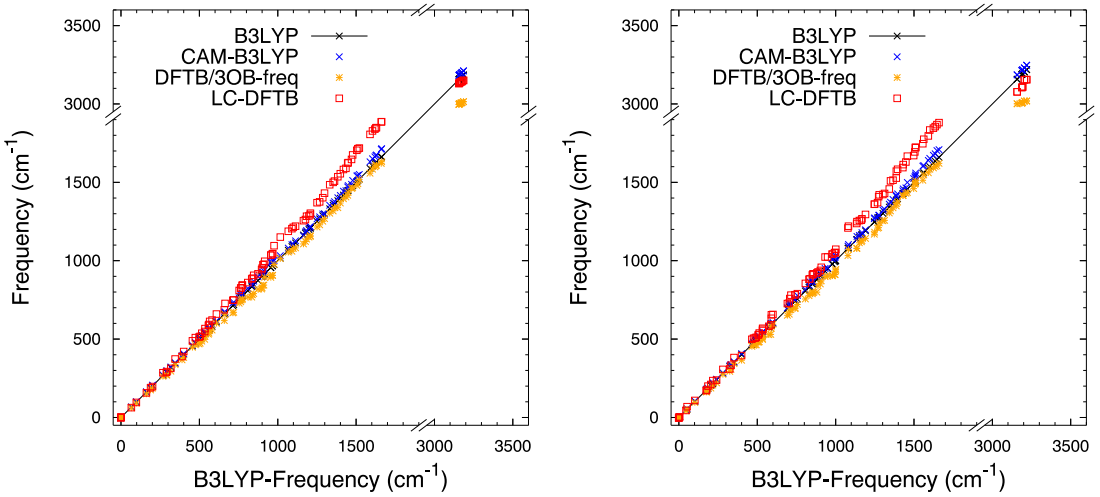
Fig. 4.23.: Studied dye molecules: (a) anthanthrene and (b) TAPP.

Vibrational spectra at 0 K and 300 K were calculated with the program hotFCHT [110] employing the time-dependent method with a graining of 1 cm^{-1} and a full width half maximum of the Lorentzian of 500 cm^{-1}. For the calculations of TAPP based on LC-DFTB2/OB2 (base) and DFTB/3OB-freq the orthogonality tolerance was increased to 0.05 (default value: 0.01).

4.6.2. Results and Discussion

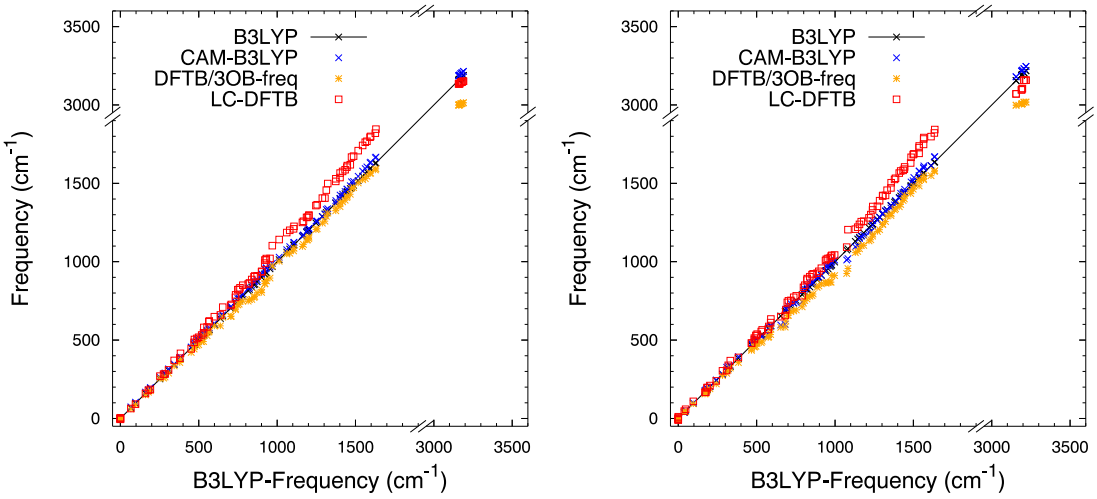
Figs. 4.24 and 4.25 show the similarity of the frequencies of the ground state and the excited state with the largest oscillator strength obtained with LC-DFTB2/OB2 (base), DFTB/3OB-freq and DFT. The B3LYP frequencies serve as exemplary DFT reference. Evidently, the resulting frequencies depend on the applied functional and LC-DFTB2/OB2 (base) is in agreement with the DFT values, though by trend slightly overestimating. An exception are the largest frequencies (> 3000 cm^{-1}) mainly corresponding to hydrogen vibrations which are lower predicted by LC-DFTB2/OB2 (base). In the case of anthanthrene, the average percental deviation to B3LYP in the excited state accounts to 6 % for LC-DFTB2/OB2 (base) and -4 % for DFTB/3OB-freq. The CAM-B3LYP frequencies of the excited state are on average 2 % larger than the B3LYP frequencies. For TAPP the excited state frequencies of B3LYP and CAM-B3LYP differ only by 1 %, the DFTB/3OB-freq frequencies are 6 % lower and the frequencies of LC-DFTB 6 % larger.

In Ref. [47] it is shown that B3LYP is overestimating the experimental C=C stretching frequency in ethene (by 70 cm^{-1}) and the stretching frequencies in benzene (by 41 cm^{-1}), while the particularly parametrized DFTB/3OB-freq method in this case provides improved results (deviations of 13 cm^{-1} and 25 cm^{-1}, respectively). Hence, it can be assumed that all methods which were compared here are overestimating the experimental frequencies at least in the ground state.



(a) Anthanthrene **(b)** TAPP

Fig. 4.24.: Correlation of ground state frequencies calculated with B3LYP, CAM-B3LYP, DFTB/3OB-freq and LC-DFTB/OB2 (base). The B3LYP frequencies are plotted on the diagonal.



(a) Anthanthrene **(b)** TAPP

Fig. 4.25.: Correlation of excited state frequencies calculated with B3LYP, CAM-B3LYP, DFTB/3OB-freq and LC-DFTB. The B3LYP frequencies are plotted on the diagonal.

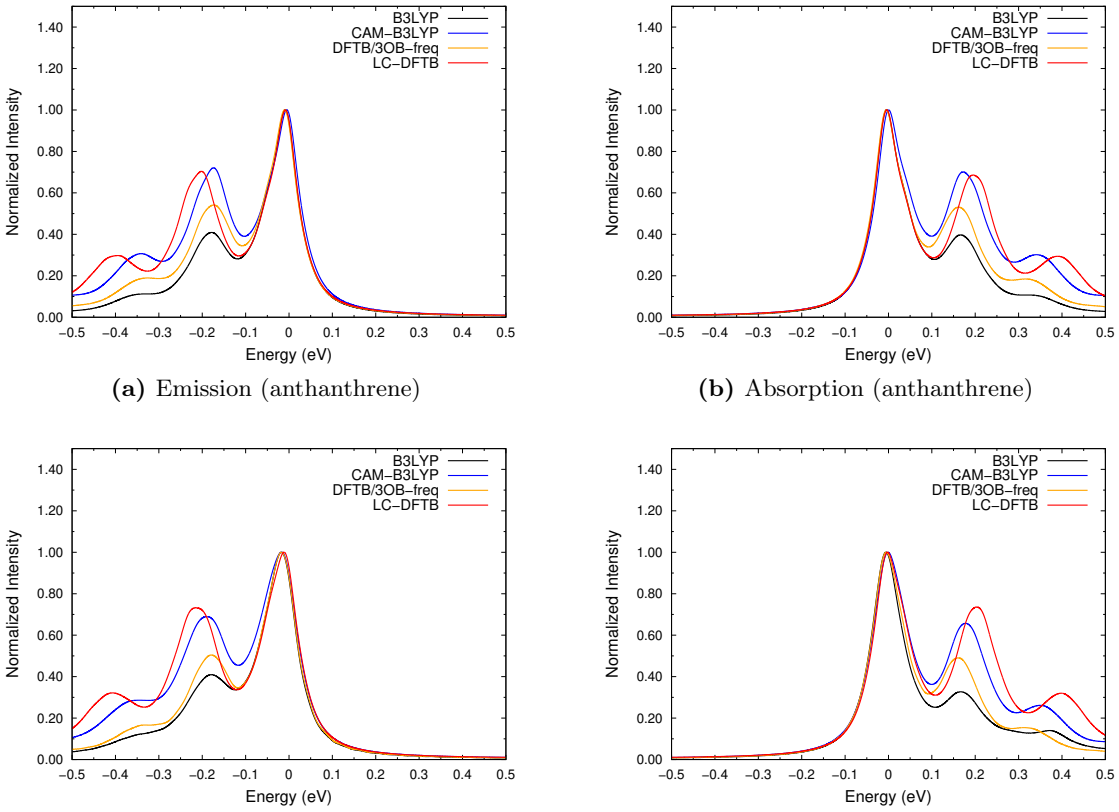
Finally, in Fig. 4.26 the vibronic spectra based on the computed Hesse matrices and frequencies are evaluated. In order to allow a convenient comparison between the theoretical methods the plots are given relative to the individual 0–0 transition. The computed spectral shapes of anthanthrene and TAPP are in agreement with the experiment [111,112].

In the anthanthrene emission spectra the three peaks of decreasing size are predicted by all methods, while the relative intensities are best reproduced by DFTB/3OB-freq. The experimental absorption spectrum of anthanthrene shows a small peak melted with the first large peak and also the second large peak contains a shoulder. This is not clearly resolved in our calculated spectra but slightly indicated in the shape of the peaks by all methods. Overall the intensities of DFTB/3OB-freq are in best agreement with the experiment.

The emission and absorption spectra of TAPP consist of three peaks of decreasing size, too. LC-DFTB2/OB2 (base) generally predicts the most pronounced peaks, which matches the experimental absorption spectrum well but reproduces the experimental emission spectrum worse.

The LC-DFTB2/OB2 (base) peaks are slightly blue-shifted compared to DFT and DFTB/3OB-freq in all computed spectra. In Ref. [111], the peak positions relative to the first peak are reported for anthanthrene. Compared to these results, the relative absorption peak positions (Tables 4.4 and 4.5) are a little too low except the LC-DFTB2/OB2 (base) ones, where the second peak fits well and the third is blue-shifted. DFTB/3OB-freq agrees quite well in the emission spectrum and the other methods overestimate the peak distances slightly. A good agreement between DFTB/3OB-freq and DFT spectra is also reported in Ref. [113].

The large peaks close to zero indicate that the 0–0 transitions are allowed. This is typically the case if ground and excited geometries are very similar, which is observed for anthanthrene and TAPP. For these molecules the Stokes shifts calculated from vertical excitation energies are significantly overestimated compared to experiment and need to be determined as difference between the respective largest absorption and emission peaks. For



(a) Emission (anthanthrene)

(b) Absorption (anthanthrene)

(c) Emission (TAPP)

(d) Absorption (TAPP)

Fig. 4.26.: Simulated vibronic progression of anthanthrene (top) and TAPP (bottom) at 300 K computed with B3LYP, CAM-B3LYP, DFTB/3OB-freq and LC-DFTB.

both molecules and with all methods compared here the predicted Stokes shifts are then smaller than 0.05 eV (exp. value for TAPP in toluene: 0.06-0.1 eV [112]).

Anthantrene and TAPP are both quite rigid molecules which is advantageous for the calculation of vibronic spectra. Molecules with flexible moieties, as for example methyls with rotating hydrogens, caused problems during the calculation and were thus not further considered. Moreover, it is more difficult to converge the geometry of such molecules into a minimum.

At the end, the efficiency of TD-LC-DFTB2 is illustrated on the example of the two large dyes anthanthrene and TAPP. The computational times needed for geometry optimization and the computation of the Hessian matrices are summarized in Tab. 4.6. The expense of the optimization is shown by means of the averaged computational time per geometry step instead of the time for the total optimization to avoid a bias due to a different number of needed cycles.

Evidently, the computational cost is dramatically reduced if TD-LC-DFTB2 is used. Excited state simulations on a significant time scale are only feasible with the semi-empirical method.

Tab. 4.4.: Relative peak positions of anthanthrene (ant.) and TAPP in eV in the absorption spectra computed with different methods at 0 K and 300 K.

	B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3
ant., 0 K	0.177	0.347	0.171	0.343	0.168	0.325	0.198	0.396
ant., 300 K	0.171	0.305	0.172	0.341	0.167	0.322	0.199	0.394
TAPP, 0 K	0.175	0.382	0.180	0.331	0.166	0.327	0.208	0.403
TAPP, 300 K	0.172	0.378	0.179	0.351	0.166	0.316	0.206	0.400

Tab. 4.5.: Relative peak positions of anthanthrene (ant.) and TAPP in eV in the emission spectra computed with different methods at 0 K and 300 K.

	B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3	Δ 1-2	Δ 1-3
ant., 0 K	0.170	0.333	0.169	0.337	0.163	0.322	0.193	0.386
ant., 300 K	0.169	0.324	0.169	0.336	0.163	0.315	0.193	0.387
TAPP, 0 K	0.162	0.313	0.168	0.351	0.160	0.320	0.196	0.397
TAPP, 300 K	0.162	-	0.172	0.328	0.161	0.315	0.203	0.396

Tab. 4.6.: CPU times for geometry optimization and semi-numerical calculation of the Hessian matrix in the excited state. For the geometry optimization averaged computation times per geometry step are given.

	averaged optimization step				Hessian matrix			
	CAM-B3LYP	B3LYP	LC-DFTB	DFTB	CAM-B3LYP	B3LYP	LC-DFTB	DFTB
ant.	4 h [a,b]	4 h [a,b]	4 s [c]	0.7 s [c]	35 d [a,b]	33 d [a,b]	29 min [b]	17 min [d]
TAPP	4 h [a,b]	5 h [a,b]	6 s [c]	0.7 s [c]	49 d [a,b]	47 d [a,b]	38 min [c]	24 min [b]

[a] CPU time estimated from 6 times the wall time obtained from parallel calculations using 6 cores
[b] Intel(R) Xeon(R) Silver 4214 CPU, 2.20 GHz
[c] Intel(R) Xeon(R) CPU E5-2630 v3, 2.40 GHz
[d] Intel(R) Xeon(R) CPU E5-2630 v4, 2.20 GHz

4.7. State Ordering

The example of TAPP showed that the ordering of the excited states can differ between the methods. The states of same character were identified for the different methods before calculation of the frequencies and vibronic spectra of TAPP. However, in the previous benchmark of the roughly 100 molecules the state ordering was not investigated further. To get an idea of the frequency of a different state ordering predicted by TD-LC-DFTB2, an extensive orbital analysis was performed retrospectively. The results are presented in the appendix. For the majority of molecules, the first excited states predicted by TD-LC-DFTB2 and the B3LYP references are of same character and thus the general conclusions drawn before do not change. In sum, eight cases were found, where TD-LC-DFTB2 predicted a different state ordering and the second excited state should be compared to the reference data instead of the first: salicylic acid, the four hydroxyquinolines, 7-azaindole, 2-hydroxypyridine and methyl-4-hydroxycinnamate. Notably, the Stokes shifts of the hydroxyquinolines, 2-hydroxypyridine and methyl-4-hydroxycinnamate significantly improved when taking into account the excited state matching the reference.

4.8. Conclusion

In sum, TD-LC-DFTB2/OB2 provided satisfying geometries in most cases and vibronic UV/vis spectra in sufficient agreement with the experiment. A combination of a OB2 parameter set for geometries and a parameter set optimized for excitation energies resulted in AEEs in very good agreement with B3LYP. The efficiency of TD-LC-DFTB2 is a big advantage since it allows the sampling of geometries which is in general not feasible with QM methods. An impression on the speed of TD-LC-DFTB2 calculations was given in subsection 4.6.2. For the study of fluorophores in complex environments, QM/MM simulations are useful since they account for dynamics and solvent effects at the same time. TD-LC-DFTB2 is a promising method to be applied in such simulations. However, before TD-LC-DFTB2 is applied in a study of a new chromophore, the predicted excited states and their ordering should be validated, for instance by comparison to the results of other QM methods.

5 Validation and Application of Excited State QM/MM Simulations

In this chapter, excited state QM/MM simulations are applied to study the fluorescence of three environmentally sensitive chromophores: Flugi-2, prodan and 4-aminophthalimide (4-AP). Reference data in the literature allows a detailed evaluation of the results of these QM/MM simulations using the TD-LC-DFTB2 method benchmarked in the previous chapter and electronic embedding. However, the findings of this chapter do not only shed light on the general performance of such simulations. Additionally, prodan and 4-AP are dyes of special interest in the next chapter which deals with optical glucose sensing.

Firstly, the properties of the dyes, their applications and previous studies on their fluorescence are summarized. Then the results of excited state QM/MM simulations of Flugi-2 are compared to the results from classical simulations reported in literature, where the dye was particularly parametrized for the excited state. Next, geometries, excitation energies and the orbital transitions of prodan and 4-AP as obtained with TD-LC-DFTB2 are compared to literature and own DFT and CC2 calculations as benchmark. Finally, these dyes are simulated in different solvents. Geometrical as well as electrostatic effects on the fluorescence are investigated.

5.1. Introduction of the Fluorophores

5.1.1. Flugi-2

Flugi-2 (Methyl-3-(cyclohexylamino)-2-(4-methoxyphenyl)imidazo[1,2-a]pyridine-7-carb-oxylate, [114]) was synthesized by Burchak *et al.* [115] as part of a combinatorial screening for the discovery of new fluorophores. The name *Flugi* was introduced in this study as abbreviation for *fluorescent Ugi* product and describes fluorophores obtained by a three component Ugi reaction. The chemical structure of Flugi-2 is given in Fig. 5.1. Flugi-2 belongs to the Flugi molecules which fluoresce with a high quantum yield in the visible range. It shows a large solvent shift in the emission spectrum in DMSO compared to the spectrum in decane (-0.26 eV) [114,115]. The emission spectra are almost featureless [115]. Besides the experimental characterization by Burchak *et al.*, Flugi-2 was computationally studied by classical MD simulations in excited state in Ref. [114].

It is expected that the fluorescence occurs from an internal charge transfer (ICT) state from the amine substituent at position 3 to the aromatic ring system. In Ref. [115] it was found that the absorption and emission wave lengths are determined by the electron donating and electron withdrawing abilities and conjugation of the substituents.
In the theoretical study [114], geometrical and electrostatic effects were analyzed to understand the colour shift of Flugi-2 from decane to DMSO. It was found that the electrostatic effect of the DMSO molecules is responsible for the red shift while geometrical differences due to DMSO causes even a slight blue shift. However, only explicit solvent interactions

show an impact on the fluorescence energies while an implicit solvent model failed to produce the red shift. In particular, the shift is attributed to a stable hydrogen bond between DMSO and the N-H group of Flugi-2. It was further found, that due to this hydrogen bond more different dihedral angles are sampled in DMSO than in gas phase.

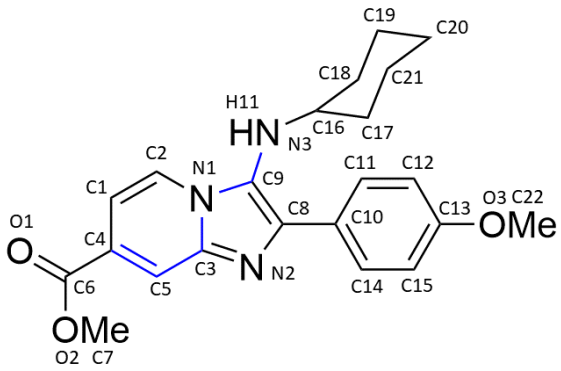


Fig. 5.1.: Chemical structure of the Flugi-2 molecule. Bonds involved in the BLA calculation are blue coloured. The studied dihedral angles are α=N2–C8–C10–C14, β=C9–N3–C16–C17 and γ=C8–C9–N3–H11.

5.1.2. Prodan

Prodan (6-Propionyl-2-(dimethylamino)naphthalene) consists of a naphthalene core substituted with an electron-withdrawing carbonyl group and an electron-donating dimethylamino group at maximal distance (Fig. 5.2). This makes it a typical candidate for charge-transfer (CT) in excited state [116]. Accordingly, remarkable Stokes shifts are observed in polar solvents [116, 117]. From cyclohexane to water, the emission spectrum shifts by almost 0.8 eV [15, 117, 118]. Prodan is a widely applied polarity probe, in particular for

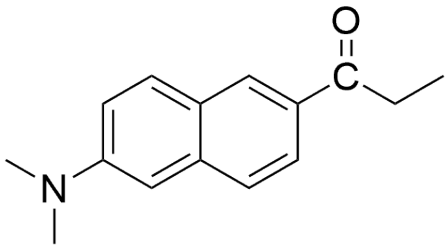


Fig. 5.2.: Chemical structure of prodan.

membranes (e.g. [119–123]). The derivatives badan and acrylodan became of interest for the development of optical glucose sensors [5–9, 124–132].

Prodan was firstly synthesized in 1979 [118] and since then it is extensively studied experimentally and theoretically. A first point of debate was the magnitude of the dipole moment. It was estimated based on the Lippert-Mataga equation [117, 118, 133], the solvent perturbation method [134] and by means of semi-empirical and density functional calculations [116, 135, 136]. Finally, Samanta and Fessenden [137] measured 10.2 D for the excited state dipole moment in benzene (9.6 D in dioxane) by microwave dielectric

absorption showing that the change of dipole moment upon excitation amounts to 4.4-5 D in aprotic solvents.

Further, the nature of the emissive excited state is discussed in literature. A planar intramolecular charge transfer state (PICT) is up to debate as well as twisted intramolecular charge transfer states resulting from rotation of either the dimethylamino or carbonyl moiety (N-TICT and O-TICT) or both (NO-TICT).

The moderate increase of dipole moment in the excited state may argue for a PICT. In refs. [138–140], the photophysical behaviour of prodan derivatives, which serve as models for the PICT, N-TICT or O-TICT states through incorporated structural constraints, is studied experimentally in different solvents. Comparison to prodan led to the conclusion that prodan emits from a PICT state. Theoretical calculations in vacuum also assign the PICT state dipole moments in the order of 10 D and significantly larger dipoles moments to the TICT states [116,141]. However, depending on the solvent and the cavity radius used in the Amos-Burrows theory to estimate the solvent effect, the N-TICT state becomes the first excited state in Ref. [116]. In a following study, where the solvent is considered by a self-consistent reaction field, the O-TICT state was suggested as emitting state instead [142]. In Ref. [141], it was found that B3LYP predicts an artificially low energy minimum for an N-twisted first excited state geometry in gas-phase as well as in solvent, due to its deficiency to describe long-range charge interactions. CIS and SAC-CI calculations instead support the assumption of a planar S1 geometry as the global minimum.

Besides characterization of the emitting state, it is aimed to understand the emission spectra of prodan in different environments. This includes an explanation for the large Stokes shifts but also for the broad and asymmetric shape which indicates dual emission [15, 143, 144]. It is undoubtedly that reorientation of the solvent molecule dipoles to the excited state dipole of prodan stabilizes the excited state of prodan and thus cause a red shift in the emission spectrum. Additionally, specific solute-solvent interactions are assumed [137, 144, 145]. The dual emission is often supposed to arise from one state in (a) its locally excited (LE) form and (b) the solvent relaxed internal charge-transfer form (ICT) which may be N- or O-twisted [146–148]. However, in recent publications [15, 149] it is stated that prodan emits from two different electronic states. This hypothesis is motivated by the findings that the amount of the species emitting at longer wavelengths (typically assigned to solvent relaxed molecules in CT state) does not increase with temperature and does also not increase with the polarity of the solvents. By multi-configurational calculations the first three excited states were found to be close in energy at the ground state equilibrium geometry. These are all planar at their respective equilibrium in vacuum and change their energetical order depending on the solvent. Instead of emission from a locally excited state and an ICT state, it is proposed that prodan emits from two of these different independent states, one of two $\pi \rightarrow \pi^*$ states and a $n \rightarrow \pi^*$ state. [15]

From a methodological point of view, several approaches are reported in literature to describe the excited prodan in solution. Optimizations or simulations employing an implicit solvent model are reported in several studies, for instance [15, 145, 150]. Additional supermolecular calculations with explicit solvent molecules were done in the TD-DFT study presented in Ref. [145]. Moreover, molecular dynamics simulations were performed with the fixed optimized geometry of the first excited state surrounded by free solvent molecules. Atomic partial charges representing the excited state electron density were assigned to the solute. Subsequently, QM calculations were performed with and without consideration of polarisation. In Ref. [15], a sequential quantum mechanics/molecular mechanics (S-

QM/MM) approach was applied using several QM methods. The polarization of prodan was computed from an average solvent electrostatic configuration (ASEC). In Ref. [151] excited prodan is described by QM using the many-body Green's function in the GW approximation and the Bethe-Salpeter equation considering solvent molecules as fixed point charges. The solvent is simulated in a classical MD simulation around prodan and relaxes according to the time evoution of the excited dye in an iterative scheme.

5.1.3. 4-Aminophthalimide

4-Aminophthalimide (4-AP) is an environmentally sensitive dye which emits from an ICT state and its emission wave length, intensity and lifetime depend on the polarity of the surrounding and particularly on the presence of specific interactions, namely hydrogen bonds [145]. The chemical structure of 4-AP is shown in Fig. 5.3. It is applied as probe in homogeneous as well as heterogeneous environments [152–156].

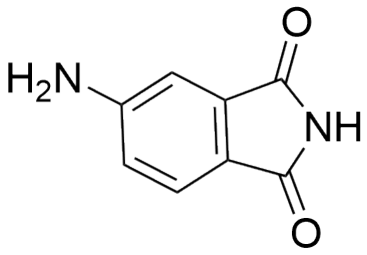


Fig. 5.3.: Chemical structure of 4-aminophthalimide.

While the absorption spectrum shows only small solvatochromic shifts, the fluorescence maxima shift significantly from aprotic to protic solvents. The red shift from acetonitrile to methanol amounts to ≈ 0.31 eV. From nonpolar diethyl ether to polar and protic water, the red shift of the emission maxima is ≈ 0.62 eV. The intensity of the fluorescence is large in both polar and nonpolar solvent with a yield around 0.7 but low in protic solvent. In water it is reduced to only 0.01. Also the lifetime is significantly shortened in protic solvents compared to aprotic solvent. In water it amounts to around 1 ns and in aprotic solvent it reaches 14-15 ns. The dipole moment in excited state is estimated from AM1 calculations and experimentally with the Lippert-Mataga equation to 8.3-9 D. [14, 157]

The individual effects of increased dipole moment and hydrogen bonding on the emission wave length were studied in Ref. [18] as well as the nature of the emitting species. The spectral properties of 4-AP were experimentally determined in different classes of solvents, which form either definitely no hydrogen bonds or act exclusively as hydrogen bond acceptor or hydrogen bond donor. It was concluded that in aprotic solvents without hydrogen bond acceptors, it is absorbed into a LE state and emitted from an ICT state. In aprotic solvents with electronegative heteroatoms and in protic solvents, the ICT state forms exciplexes with solvent molecules. If the solvent is strong protic, complexes are already formed in ground state, which can be absorbing as well. The complexes in excited state are not only deactivated by fluorescence but mostly by internal conversion, thus the quantum yield is reduced. The shifts due to non-specific solvent interactions were determined to be below 0.2 eV while the hydrogen bond contributions are significantly larger. [18]

In literature, a dispute is reported about a possible proton transfer process in water [158–160]. According to this hypothesis, the imino hydrogen would be transferred to one

of the carbonyl oxygens. This assumption was finally rouled out by studying a modified 4-AP with a N-bytyl moiety replacing the imino hydrogen [161].

5.2. Solvatochromism of Flugi-2

The solvatochroism of Flugi-2 was investigated in Ref. [114] by free MD simulations with a particularly parametrized force field for the first excited state of Flugi-2. The results from these simulations serve as reference for the QM/MM simulations performed here using a Gromacs/DFTB+ implementation. First, structural parameters of the fluorophore sampled in excited state are investigated. This includes bond lengths and dihedral angles. Second, fluorescence spectra based on the simulation snapshots are computed. As in Ref. [114], this is done in vacuum and for the dye solvated in DMSO.

5.2.1. Computational Details

The structure of Flugi-2 was optimized in vacuum with (TD-)LC-DFTB2 in ground and excited state with DFTB+ (version 20.1) [88] applying the $OB2^{0.3}$ (base) parameter set [92]. The conjugate gradient driver was employed with the default convergence criterion (force tolerance: $1 \times 10^{-4} \frac{\mathrm{Ha}}{a_0}$) and the SCC tolerance was set to 1×10^{-7} electrons. For the long-range correction the thresholded screening method was applied with a threshold of 1×10^{-16}. A frequency analysis was performed with the mode program of DFTB+ to confirm that the optimization ended in a minimum, respectively.

QM/MM simulations were performed in the ground and first excited state with the complete dye in the QM zone and explicit solvent molecules as MM part. The program consists of DFTB+ (version 20.1) embedded in GROMACS (version 2021). The QM part is computed with (TD-)LC-DFTB2 and the MM solvent (DMSO) by the same force field as used in Ref. [114], which consists of GAFF parameters and charges derived as usual from a Hartree-Fock ESP calculation and subsequent RESP fit.

The simulations were first equilibrated in ground state. After a QM/MM minimization with steepest descent to reduce the forces below 1000 kJ mol^{-1} nm^{-1}, an NVT equilibration was performed for 50 ps using a leap-frog integrator with a time step of 1 fs keeping all bonds constraint. The Nosé–Hoover thermostat [162, 163] adjusted the temperature at 300 K. Periodic boundary conditions (PBC) were applied and the Particle–Mesh Ewald (PME) method accounted for long-range electrostatics. In the following 50 ps NPT equilibration, additionally the Parrinello–Rahman barostat [164, 165] controlled the pressure with a reference value of 1.013 bar. In the productive simulation only bonds involving hydrogen were constraint. The ground state simulation has a length of 1 ns, the excited state simulation has a length of only 500 ps due to the increased computational cost.

Flugi-2 was additionally simulated in vacuum (same QM/MM program but without MM atoms). The simulations in ground and excited state started from a structure optimized in the respective state with the steep algorithm until all forces dropped below 1000 kJ mol^{-1} nm^{-1}. A leap-frog stochastic dynamics integrator was applied with a time step of 1 fs at a temperature of 300 K. The simulations in ground state and excited state were run for 500 ps.

The first geometrical parameter investigated in the following is a bond length alternation (BLA). It describes the difference between the average length of bonds with single bond character and the average length of bonds with double bond character. The choice of the

included bonds is in principle arbitrary. Here, the same bonds as in Ref. [114] are selected, which show a significant change in their length upon excitation. They are blue coloured in Fig. 5.1, while the two C–C bonds of the aromatic ring are counted as double bonds as well as the bond between C9 and the nitrogen in the aromatic ring. The other two C–N bonds are counted as single bonds.

Additionally, three dihedral angles are investigated which describe the orientation of the methoxyphenyl and cyclohexylamine moieties. They are defined in Fig. 5.1.[1] The dihedrals with corresponding fluorescence energies were clustered using the k-means algorithm in origin [167].

Fluorescence energies were computed as vertical excitation energies on the excited state geometries with OM2 and TD-LC-DFTB2. The active space of the OM2 calculations was chosen as done in Ref. [114] and consisted of 20 electrons in 20 orbitals. The mndo program was used [168]. The TD-LC-DFTB2 calculations were performed with DFTB+. Separate excitation energy calculations were performed with the OB2 (base) parameters and the parameter set optimized for excitation energies [3, 104]. The SCC tolerance was set to 1×10^{-8} electrons and the screening threshold to 1×10^{-16}. In the calculations considering the environment, the fluorophore was centered in the simulation box and the surrounding solvent molecules were represented by their point charges.

In the spectra calculation and for the dihedral clustering, the first 20 ps of the simulation in DMSO was considered as equilibration time and thus neglected. The spectra were modelled by a gaussian fit to the histogram of the vertical excitation energies computed along the trajectory, respectively. The bins had a width of 0.02 eV and Gauss functions were fitted to the bin centers in origin [167]. The intensity is computed as the number of counts in a bin times the averaged oscillator strength in the window divided by the total number of snapshots.

5.2.2. Results and Discussion

The results for the first excited state of Flugi-2 are presented in the following, while the results for the ground state are shown in the appendix.

First, the optimized and sampled geometries of (TD-)LC-DFTB2 in vacuum and in DMSO are analyzed. A BLA value of -0.059 Å is reported for the excited state structure optimized with CC2 in vacuum and a BLA value of -0.067 Å with the particularly parametrized force field [114]. The negative sign indicates a bond inversion. TD-LC-DFTB2/OB2$^{0.3}$ (base) optimization in vacuum results in a structure with a BLA of -0.062 Å. This shows that TD-LC-DFTB2 correctly predicts the bond inversion and the BLA value is also quantitatively close to CC2. For comparison, the QM/MM minimization in vacuum resulted in a BLA value of -0.049 Å which is also qualitatively correct but smaller in magnitude. QM/MM minimization in DMSO provides a very similar value of -0.046 Å. With the force field, the minimized structure in DMSO shows also a similar BLA value than in vacuum, -0.074 Å, as reported in [166]. During the excited state QM/MM simulation, the BLA values fluctuate around an average value of -0.063 Å in vacuum and -0.067 Å in DMSO (Fig. 5.4). With the force field, the BLA from the simulation in vacuum amounts to -0.072 Å and from the simulation in DMSO a value of -0.076 Å is

[1] The atom labels of C11, C12 and C14, C15 are swapped in Ref. [114] compared to Ref. [166]. It is assumed that the labelling in Ref. [166] is correct.

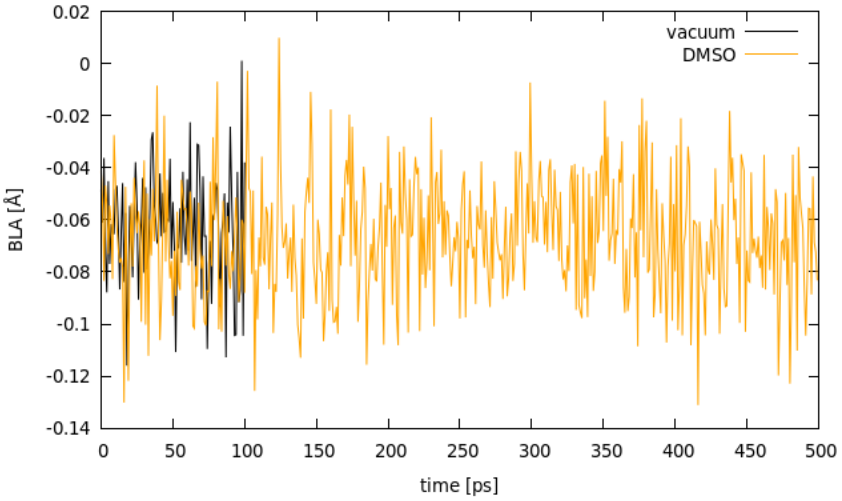


Fig. 5.4.: Bond length alternation of Flugi-2 as defined in the text during the excited state simulations in vacuum and in DMSO.

obtained [166]. This shows that both methods predict a slightly more negative BLA value in DMSO than in vacuum.

Tab. 5.1.: Cluster centers as determined from a k-means algorithm.

	cluster	vacuum	DMSO
α	1	26°, 2.83 eV	30°, 2.47 eV
	2	-21°, 2.83 eV	-29°, 2.48 eV
β	1	162°, 2.87 eV	170°, 2.55 eV
	2	-56°, 2.84 eV	-97°, 2.45 eV
	3	-116°, 2.78 eV	-123°, 2.48 eV
	4	-159°, 2.83 eV	-162°, 2.48 eV

The dihedral angles α and β sampled in the excited state simulations are directly set in relation to the fluorescence energies. In Fig. 5.5 additionally the centers obtained from a k-means clustering algorithm are shown, as it was done in Ref. [114]. The exact values are reported in Tab. 5.1.

In vacuum as well as in DMSO two clusters were found for the dihedral angle α. In both cases, they are roughly located in between of zero and ± 45 degrees and cover a range of fluorescence energies of almost 1 eV, respectively. In DMSO, the clusters are a little more separated than in vacuum. While the dihedral angles are basically the same in vacuum and DMSO, the fluorescence energies at the cluster centers mirrow the solvent shift. Hence, it can be concluded that the α dihedral is not an important factor determining the fluorescence energy.

The two clusters are also reported in Ref. [114]. However, in Ref. [114] two additional

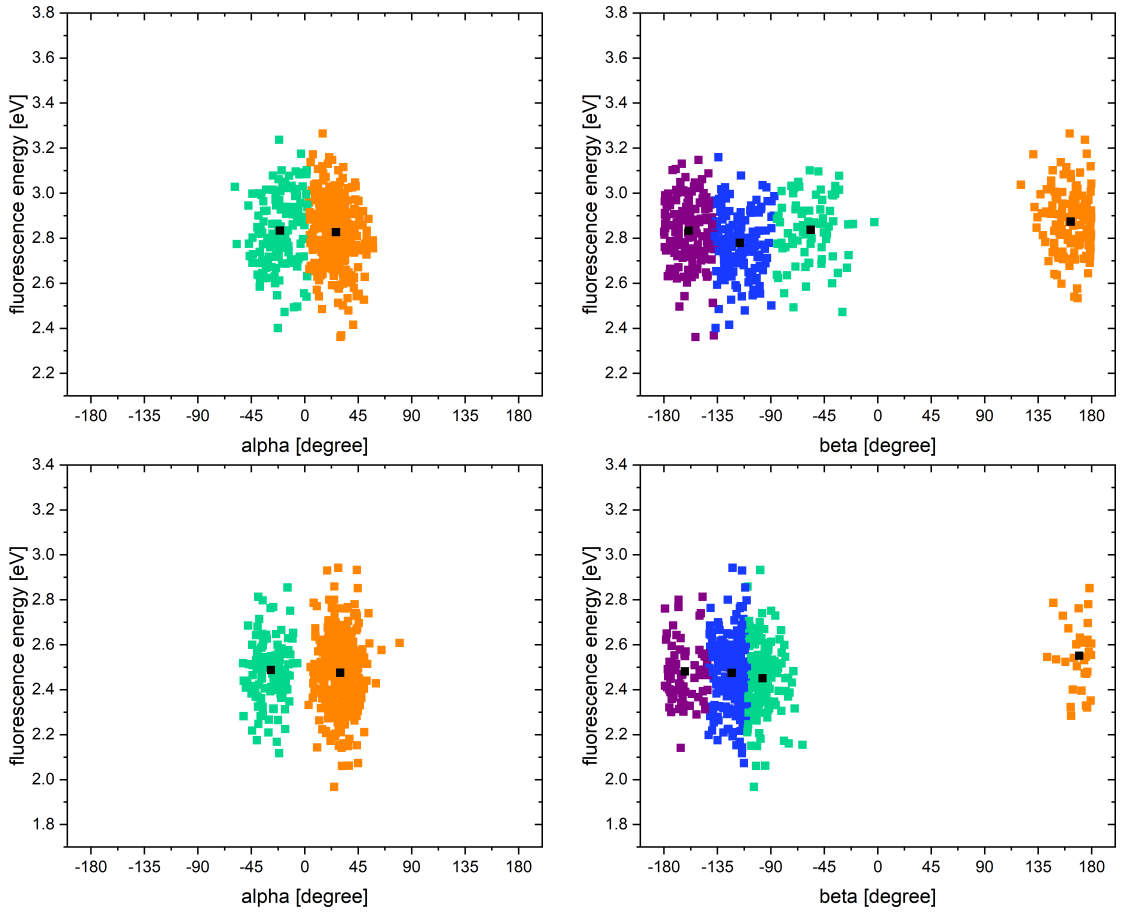


Fig. 5.5.: Dihedral angles sampled in vacuum (first row) and in DMSO (second row) and the fluorescence energies obtained with OM2 on the respective geometries. The black points indicate the center of the respective clusters.

clusters are presented before ±180°, respectively. It is possible that these clusters were not sampled in the QM/MM simulation due to the much shorter simulation time. Anyway, a rotation of methoxyphenyl moiety to sample the other two clusters would results in basically the same geometry, as the only difference can arise from the orientation of the methoxy moiety. However, this moiety is quite far from the core and is flexible itself, such that it may reorient, too. Indeed it is even remarkable that already two clusters are found with the QM/MM simulation in both cases, since the force field simulation remained for several nanoseconds within one cluster (Fig. 6 in Ref. [114]) [2]. The barrier for the flip of the methoxyphenyl moiety from one side ($\alpha \approx +25°$) to the other side ($\alpha \approx -25°$) is thus lower in the QM/MM simulation. The times at which the flips occur are visible in Fig. 5.6, which shows the dihedrals along the trajectories. It is also noteworthy, that the fluorescence energies obtained on the QM/MM geometries are red shifted compared to the values from the force field geometries, in vacuum as well as in DMSO (roughly 0.2-0.3 eV). In DMSO this shift is a little larger than in vacuum.

[2] In this figure, however, the dihedral α does not visit at any time the reported other two clusters close to±180°. Since the complete 20 ns of simulation time is shown there, an error is assumed in either the cluster analysis or that Fig. 6.

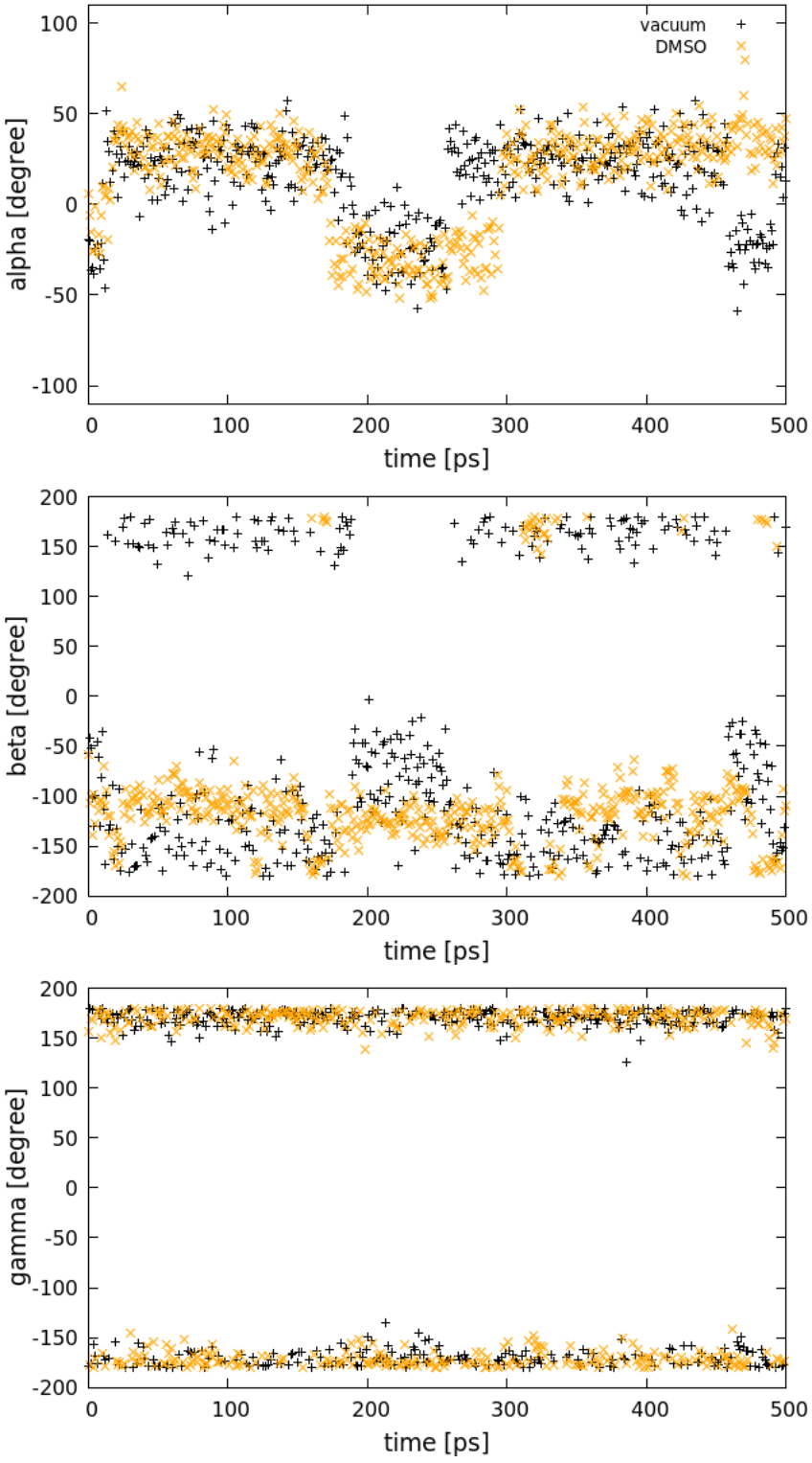


Fig. 5.6.: Dihedral angles sampled in vacuum (black) and DMSO (orange).

For the dihedral β in vacuum and in DMSO four clusters were found (Fig. 5.5, right coloumn). In DMSO, however, the dihedral angles are shifted to more negative values, such that the green and blue clusters could also be summarized in only one cluster. In vacuum as well as in DMSO, the area between roughly 0 and +135° is not sampled. In the force field simulation in vacuum instead a cluster is found in this area. Again, this cluster may be missing here due to the limited simulation time. However, it would be interesting to find out if the solvent prevents sampling in this area as suggested in Ref. [166]. Further, the force field sampled β dihedrals between 0 and −45° in DMSO which are also not observed in the QM/MM simulation. The fluorescence energies of the cluster centers here show a difference of 0.1 eV (0.09 eV) between the lowest and the largest in DMSO (vacuum) (Tab. 5.1). In Ref. [114], a difference of 0.2 eV is reported in DMSO and 0.1 eV in vacuum, but the trend between the clusters is not the same as found here.

The dihedral γ only fluctuated around 180° in both vacuum and DMSO (Fig. 5.6) in the QM/MM simulations. In Ref. [114] in contrast a completely different picture is presented. In gas phase, most of the time the dihedral fluctuates between 90° and 160°, mainly being around ≈ 150°. For a short time (≈ 1 ns) the negative side is sampled. During the flip to the other side and back, the dihedral was shortly staying in between of 0° and −25°. In DMSO, γ also stayed the main time around 150°, but the area between 20° and 45° was visited several times, too.

In sum, the sampled α dihedrals are in agreement with the force field results. The β dihedrals are slightly different but the differences may be only a results of the shorter QM/MM simulation time. The sampled γ dihedral is not in agreement with the force field and this needs further investigation. A clear, significant relation between the sampled dihedrals α and β and the corresponding fluorescence energies was not found as it was not in Ref. [114].

In Ref. [114], a stable hydrogen bond between N3-H11 of Flugi-2 (Fig. 5.1) and the oxygen of a DMSO molecule was detected and found responsible for the sampling of different dihedral angles in DMSO and vacuum. It may further have an important direct impact on the fluorescence energies since an implicit solvent model was not able to reproduce the solvent shift. Along the QM/MM trajectory, this hydrogen bond is present in only 71 % of the frames, if the default criteria of gromacs are applied (max. distance between donor and acceptor: 3.5 Å, max. angle hydrogen-donor-acceptor: 30°). Neglecting the angle criterium, in 98 % of the frames at least one DMSO oxygen is in a radius of 3.5 Å of the donor N3. The DMSO molecules building the hydrogen bond at the beginning stayed in the proximity for around 200 ps frequently forming the hydrogen bond (Fig. 5.7).

The reason for the less stable hydrogen bond are probably the much more moderate atomic partial charges in the QM/MM simulation. In Tab. 5.2, the average atomic partial charges and their standard deviations are reported and compared to the (fixed) atomic partial charges used in the force field for Flugi-2 in the excited state. While the partial charges of N3 and H11 in the force field are −1.235 and 0.515, they are only −0.166 and 0.253 in the QM/MM simulation.

Tab. 5.2.: Excited state atomic partial charges in the QM/MM simulation compared to the atomic partial charges of the force field used in Ref. [114]. The atomic partial charges in the QM/MM simulation are determined from a Mulliken population analysis. The averages over 500 snapshots spaced by 1 ps are shown. The standard deviation is reported in parentheses, respectively.

Atom	QM/MM	Force field [114]
C1	-0.130 (0.021)	−0.006
C2	-0.120 (0.029)	−0.440
N1	0.087 (0.026)	0.150
C3	0.302 (0.020)	0.564
C4	-0.066 (0.019)	−0.545
C5	-0.353 (0.020)	−0.384
C6	0.686 (0.011)	1.063
O1	-0.627 (0.024)	−0.645
O2	-0.348 (0.023)	−0.318
C7	0.065 (0.012)	−0.258
N2	-0.442 (0.016)	−0.669
C8	0.024 (0.023)	0.150
C9	0.211 (0.024)	0.674
C10	-0.003 (0.014)	−0.797
C11	-0.069 (0.013)	0.616
C12	-0.194 (0.018)	−1.311
C13	0.327 (0.016)	1.248
C14	-0.100 (0.016)	0.941
C15	-0.182 (0.018)	−1.100
N3	-0.166 (0.024)	−1.235
H11	0.253 (0.013)	0.515
C16	0.124 (0.013)	0.107
C17	-0.132 (0.010)	1.131
C18	-0.133 (0.011)	0.685
C19	-0.106 (0.010)	−0.502
C20	-0.109 (0.011)	0.754
C21	-0.107 (0.010)	−0.637
O3	-0.324 (0.020)	−0.290
C22	0.059 (0.011)	−0.620

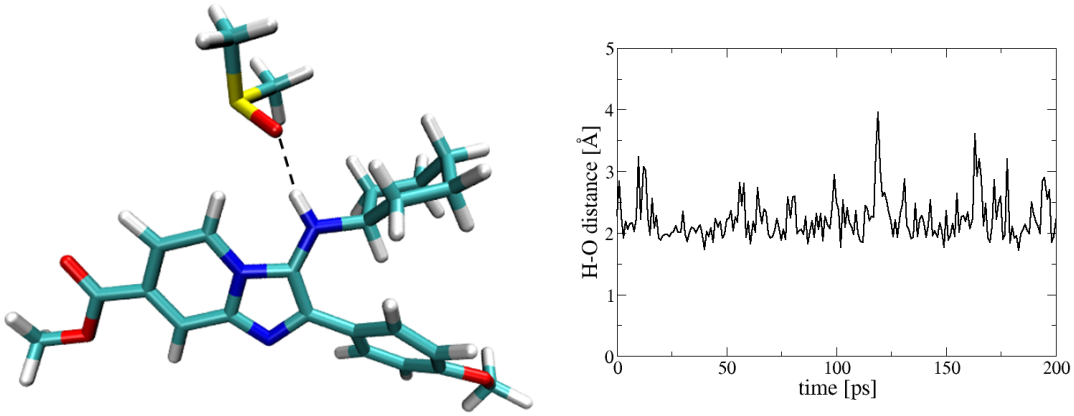


Fig. 5.7.: Representative snapshot from the QM/MM trajectory with α=41°, β=-117° and γ=170°. N3-H11 is forming a hydrogen bond with a DMSO molecule. The distance of H11 and the DMSO oxygen is shown on the right. It is frequently building a hydrogen bond during the first 200 ps of the simulation, then it is replaced by another DMSO molecule.

Fig. 5.8 shows histograms of the fluorescence energies weighted by their oscillator strengths computed with OM2 along the simulation in DMSO and the simulation in vacuum. The fluorescence energies of the snapshots from the DMSO trajectory were calculated once considering the solvent as point charges and once without environment. This allows to distinguish between geometrical and electrostatic effects of the solvent on the fluorescence energies. According to the results in Fig. 5.8 the solvatochromic shift is a pure electrostatic effect, as the histogram from the DMSO trajectory without point charges in the fluorescence calculation is almost identical to the histogram from the vacuum simulation and has its maximum at the same fluorescence energy. The maxima of the Gauss fits are reported in Tab. 5.3. The solvatochromic shift between the DMSO and vacuum amounts to -0.36 eV. This value is slightly overestimated compared to the experiment, where a shift of -0.26 eV was measured between decane and DMSO [114, 115]. The shift obtained with the particularly parametrized classical force field amounted to 0.21 eV, which is slightly underestimated. Moreover, a small blue shift of 0.09 eV was observed from the histogram of the DMSO trajectory without point charges compared to the histogram of the vacuum simulation [114].

It was additionally checked if TD-LC-DFTB2 predicts the same response to the electrostatics determined by the solvent and to the different geometries as observed with OM2. The Gaussian fits to the obtained histograms are shown in Fig. 5.9. Tab. 5.3 lists the detailed values of the fluorescence energies and the shifts due to the total solvent effect and the solvent effect on the sampled geometries. The TD-LC-DFTB2 shifts are quite similar to the OM2 shifts. With -0.35 eV for the OB2 (base) parameter set and -0.33 eV for the electronical, optimized paramter set the total solvent shifts are even closer to the experimental value of -0.26 eV. A small red shift is predicted by TD-LC-DFTB2 between the vacuum calculations along the DMSO trajector´y and the vacuum trajectory. This is in contrast to the OM2 results of this work, which predict no shift at all and also contrary to the force field/OM2 results from [114], which state a blue shift of 0.09 eV. The shift

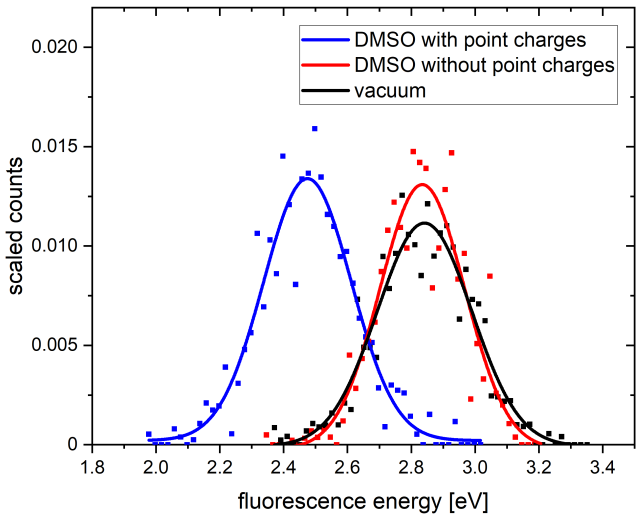
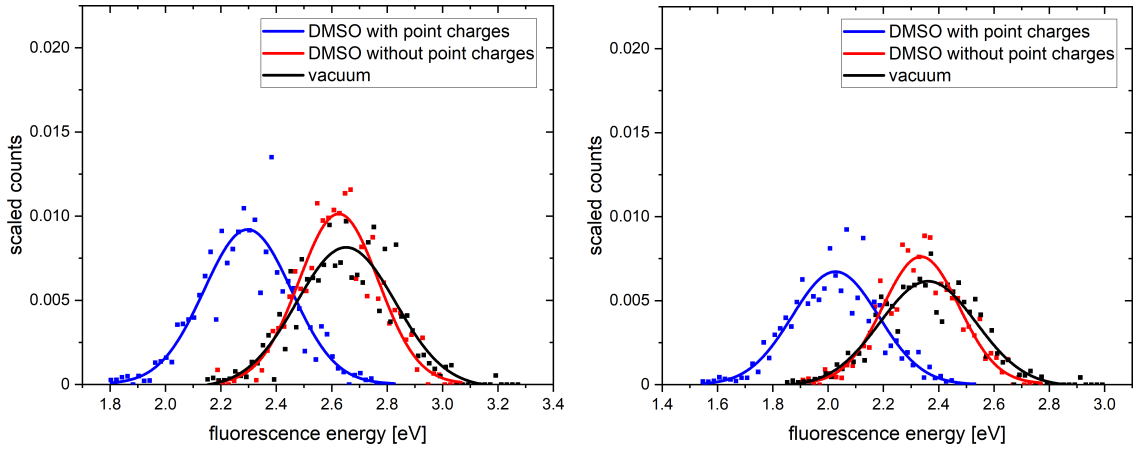


Fig. 5.8.: Gaussian fits to histograms of the fluorescence energies computed with OM2. The intensity is determined from the counts per window and their oscillator strengths. The excitation energies along the DMSO trajectory computed without point charges are almost identical to the ones obtained in vacuum.

resulting from the electrostatics corresponds to the difference between the results obtained with and without point charges along the DMSO trajectory. With OM2 it is equal to the total shift (-0.36 eV), with TD-LC-DFTB2 it is -0.33 eV (-0.31 eV) with the OB2 (opt. electronical) parameters. In Ref. [114] the shift arising from the point charges amounts to -0.30 eV computed with OM2. The widths of the Gaussians fitted to the histograms are always a little larger in the case of TD-LC-DFTB2 compared to OM2. However, in both cases, the widths of the calculations along the DMSO trajectory without point charges are the smallest and the widths of the vacuum results are the largest.

Further it is remarkable that the absolute fluorescence energies obtained with TD-LC-DFTB2 are smaller than the ones obtained with OM2. This is particularly conspicuous for the electronical parameters optimized for excitation energies. Their fluorescence energy maxima are up to 0.5 eV red shifted compared to OM2. The OM2 fluorescence energies computed along the TD-LC-DFTB2 excited state trajectory in turn are already red shifted in the order of 0.2 eV compared to the OM2 results obtained along the force field trajectory as reported in Ref. [114]. The experimental fluorescence maxima in decane and DMSO are at 2.61 eV and 2.35 eV [114,115]. Thus, regarding the absolute values, TD-LC-DFTB2/OB2 (base) comes very close to the experiment and in general, the TD-LC-DFTB2 excited state geometries are superior to the force field geometries in this respect. Finally, it is noted that the oscillator strengths obtained with TD-LC-DFTB2 are slightly smaller which results in a systematic lower intensity in all of its spectra compared to OM2. The relative intensitites of the respective three spectra computed with each method are the same except that the spectra for DMSO with point charges computed with TD-LC-DFTB2 have an intensity lower than that of DMSO without point charges, while with OM2 it is a little larger.

In sum, the differences in the shifts resulting from geometrical differences and electro-



(a) OB2 (base) parameters **(b)** pure electronical parameters

Fig. 5.9.: Gaussian fits to histograms of the fluorescence energies computed with TD-LC-DFTB2. The intensity is determined from the counts per window and their oscillator strengths.

static effects are found to be small between OM2 and TD-LC-DFTB2 with both paramter sets. Thus, based on these data TD-LC-DFTB2 is assessed as suitable method for the study of environmental influences on fluorophores. Differences in the absolute positions of the spectra were detected, but they play a minor role compared to the relative energies. Here, TD-LC-DFTB2/OB2 (base) came closest to the experiment. The differences in the intensities resulting from the oscillator strengths predicted by TD-LC-DFTB2 and OM2 are small. In particular, the heights of the three different spectra computed with each method are always in the same order, which in a study of environmental effects again is more important than the absolute values.

5.2.3. Conclusion and Outlook

TD-LC-DFBT2/OB2 (base) was assessed as QM method in a QM/MM simulation with a QM molecule in the excited state. A previous study on the same fluorophore, Flugi-2, using a force field particularly parametrized for the excited state [114] served as reference as well as the experiment [115]. The results of the new excited state QM/MM simulations are of similar quality compared to the experiment as the results obtained with the special force field. This is encouraging since the new approach saves the tedious parametrization of a dye in the excited state. Further, the issue of fixed force field parameters disappears. In particular, the atomic partial charges could be sensitive to the environment. In a QM/MM simulation in contrast to the classical MD simulation with the force field, the polarization of the fluorophore from the environment is captured during the simulation. However, the computational cost of such a QM/MM simulation is larger than that of a classical simulation and thus the possbile simulation time decreased to only few hundreds of picoseconds. This in turn means, that important conformations could be missed in one single QM/MM simulation.

The experimentally observed red shift in the emission spectrum of Flugi-2 in DMSO origins from the direct electrostatic effect from DMSO. Indirect effects on the geometries in

Tab. 5.3.: Maxima and standard deviations of the gaussian fits to the histograms obtained from the fluorescence energies computed with OM2 and TD-LC-DFTB2 with different paramter sets. The shifts between the maxima from the calculations along the solvent trajectory compared to the respective maxima in vacuum are given.

		DMSO		vacuum
		with point charges	without point charges	
OM2	max.	2.48 eV	2.84 eV	2.84 eV
	st. dev.	0.14 eV	0.13 eV	0.15 eV
	shift	-0.36 eV	0.00 eV	
TD-LC-DFTB2 OB2 (base)	max.	2.30 eV	2.63 eV	2.65 eV
	st. dev.	0.16 eV	0.14 eV	0.18 eV
	shift	-0.35 eV	-0.02 eV	
TD-LC-DFTB2 el. param.	max.	2.03 eV	2.34 eV	2.36 eV
	st. dev.	0.16 eV	0.14 eV	0.17 eV
	shift	-0.33 eV	-0.02 eV	

the solvent are small. These main findings from the classical simulation could be confirmed, but a more detailed study is necessary to find the reasons for the different deviations from the experiment and to evaluate which of the individual factors determining the spectra are accurately predicted with TD-LC-DFTB2. For example, the strength of the hydrogen bond is differently predicted by the classical force field and the QM/MM simulation. Furthermore, the dihedral angles, in particular γ, differ in both trajectories. In addition, a question to answer regarding the solvatochromism of Flugi-2 is why the implicit solvation model failed. This could likely be related to the hydrogen bond but also with the distribution of the other solvent molecules around the excited molecule.

While TD-LC-DFTB2 showed to be a suitable method for simulation as well as excitation enegy calculation in the case of Flugi-2, it needs further investigation on other fluorophores.

5.3. Solvatochromism of Prodan

As outlined in the introduction of this fluorophore, the states and conformers involved in the absorption and emission processes are still subject of discussion. It is thus started with an analysis of the excited states predicted by TD-LC-DFTB2 and reference QM methods on optimized geometries in vacuum. The results are compared to literature to identify the target state for the QM/MM simulations. A simulation of this state in vacuum turned out to be challenging, as will be described afterwards. Then the results of QM/MM simulations in solvent are reported. It is concluded with a final evaluation of the performance of TD-LC-DFTB2 in the study of prodan and a discussion about possible ways to continue.

5.3.1. Computational Details

Prodan was optimized with (TD-)LC-DFTB2/OB2$^{0.3}$ (base) in the ground and excited state in vacuum using DFTB+ (version 20.1). The conjugate gradient algorithm was chosen with the default convergence tolerance of 1×10^{-4} $\frac{\text{Ha}}{a_0}$ for the forces. The SCC tolerance accounted to 1×10^{-7} electrons and the thresholded screening method was applied with a threshold of 1×10^{-16} and a range-separation parameter of ω=0.3 a_0^{-1}. A subsequent frequency analysis was performed using the mode program of DFTB+, respectively.

For comparison, the dye was optimized with CC2 in ground and first excited state using turbomole (version 7.3) [169,170]. The aug-cc-pVTZ basis set [171] and its auxiliary basis for the RI approximation were applied. Core orbitals, the 17 lowest occupied MOs, were frozen. The density and the SCF energy convergence thresholds both amounted to 1×10^{-7}. No frequency analyses were performed due to the high computational cost.

Additionally, optimizations in ground and first excited state were performed with the ORCA program package (version 4.0.0.2) at CAM-B3LYP/def2-TZVP level of theory. The TightOpt keyword was employed, which sets the energy convergence tolerance to 1×10^{-6} Ha and the maximum gradient tolerance to 1×10^{-4} Ha. The RIJCOSX approximations were used and the def2/J auxiliary basis set. The energy and residual convergence tolerance for the excited state amounted to 1×10^{-6} Ha. The excited state optimization started from the ground state optimized structure. An excited state optimization at the same level of theory was performed previously by Beatrix M. Bold but with a slightly different start structure. Frequency analyses confirmed in both cases that a minimum was found.

Prodan was simulated in vacuum as described above for Flugi-2: a leap-frog stochastic dynamics integrator was used with a time step of 1 fs at 300 K. The simulations were run for 500 ps in the ground state, the first excited state and the second excited state, respectively.

In addition, ground and excited state QM/MM simulations were performed with prodan solvated in DMSO and in water, respectively. The complete dye constituted the QM part computed with TD-LC-DFTB2/OB2 (base) and the surrounding solvent molecules were described by a force field. For DMSO, again the force field from Ref. [114] was used. Water was described by the TIP3P water model. The QM/MM implementation was the same as used for Flugi-2, a combination of Gromacs and DFTB+. The system was equilibrated in ground state with the same procedure as described above. In the productive simulations, the timestep was 1 fs and bonds with hydrogen were kept constraint in the MM part. The Nosé–Hoover thermostat [162,163] maintained a temperature of 300 K and the Parrinello–Rahman barostat [164,165] controlled the pressure with a reference value of 1.013 bar. PBC were applied and the PME method accounted for long-range electrostatics. The coordinate precision in the output trajectory was increased to 1×10^{-6} nm. The ground state and first excited state were simulated for 500 ps, respectively. An additional simulation was performed in the second excited state in water, also for 500 ps.

Excitation energies were computed with TD-LC-DFTB2/OB2 (base) and TD-LC-DFTB2 with the electronical parameter set optimized for excitation energies [3,104] using DFTB+. The SCC tolerance was set to 1×10^{-8} and the screening threshold to 1×10^{-16}. In the calculations considering the environment, the fluorophore was centered in the simulation box and the surrounding solvent molecules were represented by their point charges.

Excitation energies from DFT were computed with ORCA at B3LYP/def-TZVP and

CAM-B3LYP/def2-TZVP level. The TightSCF keyword was applied and the RIJCOSX approximation with the def2/J auxiliary basis set. CC2 excitation energies were calculated with the aug-cc-pVTZ basis set using turbomole. The RI approximation was used and core orbitals were frozen.

Plots of orbitals from DFT calculations were created with the orca_plot program and plots of LC-DFTB2 orbitals were created with the waveplot program of DFTB+. The orbitals were visualized with VMD [100]. The isovalue was set to ±0.05, respectively.

5.3.2. State Characterization

The optimized geometries in ground and first excited state, the vertical excitation energies, oscillator strengths and orbital transitions are investigated. Table 5.4 shows a comparison of geometrical parameters describing the optimized structures in ground state. These are the BLA of the naphthalene core, the dihedral angles between the dimethylamino moiety and the naphthalene core as well as the carbonyl moiety and the naphthalene core. Further, the bond lengths between the nitrogen and the neighbouring aryl carbon, between the carbon and the oxygen of the carbonyl moiety and between the same carbon and the neighbouring aryl carbon are shown. Evidently, the methods provide similar ground state structures. The carbonyl moiety is almost in plane with the ring system and the dimethylamino moiety has a slight pyramidal shape. Prodan fluctuates also in the ground state QM/MM simulations around the planar structure (see appendix).

Tab. 5.4.: Geometrical parameters of prodan measured on the optimized geometries in ground state. The bond length alternation is abbreviated by BLA and includes all bonds in the naphthalene core. CAM-B3LYP is abbreviated by CAM. The OB2 (base) parameters were used in the LC-DFTB2 optimization.

	BLA	dih(N-Ar)	dih(Ar-CO)	d(N-C)	d(C=O)	d(C-CO)
CC2	0.024 Å	12.28°	-0.10°	1.387 Å	1.236 Å	1.486 Å
CAM	0.037 Å	10.29°	6.17°	1.378 Å	1.211 Å	1.488 Å
LC-DFTB	0.034 Å	10.34°	-5.59°	1.393 Å	1.227 Å	1.495 Å

Table 5.5 reports the first three excitation energies together with the respective oscillator strengths on these ground state geometries for all methods. These three excitation energies lie close together according to CAM-B3LYP and CC2 (within 0.2 eV). TD-LC-DFTB2 instead predicts larger gaps between the excitation energies. The second excitation energy is 0.24 eV larger than the first excitation energy and the third excitation energy is 0.61 eV larger than the second excitation energy. Additionally, a different state ordering between the methods is indicated by the oscillator strengths. This is not surprising because the states are close in energy. From the CAM-B3LYP or CC2 results it can be expected that all of these states are populated upon excitation. However, the transition with the small oscillator strength is less likely than the other two transitions.

It was checked, if the excitation energies differ mainly because of the different geometries or because of the different methods applied to compute them. Excitation energies were thus computed with different methods (TD-LC-DFTB2, B3LYP, CAM-B3LYP) using the TD-LC-DFTB2 geometry and the CAM-B3LYP geometry, respectively. The detailed values are

given in the appendix in Tables A.15 and A.16. All three excitation energies of all methods are systematically increased by roughly 0.15 eV when changing from the TD-LC-DFTB2 geometry to the CAM-B3LYP geometry. The oscillator strengths change only slightly, except for the first two excitations of CAM-B3LYP. These excitations change their order. It is concluded that the geometry of TD-LC-DFTB2 in ground state is indeed reasonable but the excitation energies predicted by TD-LC-DFTB2 differ from the other methods. However, also B3LYP, CAM-B3LYP and CC2 give no consistent picture.

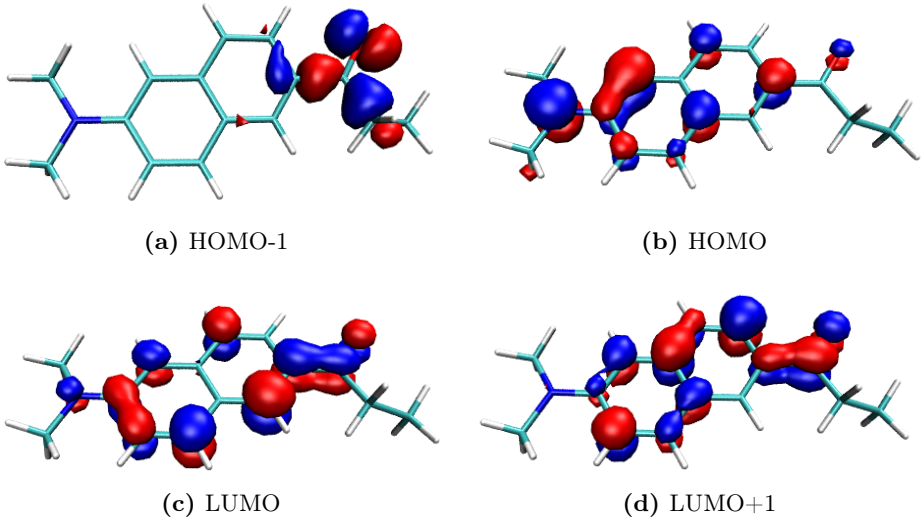
The experimental absorption spectrum is broad. In cyclohexane it shows a peak at $\approx$ 3.6 eV and two shoulders, one at $\approx$ 3.7 eV on the blue side and $\approx$ 3.3 eV on the red edge [15]. The latter is found at an intensity of less than half of the intensity of the peak. This can be explained by absorption into different states but could also indicate that different species like planar and non-planar conformers are absorbing. The latter is assumed for the similar dye badan in Ref. [22]. The vertical excitation energies in Tab. 5.5 do not agree well with the experimental values.

Tab. 5.5.: Excitation energies of prodan in vacuum computed on optimized ground state geometries. The values are in eV, oscillator strengths are shown in parentheses.

	TD-LC-DFTB2/OB2 (base)	CAM-B3LYP	CC2
S1	3.69 (4.6×10^{-5})	4.13 (0.049)	3.72 (0.22)
S2	3.93 (0.23)	4.15 (0.13)	3.84 (5.5×10^{-5})
S3	4.54 (0.076)	4.32 (0.28)	3.92 (0.28)

An orbital analysis reveals that the first excited state of TD-LC-DFTB2 in vacuum can be described mainly by the HOMO-1 to LUMO transition, the second excited state by the HOMO to LUMO transition and the third excited state by the HOMO to LUMO+1 transition. The orbitals are depicted in Fig. 5.10. Hence, the first excited state corresponds to an $n\rightarrow\pi^*$ state and the other two states are $\pi\rightarrow\pi^*$ states. The first excited state of CAM-B3LYP is the same $n\rightarrow\pi^*$ state but the n-orbital corresponds to the HOMO-2 in that case. The higher excited states S2 and S3 are mainly described by the HOMO to LUMO transition but there are relatively large contributions from other orbital transitions, too. The HOMO and LUMO of CAM-B3LYP agree quite well with the LC-DFTB2 counterparts (Fig. 5.10).

Pictures of orbitals involved in the transitions to the first three excited states are also presented in Ref. [15]. These were obtained with CASPT2. The HOMO-3 shown in Ref. [15] corresponds to the HOMO-1 of TD-LC-DFTB2. The HOMO and LUMO have in common with the TD-LC-DFTB2 HOMO and LUMO that they are π orbitals with more density on the dimethylamino (carbonyl moiety) but they show differences in the ring system. The LUMO+1 is only distributed over the rings. The first excited state predicted by CASPT2 is described by a combination of its HOMO to LUMO and HOMO-1 to LUMO transition, the second excited state is the $n\rightarrow\pi^*$ state and the third excited state corresponds mainly to the HOMO to LUMO transition and is thus a $\pi\rightarrow\pi^*$ state [15]. Quasiparticle orbitals from GW-BSE calculations are shown in Ref. [151]. These are similar to the orbitals shown in Ref. [15] but have a different order. The largest difference between these orbitals and the TD-LC-DFTB2 orbitals is found on the ring system of the HOMO.



(a) HOMO-1 **(b)** HOMO

(c) LUMO **(d)** LUMO+1

Fig. 5.10.: Orbitals of prodan obtained with TD-LC-DFTB in vacuum. The first excited state is characterized by the transition from HOMO-1 to LUMO.

In the next step, the geometry of prodan was optimized in the first excited state. Table 5.6 reports the geometrical parameters of the resulting optimized geometries. The difference between the two CAM-B3LYP variants is only the start structure. The comparison of the first excited state geometries reveals similarities between CC2 and CAM-B3LYP variant a on one hand and TD-LC-DFTB2/OB2 (base) and CAM-B3LYP variant b on the other hand. This indicates that the minima of two different states were obtained. The energy of excited prodan in the minimum found by CAM-B3LYP variant b is about 0.12 eV lower than the energy of excited prodan in the minimum of CAM-B3LYP variant a. In particular, the BLA value is clearly different in the different states. Interestingly, only CC2 predicts a complete planar excited state geometry with the dihedral angles very close to zero.

Tab. 5.6.: Geometrical parameters of prodan measured on the first excited state optimized geometries. The bond length alternation is abbreviated by BLA and includes all bonds in the naphthalene core. A negative value indicates a bond inversion compared to the ground state. CAM-B3LYP is abbreviated by CAM. The OB2 (base) parameters were used in the TD-LC-DFTB2 optimization.

	BLA	dih(N-Ar)	dih(Ar-CO)	d(N-C)	d(C=O)	d(C-CO)
CC2	-0.014 Å	0.28°	-0.01°	1.373 Å	1.245 Å	1.475 Å
CAM[a]	-0.005 Å	10.31°	6.48°	1.361 Å	1.214 Å	1.485 Å
CAM[b]	0.036 Å	7.85°	3.14°	1.387 Å	1.288 Å	1.417 Å
LC-DFTB	0.029 Å	13.45°	0.14°	1.405 Å	1.292 Å	1.412 Å

[a] structure 1 from Beatrix M. Bold
[b] structure 2 from optimization starting in ground state minimum

The orbital transitions describing the first excited state were again analyzed, this time on the geometries optimized in the first excited state. In the case of TD-LC-DFTB2, the first excited state is still described by the HOMO-1 to LUMO transition with the orbitals depicted in Fig. 5.10. This $n\rightarrow\pi^*$ state is also the first excited state of CAM-B3LYP variant b. The first excited state of CAM-B3LYP variant a instead is a $\pi\rightarrow\pi^*$ state. This indicates that the potential energy surfaces of this $\pi\rightarrow\pi^*$ state and the $n\rightarrow\pi^*$ state are crossing in between of the different start structures of the CAM-B3LYP optimizations. The $\pi\rightarrow\pi^*$ state is mainly described by the HOMO to LUMO transition (Fig. 5.10).

The dipole moments of the optimized excited states in comparison to the ground state dipole moments as well as the fluorescence energies and the oscillator strengths are summarized in Tab. 5.7. Again, clear differences between the two kinds of states become visible. The $n\rightarrow\pi^*$ state has a lower dipole moment than the ground state while the $\pi\rightarrow\pi^*$ state has an increased dipole moment.

The excited state dipole moment in benzene was experimentally determined to be around 10 D [137]. This indicates, as well as the low oscillator strengths, that the $n\rightarrow\pi^*$ state is not relevant for the emission in nonpolar environment.

Tab. 5.7.: Comparison of first excited state properties of prodan in vacuum. The methods were consistently used for optimization and computation of the properties, respectively.

method	dipole ground	dipole exc.	osc. strength	fluoresc. E
CC2	5.8 D	12.0 D	0.21	3.29 eV
CAM-B3LYP[a]	5.8 D	9.8 D	0.19	3.81 eV
CAM-B3LYP[b]	5.8 D	2.6 D	1.5×10^{-4}	3.53 eV
LC-DFTB2	5.9 D	2.0 D	3.7×10^{-5}	2.98 eV

[a] structure 1 from Beatrix M. Bold
[b] structure 2 from optimization starting in ground state minimum

It is thus focussed on the $\pi\rightarrow\pi^*$ states now. On the optimized ground state minimum, the dipole moment of the second excited state is 11.1 D and the dipole moment of the third excited state is 10.3 D according to TD-LC-DFTB2. Thus, both values are close to the experiment. The third excited state was optimized and the geometry was compared to the CC2 and CAM-B3LYP variant a first excited state structures. However, they do not agree well (Tab. 5.8). The orbital transition of the second excited state of TD-LC-DFTB2 instead agrees with the transition of CAM-B3LYP variant a describing the first excited state. Another argument against S3 as the emitting state is the larger energy compared to S2. Furthermore, the oscillator strength of S2 (0.23) is larger than the oscillator strength of S3 (0.076), see Tab. 5.5.

5.3.3. State Following Approach

Having identified the second excited state as (main) emissive state, it was tried to optimize and simulate in the second excited state. However, these attempts were not succesful, because the potential energy surfaces of this state and the $n\rightarrow\pi^*$ state crossed. The optimization in S2 did not converge, although it was run for 2000 steps. The maximal force component frequently hopped between different values. The point of the S1/S2 intersection

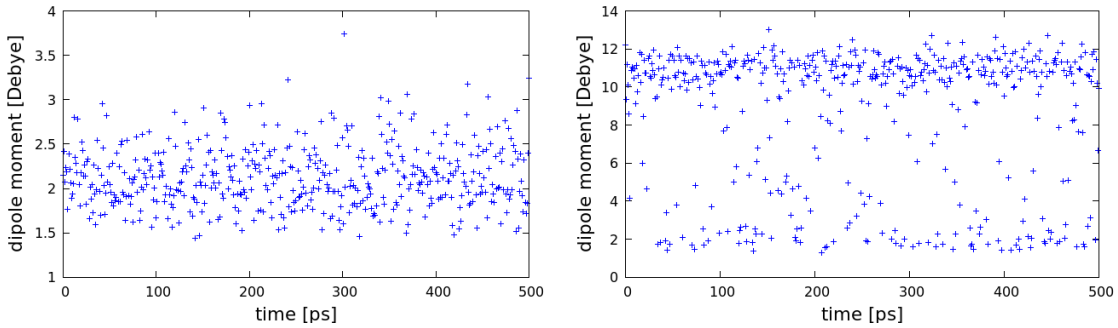
Tab. 5.8.: Geometrical parameters of prodan measured on the optimized geometries of the low-lying $\pi \rightarrow \pi^*$ states. The bond length alternation is abbreviated by BLA and includes all bonds in the naphthalene core. A negative value indicates a bond inversion compared to the ground state. The OB2 (base) parameters were used in the TD-LC-DFTB2 optimization.

	BLA	dih(N-Ar)	dih(Ar-CO)	d(N-C)	d(C=O)	d(C-CO)
CC2 (S1)	-0.014 Å	0.28°	-0.01°	1.373 Å	1.245 Å	1.475 Å
CAM-B3LYP[a] (S1)	-0.005 Å	10.31°	6.48°	1.361 Å	1.214 Å	1.485 Å
LC-DFTB2 (S3)	0.007 Å	5.79°	-1.60°	1.394 Å	1.247 Å	1.466 Å

[a] structure 1 from Beatrix M. Bold

is frequently crossed also during the simulation in S2. This is visualized in Fig. 5.11, which shows the dipole moments along a simulation in S1 and a simulation in S2 in vacuum. It seems that the sampling area of the simulation in S1 is far enough from the conical intersection, so that there is no state crossing observed. Calculation of the dipole moments of the second excited states using the geometries from the S1 simulation were all in the expected area around 10-12 D. However, since the minima of the potential energy surfaces of the different excited states have different locations, it is possible, that while sampling in S2 the intersection point is frequently visited, as observed here. The dipole moment of the third excited state is around 10 D and thus close to the one of the second excited state. To make sure, that S2 and S3 are not additionally crossing, the excitation energies, oscillator strengths and orbital transitions of the third excited state were inspected along the trajectory of the S2 simulation. The third excitation energies are well separated from the first and second excitation energies, and also the oscillator strengths and orbital transitions do not indicate a further crossing.

Due to the issue of crossing states, a more advanced approach is needed to optimize the correct state and to sample in the correct state. There are state-tracking procedures available in quantum chemistry packages, which use the overlap of configuration interaction vectors or the overlap of attachment-detachment transition densities as criterium to identify



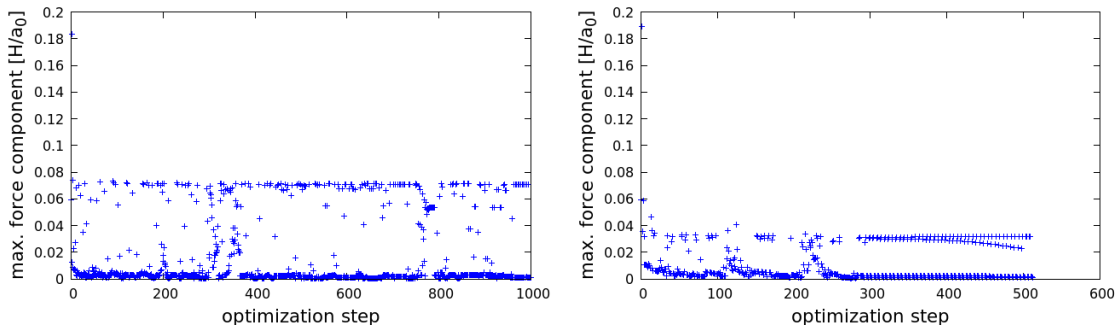
(a) S1 **(b)** S2

Fig. 5.11.: Excited state dipole moments along a trajectory in S1 and S2 in vacuum. The simulation in S2 partly samples the same state as the S1 simulation.

from all computed states at the current step the one which correpsonds to the reference state [172]. Alternatively, the overlap between natural transition orbitals (NTOs) can be utilized, as described in Ref. [172]. For this project, Thomas Niehaus implemented a state-tracking algorithm into DFTB+, which is similar to this NTO overlap approach [173]. The target state is determined every step according to the normalized overlap between eigenstates of the difference density matrix.

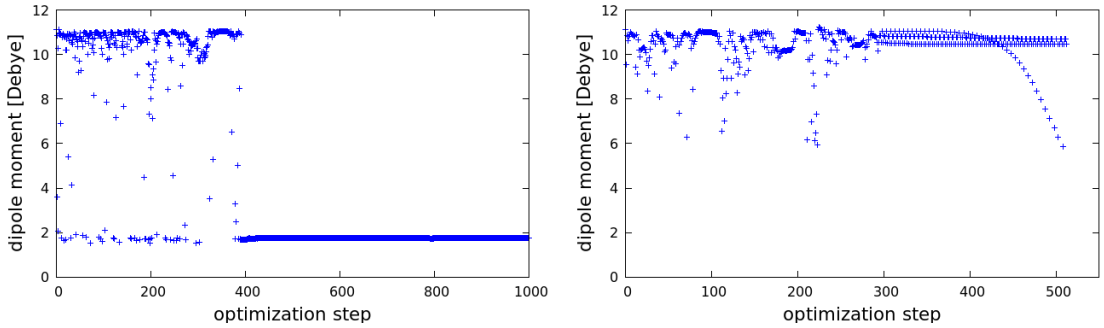
A test optimization of prodan showed that this state-tracking generally works since several correct state switches were performed. However, it was noticed that a necessary state switch was not performed later and finally still the wrong state was optimized. This happened because the overlap of the target state with the correct excited state and the wrong excited state in this optimization step were very similar. To avoid a wrong decision of the program in such a case, Thomas Niehaus implemented an error stop if the overlap to two states is more similar than 0.05 [173]. The optimization was performed again and when such a point was reached, the geometry was manually changed or the optimization was restarted with a different step size or another intergrator. Unfortunately, it was not succesful. To reduce the amount of these error stops, the similarity threshold of 0.05 was halved to 0.025. At the same time, Tomáš Kubař implemented a further comparison of the current states with the states of the next to last step in case that the overlaps with the last steps are not robust enough to find the target state [174]. Now, only if the overlaps of the current excited states with the target state from the last step and the next to last step are both too similar ($\Delta < 0.025$), the optimization stops.

Nonetheless, it was not possible to optimize the target state of prodan with TD-LC-DFTB2/OB2 (base). In Fig. 5.13, the maximum force component, which indicates if the optimization converges, is plotted versus the optimization steps for a vacuum optimization of prodan with and without the use of the state-following approach. Here, none of the optimizations found a minimum. Fig. 5.12 shows the dipole moments along both optimizations. The dipole moments were chosen since they allow an easy identification of the two different states. It is seen on the left hand side, that in the first half of the optimization the states switch several times. Then, only the wrong state is optimized. On the right hand side, the optimization stays in the correct state.



(a) opt of S2 **(b)** opt with state-tracking

Fig. 5.13.: Maximum force components during the optimizations. Neither the optimization of S2 nor the optimization with state-tracking reached the minium of the target state.



(a) opt of S2 **(b)** opt with state-tracking

Fig. 5.12.: Excited state dipole moments along the optimization steps of an optimization in S2 and an optimization following the target state, which corresponds to S2 at the beginning. A dipole moment in the region of 10-12 D indicates the target state, the low dipole moments indicate a different state. The optimization without state-tracking resulted in the optimized geometry of the wrong state. The state-following approach prevented this problem, but still was not succesful.

5.3.4. Solvated Prodan

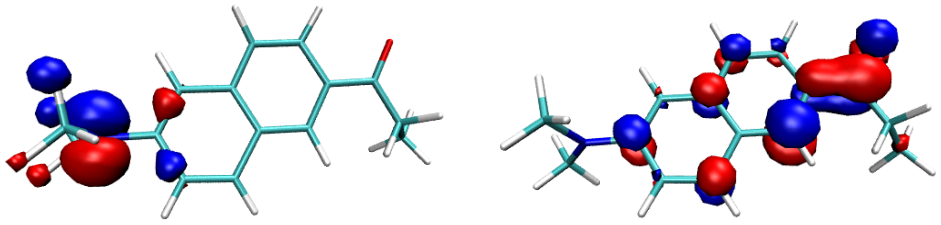
QM/MM simulations were performed to shed light on the solvent effect on the fluorescence of prodan. The polar aprotic solvent DMSO and water as polar protic solvent were chosen and compared to isolated prodan simulated in vacuum. In both solvents, TD-LC-DFTB2/OB2$^{0.3}$ (base) predicts a different state than in vacuum as first excited state. The corresponding orbital transition is shown in Fig. 5.14 exemplary from the conformation after 300 ps of the simulation in water. It is now the HOMO to LUMO transition, with the same LUMO as in vacuum and a HOMO localized on the nitrogen. The predicted dipole moment of this state is huge and the fluorescence energies are small. Fig. 5.15 shows the values computed with TD-LC-DFTB2/OB2$^{0.3}$ (base) along the trajectory in water. There the dipole moments fluctuate around 24 D, in DMSO the value is with around 20 D a little lower. The fluorescence energies in DMSO are slightly larger ($\approx$ 1.4 eV), compared to the fluorescence energies in water ($\approx$ 1.2 eV). In both solvents, the dimethylamino moiety is twisted around 90° compared to the ring plane. The oscillator strengths are very small, most times below 0.01.

A large dipole moment of 24 D in water was also found in Ref. [151] taking polarisation into account. However, though also reporting a charge-transfer state from the nitrogen side to the oxygen side, the main transition describing the excitation proceeds between two π orbitals (similar to Fig. 5.10 b and c) and no n-orbital as predicted by TD-LC-DFTB2 (Fig. 5.14). The twisted geometry of prodan disagrees with the TD-LC-DFT study of Marini *et al.* [145] on prodan in water and the experimental findings reported in refs. [138, 139] for prodan in DMSO. Further, the small oscillator strengths argue against this state as relevant state in the fluorescence of prodan.

The second excited state computed on these S1 geometries in DMSO and water corresponds mainly to the HOMO to LUMO+1 transition. The oscillator strengths are similarly small as for the transition in the first excited state but this could be traced back to the twisted geometry. A trial simulation in S2 was performed in water and indeed, a differ-

ent picture was obtained. The sampled geometries are planar and belong to two different states which are frequently crossing during the simulation. The slightly more dominant state is described by a HOMO-1 to LUMO transition. These orbitals are the same as in vacuum (Fig. 5.10). The second state is a $\pi \rightarrow \pi^*$ state described by the HOMO to LUMO transition. The orbitals are very similar to the orbitals in vacuum (Fig. 5.10). The $\pi \rightarrow \pi^*$ state has a dipole in the order of 15 D, the $n \rightarrow \pi^*$ state has a dipole in the order of 5 D.

To sum up, again it turns out to be difficult to sample the emissive state. The $\pi \rightarrow \pi^*$ state comes close to the states in water described in literature [141, 151]. However, only a fraction of the ensemble in this state could be sampled.



(a) HOMO **(b)** LUMO

Fig. 5.14.: HOMO and LUMO of prodan in DMSO and water as obtained in the simulation of S1. This transition describes the first excited state predicted by TD-LC-DFTB2/OB2$^{0.3}$ (base) in DMSO and water.

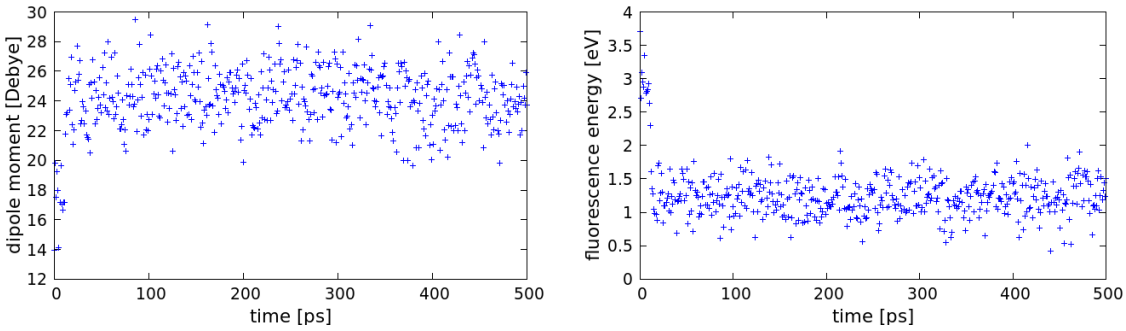


Fig. 5.15.: Excited state dipole moments and fluorescence energies computed along a trajectory of prodan in first excited state solvated in water. TD-LC-DFTB2/OB2$^{0.3}$ (base) was used and the environment considered in the calculation as point charges.

5.3.5. Discussion and Outlook

In contrast to previous studies of excited prodan in solution, in the QM/MM simulations performed for this work the dye was not kept fix in its optimized geometry but was free to move as well as the explicit solvent molecules around. This was possible due to the low computational cost of the semi-empirical TD-LC-DFTB2 method.

However, TD-LC-DFTB2 predicts a different ordering of the excited states. Moreover, the target excited states are so close in energy to other excited states that they could not be sampled separately.

In sum, TD-LC-DFTB2 predicts a dark $n\rightarrow\pi^*$ state as first excited state in vacuum. The second excited state instead has a dipole moment in agreement with the experiment and a significant oscillator strength. It was thus tried to simulate this state but during the simulation frequent state crossing occured. To avoid this, a state-following approach was tried. However, the tracking failed in several cases.

In the polar solvents DMSO and water, a twisted state was predicted as first excited state, which is unlikely correct. A simulation in the second excited state sampled a dark $n\rightarrow\pi^*$ and a $\pi\rightarrow\pi^*$ state. The latter could be the correct emissive state.

In Ref. [15] it is proposed that prodan emits also from the $n\rightarrow\pi^*$ state (Fig. 5.10 (a),(c)) with the low dipole moment. Thus, TD-LC-DFTB2/OB2 (base) is probably not completely wrong. Also, the state-crossing may be correct. The decisive question is whether the relation between the contributions of the two states can be correctly predicted. Since a simulation in the first excited state in vacuum only samples the $n\rightarrow\pi^*$ state, such a simulation is not helpful. Further, there is no reason to expect that a simulation in the second excited state would instead provide the real ensemble of states, that is the real proportion of both states. The oscillator strength of the transition from the $n\rightarrow\pi^*$ state is low and thus its impact on the final spectra is expected to be small. However, parts of the $\pi\rightarrow\pi^*$ state ensemble may not be present in reality or are missed because of the state switch. For this reason, a subsequent selection of the snapshots belonging to the $\pi\rightarrow\pi^*$ is not performed.

If it is assumed that the conical intersection is at a point far enough from the minima of the two emitting states and only close in the prediction of TD-LC-DFTB2, it would instead be meaningful to simulate the states separately. Further, the emission from both states is only a hypothesis. In order to be able to simulate only one state, improvement of the state-tracking approach for TD-LC-DFTB2 is desirable. Alternatively, a machine learning approach seems promising for such cases. A neural network can be trained with data from the target state only, such that it will predict forces or other properties of solely this state. Moreover, machine learned models can be very fast in their predictions which is an advantage compared to QM methods, even semi-empirical ones. Examples for neural network based simulations were recently reported in literature [175–178]. Though, the generation of the training data can be a critical point, since the training set has to be of sufficient size and quality. Taking the environment into account is an additional challenge because a succinct representation has to be found.

5.4. Solvatochromism of 4-Aminophthalimide

The study of 4-AP starts with an analysis of the excited states predicted by TD-LC-DFTB2 compared to CAM-B3LYP and CC2 references. Then the results of the simulations in the emissive state in vacuum, DMSO and water are presented. They are afterwards discussed and compared to literature.

5.4.1. Computational Details

The geometry of 4-aminophthalimide was optimized with (TD-)LC-DFTB/OB2 (base) in ground and the first and third excited state in vacuum using DFTB+ (version 20.1). As driver, conjugate gradient was chosen with the default convergence tolerance. The SCC tolerance accounted to 1×10^{-7} electrons and the thresholded screening method was ap-

plied with a threshold of 1×10^{-16} and a range-separation parameter of ω=0.3 a_0^{-1}. A subsequent frequency analysis was performed using the mode program of DFTB+, respectively.

As reference, the fluorophore was additionally optimized with CC2 in ground and first excited state with turbomole (version 7.3). The aug-cc2-pVTZ basis set was employed with its auxiliary basis for the RI approximation. The lowest 12 occupied MOs were frozen in the calculations. The density convergence threshold and the SCF energy convergence threshold were 1×10^{-7}, respectively.

The LC-DFT optimization with the CAM-B3LYP functional and def2-TVZP basis set was performed with the RIJCOSX approximations with def2/J auxiliary basis and the TightOpt keyword in ground and first excited state. The energy and residual convergence tolerance in the TD-DFT calculation amounted to 1×10^{-6} Ha. The ORCA program package (version 4.0.0.2) was used.

QM/MM simulations were performed in vacuum, DMSO and water as described for prodan. The first excited state and the third excited state were simulated, respectively.

The computational details for the calculation of excitation energies and the creation of orbital plots are also the same as for prodan.

5.4.2. State Characterization

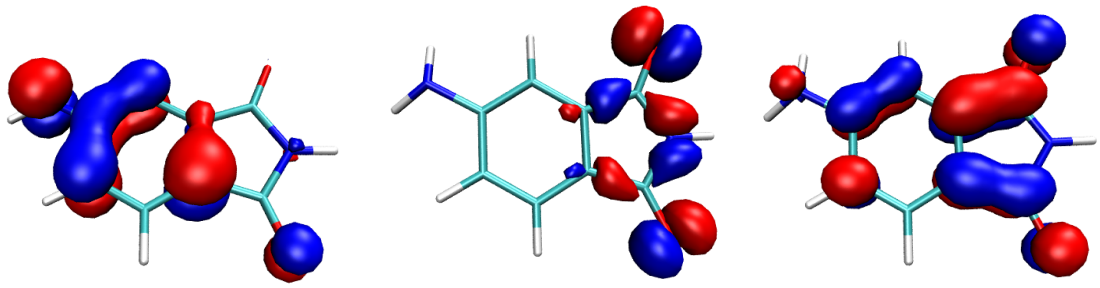
The excited states predicted by TD-LC-DFTB2 are compared to CC2 and CAM-B3LYP results. Excitation energies were computed on the ground and relaxed first excited state geometries, respectively.

The results from the ground state structures are shown in the appendix in Tables. A.18 and A.19. Similarly to prodan, the excitation energies computed with the LC-DFTB2 ground state geometry are systematically about 0.2 eV lower than their counterparts from the CAM-B3LYP ground state geometry for all applied methods. The first two excitation energies are also close in energy if computed with B3LYP or CAM-B3LYP. A comparison of the oscillator strengths indicates a different order of the states from TD-LC-DFTB2 compared to the states from DFT.

This is also observed with the first excited state geometries. A view on the oscillator strengths and orbital transitions describing the excited states reveals a different state ordering between the methods. The relevant orbitals are depicted in Fig. 5.16 as obtained with TD-LC-DFTB2. The order of the orbitals differs also between the methods. In Fig. 5.16, the orbitals are denoted according to their ordering from TD-LC-DFTB2 at the ground state geometry. The first excited state predicted by CAM-B3LYP using the CAM-B3LYP first excited state geometry is mainly described by the HOMO to LUMO transition, but the HOMO corresponds to HOMO-1 in Fig. 5.16. The oscillator strength of this transition amounts to ≈ 0.08. The first excited state predicted by CAM-B3LYP corresponds to the third excited state of TD-LC-DFTB2/OB2 (base) on the TD-LC-DFTB2 ground or first excited state geometry. This state has a significant oscillator strength and is expected to be the emissive state. It is treated as the target state in the following.

The dipole moments of the ground and target excited state, the oscillator strengths of the transition in the target state and the respective fluorescence energies are summarized in Tab. 5.9 as obtained with CC2, CAM-B3LYP and TD-LC-DFTB2/OB2 (base) on the geometries optimized with the same methods, respectively.

The target excited state is characterized further by an analysis of changes in the bond lengths in comparison to the optimized ground state geometries. The bonds investigated



(a) HOMO-1 **(b)** HOMO **(c)** LUMO

Fig. 5.16.: Orbitals of 4-AP obtained with TD-LC-DFTB2/OB2 (base) in vacuum. The first excited state is described by the HOMO to LUMO transition and the third excitated state by the HOMO-1 to LUMO transition.

Tab. 5.9.: Properties of 4-AP in vacuum: dipole moments in ground and target excited state, oscillator strengths and vertical excitation energies on the geometries optimized in the target excited state (fluorescence energies). The target excited state corresponds to the first excited state in the case of CC2 and CAM-B3LYP and the third excited state in the case of TD-LC-DFTB2. For (TD-)LC-DFTB2 the OB2 (base) parameters were used.

method	dipole ground	dipole exc.	osc. strength	fluoresc. E
CC2	4.7 D	9.7 D	0.06	3.31 eV
CAM-B3LYP	4.9 D	9.6 D	0.08	3.66 eV
LC-DFTB2	4.5 D	9.5 D	0.07	3.82 eV

here are the same as in a similar analysis in the Supporting Information of Ref. [145] using CAM-B3LYP. Fig. 5.17 compares the bond length differences for TD-LC-DFTB2/OB2 (base), LC-DFT/CAM-B3LYP and CC2. The methods agree qualitatively except for the bond labelled b2, which however does not change much anyway. The most pronounced change is a strong decrease of the C-CO bond length in the excited state.

5.4.3. Dynamics in Vacuum

Simulations in vacuum were performed in the first excited state (S1) and in the third excited state (S3) which was identified as the target state on the optimized geometry. None of the two simulations run into a conical intersection so that both states were separately sampled in each simulation.

The simulation in S1 sampled a state with a low dipole moment fluctuating around 2.5 D. It can be described by the HOMO to LUMO transition with the orbitals depicted in Fig. 5.16. The oscillator strengths are below 0.001.

In the simulation of S3, the orbital transition corresponds either to a HOMO to LUMO or a HOMO-1 to LUMO transition. The LUMO corresponds to the one shown in Fig. 5.16 (c). The HOMO or HOMO-1 looks a bit like a mixture of the orbitals in Fig. 5.16 (a) and (b), respectively. The orbitals involved in the transition are visualized in Fig. 5.18. The

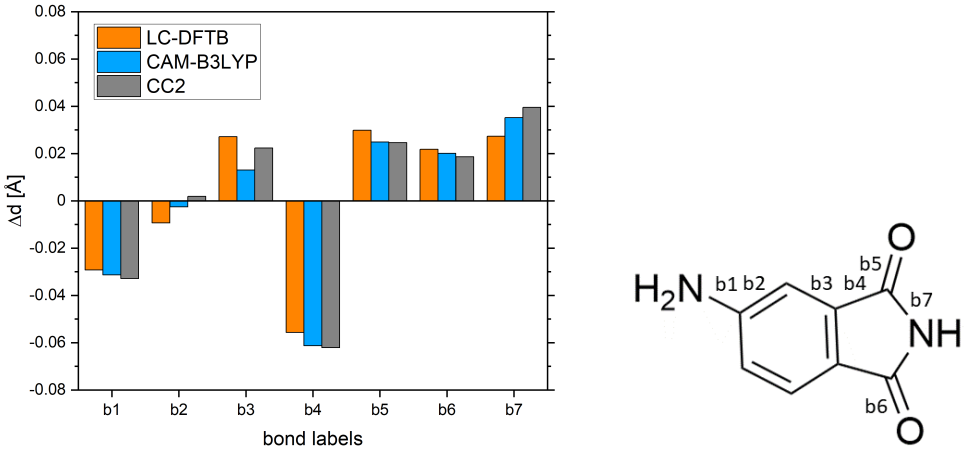
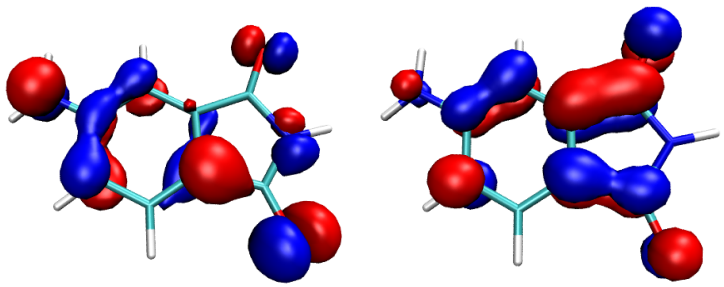


Fig. 5.17.: Bond length changes upon excitation. The C-CO bond is strongly decreased in the excited state.



(a) HOMO/HOMO-1 **(b)** LUMO

Fig. 5.18.: Orbitals of 4-AP obtained with TD-LC-DFTB2/OB2 (base) in vacuum describing the simulated third excited state. The picture was made exemplary on the snapshot after 300 ps of simulation. The HOMO and HOMO-1 frequently change their order.

dipole moment fluctuates around 9 D and the oscillator strengths are mostly between 0.05 and 0.09, that is, significantly larger than the oscillator strengths of the transitions in S1 or S2 (Fig. 5.19).

5.4.4. Solvated 4-AP

4-AP shows a different behaviour solvated in polar DMSO compared to vacuum. In the simulation in S1, the first excited state dipole moment is on average 3.6 D which is a little larger than in vacuum. This is an expected effect due to the polar environment. The oscillator strengths are still very small. The simulation in S3, however, ran several times over a conical intersection. This becomes visible for instance on the jumping dipole moments and oscillator strengths (Fig. 5.20). The transition describing S3 is either from HOMO-2 to LUMO or from HOMO to LUMO. The HOMO to LUMO transition corresponds to the larger oscillator strengths and dipole moments. In the areas with dipoles of 7-8.3 D and oscillator strengths of 0.035-0.6, both transitions occur. The HOMO corresponds to the

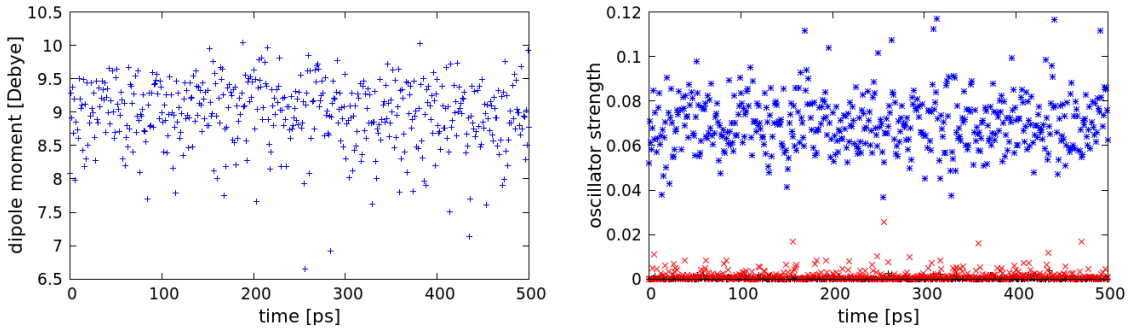


Fig. 5.19.: Dipole moments (left) and oscillator strengths (right) computed with TD-LC-DFTB2 along the S3 trajectory of 4-AP in vacuum. Left: S1-black, S2-red, S3-blue.

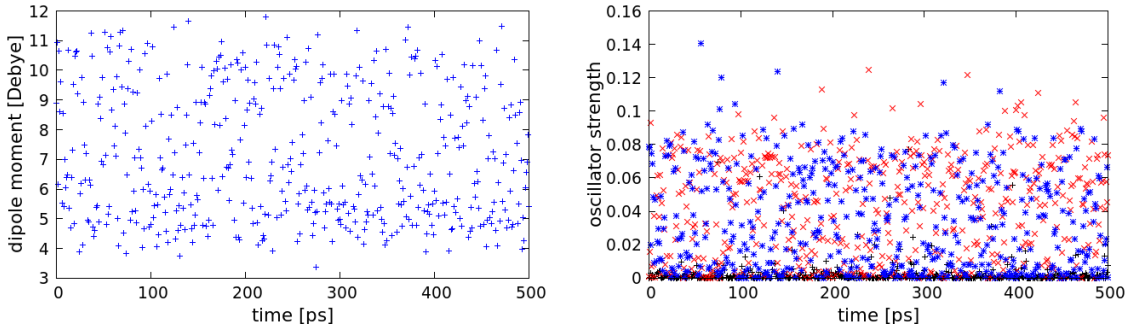


Fig. 5.20.: Dipole moments (left) and oscillator strengths (right) computed with TD-LC-DFTB2 along the S3 trajectory of 4-AP in DMSO. Left: S1-black, S2-red, S3-blue.

orbital shown in Fig. 5.16 (a) and the LUMO corresponds to 5.16 (c).

4-AP is able to form hydrogen bonds with the oxygen of DMSO via the NH2 moiety or the NH included in the five-membered ring. 87 % of the frames in the correct state show a hydrogen bond with the latter and 93 % of these frames show at least one hydrogen bond with the former. In 53 % of the frames, two hydrogen bonds were found between DMSO and the NH2 moiety. The numbers are very similar for the other state. This is not surprising since the states switch faster than the hydrogen bonds form and break.

In water, TD-LC-DFTB2 samples the transition between the HOMO and LUMO corresponding to the orbitals depicted in Fig. 5.16 (a) and (c) as first excited state. This transition has also a clearly larger oscillator strength than the second and third transition. The dipole moment increased to approximately 16 D. The values along the trajectory are shown in Fig. 5.21.

4-AP forms hydrogen bonds with water. The amino moiety and the amine in the ring of 4-AP serve as hydrogen bond donors and the oxygen atoms are additional acceptors. In the first excited state QM/MM simulation, in total eight hydrogen bonds are counted on average at the same time. The maximal number observed are eleven. The most hydrogen bonds are formed with the oxygens, which are six on average. That means each oxygen atoms interacts with three water molecules. The amino group is involved in one to two hydrogen bonds (average: 1.5), very similarly to DMSO. The amine donor forms one

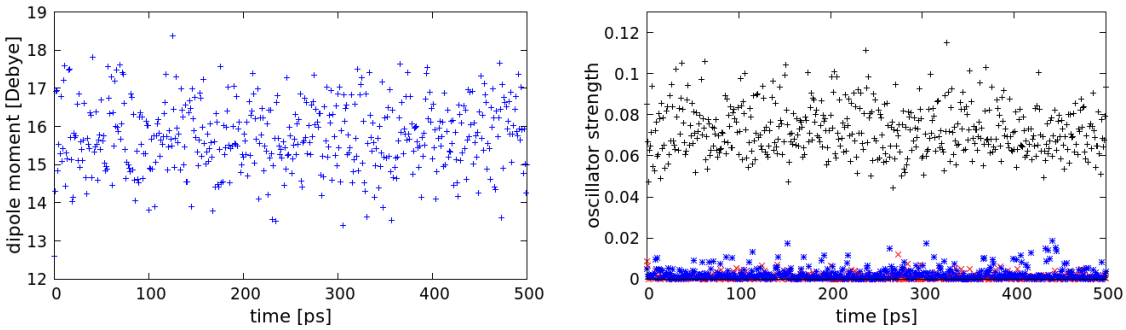


Fig. 5.21.: Dipole moments (left) and oscillator strengths (right) computed with TD-LC-DFTB2 along the S1 trajectory of 4-AP in water. Left: S1-black, S2-red, S3-blue.

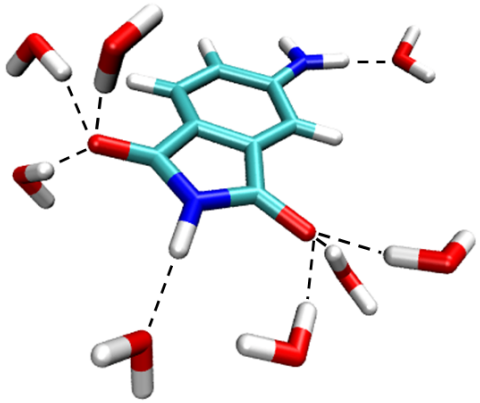
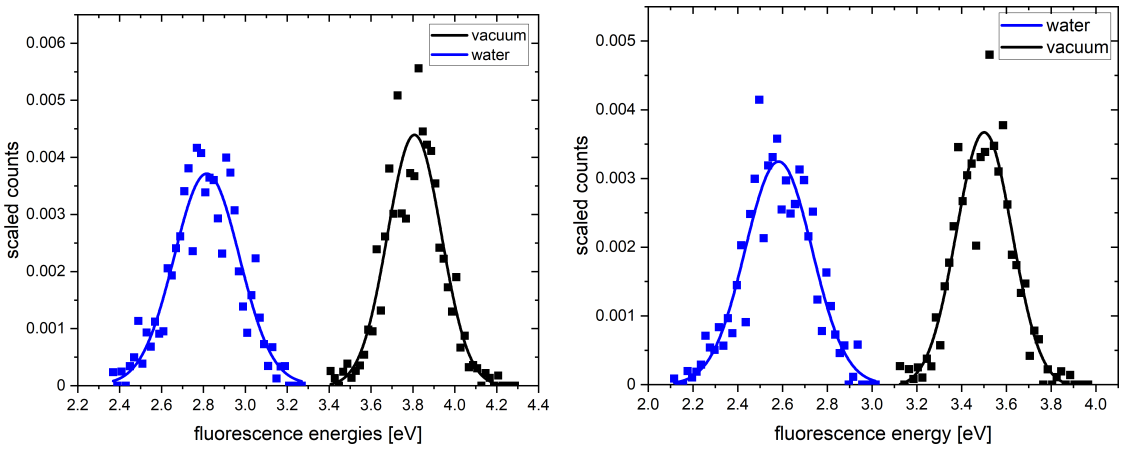


Fig. 5.22.: Arrangement of water molecules forming hydrogen bonds with the donor and acceptor groups of excited 4-AP.

hydrogen bond in 64 % of the frames. This is less often than in DMSO. Fig. 5.22 shows the arrangement of water molecules forming hydrogen bonds with 4-AP in a representative snapshot.

Finally, the fluorescence spectra in vacuum and water are modelled from histograms of the excitation energies computed along the snapshots of the trajectory. The geometries from the simulation in S3 and the third excitation energies were used to get the vacuum results and the simulation in S1 and the first excitation energies for the results in water. The histograms and a Gaussian fit, respectively, are shown in Fig. 5.23. The excitation energies were computed with the OB2 (base) parameters (Fig. 5.23 (a)) which were also used in the simulations and additionally with the pure electronical parameter set optimized for excitation energies (Fig. 5.23 (b)). With the latter, the state ordering is different in vacuum compared to the OB2 (base) parameters. Hence, the excitation energies used in the histogram were selected such that they correspond to the sampled state. These were the excitation energies with the largest oscillator strengths of the first three, respectively, described by the HOMO to LUMO transition. They corresponded mostly to the second excitation energy computed on the respective snapshot geometry. The solvatochromic shifts between vacuum and water amount to -0.991 eV (OB2 parameters) and -0.918 eV (el. parameters).



(a) OB2 (base) parameters **(b)** pure electronical parameters

Fig. 5.23.: Gaussian fits to histograms of the fluorescence energies computed with TD-LC-DFTB2. The intensity is determined from the counts per window and their oscillator strengths.

5.4.5. Discussion and Outlook

The third excited state from TD-LC-DFTB2 was identified as the emitting state by an analysis of the excited states in vacuum, corresponding to the first excited state of CC2 and LC-DFT/CAM-B3LYP. The optimized structure in S3 agrees quite well with the higher level methods regarding dipole moments, oscillator strengths and bond length changes upon excitation. The calculated ground and excited state dipole moments are in the range of reported experimentally and computationally derived values [145, 157, 179]. The bond length changes are also in qualitative agreement with Ref. [145]. However, smaller differences are observed even between their CAM-B3LYP results and the CAM-B3LYP results of this work. For example, the reported change of the bond length b1 is only -0.02 Å (here: -0.03 Å) and the change of the bond length b2 is there around -0.01 Å (here: -0.003 Å), Fig. 5.17. Reasons can be the different basis set and different quantum chemistry program used in Ref. [145]. This shows that the differences between TD-LC-DFTB2 and CAM-B3LYP are probably uncritical since they are of the same order.

In water, the emissive state was predicted as first excited state by TD-LC-DFTB2. This can be explained by a stronger stabilisation of this charge-transfer state in the polar water in comparison to the other states which in vacuum were predicted as first and second excited states. The dipole moment of the emissive state in water increased by 7 D compared to the value in vacuum indicating a stronger charge separation in water. The excited state dipole moment in water ($\approx$ 16 D) is a little larger than the one found in Ref. [145] with a PCM model and CAM-B3LYP (15 D) and a little lower as determined experimentally in Ref. [179] in protic solvents (18 D). Since the former does not consider hydrogen bonds, the dipole moment found here is reasonable but may still be slightly underestimated. The hydrogen bond pattern observed in the excited state QM/MM simulation differs to the one of Ref. [145], which proposed only two water molecules per oxygen instead of three. On the other hand, they consider two hydrogen bonds on the amino moiety while in the QM/MM simulation often only one hydrogen bond is present there. The effect of the dif-

ferent hydrogen bond patterns on the spectra and the stability of the respective hydrogen bonds are interesting points for further study.

Excited 4-AP could be simulated in vacuum and water, but not in DMSO because the excited states predicted by TD-LC-DFTB2 were too close in energy and frequently crossing. The solvent shift thus could only be determined between vacuum and water. The experimentally measured emission of 4-AP in different solvents is reported in literature [152,157,179,180]. The emisison in the very nonpolar hexane is taken as reference for the computed results in vacuum. There, the emission maximum is located at 25 965 cm^{-1} (385 nm, 3.219 eV) [180]. It is shifted in water to 18 520 cm^{-1} (540 nm, 2.296 eV) [152,180]. Thus, the experimental solvent shift from hexane to water amounts to -0.923 eV. This is very close to the results obtained with TD-LC-DFTB2 and the OB2 (base) parameters for simulation and the optimized electronical parameters for excitation energy calculation (-0.918 eV). If only the OB2 (base) parameters are used then the shift (-0.991 eV) is overestimated by 0.068 eV, which are $\approx$ 7 % of the experimental shift. The absolute values of the computed fluorescence energies are red shifted with both parameter sets, but more so with the OB2 (base) parameters.

In further studies, the state ordering in additional solvents can be investigated. As for prodan, a way needs to be found to simulate the dye in the proximity of the conical intersections. For instance, state tracking or the replacement of TD-LC-DFTB2 in the simulation by a machine learning model could be options. However, in the latter approach the consideration of the environment could become a challenge, as discussed already for prodan.

4-AP is a promising dye for optical glucose sensing as described in chapter 6 on the example of badan. A derivative of 4-AP with a flexible linker to the protein was synthesized by Samantha Wörner [181]. The properties of this modified 4-AP can be studied as next step, first in vacuum and pure water and then in the protein environment. A similar behaviour as observed for 4-AP is expected for this derivative. Hence, a sufficiently polar environment of the dye attached to the protein will be decisive for a successful excited state QM/MM simulation.

5.5. Concluding Remarks

In this chapter, TD-LC-DFTB2 was applied in excited state QM/MM simulations for the study of three different environmentally sensitive fluorophores: Flugi-2, prodan and 4-AP. Reasonable results were obtained for Flugi-2 without any problems. On the contrary, it was not possible to simulate the emissive state of prodan with TD-LC-DFTB2. In vacuum, state crossing occured and in the solvents the correct state was difficult to identify. The results for 4-AP lie in between of these extreme cases. It was possible to simulate the emissive state in vacuum and water, but not in DMSO due to state crossing. With increasing polarity of the environment, the emissive state moves from the third excited state in vacuum to the first excited state in water. Hence, TD-LC-DFTB2 is a helpful method in excited state QM/MM simulations if it is able to describe the excited state of the fluorophore correctly. This needs to be carefully checked for the individual fluorophore of interest.

6 Understanding Optical Glucose Sensing: Glucose/Galactose Binding Protein Combined with Badan

In parts reproduced from
Z. Pang, M. Sokolov, T. Kubař and M. Elstner. Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model.
Phys. Chem. Chem. Phys., 24: 2441-2453, 2022
with permission from the PCCP Owner Societies
DOI:10.1039/D1CP03733A

Diabetes is a disease worldwide millions of people suffer from. The blood glucose level of a patient is not properly regulated by his body and need to be controlled frequently. A well-known method to do this is the finger-prick method. However, it is on one hand uncomfortable and painful and on the other hand cannot display the glucose concentration continuosly in real-time. Sensors based on glucose oxidase are available which are able to monitor the blood glucose concentration continuously. However, they show deficiencies in accuracy and require a frequent calibration. [182]

A promising alternative are fluorescence based glucose sensors [125]. The Glucose/Galactose Binding Protein (GGBP) found much interest in the development of such sensors. This protein takes a specific closed conformation when binding a glucose molecule, while in its *apo* form it shows open conformations. The basic idea for the optical sensor is to utilize this conformational change to change the environment of an attached fluorophore such that this dye gives an optical signal upon glucose binding [6, 9, 128]. This is schematically shown in Fig. 6.1. Since the galactose concentration in blood is much smaller than the glucose concentration, no falsification from galactose binding is expected [6].

The decisive factors for the usefulness of such a sensor are the amount of fluorescence change caused by glucose binding and the strength of the binding which is quantified by the dissociation constant k_D, see Eq. 6.1. The former depends on the choice of the fluorophore as well as on its specific position in the protein. The latter determines the operating range of the sensor. [5]

$$k_D = \frac{[\text{free glucose}][\text{apo protein}]}{[\text{complex}]} \tag{6.1}$$

If the glucose concentration in the sample and the k_D have similar values, the signal response is optimal because a sufficient amount of protein is available to bind new glucose molecules when the sugar concentration increases, and a sufficient amount of protein still produces a signal in the case when the sugar concentration decreases.

In the following subchapters, the biological system is described first and and the studies targeting the conformational changes and binding mechanism of GGBP are summarized.

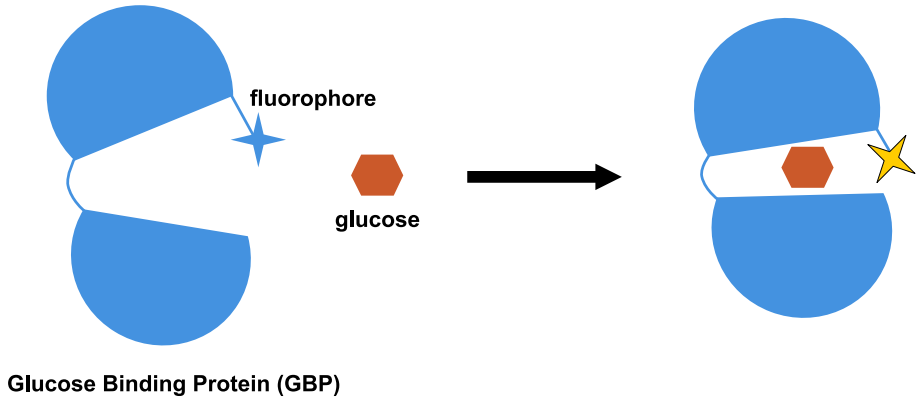


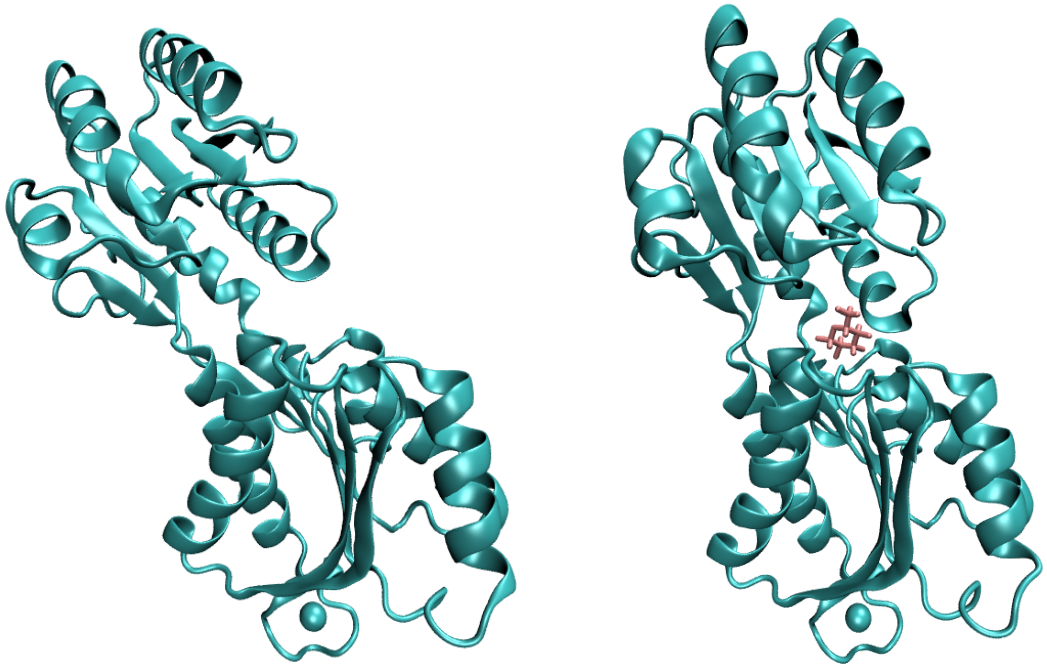
Fig. 6.1.: Working principle of an optical glucose sensor based on the Glucose Binding protein.

Afterwards, an overview is given on the large amount of experimental studies dealing with the development of such kinds of optical glucose sensors. This then leads to the motivation of the work presented here, which is the theoretical description of such an optical sensor and to understand its underlying mechanism.

6.1. The Biological System

GGBP is a water soluble, periplasmic protein of gram-negative bacteria. It consists of two similar domains in Rossman fold which are connected by three hinge strands, which is typical for a periplasmic binding protein (PBP). In the bacteria it is involved in chemotaxis and mediates the D-glucose and D-galactose transport by the Mg1ABC transporter. A calcium ion stabilizes the loop region in the C-terminal domain. [183–185] Like other PBPs, it can take open and closed conformations and the binding is thought to work like a Venus flytrap [183] (see Fig. 6.1). The protein without ligand (*apo* form) is expected to be mostly open and with ligand (*holo* form) the protein is closed. Several crystal structures were resolved: *Apo* and *holo* closed GGBP conformations from *Salmonella typhimurium* [186–188] and *apo* open and *holo* closed GGBP conformations from *E. coli* [183, 185, 189]. Fig. 6.2 shows the crystal structure of *apo* open GGBP (a) and a crystal structure of *holo* closed GGBP with glucose in the binding pocket (b) from *E. coli*.

The β-isomers of D-glucose and D-galactose are selectively and tightly bound by GGBP with a k_D of 0.2 and 0.4 μM, respectively [80]. In the following it is focussed only on the glucose. The sugar interacts with several amino acids via hydrogen bonds. Among them are three aspartates and one arginine which form particularly strong hydrogen bonds due to their charged side chains. Additionally, the sugar is sandwiched between two aromatic residues, phenylalanine and tryptophan. A picture of the binding pocket and the hydrogen bonds between the protein and the sugar is shown in Fig. 6.3. The hydrogen bond network even extends to a second shell of amino acids stabilizing the amino acids which interact with glucose, as shown in Ref. [185].



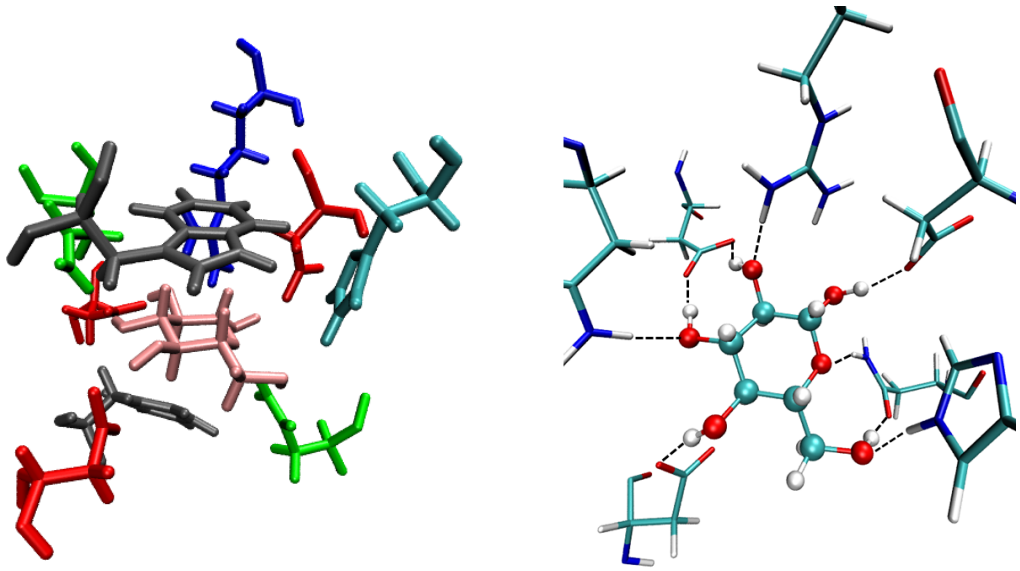
(a) *Apo* GGBP (PDB-ID: 2FW0). **(b)** *Holo* GGBP (PDB-ID: 2FVY).

Fig. 6.2.: Crystal structures of *apo* and *holo* GGBP [183]. The *apo* conformer is more opened than the *holo* conformer. A calcium ion stabilizes the loop region in the C-terminal domain (bottom). The sugar in the binding pocket is pink coloured.

Conformational Changes and Binding Mechanism

In an experimental study [190], *apo* GGBP was found to take several different conformations, open and closed ones. According to them, the glucose binding process follows a population shift mechanism and stabilizes with sugar in the closed conformation. This is different to the induced fit mechanism found in the same study for the structurally similar ribose binding protein (RBP). Mutation of two hinge residues in GGBP and RBP, respectively, was sufficient to interchange their binding mechanisms.

The conformations and the energetical landscape of GGBP were investigated further in a combined NMR and MD study [191]. The *apo* protein was found to take 68 % open conformations and 32 % closed conformations. In accordance to Ref. [190], the ligand binding is stated to shift the equilibrium to the closed conformation similar as in a conformational selection mechanism. The energy profile of the protein conformations was constructed by means of two steered MD simulations. The transition state between *apo* open and *holo* closed was estimated to be $\approx$ 3.5 kcal mol^{-1} higher than the open conformation. The energy difference between *apo* open and *apo* closed was estimated to $\approx$ 6 kcal mol^{-1}. The conformational changes were described by an angle quantifying the amount of vertical opening/closing and a dihedral representing the amount of twisting of the two domains in the opening/closing motion. Interestingly, *apo* conformations were detected which are much more opened than the *apo* open crystal structure.



(a) Amino acids coordinating glucose **(b)** H-bonds with glucose

Fig. 6.3.: Binding pocket of GGBP. (a) The sugar (pink) is sandwiched between two aromatic residues (gray) and hold by hydrogen bonds with aspartates (red), asparagine (green), arginine (blue) and histidine (turquoise). (b) Nine hydrogen bonds hold the glucose in the binding pocket.

6.2. Experimentally Tested Glucose Sensors

Engineered glucose sensors based on GGBP combined with a fluorophore were firstly presented by Marvin and Hellinga in 1998 [128]. Several fluorophore positions in the protein were scanned based on an inspection of the crystal structure. A large number of studies from various groups followed, testing different fluorophores as well es different positions in the protein [6–9]. In 2008, Ge *et al.* [129] showed the potential of such sensors for real-time monitoring of the sugar in microdialysis. In the mean time, a few promising combinations of fluorophores and their specific position in the protein were further applied to develop devices for practical usage [125, 127, 130, 132].

Among the tested dyes are badan and the related acrylodan, 1,5-IAEDANS, MDCC, NBD with two kinds of linkers, fluorescein, pyrene and the FRET systems Alexa Fluor 488/555 with QSY 7 and acrylodan with rhodamine. The tried positions at the protein are also very different. Since GGBP does not contain any cysteine residue, it is possible to construct a well-defined sensor by mutation of a selected residue to cysteine. A fluorophore with a thiol-reactive moiety can then be attached specifically at the desired position. Locations close to the binding site but not involved in sugar binding are called *peristeric*. There, the fluorophores are expected to experience large environmental changes when glucose binds. *Allosteric* positions are far away from the binding site. This requires a concerted local change at that position upon binding but on the other hand does not alter the natural binding through steric hindrance. *Endosteric* attachment sites in contrast allow a direct nonbonded interaction between the bound sugar and the fluorophore. [7, 128]

Tab. 6.1 gives an overview of tested dyes, positions and their classification and the qualitative outcome. Intensity changes of more than 50 % between fully saturated proteins

and absence of glucose are counted, while smaller changes are neglected. The changes can be to higher or to lower fluorescence intensity upon glucose addition. It is evident that the optical response is very sensitive to the linking position of a dye. Badan at position H152C shows a large signal change but none at position M182C. Furthermore, there is no guarantee that a position working for one dye will be suitable for another dye. For instance, the fluorescence intensity of badan attached at H152C increases by 300 % [9] but the very similar dye acrylodan gives a poor response to glucose binding at the same position (no change [6,128], ≈ 70% increase [7]).

In addition or alternatively to an intensity change also a change in the emission wave length or fluorescence lifetime can be used as optical signal. This can be advantageous in practice since a measured intensity can be falsified by scattering, reduced by the absorbance of a sample and it depends on the fluorophore concentration [13,125].

Tab. 6.1.: Tested glucose sensors based on GGBP. The mutations show the dye positions. *e*, *p* and *a* classify these positions into endosteric, peristeric and allosteric. Only intensity changes of at least 50 % between *apo* proteins and fully saturated proteins are counted.

dye	position		intensity change?	reference
Badan	H152C	e	yes	[9]
	M182C	e	no	[9]
Acrylodan	Y10C	e	no	[7]
	H152C	e	no/yes	[6,7,128]
	M182C	e	no	[6,8]
	W183C	e	yes	[7]
	N15C	p[1]	no	[7,128]
	E93C	p	yes	[7]
	G148C	p	no	[6]
	E149C	p	yes	[7]
	L255C	a	yes	[7,128]
	D257C	a	no	[7,128]
	P294C	a	no	[128]
	V296C	a	no	[128]
MDCC	H152C	e	no	[6]
	M182C	e	no	[6]
	G148C	p	no	[6]
IAEDANS	H152C	e	no	[6]
	M182C	e	no	[6]
	G148C	p	no	[6]

Tab. 6.1.: Tested glucose sensors based on GGBP, continued.

dye	position		intensity change?	reference
IANBD ester	H152C	e	no/yes	[6, 128]
	M182C	e	no	[6]
	N15C	p[1]	no	[128]
	G148C	p	no	[6]
	L255C	a	no	[128]
	D257C	a	yes	[128]
	P294C	a	no	[128]
	V296C	a	no	[128]
IANBD amide	Y10C	e	no	[7]
	H152C	e	yes	[7]
	W183C	e	no	[7]
	N15C	p[1]	no	[7]
	E93C	p	yes	[7]
	E149C	p	yes	[7]
	L255C	a	no	[7]
	D257C	a	yes	[7]
	V296C	a	yes	[7]
Fluorescein	Y10C	e	yes	[7]
	H152C	e	yes	[7]
	W183C	e	no	[7]
	N15C	p[1]	no	[7]
	E93C	p	no	[7]
	E149C	p	yes	[7]
	L255C	a	no	[7]
	D257C	a	no	[7]
	V296C	a	no	[7]
Pyrene	Y10C	e	no	[7]
	H152C	e	yes	[7]
	W183C	e	no	[7]
	N15C	p[1]	yes	[7]
	E93C	p	no	[7]
	E149C	p	yes	[7]
	L255C	a	yes	[7]
	D257C	a	no	[7]
	V296C	a	no	[7]

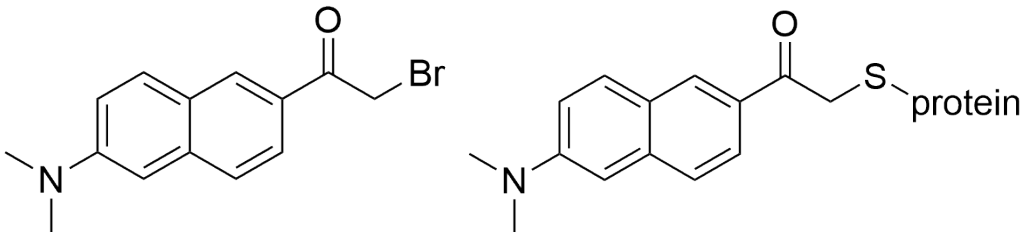
[1] in Ref. [7] classified as endosteric

It is focussed on the fluorophore badan (6-bromoacetyl-2-dimethylaminonaphthalene) in this work. The chemical structure of the dye is depicted in Fig. 6.4. Position H152C was found to provide a 300 % increase in fluorescence intensity as response to glucose binding, but the k_D of this single mutant is in the order of 0.002 mM which is too small for the measurement of blood glucose [9]. Blood glucose concentrations in humans with diabetes can raise up to 30 mM [182]. The GGBP sensor was further developed to the GGBP triple mutant H152C/A213R/L238S with badan attached at position H152C. The fluorescence intensity doubles upon glucose addition and with the two additional mutations an appropriate dissociation constant of 11 mM in phosphate-buffered saline could be achieved [5]. This is in the order of human blood glucose concentrations. The high-k_D triple mutant has been immobilised on a solid surface and tested in vitro using the lifetime as signal [125]. A picture of the triple mutant is shown in Fig. 6.5.

Very recently, an experimental study was published which explicitely aimed for an understanding of the fluorescence properties of badan attached to GGBP at position H152C (single mutant) [19]. To get a reference for the fluorescence properties of the attached badan molecule enveloped by solvent, a new mutant was created with badan at position W284C, which is on an outer edge of the binding pocket. Further, it was shed light on the influence of tryptophane 183, which is a known fluorescence quencher of badan and in the proximity of position H152C. Therefore, two double mutants were investigated: H152C-Bad/W183F and H152C-Bad/W183A. In summary, two isomers were found to contribute to the fluorescence spectra. These were proposed as molecules in PICT and TICT state. It is stated, that the solvent impacts the relaxation from the PICT to the TICT state and quenches fluorescence via hydrogen bonds. Furthermore, tryptophane was identified as quencher, which is blocked by glucose in the *holo* protein. [19]

6.3. Motivation

As illustrated by the large number of tested dyes and positions, the fluorophores and their positions in the GGBP were found almost empirically by testing a variaty of combinations. Trial attachment sites for environmentally sensitive dyes were selected based on crystal structure inspection or comparison to similar binding proteins [7, 128]. It is generally expected to achieve the best signal changes when the dye is located at a protein position which experiences a large conformational change upon glucose binding. However, the conformations of the fluorophores at their positions, their specific environment in the different



(a) Free badan **(b)** Badan attached to cysteine residue of the protein

Fig. 6.4.: Chemical structure of badan.

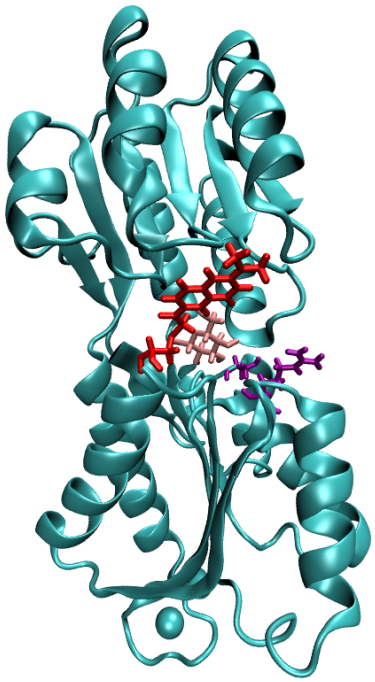


Fig. 6.5.: Triple mutant H152C-badan/A213R/L238S. Badan is shown in red, the other two mutations at the edge of the binding pocket are shown in purple and the bound glucose molecule is shown in pink.

protein conformations and their interactions with the protein, solvent and maybe also the sugar are not known. These factors determine the optical properties of the fluorophores. Hence, knowledge about them would be helpful not only to understand the already developed sensors but also in the design of new sensors of this kind.

Here the GGBP triple mutant H152C/A213R/L238S with badan at position H152C [5] is studied in detail. This mutant is not only interesting because of the significant optical signal but also due to the mutations increasing the k_D to 11 mM, which is appropriate to detect blood glucose. In the publication presenting the GGBP-Badan triple mutant [5], the change of fluorescence intensity is suggested to be caused by a more hydrophobic environment of the dye in the *holo* protein. It is imagined that Badan is enveloped by the protein in the closed conformation with bound sugar. This hypothesis is tested in this study.

Classical MD simulations are a suitable method to investigate conformational changes of such smaller biomolecular systems on a time scale of a few hundreds of nanoseconds. This is sufficient to sample the important fluorophore conformations and to get an idea about the interactions of the fluorophore and its nearest environment. Since the sensor is based on fluorescence, a classical MD simulation can only provide an idea on the behaviour of the dye and its conformational freedom. A simulation of the excited dye is needed to sample the excited state geometries and to get an accurate picture of the interactions, for instance the strength of hydrogen bonds. Excited state QM/MM simulations as discussed in the last chapter were tried for this purpose.

While there are a few studies on the conformations of the wildtype GGBP (*vide supra*), there is no knowledge about the triple mutant with badan. However, the conformational space of the protein cannot be sampled sufficiently with classical MD simulations. For this reason, metadynamics simulations were emloyed by Ziwei Pang to shed light on the free energy surfaces of the wildtype and the triple mutant. The protein conformations sampled in the classical MD simulations will be compared to these results.

Although not mentioned in previous theoretical studies on GGBP, it turned out in first test simulations that the charges from the glycam force field, which are typically applied for glucose, are not suitable so simulate the sugar in the binding pocket of GGBP. It was observed that the wildtype protein with glucose in the binding pocket did not remain closed during the simulation time of 1 μs, which is unrealistic remembering that the dissociation constant is only 0.2 μM. Since the binding pocket is highly polar including even charged amino acids, it is very likely that polarisation needs to be taken into account. This motivates the derivation of a new charge set for a proper description of the bound glucose.

6.4. Computational Details

For this study, the wildtype protein and the triple mutant H152C-badan/A213R/L238S were simulated with different charge sets for the glucose molecule. First, the details about the derivation of different charge models are reported and then the preparation and conditions of the classical MD simulations and QM/MM simulations.

6.4.1. Charge Models for Glucose

New partial charges for β-D-glucose were derived to include polarisation effects. The following procedure was applied separately to the wildtype GGBP and the GGBP-badan triple mutant and is intended to describe the glucose molecule in the binding pocket. First, the system consisting of the solvated closed protein with glucose inside the binding pocket was energy minimized with Gromacs using the glycam 06j partial charges for the sugar. The resulting geometry was visualized with VMD and the coordinates of the atoms within 5 Å of the glucose molecule were saved. The atomic partial charges of these atoms were assigned to the saved positions and included in the calculation of the electrostatic potential (ESP) of the glucose molecule. This calculation was done at the HF level with the triple-zeta 6-311G* basis set using Gaussian [192]. Finally, the new partial charges were obtained from a restrained electrostatic potential (RESP) fit using Antechamber [193]. In the following, the label BPC (binding pocket polarized charges) refers to this charge set.

Since this method applies a hard cut-off, the impact of the applied 5 Å radius was checked by comparing to the results from radii of 4, 6, 7, 8 Å. Additionally, the charges were derived from a point charge file including complete residues. In the first variant, all residues were considered with at least one atom within a radius of 4 Å to glucose. In the second variant this radius was set to 5 Å.

In another attempt, the charges were derived from a QM/MM MD simulation. The glucose molecule located in the binding pocket of the wild-type protein was set as the QM

region, which was described with the semi-empirical DFTB/3OB method. The protein and the solvent were considered as MM part and were described by a force field. Periodic boundary conditions (PBC) were applied and long-range electrostatics was treated with the particle–mesh Ewald method (PME) [194]. This simulation was performed with Gromacs interfaced to the DFTB+ program [88, 195]. First, a QM/MM steepest descents minimization was performed until all forces were lower than 1000 kJ mol^{-1} nm^{-1}. Then, the QM/MM MD simulation with a time step of 1 fs was carried out at a temperature of 300 K and a pressure of 1.013 bar maintained by the Nosé–Hoover thermostat and the Parrinello–Rahman barostat, respectively. The simulation was run for 1.1 ns, of which the first 0.1 ns were cut as equilibration time. The DFTB Mulliken charges were recorded every 10 fs over the final 1 ns, resulting in 100,000 charge values per atom, which were finally averaged.

Additionally, the charges of glucose dissolved in pure water were determined from another QM/MM simulation. A glucose molecule was placed in a cubic box of size $2.7 \times 2.7 \times 2.7$ $Å^3$. The glucose corresponded to the QM region, which was treated with DFTB/3OB, and the water in the MM region was represented by the TIP3P model. PBC and PME were applied. The system was minimized with steepest descents with a force threshold of 1000 kJ mol^{-1} nm^{-1}. Then, MD equilibration was performed with a time step of 1 fs and all bond lengths constrained. The temperature was adjusted to 300 K in an NVT simulation of 50 ps performed with the Nosé–Hoover thermostat, followed by an NPT simulation of 50 ps at 300 K and 1.013 bar using the Parrinello–Rahman barostat additionally. The production simulation was performed without contraints for 1.1 ns, of which the first 0.1 ns were discarded. Every 10 fs a snapshot was taken containing the DFTB Mulliken charges of the atoms. The charges were recorded every 10 fs and were averaged to yield the final atomic charges.

An alternative charge set was derived by Ziwei Pang, which is intended to represent the situation of a glucose molecule in water. The geometry of the glucose molecule was optimized at the B3LYP/6-31G* level in the presence of implicit water represented by the polarizable continuum model (PCM) [196] in Gaussian. Then, the ESP was computed at the HF/6-31G* level still considering implicit water, and these water polarized charges (WPC) were subsequently obtained from a RESP fit with Antechamber.

6.4.2. Molecular Dynamics Simulations

The wildtype protein and the triple mutant H152C-badan/A213R/L238S were simulated with glucose initially in the binding pocket and without any glucose in the system, respectively. The simulated systems were already prepared by Violetta Schneider starting from the crystal structures of an *holo* closed GGBP (PDB-ID: 2FVY) and an *apo* open GGBP (PDB-ID:2FW0) [183]. Mutants were obtained by replacing the respective residues in these crystal structures. The proteins were solvated in a dodecahedral TIP3P water box with a minimum distance of 20 Å to the edges of the box. Six sodium ions were added to neutralize the system, respectively. The protein was described by the Amber14SB [197] force field, the attached fluorophore badan by the GAFF force field [198] and the counterions by Joung-Cheatham parameters [199]. In the first simulation set, the glucose was described by the glycam force field.

For equilibration, first a steepest descents minimization was performed reducing all forces below 100 kJ mol^{-1} Å^{-1}, followed by a conjugate gradient minimization reducing all forces further below 50 kJ mol^{-1} Å^{-1}. In the next step, the system was heated to 298 K during an NVT MD simulation of 1 ns with a time step of 2 fs using the Bussi thermostat [200]. The lengths of all bonds were kept constrained to their respective equilibrium values. Subsequently, an NPT simulation of 1 ns with a time step of 2 fs was performed at a temperature of 298 K maintained by the Nosé–Hoover thermostat [162, 163] and a pressure of 1 bar controlled by the Parrinello–Rahman barostat [164, 165].

Finally, the system was equilibrated for 10 ns keeping only bonds involving hydrogen atoms constrained. Identical settings were used to carry out the actual production simulations of 1 μs.

The equilibration and productive simulations were all performed using Gromacs [201–207]. PBC were applied and long-range electrostatics was described by the PME method.

The simulations of the wild-type and triple mutant with badan were repeated under same conditions (but without the extra 10 ns equilibration) using the BPC and WPC charge models for the glucose molecule and bonded parameters from the GAFF force field by Ziwei Pang and me with equal contribution. Additionally, the *apo* and *holo* (BPC) triple mutant with badan were simulated under the same conditions but with position restraints of 1000 kJ mol^{-1} nm^2 on the protein atoms during the NVT and NPT equilibration.

6.4.3. QM/MM Simulations

Ground and first excited state QM/MM simulations were performed with badan in the QM part and the solvated protein in the MM part. The same implementation (Gromacs/DFTB+) was used as in the previous chapter. Two variants for the QM part computed with LC-DFTB2 were tested. In the first variant, the link atom was set on the C$-$S bond and the OB2 (base) parameter set was used. In the second variant, the sulfur was included in the QM zone and the link atom was set on the $C_\beta - C_\alpha$ bond. The new parameter set from Ref. [91] was used which contains repulsive parameters for sulfur. The range-separation parameter was ω=0.3 a_0^{-1} in both variants.

Three representative snapshots were selected from the classical MD simulations as start strucutres, each with badan in a different conformation and environment. The system was shortly equilibrated for 20 ps at 300 K with the Nosé–Hoover thermostat [162, 163] in ground state. The time step was 1 fs and all bonds were kept constraint. The subsequent productive simulations with the same time step sampled an NPT ensemble employing the Parrinello–Rahman barostat [164, 165] for a pressure of 1.013 bar. Only bonds involving hydrogen were constrained and PBC and PME were used. The simulation length is 500 ps, respectively.

Vertical excitation energies were computed with TD-LC-DFTB2 along the ground state QM/MM trajectories using DFTB+ (version 20.1). The SCC tolerance was 1×10^{-8} electrons and the screening threshold was 1×10^{-16}. The calculated QM part was the same as in the simulations and the remaining atoms of the system were considered as point charges. Gaussian fits to histograms of the vertical excitation energies provided the absorption maxima. The transition with the larger oscillator strength from the first two transitions was considered, respectively. The first 20 ps were counted as additional equilibration time and not used for the histograms.

6.5. Results and Discussion

Firstly, the newly derived sets of polarised charges for the glucose molecule are evaluated. Then the simulated conformations of GGBP are discussed. Finally, the conformations of badan and its interactions are studied in detail.

6.5.1. Glucose Charges

The atomic partial charges derived for the glucose molecule in the binding pocket of the wildtype for different cut-off radii for the environmental point charges and alternatively for the point charges of complete residues in the proximity of glucose were examined first. The glucose charges are shown in Fig. A.1 and Tab. A.20 in the appendix. Obviously, the amount of environment considered in the calculation of the ESP plays a minor role in the range between 4 to 8 Å. The charges of the same atoms are very similar and do not seem to converge with an increase of the cut-off radius. From that data it seems to be justified to take the environmental point charges within a radius of 5 Å to polarise the glucose.

Fig. 6.6 displays the glucose molecule with the names of the atoms. The polarized glucose charges obtained considering the environment within 5 Å in the binding pocket are compared to the charges derived with different approaches in Tab. 6.2 and Fig. 6.7. Additionally, the deviations of the atomic partial charges from the various new charge models to the glycam force field are visualized in Fig. 6.8. They were calculated as difference between the absolute values of the partial charges, respectively. Hence, a positive value means, that the absolute partial charge increased compared to the glycam force field and a negative value indicates a decrease. Clear trends are observed between the methods for the different groups of glucose atoms: oxygens, hydrogens bonded to oxygen, carbons and aliphatic hydrogens.

In the BPC models, the absolute values of the atomic partial charges of the hydrogens in the hydroxyl groups and most oxygens are larger than in the glycam force field or the other charge models. This is an expected result due to the polarisation from the polar binding pocket. The largest difference between the BPC model for the wildtype and the BPC model for the triple mutant with badan is found for the ring oxygen of glucose (O2).

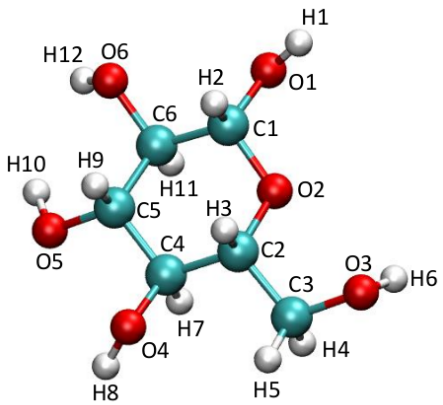


Fig. 6.6.: Glucose molecule with names of the atoms.

Tab. 6.2.: Atomic partial charges computed for the glucose molecule with various methods.

atom	BPC wt	BPC triple mut. Bad.	WPC[1]	QM/MM wt pocket	QM/MM water	Glycam
H1	0.540	0.544	0.483	0.437	0.413	0.445
O1	−0.747	−0.748	−0.708	−0.619	−0.625	−0.639
C1	0.347	0.257	0.414	0.464	0.451	0.384
H2	0.098	0.119	0.040	0.007	0.029	0.000
O2	−0.542	−0.462	−0.487	−0.499	−0.486	−0.471
C2	0.124	0.022	0.113	0.144	0.155	0.225
H3	0.140	0.126	0.089	0.056	0.056	0.000
C3	0.245	0.105	0.255	0.137	0.131	0.282
H4	0.031	0.096	0.047	0.045	0.058	0.000
H5	0.031	0.096	0.047	0.059	0.055	0.000
O3	−0.835	−0.805	−0.730	−0.640	−0.626	−0.688
H6	0.533	0.534	0.450	0.419	0.406	0.424
C4	0.052	0.063	0.098	0.166	0.157	0.276
H7	0.110	0.125	0.103	0.045	0.064	0.000
O4	−0.717	−0.685	−0.748	−0.609	−0.626	−0.714
H8	0.516	0.518	0.500	0.421	0.407	0.440
C5	0.156	0.112	0.219	0.176	0.160	0.284
H9	0.138	0.160	0.078	0.053	0.062	0.000
O5	−0.806	−0.832	−0.741	−0.690	−0.624	−0.709
H10	0.522	0.533	0.487	0.449	0.406	0.432
C6	0.248	0.318	0.138	0.129	0.120	0.310
H11	0.095	0.095	0.103	0.080	0.073	0.000
O6	−0.858	−0.868	−0.724	−0.684	−0.623	−0.718
H12	0.581	0.575	0.474	0.455	0.407	0.437

[1] done by Ziwei Pang

The mutated residue His152 is located in the proximity of this oxygen. In the wildtype, the histidine is weakly interacting with its N-H group. The resulting polarisation is missing in the triple mutant.

The new water charge model shows the same trend but the differences to the glycam force field are more moderate. It was expected to obtain values close to the glycam force field where the charges were also derived in aqueous solution. In contrast to the BPC model, the charges derived from the QM/MM simulation in the binding pocket do not display the polarization effect. Charges of atoms involved in hydrogen bonds have even lower absolute values than the corresponding charges of the glycam force field. A comparison to the QM/MM charge set derived for water reveals only small differences between both QM/MM charge sets.

The results indicate that the BPC charge models are most suitable to represent the situation in the binding pocket. Since the glucose molecule is tightly bound by the amino acid network, its conformational freedom is strongly limited and a charge model based on a single geometry with point charges from a single conformation of the atoms within 5 Å is reasonable. The coordinates for the wild-type were obtained from the solvated crystal structure after an energy minimization which accounted for a representative structure. In the case of the triple mutant with badan this is not that clear since the dye is flexible on one hand and the bound glucose state may be slightly different than in the wild-type. However, these effects are expected to be small and the charge model proves later to be satisfying.

The QM/MM simulations with electronic embedding should capture the effect of different QM and MM conformations together with the polarization effect of the environment on the electron density in the QM zone. The atomic partial charges were obtained from the electron density by means of a Mulliken population analysis. Such a population analysis is computationally cheap but shows to be unsuitable for the purpose here.

The charges are fix during the complete simulation time. This means, that the glucose with BPC charges may be overpolarized in water after an unbinding event and then be better described by for instance the WPC charges. The size of this effect was investigated

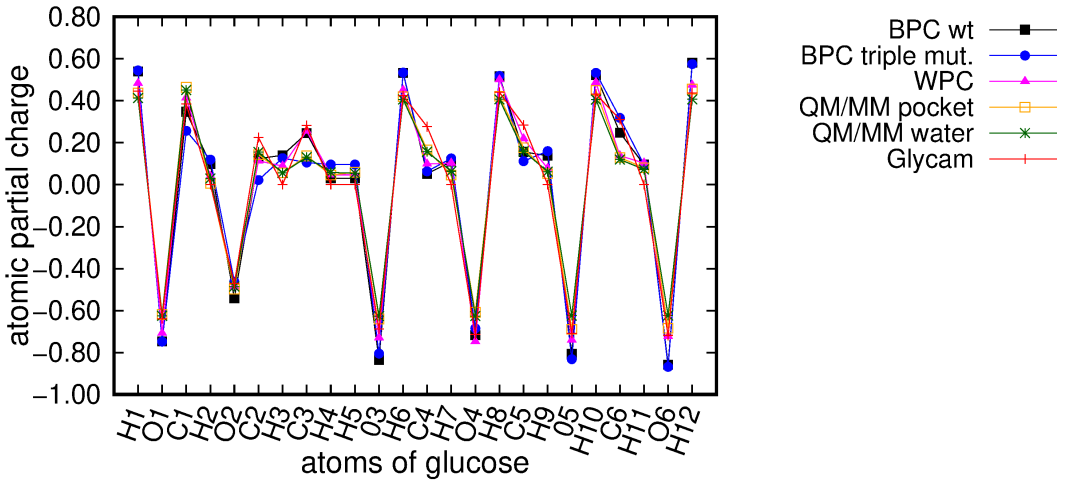


Fig. 6.7.: Partial charges of glucose atoms derived with different methods detailed in the text.

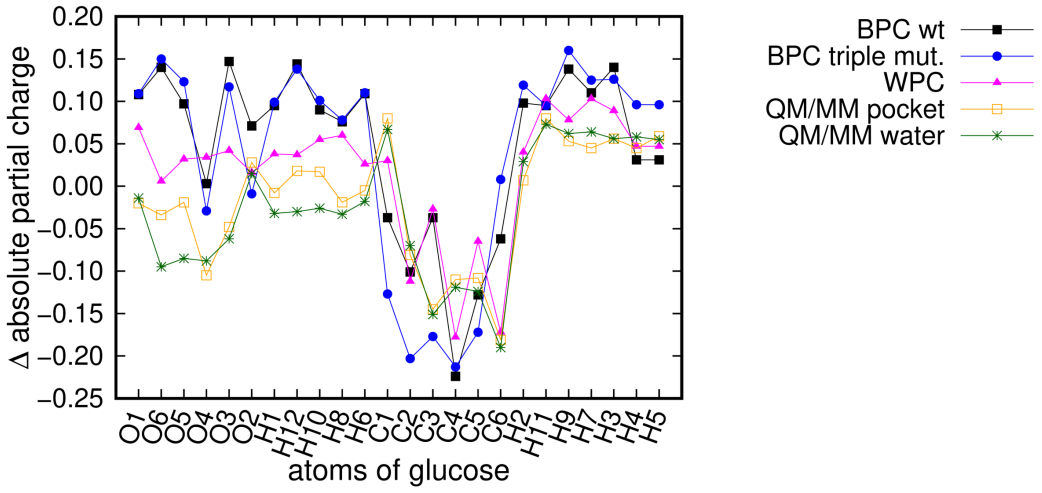


Fig. 6.8.: Difference between the absolute values of newly derived partial charges for glucose and the absolute values of the Glycam charges. The atoms are sorted for an easy recognition of the trends of the various methods.

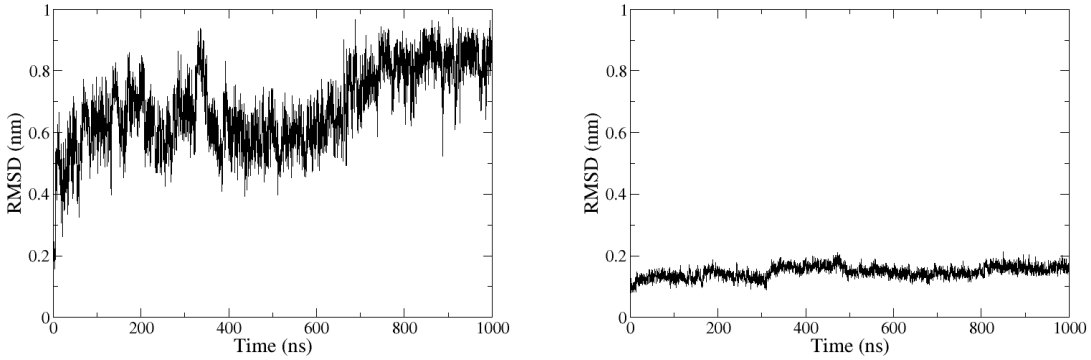
by Ziwei Pang in the context of a free energy study to GGBP. For the primary goal here, which is the study of stable protein conformations and the badan conformations in dependence of the glucose locations and protein conformations, it is less important than a stable closed state. The current polarization could in principle be described by a polarizable force field. However, this would be computationally too demanding for the desired simulation time.

6.5.2. Protein Conformations

The application of the newly derived BPC model led to a stable closed wild-type protein as expected from the experimental k_D value, which corresponds to a free energy difference of approximately 9.2 kcal mol^{-1} between the *holo* closed and *apo* open protein. Starting from the *holo* closed conformation, such a difference plus the energy of the barrier is almost impossible to be overcome within 1 μs. In Fig. 6.9, the root mean square deviations (RMSD) of the protein backbone with respect to the initial closed crystal structure are compared for a simulation with glycam glucose charges and the BPC charge model. While the RMSD in the first case increases very fast to almost 1 nm and shows large fluctuations, in the second case it remains below 0.3 nm and fluctuates only slightly.

This result shows that the change of the charges has the desired effect. However, it needs to be mentioned that the almost direct opening of the protein while using glycam charges is an extreme example. In another simulation, the RMSD remained below 0.5 nm and the glucose did not leave the protein within the simulated microsecond.

In Ref. [80] it is reported that in all of their four test simulations starting in *holo* closed form the protein opened within 100 ns. In Ref. [191], the closed state was also only simulated for 100 ns, although without reporting stability problems. Thus, a microsecond long simulation in the *holo* closed state is a new achievement enabled by the derived polarized glucose charges. The found stable state in solution can now be compared to the crystal structure and the states observed for the triple mutant with badan in solution.

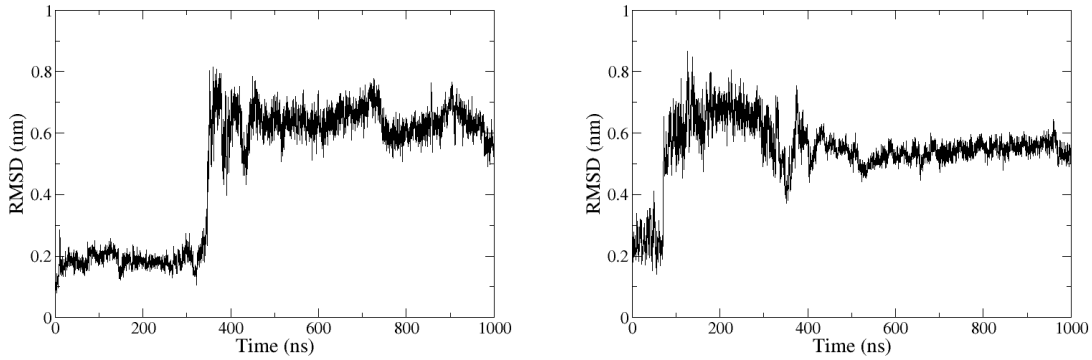


(a) glycam charges **(b)** BPC

Fig. 6.9.: Root mean square deviations of the wild-type protein backbone during the simulation with respect to the closed crystal structure. In the simulation with the glycam force field describing the glucose, the protein quickly opens and the glucose leaves the binding pocket. In the simulation applying the BPC charge model the protein is very stable in its closed state with the glucose in the binding pocket.

The triple mutant with badan has a larger k_D value and thus binds the glucose weaker than the wild-type. The free energy difference between the *apo* and *holo* state is approximately 2.7 kcal mol^{-1}. Depending on the size of the energy barrier, it may be possible to observe an unbinding process within 1 μs. The RMSDs of a simulation with glycam charges and of the simulation with BPC charges (with position restraints during the equilibration) are shown in Fig. 6.10. Evidently, the protein undergoes large conformational changes in both cases. In the simulation with glycam charges, the glucose leaves the binding pocket during the conformational change after about 350 ns. In the simulation with BPC charges, the glucose remains in the binding pocket. In both simulations, the closed protein firstly opens and then takes a half-closed (simulation with glycam charges) or differently closed (simulation with BPC charges) conformation. From these simulations alone it is not possible to judge whether the BPC are superior to the glycam charges. However, it is also not clear if the start from the crystal structure of the wildtype is a stable structure of the triple mutant which could lead to a too fast opening or glucose unbinding.

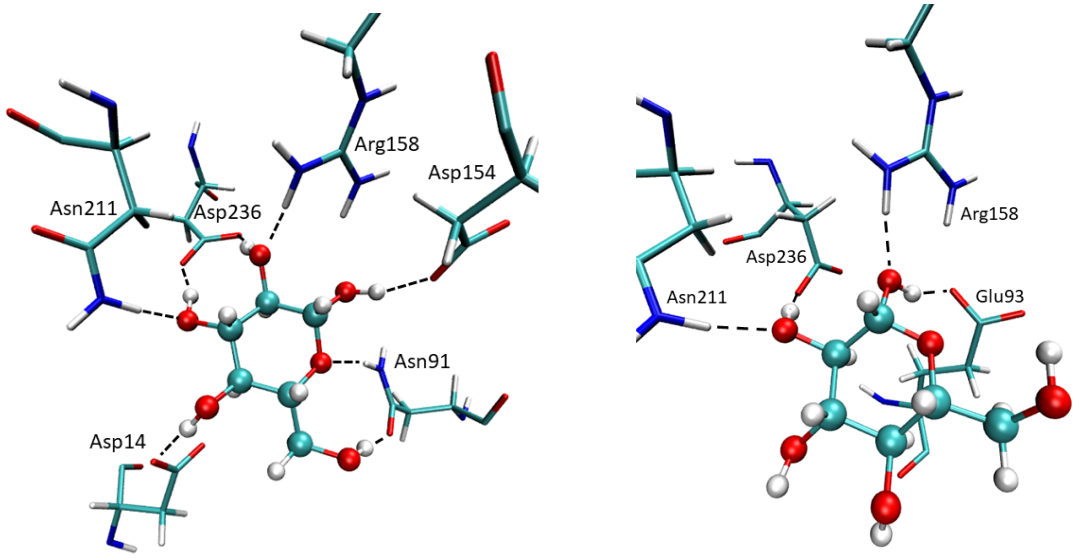
The hydrogen bonds between glucose and the protein in the simulations with BPC and glycam charges were analysed. The hydrogen bond pattern of the wildtype is shown in Fig. 6.11 (a) but without the His152 which is mutated to a cysteine in the triple mutant. The mutation A213R brings an arginine in the proximity of glucose, which can form a hydrogen bond. In the simulation with glycam charges, it is observed that Asn66 forms a hydrogen bond with the same oxygen of glucose as His152 in the wildtype. The hydrogen bond with Arg158 becomes very weak and also the hydrogen bond with Arg213 is unstable within the first 100 ns. When the protein opens, these hydrogen bonds break as already mentioned. It is different in the simulation with BPC charges. Hydrogen bonds between the glucose and Asp236, Arg158 and Asn211 are found after 50 ns. They are strong enough



(a) glycam charges **(b)** BPC

Fig. 6.10.: Root mean square deviations of the backbone of the triple mutant with badan during the simulation with respect to the closed crystal structure.

to hold the glucose for roughly 300 ns when the protein is in open state. Especially Asp236 interacts persistently with glucose. Glu93 additionally forms a strong hydrogen bond to glucose in the final stable closed state of the triple mutant (Fig. 6.11 (b)). The same hydrogen bond pattern is found in the second simulation using BPC. There additional hydrogen bonds between Arg213 and glucose are observed.



(a) wildtype structure without His152 **(b)** structure found in sim. with BPC

Fig. 6.11.: Hydrogen bonds between glucose and the triple mutant.

The charges are important to get a reasonable free energy from metadynamics simulations. These were performed and analysed by Ziwei Pang and support the use of the BPC charges. The minimum of the closed triple mutant is indeed a little shifted compared to the minimum of the closed wildtype. Interestingly, an extreme open structure corresponds to a minimum of the triple mutant as well as the wildtype. [208]

A hinge and a twist angle were used as CVs in the metadynamics simulations based on refs. [190, 191]. The hinge angle θ describes the opening–closing motion and is defined by the centers of mass of the N-terminal domain (residues 3–108 and 258–291), the hinge region (residues 109–111, 255–258 and 292–296) and the C-terminal domain (residues 112–254 and 297-306). The torsion angle φ describes the twisting motion and is defined by the centers of mass of the N-terminal domain, the N-terminal domain base (residues 109, 258 and 292), the C-terminal domain base (residues 111, 255 and 296) and the C-terminal domain (Fig. 6.12).

The RMSD gives a fast impression on the stability of the protein structure but it is not useful to distinguish various conformations. The larger the RMSD value to the more different conformations it could correspond. For this reason, the observed protein conformations were characterized based on the θ and φ angles. They were measured by means of the *driver* tool from plumed [209].

The values of the crystal structures [183] are reported in Tab. 6.3. Evaluating the simulations in water reveals differences between the artificial crystal structures and the real structures and their fluctuations in solution. This analysis was mainly done by Ziwei Pang and is found in Ref. [208]. Here the main points are summarized for completeness.

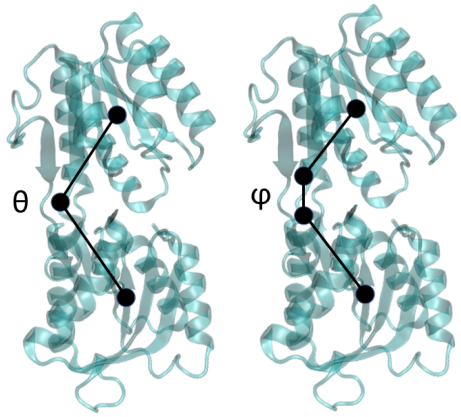


Fig. 6.12.: Hinge and twist angle describing the conformational changes of GGBP.

Tab. 6.3.: Hinge and twist angles of the crystal structures.

descriptor	*holo* closed	*apo* (half-)open
θ	122°	143°
φ	65°	90°

The θ angle in the wild-type simulation with BPC remains quite close to the crystal value but the φ value shows fluctuations to lower values (around 20-30°) which indicate a twist compared to the crystal structure. Without glucose both angles clearly increase: θ fluctuates around 150°, φ becomes larger than 100° up to more than 170°. This indicates that the *apo* crystal structure has a more intermediate, half-open conformation and thus not represent the real open form. This is possibly due to the presence of crystallization reagents, i. e. citrate and sodium ions which stabilize a more closed form [183, 191]. As already mentioned, also the metadynamics simulations indicate a significantly more open structure as minimum. Besides, in Ref. [191] it is discussed that the *apo* protein appears to about 30 % in half-closed conformations but otherwise is significantly more opened.

In the several free MD simulations of the triple mutant with badan using the different charge sets for glucose, different metastable closed, half-opened and widely opened states are observed. These are discussed in the next subchapter together with the conformations of badan.

There are two simulations with BPC and two simulations of the *apo* triple mutant with badan. One of them is started after an equilibration with position restraints and the other one from the usual equilibration without position restraints, respectively. This was done to detect possible artifacts from equilibration. The direct comparison between the respective simulations showed that the position restraints only delayed the opening motion of the protein a little. However, in the other two simulations with glycam and WPC charges also no direct opening is observed. Hence, there is probably no equilibration artefact. In the following analysis, they are treated as equivalent. The simulation with position restraints in the equilibration is referred to as simulation 1 and the other simulation as simulation 2, respectively.

6.5.3. Badan Conformations and Interactions

The badan conformations are analyzed to gain some hints on the working principle of the glucose sensor. It is of particular interest if or how often the dye is oriented inside the binding pocket and if there are differences in its conformation depending on the protein conformation or presence of glucose.

Although the BPC are the most reliable charges, the simulations with glycam charges or WPC were evaluated as well. It is not expected that the charges have a notable influence on the badan conformations. In principle, they effect the probability of an interaction between badan and the glucose but no hydrogen bonding between glucose and badan is observed. They may further have an indirect impact on the badan conformations since the binding of glucose influences the protein conformations. However, since the analysis is performed always in relation to the current protein conformation and glucose position, this is not an issue. The frequency of a certain constellation can be taken into account afterwards.

Several different badan orientations are observed in the simulations, which are determined by three flexible dihedral angles between the naphthalene core of the dye and the cysteine side chain of the protein. These dihedrals (C3-C4-S1-C14, C4-S1-C14-C7, S1-C14-C7-C8, see Fig. 6.13) were plotted versus simulation time in Fig. 6.15, together with the angle θ and dihedral φ indicating the protein conformation (Fig. 6.12). This allows the identification of badan conformations which are stable for a significant time in a certain protein conformation. At the beginning, badan was located inside the binding pocket.

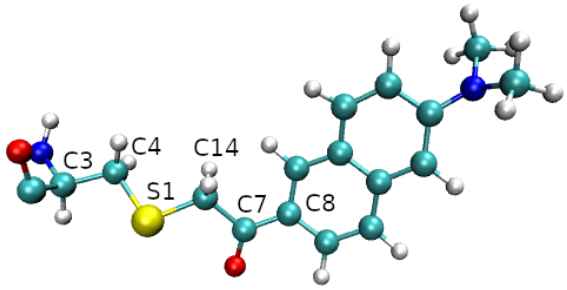


Fig. 6.13.: Cysteine with attached badan. The atoms involved in the three studied dihedral angles are labelled.

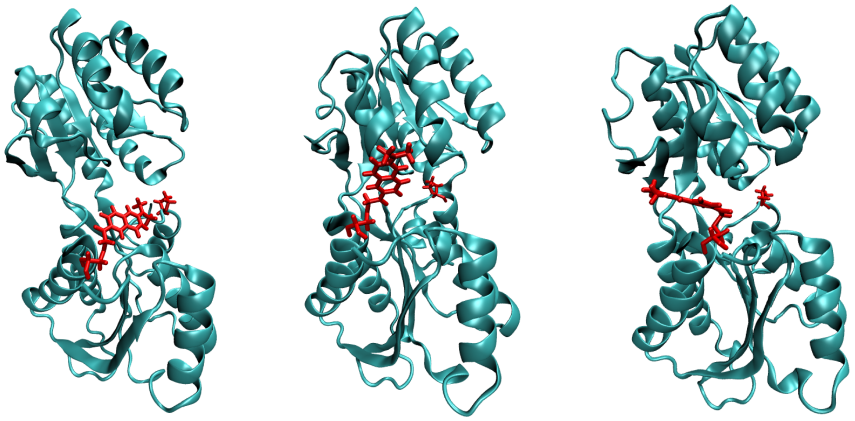
In both simulations with BPC, the protein and badan take conformations which are stable for several hundreds of nanoseconds, in one simulation after 500 ns, in the other simulation after 100 ns (Fig. 6.15 (a),(b)). Both protein states correspond to a twisted closed state with the glucose in the binding pocket. The badan dihedrals fluctuate around the same values ($\approx 100°, -55°$ and $-70°$) and the dye points outside the protein away from the binding pocket (Fig. 6.14 (c)). In the periods where the protein is in an open state, badan is quite flexible as seen on the frequently changing dihedral angles, respectively.

The stable badan conformation in the BPC simulations is also found in the glycam simulation and persists there for more than 300 ns (Fig. 6.15 (c)) with short exceptions. In this case the protein is in a closed state similar to the closed crystal structure. Badan changes its conformation in this simulation after the protein opens and the glucose leaves. The protein stabilizes after some time in a twisted half-open conformation and badan at the same time stabilizes in a straight (vertical) conformation (Fig. 6.14 (b)). This badan conformation is also frequently visited in the simulation with WPC, while the protein is stable in a less twisted closed state (Fig. 6.15 (d)). In this simulation the glucose left the binding pocket after around 200 ns when the protein opened the first time. Until there, badan remained in a stable conformation inside the pocket. In the last 150 ns another stable conformation was observed with badan inside the pocket of the open protein.

The two simulations without glucose differ in their protein conformations as seen on the φ dihedral very clearly. In simulation 1, after about 250 ns the protein stabilizes in a very twisted state (Fig. 6.15 (e)). Badan changes its conformations frequently and finally stabilizes after more than 700 ns simulation time in a conformation inside the binding pocket.

In the other simulation without glucose, the protein instead takes a half-opened conformation in roughly the first half of the simulation time and a conformation very close to the closed crystal in roughly the second half of the simulation time (Fig. 6.15(f)). In the first simulation part, badan is straight (Fig. 6.14 (b)) and in the second part, it is in the binding pocket (Fig. 6.14 (a)).

In the *apo* simulations it is observed that the dye comes close to a tryptophane residue (Trp183) inside the binding pocket. A snapshot from each simulation of the *apo* protein is shown in Fig. 6.16 visualizing badan and Trp183. This it noteworthy, since in Ref. [210] it was found that tryptophane can quench the fluorescence of badan in micelles and cytochrome P450 proteins. This quenching is explained by an electron transfer from



(a) Badan inside of the binding pocket, distance to Pro239: 10 Å.

(b) Badan outside, distance to PRO239: 25 Å

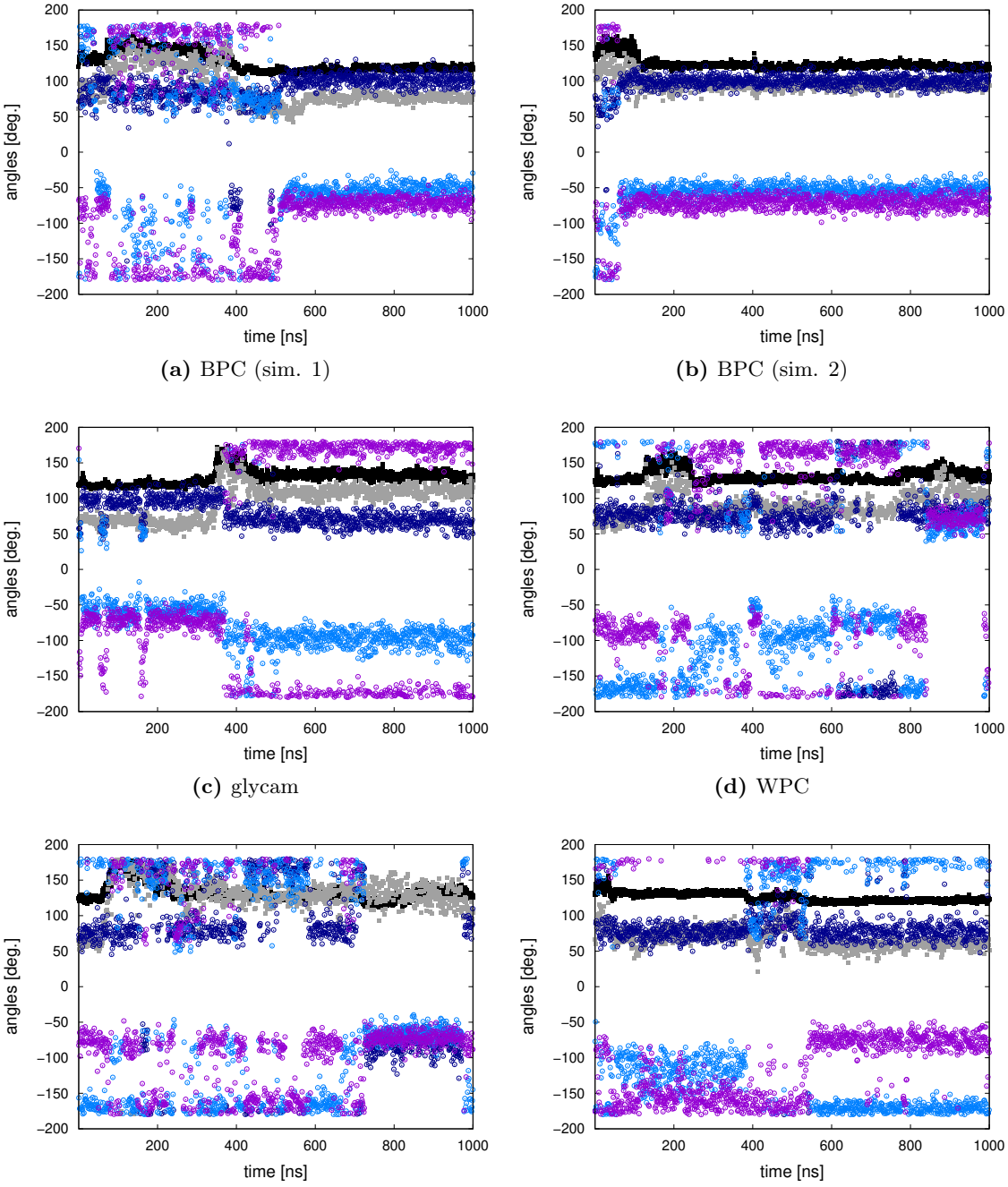
(c) Badan outside, distance to PRO239: 26 Å

Fig. 6.14.: Representative badan orientations. Badan and Pro239 are red coloured. The distance between the nitrogen of the Dimethylamino moiety of Badan to the nitrogen of Pro239 is used to distinguish badan conformations inside the binding pocket and outside of the binding pocket.

tryptophane to excited badan in the same study. A Trp-badan complex is assumed and surrounding water was found to be a precondition for the electron transfer reaction. These criteria are fulfilled according to the simulations of GGBP with badan. In the recent study on the single mutant H152C-badan quenching from Trp183 is also proposed based on a comparison to the mutant H152C-badan/W183F [19].

This interesting finding was investigated further by measuring the distance between badan and Trp183. Fig. 6.17 shows the distances between the centers of the ring systems. Badan and Trp183 come close to each other up to 4 Å. The distance increases to about 10 Å when the dye is outside the protein.

It is notable that the dye goes frequently inside the binding pocket when there is no glucose. In a next step, it is thus investigated how often the dye is inside the pocket when there is glucose and how often it is outside. Besides the possible quenching of Trp, this is further of interest since badan is known to be environmentally sensitive. All possible orientations of badan were observed among the stable conformations in the free MD simulations: inside the binding pocket, straight along the two domains of the (open) protein and outside pointing away from the binding pocket. However, the dihedrals do not easily display these three situations. On one hand, all three of them need to be considered to describe one conformation and on the other hand even different dihedral combinations describe the same badan orientation. This becomes obvious when the distances to Trp183 in Fig. 6.17 and the dihedrals in Fig. 6.15 (e) and (f) are compared. In the further evaluation the dihedrals are thus replaced by a simpler measure.



(a) BPC (sim. 1)

(b) BPC (sim. 2)

(c) glycam

(d) WPC

(e) *apo* (sim. 1)

(f) *apo* (sim. 2)

Fig. 6.15.: Angles describing the protein and badan conformations in the various free MD simulations. Black: θ angle describing the opening/closing motion of the protein; gray: φ dihedral describing the twisting motion of the protein; dark-blue: badan dihedral C3-C4-S1-C14; light-blue: badan dihedral C4-S1-C14-C7; purple: badan dihedral S1-C14-C7-C8.

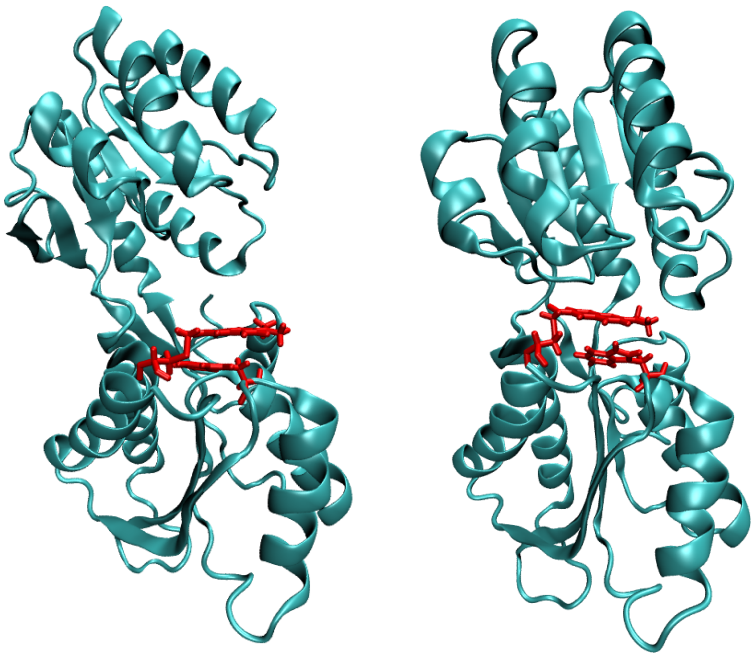
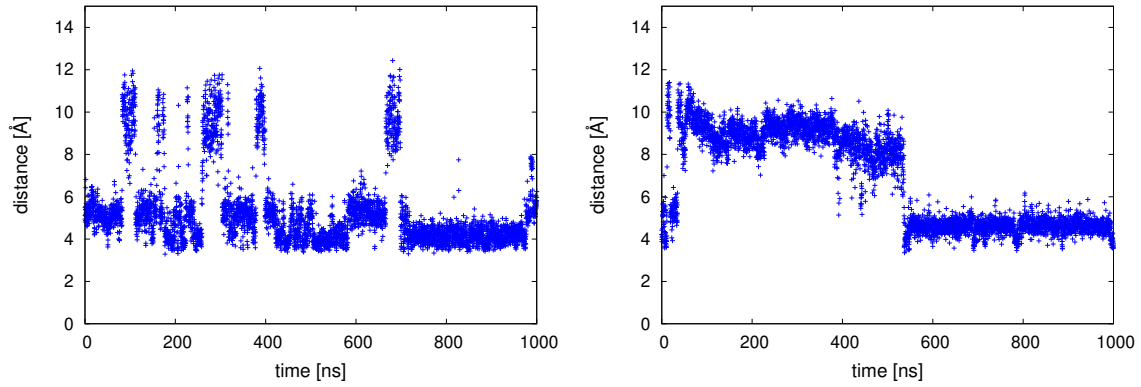


Fig. 6.16.: Stable badan conformations found in the *apo* simulations, where the dye is close to Trp183. This may lead to fluorescence quenching.



(a) *apo* (sim 1) **(b)** *apo* (sim 2)

Fig. 6.17.: Distance between badan and Trp183 along the trajectories of the *apo* protein.

How often is badan inside the pocket when glucose is there?

A semi-quantitative analysis of the badan orientation in relation to the protein conformations is presented. Two-dimensional histograms were created to visualize the amount of snapshots with badan inside or outside of the protein in the diverse protein conformations. The evaluation is done separately for the case of glucose in the binding pocket and the case of an empty binding pocket.

Since the glucose is able to move in the binding pocket or to halfly unbind while remaining sticked to few amino acids on only one domain, there is need to clarify what is meant by *glucose in* and *glucose out*. The glucose is considered as *in* here, if the distance of its center of mass and the center of mass of the binding pocket is not larger than 10 Å. The center of the pocket is defined by the center of mass of Asp14, Asn91, the atoms of Cys152 until C_β, Asp154, Arg158, Asn211, Asp236 and Asn256. This allows the glucose to rotate in the pocket, which it does frequently in the triple mutant, while keeping the value of the measured distance relatively constant. This would not be the case if simply the length of one hydrogen bond between glucose and the protein would be measured. The chosen measure, however, does not necessarily show how tightly the sugar is bound but the applied definition aims primarily on only locating the sugar in the area of the binding pocket.

It would be advantagous to describe the badan conformations with one single collective variable (CV). Since it is most important to know if the dye goes inside the protein, the distance between the nitrogen of the amino group at the outer end of badan and the nitrogen of a proline located on the other side of the binding pocket is chosen as measure. This distance becomes smaller if the dye comes inside the pocket and is large when it is outside.

This is a relatively simple test to decide if badan is inside or outside. However, it cannot clearly distinguish between the different conformations outside or at the cleft of the binding pocket. To achieve this, a second distance could be measured to an amino acid on the other domain, which could specifically detect the straight conformation at the edge of the binding pocket. Representative badan conformations and the corresponding distances to Pro239 are visualized in Fig. 6.14.

As for the badan conformation, one single CV is needed to represent the protein conformation. For this, a line was defined by the centers of the two minima from the metadynamics simulation of Ziwei Pang [208]. These have the following coordinates, if θ is on the x-axis and φ on the y-axis with unit degree, respectively: A(113/50), B(168/171). The equation of the line reads:

$$\varphi = 2.2 \cdot \theta - 198.6.$$

The points in the θ-φ space are projected on this line. The intersection of the normal with the y-axis is computed for every point $P(\theta_i/\varphi_i)$ according to

$$c_i = \varphi_i + \frac{1}{2.2} \cdot \theta_i$$

With c_i the projected θ' values can be obtained by

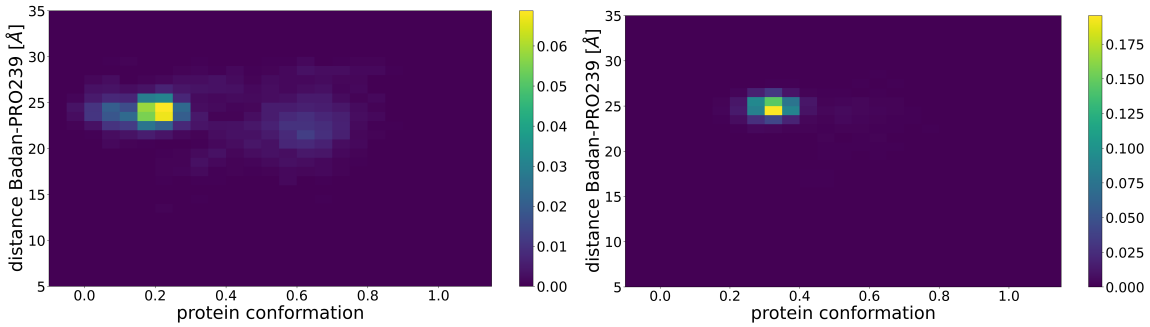
$$\theta' = \frac{c_i + 198.6}{2.2 + \frac{1}{2.2}}$$

and finally scaled arbitrarily on the new axis. The scale is chosen such that the center of minimum A corresponds to 0 and the center of minimum B corresponds to 1. The CV values are thus computed by

$$CV^{\text{final}} = \frac{\theta' - 113}{55}.$$

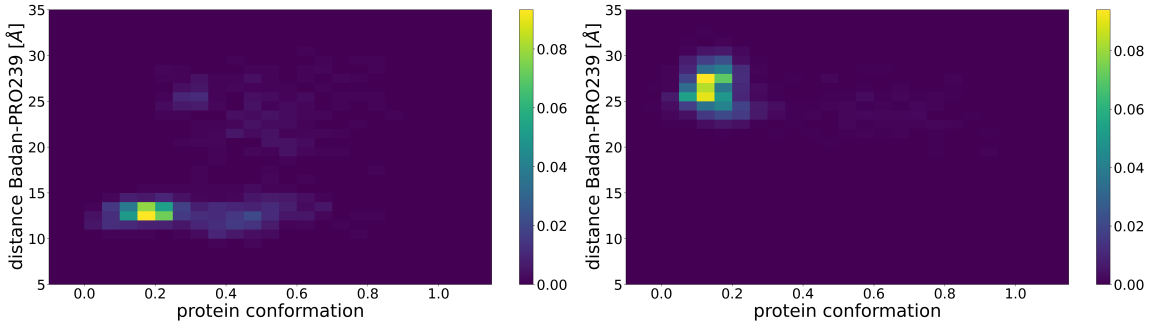
2D histograms were made with python 3.9 (numpy function histogram2d), with 25 bins on the x and y axes, respectively. The x axis shows the protein conformation as described above and the y axis shows if badan is inside the binding pocket or somewhere outside. The counts of the different areas were divided by the total number of snapshots used in the histogram.

The histograms disprove the hypothesis that badan is generally enveloped by the protein when a glucose molecule is bound, as hypothesized in [5]. In Fig. 6.18 (a),(b),(c) it becomes clear that the dye is outside of the binding pocket when the glucose is inside. On the contrary, in the simulation with WPC (Fig. 6.18 (c)) badan and glucose are inside the



(a) BPC (sim. 1)

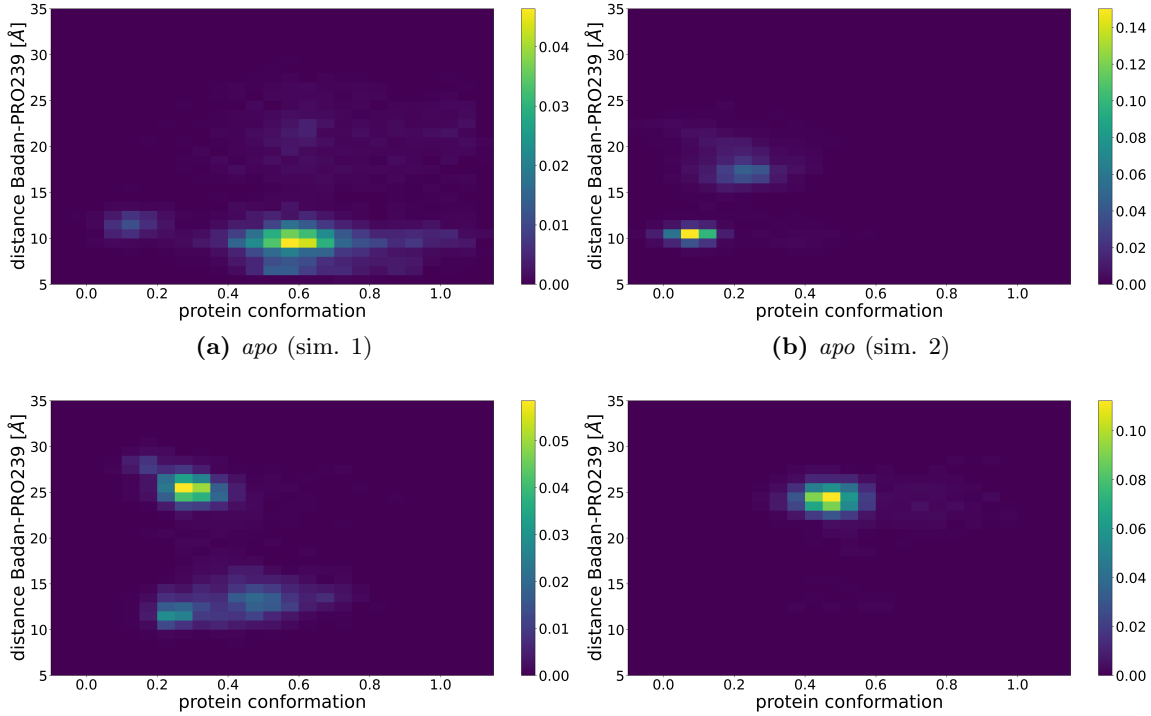
(b) BPC (sim. 2)



(c) WPC

(d) glycam

Fig. 6.18.: 2D histograms of protein conformation and badan orientation with glucose in the binding pocket. The x-axis represents the line connecting minima A and B from the metadynamics simulations. It is scaled such that 0 corresponds to A and 1 corresponds to B. The y-axis shows the distance between the top of badan and Pro239. A distance smaller than 15 Å indicates that badan is inside the binding pocket and a larger distance means that badan is outside. The counts in the histogram bins are divided by the total number of snapshots, respectively.



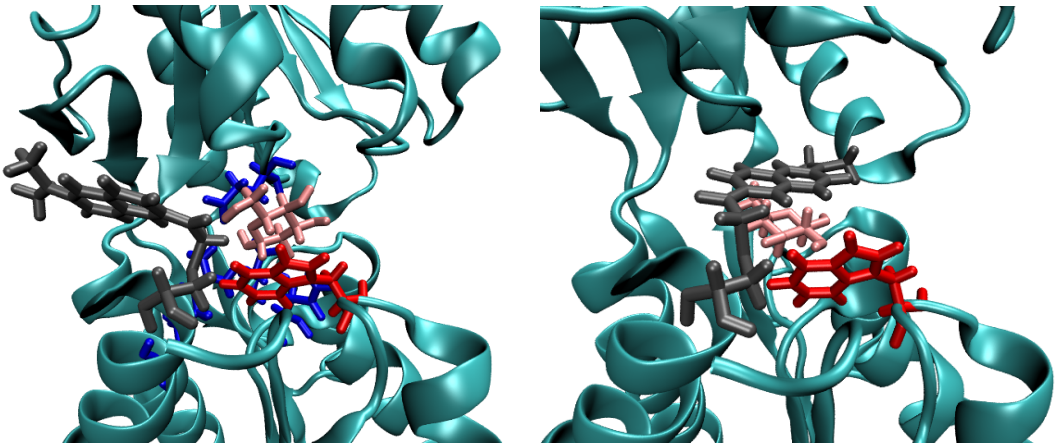
(a) *apo* (sim. 1) **(b)** *apo* (sim. 2)

(c) WPC **(d)** glycam

Fig. 6.19.: 2D histograms of protein conformation and badan orientation without glucose in the binding pocket. The x-axis represents the line connecting minima A and B from the metadynamics simulations. It is scaled such that 0 corresponds to A and 1 corresponds to B. The y-axis shows the distance between the top of badan and Pro239. A distance smaller than 15 Å indicates that badan is inside the binding pocket and a larger distance means that badan is outside. The counts in the histogram bins are divided by the total number of snapshots, respectively.

binding pocket at the same time. However, it needs to be considered that the simulations started with the dye in the pocket and the glucose left the pocket after around 200 ns. This means, that it is not clear if badan would stay inside for a longer time when the glucose remained there. In the same simulation, after the glucose left, the fluorophore takes conformations inside and outside of the binding pocket, roughly equally distributed (Fig. 6.19 (c)). In the simulations without glucose (Fig. 6.19 (a),(b)), badan is oriented inside the binding pocket.

These results show that all variants are possible: glucose and badan inside the pocket at the same time, glucose inside but badan outside, badan inside when no glucose is there and an empty binding pocket with badan oriented outside. However, it is indicated that badan is less likely inside at the same time as glucose, since Fig. 6.18 (c) may not be representative. In the case that no glucose is in the binding pocket, badan can take stable conformations inside or outside of the pocket.



(a) BPC **(b)** WPC

Fig. 6.20.: Arrangement of Trp183 (red), glucose (pink) and badan (grey). The amino acids forming stable hydrogen bonds with glucose are shown in blue.

The evaluated simulations started from the crystal structures and not the minima found in the metadynamics simulations. However, the sampled closed states do not deviate much from the minimum of the *holo* protein (see Figs. 6.18 and 6.19). The open states visited in the simulations are not as open as the minimum of the metadynamics simulations. This minimum is reached only shortly in the *apo* simulation 1 (Fig. 6.19 (a). Nonetheless, the results are expected to be meaningful since the protein is sufficiently open to allow a free movement of the dye already on half way between open and closed. The *apo* crystal structure of the wildtype for instance is located at roughly 0.4 on the histogram scale.

Trp183 is one of the aromatic amino acids sandwiching the glucose in the closed wildtype crystal structure (Fig. 6.3 (a)). Thus, it is not surprising that glucose and badan may compete about the space behind Trp183 in the binding pocket. Fig. 6.20 visualizes the binding pocket with Trp183, glucose as found during the BPC simulations and badan. It is evident, that Trp183 is still close to the glucose molecule and likely it supports the glucose binding by sterically keeping the glucose in its position. There may be also a carbohydrate-aromatic interaction [211].

The simulation with WPC is the only example for glucose and badan being in the binding pocket at the same time. A picture of this situation is found in Fig. 6.20. Evidently, the glucose molecule is deeper in the binding pocket in this case making some space for badan above Trp183. Compared to the BPC simulation, the glucose is hold by fewer hydrogen bonds which are frequently exchanging.

Environmental Changes in the Binding Pocket

After analysis of the orientation of badan it remains to characterize the surrounding of the dye inside the binding pocket. A possible interaction with tryptophane was already mentioned. But how shielded is badan in the pocket from the water solvent? And how hydrophobic is the pocket at all? The number of polar amino acids binding the sugar adumbrates that the environment of the dye in the pocket could be similarly polar as the

environment outside. On the other hand, the dye has no particularly long linker, so it should also feel the protein if it is oriented outside the pocket. Further, it is evident that water is able to enter the binding pocket from simply watching the simulation trajectories. Hence, the difference in polarity between inside and outside is not clear at all.

In the following, some quantities are presented which are directly obtained from the classical MD simulation. The number of water molecules in different radii around badan, the number of hydrogen bonds with water but also with the protein and glucose. The analysis was done with gromacs. Subsequently, the results of ground state and trial excited state QM/MM simulations are summarized.

Classical MD simulation

Tab. 6.4 reports how many water molecules encompass badan inside and outside of the binding pocket on average. The BPC simulations contain glucose in the binding pocket in contrast to the *apo* simulations. Only two frames of BPC simulation 2 have badan inside the pocket. Thus, no values are shown for this case. The numbers show that there is almost as much water around badan if it is oriented inside the binding pocket as if it is outside. It is also evident that there is no big difference arising from the presence or absence of glucose. However, the number of frames with badan inside in the BPC simulation 1 is also very small (17) and hence should not be overinterpreted. Interestingly, in the *apo* simulation 2 the numbers are a little smaller compared to the other simulations.

Tab. 6.4.: Average number of water molecules surrounding badan. Water molecules were counted if the distance of their center of mass to any atom of badan (only dye core without cysteine linker) was smaller than the chosen cut-offs.

Badan pos.	simulation	3 Å	5 Å	10 Å
in	BPC (sim. 1)	14	41	181
	apo (sim. 1)	15	44	193
	apo (sim. 2)	12	29	121
out	BPC (sim. 1)	16	49	224
	BPC (sim. 2)	15	47	216
	apo (sim. 1)	18	56	254
	apo (sim. 2)	13	37	173

The carbonyl oxygen of badan is a typical hydrogen bond acceptor, the nitrogen could serve as a weaker acceptor. Tab. 6.5 gives an impression on how often badan forms hydrogen bonds with water, the protein or glucose. The hydrogen bonds were counted applying the default criteria of gromacs which are a maximum distance between donor and acceptor (heavy atoms) of 3.5 Å and a maximum angle between hydrogen-donor-acceptor of 30°. Again, the BPC simulations have no representative number of frames with badan inside but for badan outside they witness frequent hydrogen bonding with water as expected. Remarkably, this is also found for the *apo* simulations and badan inside. Interestingly, the *apo* simulations report a significant number of hydrogen bonds between badan outside

and the protein. In simulation 2, in 73 % of the simulation time with badan outside the dye is specifically interacting with the protein. An analysis of the trajectory reveals that a hydrogen bond between the backbone N-H group of Asp154 and the carbonyl group of badan is stable for about 300 ns. Badan remains oriented similar to Fig. 6.14 (b) during that time. A few times, badan and the NH in the side chain of Trp183 form a hydrogen bond. The same is observed in *apo* simulation 1 but the hydrogen bond to Asp154 was only stable for few nanoseconds. These two hydrogen bonds are further the most important ones with the protein in the BPC simulations, too.

In sum, outside of the pocket there are slightly fewer hydrogen bonds to water than inside showing that water easily reachs the dye in the pocket. Comparing the percentage of frames with no hydrogen bonds, the numbers for glucose in the pocket and badan outside are similar to no glucose and badan inside. The most hydrogen bonds in total are observed in the *apo* simulations with badan outside.

Tab. 6.5.: Percentage of frames with hydrogen bonds (one or more) between badan and water, badan and protein and badan and glucose. It is possible that the dye forms a hydrogen with water and with the protein at the same time.

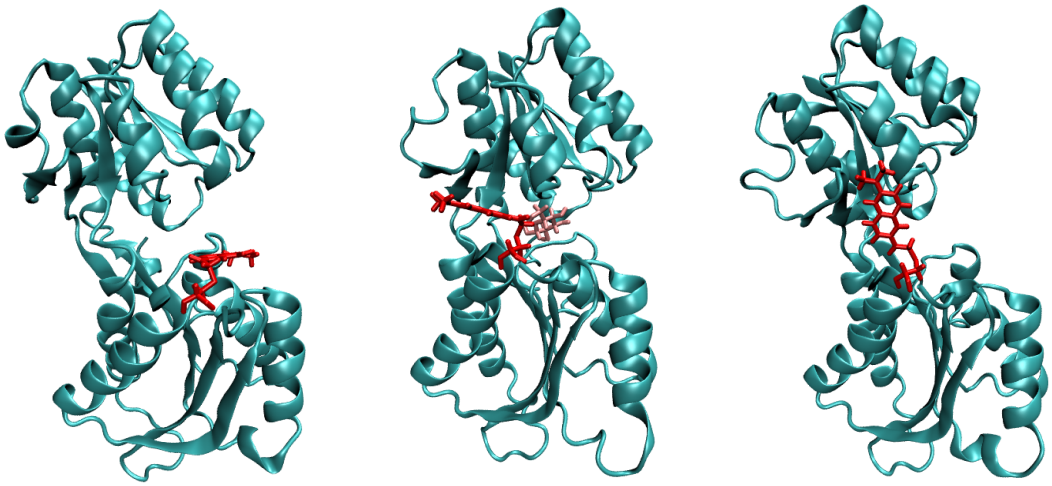
Badan pos.	simulation	water	protein	glucose	none
in	BPC (sim. 1)	88 %	12 %	0 %	12 %
	apo (sim. 1)	81 %	2 %	-	18 %
	apo (sim. 2)	85 %	4 %	-	13 %
out	BPC (sim. 1)	71 %	16 %	1 %	16 %
	BPC (sim. 2)	76 %	6 %	0 %	18 %
	apo (sim. 1)	58 %	40 %	-	10 %
	apo (sim. 2)	30 %	73 %	-	4 %

QM/MM simulations

Excited state QM/MM simulations of badan in the emitting state can shed light on the specific interactions in the excited state. Moreover, fluorescence spectra can be computed based on histograms of vertical excitation energies computed along the trajectory. In these calculations, the environment can be considered as point charges. Such spectra can then reveal changes in the complex electrostatics around badan in its different conformations. The experimental fluorescence spectra obtained in the presence and absence of glucose are slightly shifted to each other. Thus, large changes in the polarity of the environment of badan are not expected but a small effect may be observed.

The start conformations of the simulations are (i) badan inside, (ii) badan outside pointing away from the binding pocket and (iii) badan outside straight from the C-domain to the N-domain (Fig. 6.21). From each of these conformations two simulations were performed differing in the size of the QM part and the LC-DFTB2 parameters applied (see comp. details).

Badan is very similar to prodan studied in the previous chapter. Not surprisingly,



(a) Badan inside, sim. (i) **(b)** Badan outside, sim. (ii) **(c)** Badan outside, sim. (iii)

Fig. 6.21.: Start structures of the QM/MM simulations.

badan attached to GGBP behaves in this aqueous environment like prodan in water in the excited state QM/MM simulations. The dimethylamino moiety is twisted by 90° in the first excited state in all of the three conformations and indenpendently of the parameter set used. However, it is interesting that this twist only happens after ≈ 100 ps in the simulation with badan in conformation (iii) and the OB2 (base) parameters but almost immediately in all other simulations. It seems that the dimethylamino moiety is so close to the protein that a twist is sterically hindered and it needs this time to move slightly away. The computed excitation energies show a large jump upon the twist from ≈ 3 eV to only ≈ 1.5 eV. At the same time also the oscillator strengths become significantly smaller.

Although the photophysical properties of badan are still explained with planar and twisted ICT states in a recent publication [19], it seems unlikely that there is a twisted state if the studies on prodan are recalled (chapter 5). In Ref. [19], the single mutant H152C-badan is investigated experimentally and it is assumed that a TICT state emits in the *apo* protein. It is stated that the emission from this state has a low intensity due to its solvation and formation of hydrogen bonds. Emission with higher intensity from an unsolvated PICT state instead is proposed for the *holo* protein [19]. However, this hypothesis does not agree with the picture obtained from the classical MD simulations. As discussed above, badan is solvated in the *apo* as well as in the *holo* protein. A large difference between the single mutant and the triple mutant studied here is not expected in this respect. It is thus concluded, that the sampled state does unlikely correspond to the real emitting state and no further analysis is performed on these simulations.

The absorption spectra of Badan attached to GGBP are expected to be solvent dependent as it is the case for free badan [22] and prodan [117]. However, there is no experimental reference data for the absorption spectra of the *apo* and *holo* triple mutant with badan. Nonetheless, the computed spectra can give an idea about the colour shifts between the different orientations of badan.

The results from the six ground state QM/MM simulations are shown in Tab. 6.6. Both parameter sets show the same trend in the absorption energies. The simulation with badan

inside the binding pocket (i) is in between of the variants with badan outside. Variant (ii) is red shifted compared to variant (i) and variant (iii) is blue shifted. This analysis shows that there are small colour shifts from the different local environments of badan in its different conformations. The absorption maximum of prodan in water computed with TD-LC-DFTB2/OB2 (base) is 3.67 eV (see Tab. A.14 in the appendix). This is indeed very close to the absorption maxima of badan attached to the GGBP triple mutant computed with the same parameters.

Glucose is only bound in variant (ii). In the experiment, the emission from the *holo* protein is red shifted compared to the emission from the *apo* protein. This means, the computed absorption energies do not reproduce the expected shift direction. This is very probable because only one protein conformation is considered for each badan conformation. These trial QM/MM simulations were started from stable protein conformations sampled in the classical MD simulations. The protein is open in simulation (i) and belongs to the large signal in the histogram of *apo* simulation 1 (Fig. 6.19 (a)). The protein is also relatively open in simulation (iii), see Fig. 6.19 (d). In simulation (ii) in contrast the protein is in a closed state, see Fig. 6.18 (a). To get a reliable picture, much more calculations for different protein conformations need to be done. However, this is probably not worth the effort for only absorption spectra.

Tab. 6.6.: Absorption maxima as determined from histograms of vertical excitation energies computed along the QM/MM trajectories. The values are in eV.

param. \ conform.	(i)	(ii)	(iii)
OB2 (base)	3.70	3.66	3.76
from Ref. [91]	3.72	3.70	3.80

6.6. Conclusion and Outlook

In sum, in this chapter the behaviour of badan attached to GGBP in different protein conformations with and without bound glucose could be described. For this purpose, new partial charges for the glucose molecule were determined which take care of the polarisation inside the binding pocket. Interactions of the dye were detected and its environment was characterized based on classical MD simulations. It was found that the dye is more likely inside the binding pocket when there is no glucose molecule. Inside the binding pocket, the dye is still water accessible and there were no dramatic changes in the polarity detected. These findings question a previous assumption on the working mechanism of this optical glucose sensor. However, it was discovered that the dye comes close ($\approx$ 4 Å) to a tryptophane residue (Trp183) inside the binding pocket. Since tryptophane is a known quencher of the fluorescence of badan, this finding can explain the fluorescence increase upon glucose binding. It is also in line with a recent experimental study on the single mutant H152C-badan [19].

The single mutant H152-badan could be simulated as well to find the reason for the more pronounced fluorescence increase compared to the triple mutant. Similarly, a mutant with badan at another position could be studied where no signal change is observed to get more

evidence about the proposed explanation with Trp183, for instance M182C. Moreover, the classical simulations used for the analysis of the badan conformations do not exactly sample the minima found in the metadynamics simulations. Simulations starting in these minima could be performed and analysed for a more accurate picture, especially in terms of quantity.

Further investigation is needed to determine the impact of the individual factors, namely the different dye conformations, the polarity of the environment, formed hydrogen bonds and the interaction with Trp183 on the fluorescence spectrum of badan. In particular, a quantum mechanical description of the fluorophore in excited state is necessary. The effect of the polarity of the environment on the fluorescence energy of prodan is larger in the excited state and the change in electron distribution may also change the strength of the hydrogen bonds. Excited state QM/MM simulations which are also discussed in this work are a helpful tool in the study of fluorescence properties since the dynamics of the dye in excited state can be described embedded in the complex environment of protein, ligand and solvent. However, a proper QM method needs to be found which on one hand correctly describes the excited state of badan and on the other hand has a feasible computational cost. Tests with TD-LC-DFTB2 provided the same unreliable results as found for solvated prodan. Though, there is a possibility that a higher excited state of TD-LC-DFTB2 describes the emissive state. This needs further investigation.

Finally, a sensor based on a different fluorophore changing the wavelength instead of the intensity could be desirable. 4-AP as studied in the previous chapter is a promising candidate showing large solvent shifts. There is already preliminary work on a 4-AP derivative with a flexible linker [212]. This linker contains a moiety to enable covalent linkage with a cysteine and a aliphatic chain between 4-AP and this moiety to increase the range of 4-AP. In classical MD simulations it showed interesting conformations inside and outside of the binding pocket for some of the scanned linking positions which are expected to have different fluorescence energies. However, the flexibility of this dye requires sufficient sampling in the respective protein conformations. This leads to the next challenge, finding the minima on the free energy surface. An individual charge set for the system should be derived. Then, metadynamics simulations could be helpful or at least sufficiently long free MD simulations are required. Excited state QM/MM simulations should be applied for the calculation of the final fluorescence spectra for representative conformations. Excited 4-AP was succesfully simulated in aqueous environment in a QM/MM simulation with TD-LC-DFTB2 and thus this method can probably be also applied for the derivative. If a significant colour shift is observed, the sensor may be tested experimentally.

7 Exciton Transfer in the Fenna-Matthews-Olson Complex

7.1. Introduction

Photosynthesis is a process in which light energy is converted to chemical energy. In the oxygenic photosynthesis of plants in the net equation light is used to produce carbohydrates and oxygen out of carbon dioxide and water. The reaction takes place inside the chloroplasts. Light energy is collected by pigment molecules like chlorophylls and carotenoids and transfered to two coupled photosystems: photosystem I which is of Fe-S type and photosystem II which is of quinone-type. Within each photosystem a so-called *special pair* of two chlorophylls causes a charge-separation. Via several electron transfer reactions and the flow of electrons from photosystem II through the cytochrome-bf-complex to photosystem I, NADPH synthesis and a proton gradient are achieved. This proton gradient is then used for the synthesis of ATP. Oxygen is produced as a byproduct in photosystem II. The carbohydrates are built in the subsequent Calvin cycle using the previously produced ATP and NADPH. [10, 213]

Some bacteria perform photosynthesis, too. However, with the exception of most cyanobacteria, their photosynthesis variants are anaerobic. Green sulfur bacteria for instance take carbon dioxide and hydrogen sulfide as reactants, oxidize the sulfur and produce carbohydrates and water. Sunlight is collected in chlorosomes in the cytoplasm, which contain a large amount of pigment molecules like Bacteriochlorophyll (BChl), carotenoids and quinones. The chlorosome lies on a baseplate of CsmA proteins and BChls which is the link to Fenna-Matthews-Olson (FMO) complexes. The FMO complexes transfer the collected excitation energy to the reaction centers, which are located in the cytoplasmic membrane. Green sulfur bacteria possess only type I reaction centers, which consist of six subunits and several pigments. Two stereoisomers of BChl a form the special pair and absorb light at 840 nm. The energy transfer efficiency within the chlorosome, from the chlorosome to the FMO complex and within the FMO complex is almost 100 %. However, between FMO and the reaction center the transfer efficiency is only 35-75 %. [10] A scheme of the photosynthetic apparatus is depicted in Fig. 7.1. [10, 214]

In the following, the FMO complex of green sulfur bacteria is described in detail. The fascinating efficiency of the energy transfer in this complex on one hand and the early access to the crystal structure of this relatively small pigment-protein complex on the other hand led to a large amount of both experimental and theoretical studies on this system. An overview over important studies as well as remaining open questions is presented afterwards. If not stated otherwise, the FMO complex of *Chlorobaculum tepidum* is considered. The very similar FMO complex of *Prosthecochloris aestuarii* is also extensively addressed in literature.

The FMO complex is a water-soluble, trimeric pigment-protein complex. It contains eight BChl a molecules per monomer in a disordered fashion which transfer excitation energy

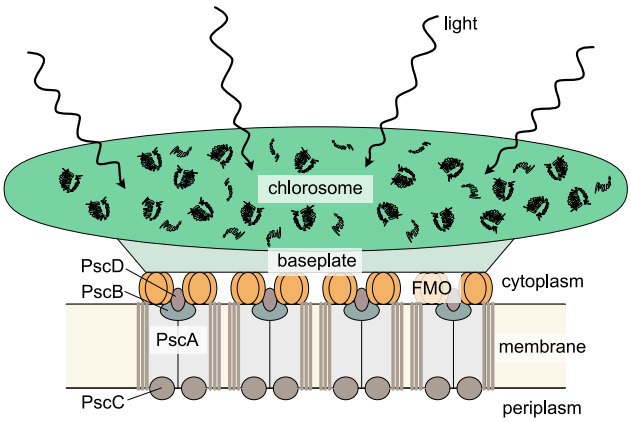


Fig. 7.1.: Location of FMO complexes between chlorosome and reaction centers. The chlorosome contains several pigment molecules collecting the sun light. The energy is then transferred via FMO complexes to the reaction centers in the cell membrane.

from the chlorosome to the reaction center. The FMO trimer and the arrangement of the seven BChl pigments enveloped by each protein monomer are shown in Fig. 7.2. The eighth BChl resides between the monomers, outside of the scaffold. The outer side of the trimer is dominated by β-sheets while around the middle and between the monomers mainly α-helices are found. The monomers are held together by salt bridges and interactions of polar residues [215].

The protein structure of FMO from *Chlorobium limicola* was resolved by X-ray crystallography almost half a century ago by Fenna and Matthews as the first chlorophyll-containing pigment-protein complex [216]. First crystal structures of the FMO proteins from *Prosthecochloris aestuarii* and *Chlorobaculum tepidum* followed 1979 and 1997, respectively. The eighth BChl a molecule, which is the only one located outside the protein scaffold between the monomers, is easily lost in the isolation and purification process and thus was discovered much later: its existence was first proposed in 2004 [217] and finally confirmed by mass-spectroscopy and HPLC analysis in 2011 [218].

The C_3 symmetry axis of the FMO complex is roughly perpendicular to the chlorosome baseplate on one side and the cytoplasm membrane with embedded reaction center on the other side [219]. The complex is oriented such that BChl 3 is the chromophore closest to the cytoplasm membrane [220]. In Ref. [221] it is proposed that two reaction centers with two FMO complexes each form a supercomplex with about 4 nm free space to neighbouring supercomplexes. In the same work, the authors estimate about 30 reaction centers and 60 FMO complexes per chlorosome. BChl 8 possibly interacts with the baseplate and both stabilizes the FMO protein at its position and acts as entrance of the excitation energy collected in the chlorosome [218]. A crystal structure of an FMO trimer connected with an almost complete reaction center was published recently [10]. It is supposed that the missing second FMO complex and PscC subunit of the reaction center were lost during sample preparation. In this crystal structure, the FMO trimer interacts with one of the two PscA subunits of the reaction center. Surprisingly, the FMO axis was found to be not perfectly perpendicular but oriented at an angle of approximately 75° to PscA. As a consequence, only two of the three BChl 8 pigments may interact with the baseplate and

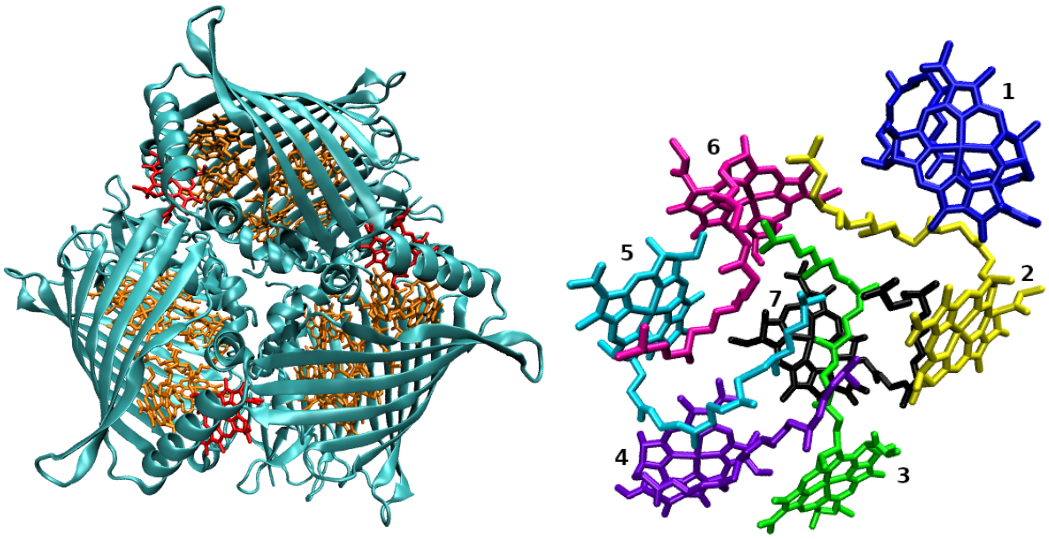


Fig. 7.2.: Left: Top view of FMO trimer[1]. The seven BChl a molecules enveloped by each monomer are orange coloured and the eigth BChl a between the monomers is shown in red, respectively. Right: Arrangement and numbering of the seven BChl a molecules enveloped by each monomer. Every molecule is shown in a distinct colour.

also only two of the BChl 3 may transfer the energy to the reaction center. The relatively large distance between the FMO pigments and the pigments in the reaction center (> 20 Å) could explain the reduced efficiency of the energy transfer compared to the previous steps. [10]

The chemical structure of BChl is shown in Fig. 7.3. The pigment consists of a Bacteriochlorin ring with 18 π-electrons and a coordinated magnesium ion in the center. Besides several smaller substituents the pigment contains a long aliphatic chain, the phytyl tail. The Mg-to-Mg distances between neighbouring pigments in the protein scaffold are in a range of approximately 11 Å (BChl 3 to BChl 4) to 18 Å (BChl 2 to BChl 3) in the crystal structure [10, 223]. BChl 7 is quite centrally located and thus shows very similar distances to all other pigments except to BChl 1, where the distance is roughly 20 Å. BChl 8 of one monomer is closer to BChl 1 of the neighbouring monomer ($\approx$ 13 Å) than to the closest BChl of the same monomer (> 23 Å). The magnesia centers in the Bacteriochlorin rings are additionally coordinated by the protein: BChl 1, 3, 4, 6 and 7 by a nitrogen from histidine and in BChl 5 by the backbone oxygen from a leucine. The magnesium atom of BChl 2 interacts with the oxygen of a water molecule.

BChl a absorbs in the areas of 400 nm (near UV), 600 nm (visible range) and 800 nm (near IR). The first absorption band around 400 nm consists of the soret bands (B bands), the other two maxima are Q bands [224]. The Q bands are named according to the direction of their transition dipole moments along the molecular axes: Q_x and Q_y (see caption Fig. 7.3) [224]. The excitation energy transfer in the FMO complex occurs via the Q_y transition (around 800 nm), which is focused on in the following. In the experimental absorption spectrum of the FMO complex of *Chlorobaculum tepidum* at 77 K strong ab-

[1] picture was made based on structure obtained from Sayan Maity [222]

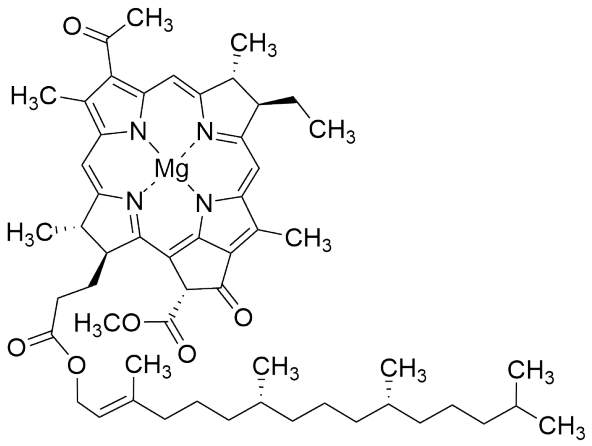


Fig. 7.3.: Chemical structure of BChl a. The molecular z-axis is perpendicular to the molecular plane, the x-axis runs on the line through the nitrogen on top right, magnesium and nitrogen bottom left and the y-axis through nitrogen top left, magnesium and nitrogen bottom right. The transitions in the absorption spectrum are indexed according to the dipole moment orientation. For the energy transfer in FMO, the Q_y transition is important with the dipole moment along the molecular y-axis.

sorption peaks are found at 806 nm and 815 nm, with the latter being a little weaker than the former [223]. These peaks are part of a broader band with an additional smaller peak at 825 nm [223, 225] and a shoulder on its blue side. At room temperature the absorption spectrum is almost featureless [225]. Measured in solution (triethylamine), BChl a absorbs at 771-775 nm depending on the temperature and ligation of magnesium [226]. Thus, the FMO protein causes a red-shift on the Q_y absorption band [225].

The excitation energy transfer mechanism through the FMO complex is explained by a funnel potential. According to a simple picture, the transfered energy would move down from pigment to pigment like descending a ladder. Individual slightly different excitation energies of the pigments forming the ladder arise from their individual surrounding determined by the protein. In this case geometrical as well as electrostatic effects play a role. These individual excitation energies of the pigments are called *site energies*. However, the pigments excitonically couple with each other. This causes a splitting into new energy levels, called exciton energies. They can be computed by diagonalizing the matrix with the site energies as the diagonal and the couplings as the off-diagonal elements (see chapter 2). In this picture, the excitation energy is transfered delocalized over several pigments. The contributions of the individual pigments to the different exciton levels shed light on the spatial pathway of the transfer. From 2D spectroscopic experiments it is known, that the excitons do not simply follow one ladder but there are distinctive pathways ending at BChl 3 [225, 227].

In principle, the excitonic energy levels are accessible from experimental absorption spectra. However, only three distinct peaks are observed at 77 K (*vide supra*) instead of seven or eight due to homogeneous and inhomogeneous broadening. Thus, higher-dimensional spectroscopic techniques are necessary to extract the excitonic energies. A set of fully ex-

perimental excitonic energies were determined by Hayes and Engel in 2011 [26] by means of 3D-third-order electronic spectroscopy, which are presented in Tab. 7.1.

Tab. 7.1.: Experimental exciton energies from Ref. [26]

Exciton	Energy ± SD in cm^{-1} [26]	converted to eV
1	12,121 ± 15	1.503
2	12,274 ± 20	1.522
3	12,350 ± 16	1.531
4	12,415 ± 12	1.539
5	12,454 ± 17	1.544
6	12,520 ± 19	1.552
7	12,606 ± 25	1.563

Knowledge about the individual site energies and couplings generating the exciton energies provides insight in the mechanism behind the efficient energy transfer. In Ref. [26] the site energies were determined from the experimental excitonic energies and couplings computed in the point-dipole approximation on the crystal structure. This required an iterative procedure with an initial guess for the site energies. A genetic algorithm was employed to find the optimal initial guess leading to the best fit of the linear absorption spectrum.

Spectral fits with the site energies as parameters to be optimized were also performed in a lot of early studies on FMO complexes [228–234]. An alternative approach is the determination of site energies by the calculation of energy shifts based on structural data. The site energy of a pigment m is split into a reference value E_0 and the shift ΔE_m due to the specific environment of pigment m: $E_m = E_0 + \Delta E_m$ [25, 234–236]. Depending on the method used for computation of ΔE_m, site energies close to the ones from a fit of optical spectra can be achieved [25].

Another ansatz to determine the site energies is the direct calculation of the excitation energies with a semi-empirical or quantum-mechanical method considering the individual pigment environment. This is done in one early study with INDO/S [237], where an unrealistically large range from the lowest to the largest site energy was obtained. In later studies, more reliable results were achieved with respect to the energy range, but the predicted energy ladders of the site energies differ in the various studies on the FMO complexes of *Prosthecochloris aestuarii* [238–240] and *Chlorobaculum tepidum* [81,241,242]. These differences may be traced back to differences in the methods used for the excitation energy calculation, the amount of included environment, the consideration or neglect of polarization effects, sampling of structures and the applied force field or use of only the crystal structure, inclusion or neglect of the eighth pigment and if a monomer or trimer is studied.

Specific protein effects influencing the site energies and thus the energy transfer in FMO were analysed in Ref. [234]. It was found that charged amino acids and ligands of the magnesium have a significant impact on the site energies. In Ref. [235], additionally an even more important factor was determined: the electric field of the backbone of two

α-helices shifts the site energy of BChl 3 to a lower energy, making it the energy sink. Furthermore, hydrogen bonds between protein and the keto or acetyl groups of the BChls were identified as contributors to the site energies. Geometrical effects and polarization effects were found to be less important [25, 235].

A further aspect of the energy transfer debated in literature is long-lived quantum coherence. This phenomenon was observed in 2D electronic spectra at 77 K and ascribed to superpositions of excitonic states [243]. The transfer mechanism proposed is in contrast to the picture of hopping excitons. The quantum coherence was reported to survive even at physiological temperature, protected by the protein [244]. However, in a later study the long-lived quantum beats persisiting for picoseconds were found to origin mainly from ground-state but also excited-state vibronic coherence [245]. The effect of these oscillations on the energy transfer is not finally clarified. In a recent computational study, a modest enhancement of the excitation energy transfer efficiency due to vibronic modes is reported [246].

7.2. Motivation

A simulation of the exciton transfer in the FMO complex can provide new insight about the transfer mechanism and the reasons for the remarkable efficiency of the energy transfer. This work includes important steps towards a direct simulation of the exciton propagation in a semi-classical approach. In such a simulation, the site energies and couplings of the BChls will be needed as Hamiltonian elements for the exciton states. TD-LC-DFTB2 is a promising method to compute these quantities due to its efficiency and the incorporated long-range correction which is expected to give a reliable description of the excited states of the BChls.

The main question of the first part is the reliability of the site energies, couplings and resulting excitonic energies predicted by TD-LC-DFTB2 compared to previous approaches. This is mainly discussed already in [81] based on site energies and couplings computed on the structures of one nanosecond from a classical MD trajectory. It was found in Ref. [81] that the site energies of the BChls as predicted from TD-LC-DFTB2 are in a reasonable range compared to the experiment but the couplings and the resulting excitonic energies are overestimated. Here, these quantities are recalculated using a new parameter set. In addition, the impacts of the amount of sampling and the amount of environment considered in the calulations is discussed in detail.

The goal of the second part is to obtain reliable site energies and couplings even faster than with a TD-LC-DFTB2 calculation. This is desirable to save computational cost during the exciton transfer simulation. The method of choice is machine learning. Neural network models, which predict site energies and couplings from the coordinates of the pigments, are trained based on TD-LC-DFTB2 results. The main point adressed is the quality of the resulting models. It is checked how the outcome depends on the training set size and some details of the provided input to get a well-functioning model for the exciton propagation.

7.3. Computational Details

One MD simulation of the FMO complex from *Chlorobaculum tepidum* (PDB-ID:3ENI [223]) was already performed for my Master thesis [247] but is further evaluated here. Only one monomer was considered with the seven BChl *a* embedded in the protein scaffold.

The system was obtained from the group of Ulrich Kleinekathöfer, already set up for simulation [248]. The CHARMM27 force field [249] and TIP3P water were used. The preparation of the productive simulation by means of the GROMACS program package was carried out in three steps. First an energy minimization was performed using the steepest descent algorithm until a maximum force of less than 100 kJ mol^{-1} nm^{-1} was achieved. Then the system was equilibrated at 300 K applying the Nosé-Hoover thermostat [162, 163] in a 1 ns long NVT simulation. Bonds to hydrogens were constrained and position restraints were set for the heavy atoms of the protein and the pigments with a force constant of 1000 kJ mol^{-1} nm^{-1} in each dimension. The time step was 1 fs, periodic boundary conditions were applied and long-range electrostatic interactions were included by means of the PME method. Finally, a 5 ns long NPT equilibration under the same conditions followed additionally employing the Parrinello-Rahman barostat [164] for controlling the pressure with a reference value of 1.013 bar. For the final productive simulation the contraints and position restraints were removed and the simulation with a length of 100 ns was carried out using the GROMACS program package (version 2016.3) [201–207].

A second productive simulation was performed starting from the equilibrated structure. In this simulation, which has a length of 50 ns, all bonds involving hydrogen were kept contrained. Moreover, this simulation was repeated saving the output trajectory with higher precision. From this simulation, coordinates with four decimal digits in Ångström instead of only two were obtained.

Excitation energies of the BChls were computed on snapshots of the trajectory with TD-LC-DFTB2 as implemented in DFTB+ using a set of parameters optimized for excitation energies [3, 104].[2] The SCC tolerance was set to 1×10^{-10} electrons and the screening threshold was 1×10^{-6}. The excitation energies were computed for isolated BChl and including the environment as point charges. The phytyl tail of the BChl was cut after the carboxylate ester and saturated with a methyl moiety (Fig. 7.4) to save computational cost. The remaining atoms of the phytyl tail were considered as MM environment, respectively. Coulomb couplings were computed from transition charges obtained from TD-LC-DFTB2 between pairs of BChl along the trajectory [81]. The same parameters were used as for the calculation of the excitation energies. The script applied was written by Julian J. Kranz. In these calculations, the phytyl tail of the BChl was truncated as well.

Neural networks [175] were trained to predict the excitation energies and the couplings between the chromophores. The script used was written by Mila Krämer. The coordinates of the BChls sampled in the simulations served as input and the site energies and couplings obtained with TD-LC-DFTB2 were the targets. The coordinates and the energies were converted into atomic units, respectively. As described above, the phytyl tail of the BChl was cut in the TD-LC-DFTB2 calculations.

Site energies and couplings were computed in vacuum for 50 000 snapshots of the simulation without constraints and the simulation with the contrained hydrogens, respectively. The snapshots were equally spaced by one picosecond. Since each snapshot provides results for seven BChls, in total 350 000 pairs of coordinates and resulting site energies were

[2] Note, that in Ref. [81] the parameters from Ref. [3] were used and here the modified parameters as described in Ref. [104]. These parameters from Ref. [104] were also used for the benchmark of the TD-LC-DFTB2 gradients [4] to compute the excitation energies in the TD-LC-DFTB2/comb. variant.

obtained with each simulation. Only the strongest couplings were taken into account, which are the ones between BChls 1-2, 2-3, 3-4, 4-5, 5-6, 6-7 and 4-7. Thus, also 350 000 coordinate-coupling pairs were gained from the snapshots of each of these two simulations. The snapsots with more precise coordinates from the third simulation were used to compute further 350 000 excitation energies, but no additional couplings.

The signs of the couplings needed to be aligned. This was done by the computation of the scalar product of the vectors containing the transition charges of each snapshot with a reference snapshot. If the scalar product for two snapshots had a different sign, then the calculated coupling was multiplied by -1. The script used was written by Philipp M. Dohmen.

Input data files for the neural network were prepared separately for site energies and couplings and separately for the three different simulations. The results from the unconstrained simulation were prepared with the coordinates of all 85 atoms from BChl without phytyl tail. The input data from the simulations with constrained hydrogens contained only the coordinates of the heavy atoms of the BChls of which there are 47 (Fig. 7.4).

The hyperparameters describing the neural network were determined within a specified range by a hyperband algorithm [250] with a factor of 30. The maximum number of epochs for the training of a single model was set to 1000. If not mentioned otherwise, the depth was allowed to be either three, four or five and the number of neurons per layer was allowed to take a value in the range of 20 to 1000 in steps of 5. Further, a choice was made between the two regularization functions L_1 and L_2. For the start step size either 1×10^{-3}, 5×10^{-4} or 1×10^{-4} could be selected. Overfitting is prevented by early stopping with a patience of 30. From the amount of training data specified in the input (*ntrain*), 90 % is used for training and 10 % for validation. The final model evaluation is done on a separate test data set which contains 50 000 input/output pairs. The batch size was 64.

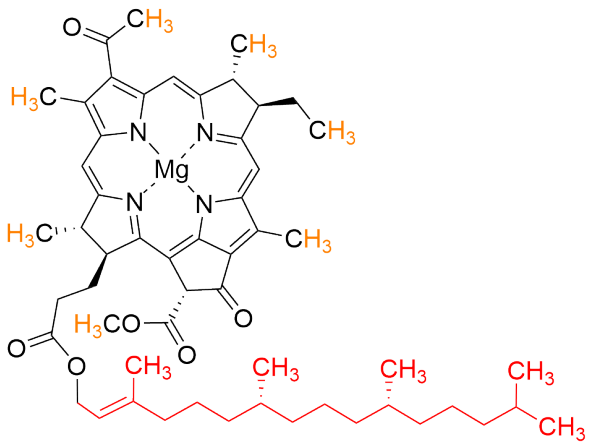


Fig. 7.4.: Chemical structure of Bacteriochlorophyll a. The phytyl tail (red) is negelcted in all QM calculations. The hydrogen atoms in the remaining molecule (orange) are neglected in some of the machine learning inputs.

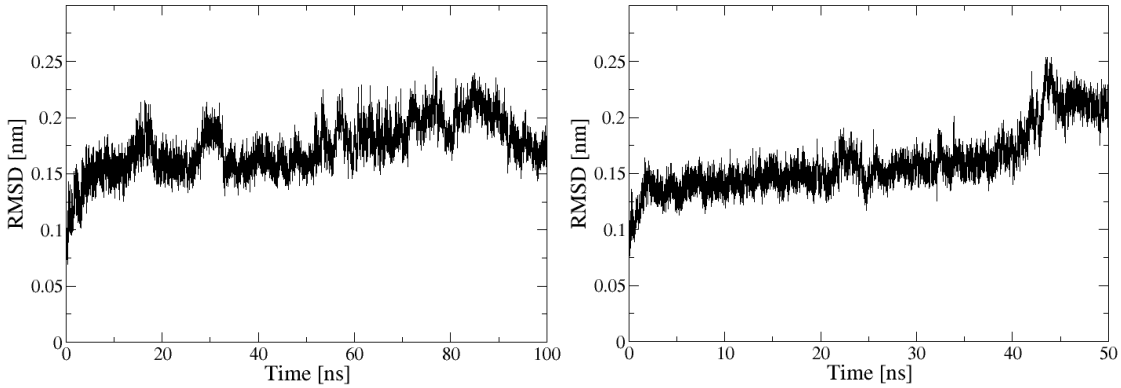
7.4. Results and Discussion

7.4.1. Site Energies and Couplings Predicted by TD-LC-DFTB2

The root mean square deviation of the backbone of the FMO protein plotted versus the simulation time (Fig. 7.5 a) shows that the complex remains relatively stable during the simulation but small structural changes appear. In Ref. [81], the 11th nanosecond was studied, which lies in a very stable period of the simulation. In the following, it is examined if there are significant differences in the site energies of other parts of the trajectory. For completeness, the RMSD of the simulation with the constraints on the bonds involving hydrogen is shown in Fig. 7.5, too. It also shows a quite stable protein along the simulation time with a smaller structural change after 40 ns.

The first and last nanoseconds of the simulation as well as the nanoseconds after 10, 15, 50 and 85 ns simulation time were selected for excitation energy calculation to detect a possible correlation to the RMSD. The excitation energies were determined from Gaussian fits to histograms of the respective nanoseconds with 1000 snapshots, respectively. The maxima of the Gaussians are plotted in Fig. 7.6 for the seven BChls with ± the standard deviation as error bars. The protein was centered in the simulation box and the complete box was included in the excitation energy calculation via point charges.

It is evident that there are only very small differences in the excitation energies of the seven pigments and their distributions overlap to a large extent. A comparison of the energy ladders obtained at different parts of the simulation reveals slight differences between the beginning of the simulation and later trajectory parts. After 10 ns (gray curve in Fig. 7.6) the excitation energies are very similar to the very beginning (black curve in Fig. 7.6), but after 15 ns the excitation energy of BChl 2 is increased by more than 10 meV (red curve in Fig. 7.6). In the time between the 16th and 51st ns also the excitation energy of BChl 3 increases in the order of 10-20 meV (black, gray and red curves vs. blue, orange and green curves in Fig. 7.6). However, even smaller differences between for instance the black and gray curve result in a different energy ladder. The range between the largest and smallest



(a) no constraints **(b)** bonds with hydrogen constrained

Fig. 7.5.: Root mean square deviation of the FMO backbone with respect to the starting structure of the simulation in nm. The RMSD values indicate small structural changes during the simulation.

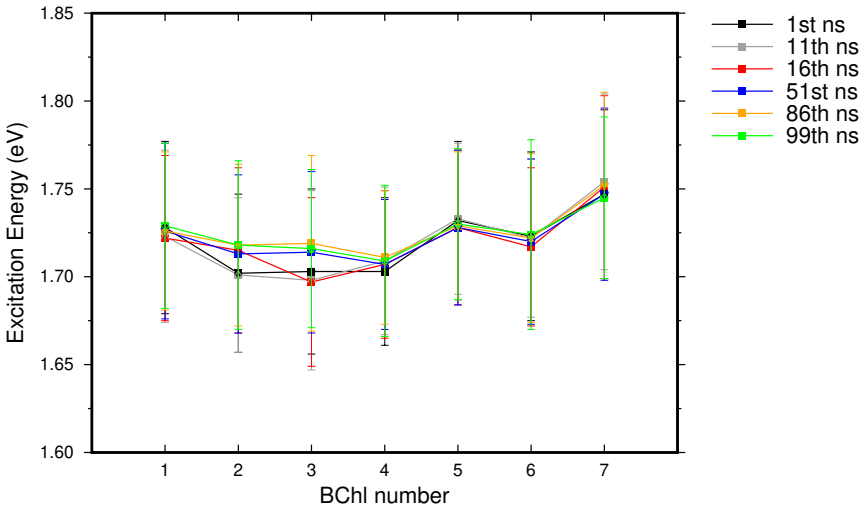


Fig. 7.6.: Excitation energies computed with TD-LC-DFTB2 along different sections of the trajectory considering the whole simulation box as point charges. The excitation energies of the pigments differ only very little. A dependency of the excitation energies of BChl 2 and 3 on the selected trajectory piece is observed.

excitation energies amounts to 0.045 eV at the first nanosecond, it is maximal (for the here computed trajectory parts) after 10 ns with 0.056 eV (in agreement with [81]) and minimal at the last nanosecond of the simulation with 0.036 eV. The site energy range obtained from a fit to optical spectra reported in Ref. [234] amounts to 0.054 eV and in Ref. [26] to 0.057 eV. Thus, the TD-LC-DFTB2 results are in agreement with this previous approach. However, polarization effects were not taken into account. In the computational approach of Ref. [242], with ZINDO/S-CIS as method for excitation energy calculation and using polarized protein-specific charges, the obtained site energy range is smaller with 0.027 eV.

In sum, the excitation energies of the different BChls are very similar and build no distinct energy ladder. During the simulation, the excitation energies of BChl 2 and 3 change more than the excitation energies of the other pigments. However, this cannot be explained by the RMSD which is similar for the 11th and 51st nanosecond and the 16th and 86th nanosecond. However, it is possible that the structural change responsible for the change in the excitation energy is too small to be visible in the RMSD. Further, it should be noted that the same RMSD value not necessarily corresponds to the same structure but possibly to a different structure which only deviates to a similar extent from the reference structure. The range of the obtained site energies is in agreement with previous studies based on fits to optical spectra.

It is possible that artifacts arise due to the fact, that the environment is not equally distributed around the respective pigment whose site energy is computed because the whole protein but not the respective individual pigments were centered in the simulation box. Moreover, there is a hard cut-off of the inlcuded point charges at the edge of the

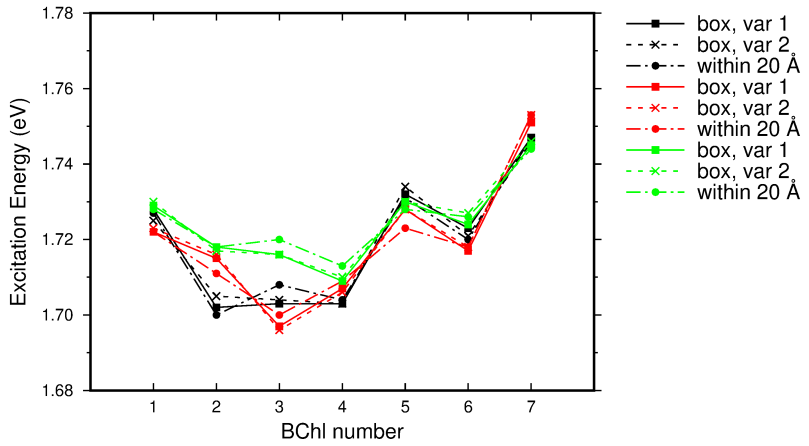


Fig. 7.7.: Excitation energies computed with TD-LC-DFTB2 using different approaches to define the point charge environment. As in Fig. 7.6, the black curve shows the results of the first nanosecond, the red curve corresponds to the 16th nanosecond and the green curve to the last nanosecond. *box, var 1* means the whole simulation box is included with the protein in the center. It is equivalent to the results in Fig. 7.6. *box, var 2* means that the whole box is included but the respective pigment is in the center and PBC wrapping is applied. *within 20 Å* means that the point charges in a sphere with a radius of 20 Å around the magnesium of the respective pigment are included. In this case, a smooth cut-off is applied. The y-axis shows a very small range of 0.1 eV to make the small differences between the approaches visible. For clarity, no error bars are shown. The standard deviations are of same order in all approaches.

simulation box. The sensitivity of the excitation energies to these issues is estimated by comparing the results of two different approaches: (i) instead of the protein, the respective pigment is located in the center of the box and PBC wrapping is applied. This means that the part of the environment, which would be outside the box if the pigment is the center, is cut on this side and again added on the opposite side of the box. The Tcl script used for this approach was written by Sayan Maity. (ii) The environment is limited within 20 Å of the magnesium of the respective pigment. In this approach, the charges are smoothed between a distance of 18 Å and 20 Å with a linear function to avoid the hard cut-off. The python scripts used for this approach were written by Beatrix M. Bold.

Fig. 7.7 shows the results of these two approaches compared to the previous results for the first (black), 16th (red) and last (green) nanosecond. The two approaches including the complete box and a hard cut-off provide very similar results. If only the environment within 20 Å is considered, the maximal difference to the initial approach amounts to 5 meV. Hence, the same picture is obtained with all of these three approaches and there is no severe artefact.

The effect of the point charge environment on the excitation energies is studied for the 1st, 11th and 16th nanoseconds of the trajectory. Fig. 7.8 shows the shift between the site energies computed with consideration of the environment and the excitation energies of the pigments in vacuum. Evidently, the range of the excitation energies in vacuum is

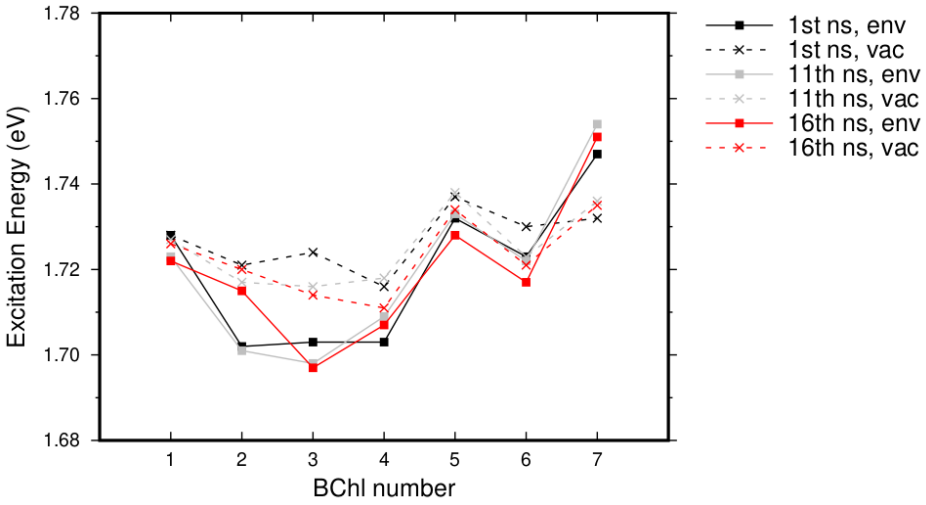


Fig. 7.8.: Excitation energies computed with TD-LC-DFTB2 with consideration of the environment as point charges and in vacuum. The range of the excitation energies in vacuum is decreased compared to the range with environment. The y-axis is zoomed to a range of only 0.1 eV to make the small differences visible.

roughly only half of the range which is obtained with environment. The absolute environmental shifts are small (max.: 0.021 eV for BChl 3 in the 1st ns). It is remarkable that the environmental shifts for pigments 2, 3 and 7 are relatively large compared to the very small shifts for pigments 1, 4, 5 and 6. However, the environmental shift for pigment 2 decreases between the 11th and 16th nanosecond of the simulation.

In contrast, the environment has a very small impact on the Coulomb couplings between the pigments. Fig. 7.9 shows as example the couplings between BChl 1 and 2 obtained with and without consideration of the environment along the 11th nanosecond of the classical trajectory. A small and relatively constant shift of less than 0.005 eV is observed due to the environment. Tab. 7.2 reports the mean coupling values with and without environment along this trajectory section for all important pigment pairs. Very similar values are presented in Ref. [81], obtained for the same structures but with different TD-LC-DFTB2 parameters.

As the environment has a negligible effect on the couplings, the sampling effect is examined in vacuum. The values of the first 50 ns are shown in Fig. 7.10. The couplings remain stable, except the ones between BChls 1 and 2 which show a larger change after around 5 ns and generally fluctuate more than the couplings between the other pigment pairs. The couplings between pigments 6 and 7 show larger fluctuations as well.

In Fig. 7.11, the strongest couplings determined in this work are compared to the results of two publications. The Coulomb couplings computed with TD-LC-DFTB2 are larger than the couplings reported in the literature, by a factor of three to four on average. The trend is similar but the coupling between BChls 6 and 7 is clearly too large predicted in relation to the other couplings. In Ref. [242], the couplings were determined simply within the point-dipole approximation using a screening factor of 0.8, which should account for the protein screening. In Ref. [240], Coulomb couplings were computed with the transition densities from TD-DFT/B3LYP and an additional term accounting for the polarizable en-

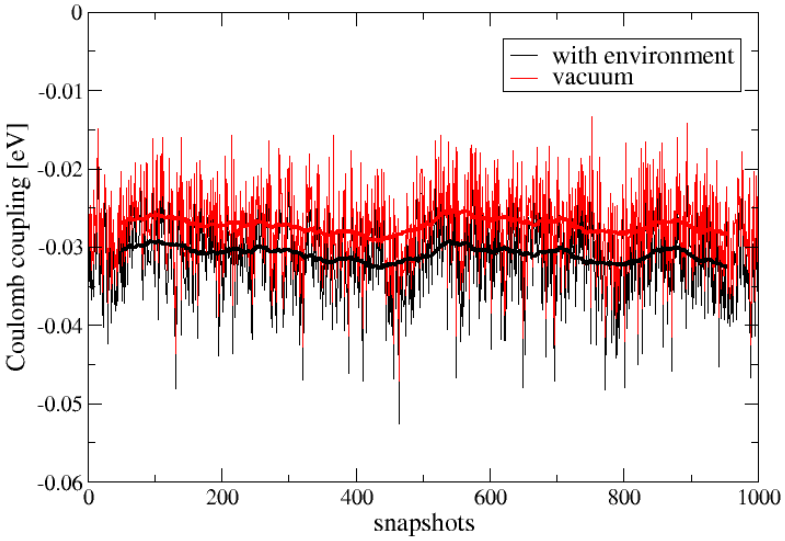


Fig. 7.9.: Coulomb couplings computed from TD-LC-DFTB2 transition charges obtained with and without consideration of the environment as point charges for pigments 1 and 2. The 11th nanosecond of the trajectory was analyzed taking a snapshot every picosecond. The thick lines show the running averages over 100 data points. The environment shifts the couplings by a small and relatively constant value.

vironment. In the latter study, also vacuum coupling values are given. Note, however, that this study was done on the FMO complex of *P. aestuarii*. Here, Coulomb couplings were computed taking the environmental effect on the transition charges into account by including point charges in their calculation. It becomes clear in Fig. 7.11 that the environmental effect is small in all cases.

The resulting excitonic energies were computed for the 11th nanosecond of the simulation. For each snapshot the Hamilton matrix was constructed with the site energies of the pigments on the diagonal and the absolute values of the couplings on the off-diagonal. The couplings between all possible pigment pairs were accounted for. The diagonalisation of the Hamilton matrices yielded seven excitonic energies for each snapshot. These excitonic energies were each averaged along the trajectory. Tab. 7.3 presents the results obtained from a Gaussian fit of the histogram of the excitonic energies. The range from the lowest to the highest excitonic energy amounts to 0.187 eV. This is clearly larger than the range found in the experiment (0.060 eV, Tab. 7.1). The computational studies of refs. [242] and [240] included the eighth BChl and are thus not suitable for a comparison of the range of the excitonic energy levels. One reason for the overestimation found here and in Ref. [81] with TD-LC-DFTB2 are surely the overestimated couplings. Since their trend is mostly correct, a possible solution could be a scaling of the TD-LC-DFTB2 couplings to improve the excitonic energies.

The excitonic energies were additionally determined taking only the strongest couplings into account, i. e. between neighbouring pigments and BChls 4 and 7. As expected, the results are very similar (Tab. 7.3).

Tab. 7.2.: Arithmetic means of the Coulomb couplings obtained with and without consideration of the environment from 1000 snapshots. The absolute values in eV (cm^{-1}) are shown, respectively.

pigment pair	with environment	without environment
1-2	0.031 (250)	0.027 (218)
2-3	0.015 (121)	0.014 (113)
3-4	0.021 (169)	0.020 (161)
4-5	0.038 (306)	0.038 (306)
5-6	0.028 (226)	0.026 (210)
6-7	0.026 (210)	0.028 (226)
4-7	0.016 (129)	0.018 (145)

Tab. 7.3.: Excitonic energies obtained as mean eigenvalues from the Hamiltonians consisting of the site energies and couplings computed with TD-LC-DFTB2. The values obtained with consideration of the strongest couplings only are given in parentheses. The 11th ns of the simulation was used.

Eigenvalue	excitonic energy [eV]
1	1.645 (1.644)
2	1.675 (1.674)
3	1.701 (1.702)
4	1.727 (1.730)
5	1.754 (1.756)
6	1.784 (1.785)
7	1.832 (1.823)

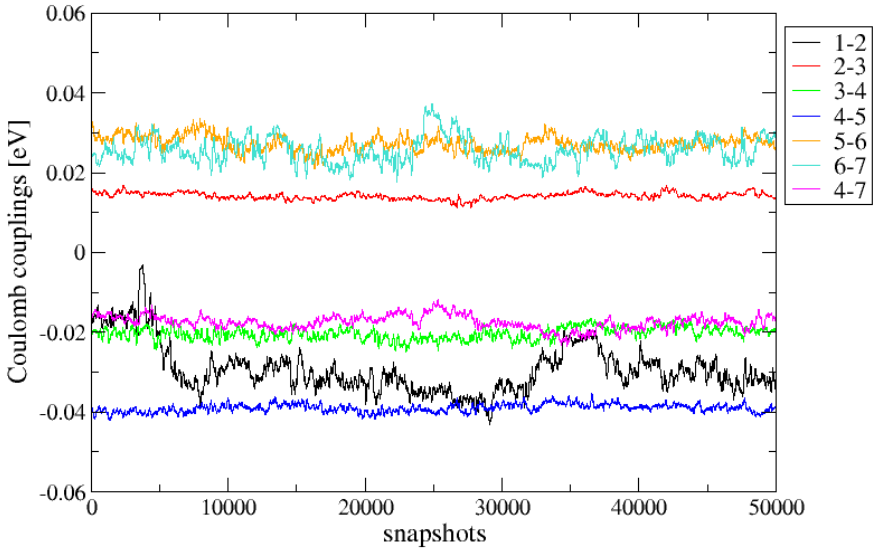


Fig. 7.10.: Coulomb couplings in vacuum obtained with TD-LC-DFTB2 along the first 50 ns of the trajectory for all important pigment pairs. The running averages over 100 data points are shown. The couplings are stable for the 50 ns except the coupling between BChls 1 and 2.

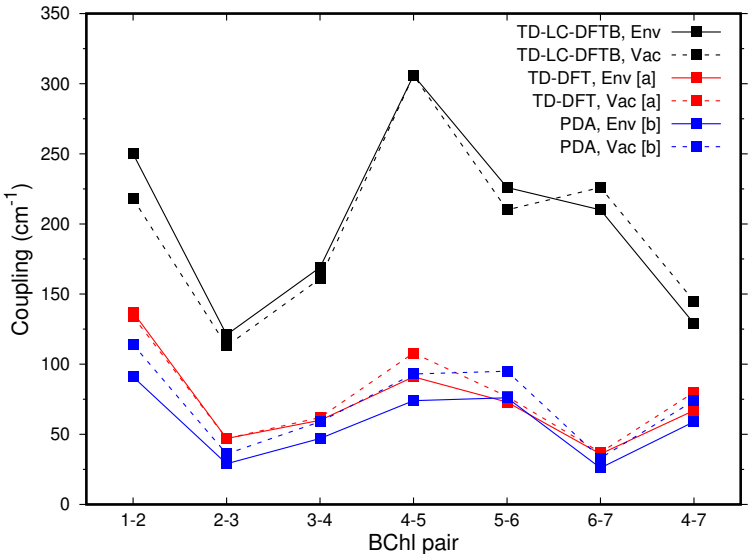


Fig. 7.11.: Absolute values of Coulomb couplings computed with TD-LC-DFTB2 compared to literature. [a] Jurinovich et al. [240], [b] Jia et al. [242]. The vacuum values within the point-dipole approximation (PDA) were obtained by dividing the values from Ref. [242] with the screening factor of 0.8 applied there.

7.4.2. Site Energies and Couplings Predicted from a Neural Network

A requirement for a simulation of exciton movement through the FMO complex is the very fast evaluation of site energies and pigment coupling. Although TD-LC-DFTB2 already is a very fast method, a neural network can be faster. The efficiency of TD-LC-DFTB2 is thus used for the creation of a large amount of training data for the neural network which is finally intended to predict the site energies and couplings during the exciton transfer simulation.

First, a few test runs were performed to receive an impression of the impact of the kind of input file (with or without hydrogen coordinates), the amount of training data and the hyperparameters. The hyperparameter search was performed in each test run, which led to appropriate final hyperparameters for the respective model training. These differ between the runs but the limits for the search were always the same, as described in section 7.3. Except from the hyperparameters, the final model depends on the random initial weights and the specific training, validation and test subsets which contain data randomly assigned out of the whole input data pool. Thus, small fluctuations in the performance for the same input pool and training conditions are normal. To get an idea of the size of these fluctuations, few runs were repeated several times.

In a first step, the dependence of the training outcomes for excitation energies on including or omitting the hydrogen coordinates of BChl a was examined. The data for the first case was obtained from the simulation without constraints and the data for the second case from the simulation keeping bonds involving hydrogen constrained. Fig. 7.12 shows learning curves for both cases together in one graph. The mean absolute errors and R^2 scores evaluated with the final model on the test set are plotted versus the amount of data used for training in Fig. 7.12. As it is expected from a systematic learning process, the prediction quality increases with the training set size in both cases. Evidently, the performance of the model dramatically improves between 1 000 and 10 000 examples in the training set. Then, additional data sligthly improves the outcome. The information about the hydrogen positions seems to be of minor importance since the curves are close to each other and cross several times. The curves are not smooth due to the typical fluctuations explained above.

Tables 7.4 and 7.5 report the R^2 metric, the MAE and the loss values for three identical runs with ntrain=10 000 and two identical runs with ntrain=300 000, respectively. It is visible that the machine learns better in these cases if the hydrogen coordinates are ignored. A reason could be that their impact on the excitation energies is very small and thus the information about their position could also hinder the learning by making the input unnecessarily complicated. This finding is pleasant from a practical point of view since the neglect of the hydrogens reduces the memory and computational demand.

To be sure that the constraints on the bond lengths with hydrogens do not narrow the excitation energy distribution, histograms and Gaussian fits of the excitation energies from both simulations were compared (Fig. 7.13). The maximum of the excitation energies from the simulation with constraints is only blue shifted by 0.002 eV and also the difference in the width of the distributions is negligibly small.

Fig. 7.14 shows scatter plots for the two extreme examples: the training set size of 900 and the training set size of 270 000. The test data used was not seen by the model before. Clearly, the fewer training data is provided, the larger the deviation of predicted values

Tab. 7.4.: Comparison of model performance with and without consideration of the hydrogen positions to predict the site energies. Amount of training data: 9 000, validation data: 1 000, test data: 50 000.

Hydrogens	run	R^2	MAE [eV]	loss
yes	1	0.704	0.017	0.383
	2	0.612	0.019	0.409
	3	0.661	0.018	0.371
no	1	0.786	0.014	0.455
	2	0.805	0.013	0.418
	3	0.804	0.014	0.431

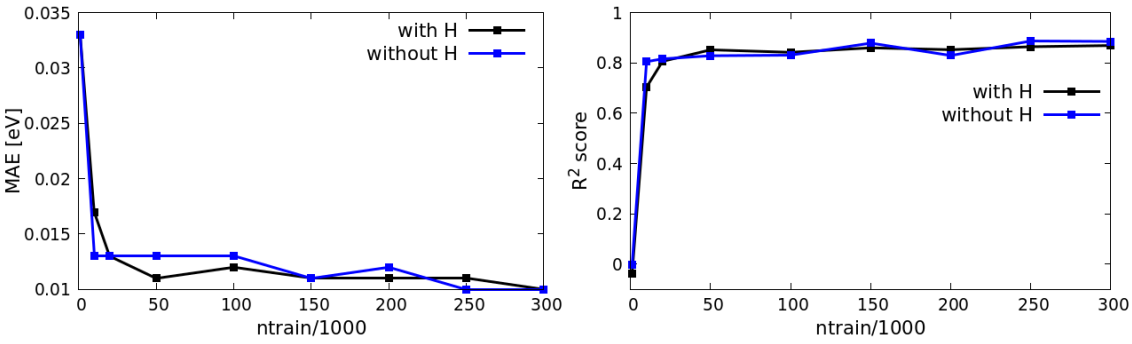
Tab. 7.5.: Comparison of model performance with and without consideration of the hydrogen positions to predict the site energies. Amount of training data: 270 000, validation data: 30 000, test data: 50 000.

Hydrogens	run	R^2	MAE [eV]	loss
yes	1	0.869	0.011	0.170
	2	0.870	0.010	0.161
no	1	0.885	0.010	0.149
	2	0.882	0.011	0.142

from the reference line. In particular, the predicted values lie in a smaller range which causes the apparent rotation of the output data compared to the reference line. Further, the results from the machine trained with less data are broader distributed.

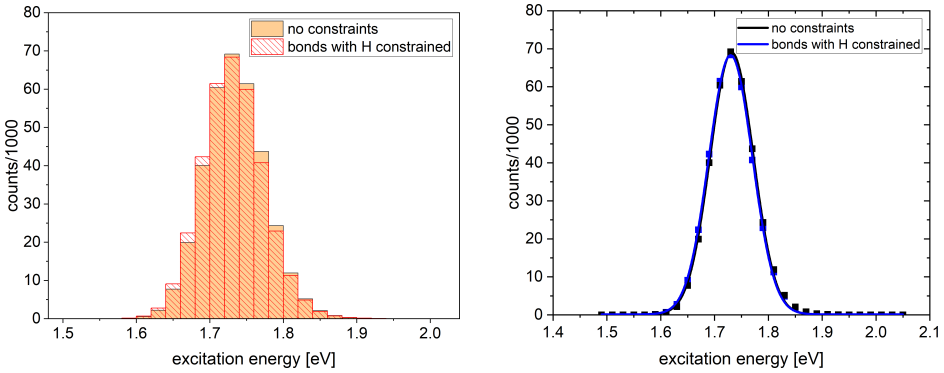
The achieved R^2 scores are sufficient but it is desirable to get them as close to one as possible. One reason for a poor training outcome can be the quality of the input data. Thus, it was tested if a representation of the geometries by more precise coordinates (rounded to 1×10^{-4} Å instead of 1×10^{-2} Å in the simulation output, five digits after conversion to bohr) can improve the R^2 score. Trainings with training set sizes of 9 000 and 270 000 resulted in R^2 scores of 0.803 and 0.883. This is comparable to the results before. Hence, it is concluded that the coordinate precision is not the limiting factor for the training success.

Except the R^2 score, the size of the MAE in relation to the width of the excitation energy distribution is an important criterium to estimate the worth of the model. The MAE obtained with the model trained with 270 000 examples is ≈ 0.011 eV. The standard deviation of the excitation energy distributions amounts to 0.04 eV, which is thus four times the MAE. The relative error determined from the MAE compared to the maximum of the Gauss fit of the excitation energy histogram, 1.730 eV, is less than 1 %. In sum, the results are acceptable.



(a) Mean absolute error **(b)** R^2 score

Fig. 7.12.: Learning curves obtained with different training set sizes with and without the hydrogen positions in the input. 10 % of ntrain was used for validation, respectively.



(a) Histograms **(b)** Gaussian fits of histograms

Fig. 7.13.: Excitation energies computed with TD-LC-DFTB2 on the sampled geometries from the simulations without constraints and with constraints on bonds involving hydrogen. The differences between the distributions are vanishing small. The plots were generated with origin [167].

The couplings between the BChl a chromophores were learned faster by the neural network, i. e. less training data was needed to achieve high R^2 scores. The learning curves are shown in Fig. 7.15 and detailed values are reported in Tab. 7.6. Already with only 900 training examples, R^2 scores larger than 0.97 are achieved. For comparison, the R^2 scores of the site energies with 900 training examples were around zero and with 270 000 training examples still below 0.9. Increasing the training set size of the couplings to 270 000 results in a small improvement of the R^2 score by roughly 0.005. The mean absolute error changes less than 1 meV from a training set size of 900 to 270 000. With its 3 meV it is a fifth of the average value for the coupling between BChls 2 and 3, which show the weakest coupling of all considered pairs.

The learning curves show a better learning with the hydrogens. However, the difference is very small and thus it is possbile to ignore the hydrogens in the training of the couplings, too.

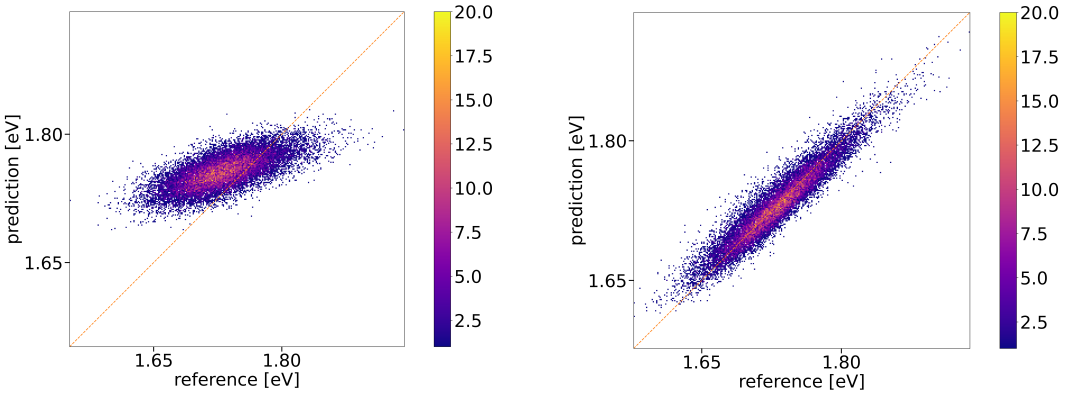


Fig. 7.14.: Scatter plots from 20 000 test data points visualizing the correlation between predicted and reference excitation energies for different training set sizes. Hydrogen positions were not considered. The colours indicate the density of the points. Left: ntrain=1 000, R^2= -0.001; right: ntrain=300 000, R^2=0.885.

Scatter plots for training set sizes of 900 and 270 000 are shown in Fig. 7.16. Separate regions with dots are seen which correspond to the coupling values from one or several of the seven distinct BChl pairs. The couplings in the training set from the simulation with constraints on bonds involving hydrogen are shown in the appendix (Fig. A.2). They are very similar as the ones shown in Fig. 7.10 except that the couplings of BChls 1 and 2 stay relatively close to the couplings of BChl 4 and 7, i. e. they take less values below -0.03 eV. Hence, the most negative cluster around -0.04 eV arises from the BChl pair 4-5 only. The other, larger cluster in the negative area origins from the pairs 3-4, 4-7 and 1-2. On the positive corner, the small cluster shows couplings from the BChl pair 2-3. The remaining cluster arises from the pairs 5-6 and 6-7.

The two bigger clusters are obviously broad because they consist of data points from several BChl pairs and the couplings of these pairs itself show some fluctuation on their averages along the trajectory. These clusters improve with more training data mainly by rotating on the reference line. Interestingly, the couplings from pair 2-3 appear as small signal well aligned with the reference line while the couplings from pair 4-5 are broadly distributed. The number of test points in each cluster, however, is comparable. This remarkable difference between the clusters of BChls 2-3 and 4-5 is not clear from Figs. A.2 and 7.10, since the average couplings of both are very stable. Indeed, the fluctuations around these averages are very different between these pairs, as shown in Fig. A.3. The couplings between BChl pair 2-3 vary less than 0.01 eV while the other coupling pairs including BChl 4-5 fluctuate in a range of roughly 0.02 eV.

The MAE in relation to the width of the clusters is not extremely different in the plot for ntrain=300 000. However, the cluster of pair 2-3 is evidently better aligned with the reference line than the other clusters in the case of the small training set size. These couplings are thus easier learned by the neural network.

In sum, the learning outcome for the couplings is satisfying, especially if it is considered that a multimodal distribution was learned. Large R^2 scores are achieved and the MAE is acceptable.

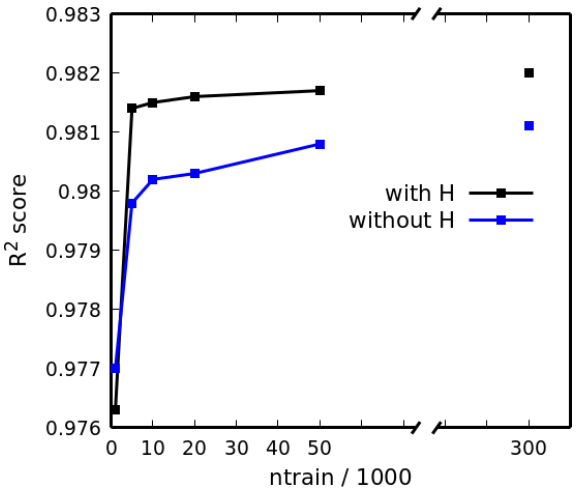


Fig. 7.15.: Learning curves for the couplings obtained with different training set sizes with and without the hydrogen positions in the input. 10 % of ntrain was used for validation, respectively. Note the zoom in the y-axis compared to the learning curves of the site energies.

7.5. Conclusion and Outlook

The excitation energies and Coulomb couplings of the BChl a chromophores in the FMO complex were computed with TD-LC-DFTB2. Their behaviour along 50 ns long trajectories from classical MD was investigated as well as their dependency on the environment. For the site energies a reasonable energy range was obtained with consideration of the environment of the pigments. It was found that the site energies of BChls 2 and 3 are particularly sensitive to small changes in their environment over the simulation time. The range of the excitation energies in vacuum is smaller but still differences between the individual pigments are observed arising from their different geometries determined by the protein scaffold. The computed Coulomb couplings are overestimated compared to previous studies. A scaling of their values may be helpful to improve the excitonic energies resulting from the site energies and couplings.

Neural networks were trained to predict the site energies and Coulomb couplings even faster than TD-LC-DFTB2. These models can save a lot of computational cost in the further study of the exciton propagation through the complex. Neural network models were obtained which are able to predict the excitation energies and couplings in vacuum with satisfying accuracy. Training data from 50 ns MD simulation were used which surely account for all important accesible geometries under the studied conditions. An open challenge is the consideration of the environment which has a nonnegligible impact on the excitation energies. As next step, it will be tested if the electrostatic potential on the BChl a atoms as additional input for the machine is sufficient to get site energies in accordance to the surrounding of the pigments. Finally, a simulation of the exciton propagation through the FMO complex will be performed applying the machine learned models.

Tab. 7.6.: Comparison of model performance for different training set sizes and with and without the information about the hydrogen positions to predict the couplings. The training with ntrain= 10 000 was repeated two times with identical setting. 10 % of the specified training set size was used for validation, respectively. Amount of test data: 50 000.

Hydrogens	ntrain	R^2	MAE [eV]	loss
yes	1 000	0.976	0.003	0.086
	10 000	0.982	0.003	0.022
	10 000	0.981	0.003	0.025
	10 000	0.981	0.003	0.025
	300 000	0.982	0.003	0.020
no	1 000	0.977	0.003	0.034
	10 000	0.980	0.003	0.027
	10 000	0.980	0.003	0.026
	10 000	0.980	0.003	0.029
	300 000	0.981	0.003	0.020

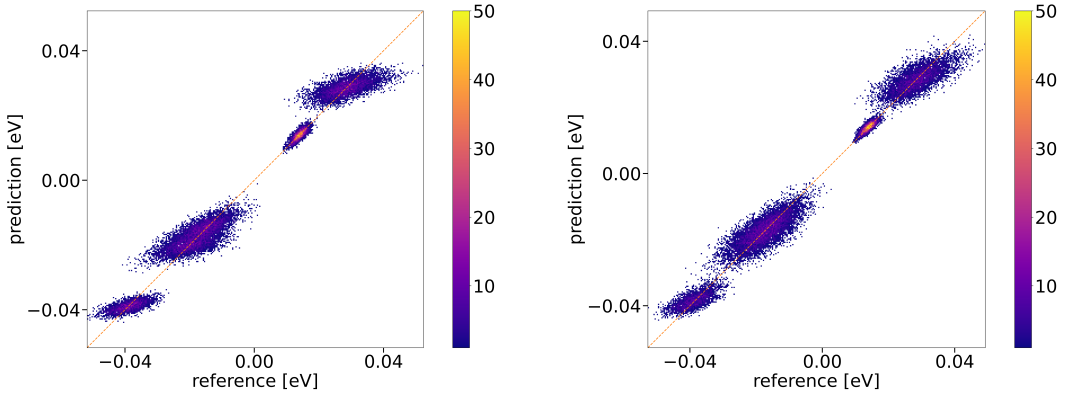


Fig. 7.16.: Scatter plots from 20 000 test data points visualizing the correlation between predicted and reference couplings for different training set sizes. Hydrogen positions were not considered. The colours indicate the density of the points. Left: ntrain=1 000, R^2= 0.977; right: ntrain=300 000, R^2=0.981.

8 | Conclusion and Outlook

The theoretical investigation of the interaction of molecules with light is challenging, since it requires a quantum mechanical description of the process. At the same time, the environment of the molecules needs to be considered. DFT is limited to system sizes of not much more than 100 atoms, high-level *ab initio* methods can only be used for even smaller molecules. The high computational cost of these methods additionally forbids a description of the dynamics of the molecules, either in ground or excited state.

In this work, possible solutions to these problems were discussed. On one hand, a relatively new semi-empirical method (TD-LC-DFTB2) was benchmarked for its ability to describe optical properties of molecules. This method is computationally much more efficient and thus saves computational time when calculating large chromophores. Further, the computation of the dynamics of dyes is computationally feasible, even in excited state. On the other hand, QM/MM methods were applied. This means, the studied systems consisting of a chromphore and its environment, which may be pure solvent or a complex biological system, is divided into a part described by a QM method and a part described in the classical picture. This approach is not new but this work contributes the detailed assessment of the efficient QM method TD-LC-DFTB2 as implemented in DFTB+. In particular, excited state QM/MM simulations with TD-LC-DFTB2 are a new and helpful method in the study of fluorescence.

Firstly, TD-LC-DFTB2 was extensively benchmarked against high-level *ab initio* and DFT methods for excited state geometries and transition energies. Trends in the bond length alternations, adiabatic excitation energies and theoretical Stokes shifts were well reproduced. The deviation of the absolute values of the adiabatic excitation energies were on average 0.44 eV with respect to CC2 but the error could be halved by using a different parameter set for the calculation of the excitation energies. The long-range correction in TD-LC-DFTB was particularly useful to obtain improved reorganization energies for polycyclic aromatic hydrocarbons compared to TD-DFTB. Reasonable vibronic UV/vis spectra could also be simulated with TD-LC-DFTB. However, some problematic cases were also detected in the benchmark. TD-LC-DFTB2 failed in some cases to describe the angles at triple bonds correctly in excited state and was not able to describe the shared proton in excited salicylic acid. Another point to be aware of is the possibility of a different ordering of the excited states predicted with TD-LC-DFTB2 compared to DFT or *ab initio* methods. This was found for some molecules in the benchmark and also for dyes studied with excited state QM/MM simulations later in this work.

In a second step, the benchmark of TD-LC-DFTB2 was extended to excited state QM/MM simulations. Three different fluorophores (Flugi-2, prodan and 4-aminophthalimide) were simulated in homogeneous environments of different polarities. Solvatochromic shifts in the emission spectra determined with TD-LC-DFTB2 come close to the experimental reference values in the case of Flugi-2 and 4-aminophthalimide. Issues were faced for prodan in general and 4-aminophthalimide in DMSO. Excited states were crossing during the simulations and thus the configurational space of the emitting state could not be

sampled reliably.

A derivative of prodan, badan, was studied in the context of optical glucose sensing in the next part of this work. Due to the issues faced for prodan, excited state QM/MM simulations of badan attached to GGBP were not trusted. However, the glucose sensor system was extensively studied by means of classical MD simulations. For the first time, such a sensor system consisting of the GGBP with a fluorophore linked to a specific position is theoretically studied. A specific charge model for the glucose molecule considering polarisation was derived for a proper description of the bound state in the wildtype and in a triple mutant of the protein with badan at position H152C. This triple mutant showed promising results in the experiment and hints were gained regarding the working mechanism of this sensor. In contrast to previous assumptions, it was found that the dye is not enveloped by the protein when a glucose molecule is bound. Instead, the reason for a higher fluorescence signal of the *holo* protein compared to the *apo* protein is likely the reduction of fluorescence quenching by a tryptophane in the binding pocket. When the glucose is bound, it hinders the dye to build a complex with the tryptophane. Taken as a whole, the study of this sensor showed that even if the working principle of this specific sensor is understood, it remains a challenge to design a new sensor with a different dye. Nonetheless, the gained knowledge is a first step in that direction.

In the last part of this work, a pigment protein complex was studied which obtained interest due to its highly efficient energy transfer. The BChls embedded in this FMO complex build excitonic states, which can be computeted from the site energies of the individual pigments and their couplings. These quantities were computed with TD-LC-DFTB2 in a QM/MM fashion, i. e. the environment was included in the QM calculations as point charges. In addition to previous work, the behaviour of the site energies and couplings was investigated on a larger time scale in this thesis. Only small changes were observed along the trajectory but these are enough to change the energy ladder of the site energies. A direct simulation of the exciton propagation through the FMO complex is desirable for the further study. Although TD-LC-DFTB2 is a very efficient method, the computation of the site energies and couplings during such a simulation will slow it down. For this reason, neural networks were trained to predict the site energies and couplings from the geometries of the BChls. TD-LC-DFTB2 results were used as training data. The final models show satisfying R^2 scores of 0.885 (site energies) and 0.981 (couplings). These results were achieved without considering the coordinates of the hydrogens, which reduces the computational cost further without sacrificing quality. However, the site energies and couplings were only learned in vacuum. This is reasonable for the couplings which are only slightly shifted by the environment but a too crude approximation for the site energies. The range between the lowest and largest site energy and the order of the site energies of individual pigments depend clearly on the environment. As a next step, it should be tried to learn a model which can predict the site energies in the environment. For instance, the electrostatic potential on the atoms of the pigments could be used as representation of the environment.

In sum, TD-LC-DFTB2 was found to be a suitable method for the study of fluorescence, for example by means of excited state QM/MM simulations. However, not all molecules are properly described and thus it should not be used blindly. Some issues may be solved by a reparametrization, which however can be involved. The feasible simulation time in excited state is in the order of few hundreds of picoseconds. It could also be desirable to increase this time scale further, for example for large fluorophores with high configurational freedom

or fluorophores in a complex environment with slower processes. In future work, it could be tried to switch to machine learned models for the prediction of excited state forces. This would save computational time during a simulation. Possibly, a machine learned approach could also be helpful to avoid unwanted state crossing in a simulation. If the training data consists solely of training examples for one specific state, then a subsequent simulation using the machine learned model will only sample this state. However, it remains a challenge to obtain the relevant training data. TD-LC-DFTB2 is useful for the generation of training data in general due to its efficiency but it must also show a sufficient accuracy for the specific case.

A | Appendix

A.1. TD-LC-DFTB2 Benchmark

A.1.1. Tables of Values

The detailed values of the bond length alternations, the adiabatic excitation energies, the excited state reorganization energies, the Stokes shifts and the vibrational frequencies are reported in the following tables.

Tab. A.1.: Bond length alternation in the ground and excited state structures. The B3LYP and the CC2 optimized structures were taken from Ref. [89]. (TD-)DFTB/3OB is abbreviated by DFTB and (TD-)LC-DFTB2/OB2$^{0.3}$ (base) by LC-DFTB.

	ground state				excited state			
molecule	B3LYP	DFTB	LC-DFTB	CC2	B3LYP	DFTB	LC-DFTB	CC2
Tetraphenylporphine	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Porphycene	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0023	0.0000	0.0000
Porphine	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Chlorin	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tetracene	0.0349	0.0307	0.0346	0.0275	0.0071	0.0074	0.0089	0.0025
Perylene	0.0345	0.0305	0.0375	0.0259	0.0014	0.0036	0.0000	−0.0075
Phenanthrene	0.0319	0.0302	0.0347	0.0264	−0.0025	−0.0010	0.0021	−0.0099
Fluorene	0.0166	0.0139	0.0153	0.0138	0.0057	0.0009	−0.0005	0.0047
Carbazole	0.0020	0.0025	0.0044	0.0004	−0.0263	−0.0232	−0.0244	−0.0314
Dibenzofuran	0.0007	−0.0024	−0.0011	0.0005	−0.0265	−0.0258	−0.0230	−0.0305
Pyrrolo[3.2-h]quinoline	0.0229	0.0199	0.0258	0.0150	−0.0014	0.0011	−0.0011	−0.0207
Tryptamine-A-ph	0.0115	0.0087	0.0125	0.0053	−0.0188	-	−0.0332	−0.0305
Tryptamine-A-py	0.0113	0.0086	0.0124	0.0052	−0.0183	-	−0.0329	−0.0301
Tryptamine-A-up	0.0113	0.0086	0.0125	0.0051	−0.0170	−0.0138	−0.0327	−0.0305
Tryptamine-Ph-out	0.0114	0.0086	0.0125	0.0051	−0.0195	−0.0127	−0.0330	−0.0309
Tryptamine-Ph-up	0.0110	0.0085	0.0125	0.0047	−0.0161	−0.0133	−0.0311	−0.0299
Tryptamine-Py-out	0.0112	0.0086	0.0125	0.0050	−0.0182	−0.0147	−0.0324	−0.0292
Tryptamine-Py-up	0.0114	0.0087	0.0125	0.0052	−0.0173	−0.0131	−0.0329	−0.0309
2.4.6.8-Decatetraene	0.1157	0.1067	0.1243	0.1066	0.0475	0.0482	0.0495	0.0343
5-MO-salicylic acid	0.0180	0.0146	0.0138	0.0142	−0.0217	−0.0234	−0.0217	−0.0369
Salicylic acid	0.0090	0.0091	0.0089	0.0071	−0.0440	0.0003	−0.0025	−0.0741
3P-1-propionic acid, gauche	0.0008	0.0002	0.0012	0.0010	0.0015	0.0020	−0.0022	−0.0006
3P-1-propionic acid, anti	0.0005	0.0003	0.0000	0.0002	−0.0003	0.0020	−0.0013	0.0000
1-Naphthol, cis	0.0318	0.0300	0.0352	0.0247	−0.0157	−0.0068	−0.0231	−0.0267
1-Naphthol, trans	0.0327	0.0302	0.0352	0.0260	−0.0235	−0.0146	−0.0272	−0.0121
2-Naphthol, cis	0.0327	0.0307	0.0360	0.0261	−0.0199	−0.0169	−0.0231	−0.0158
2-Naphthol, trans	0.0322	0.0307	0.0361	0.0254	−0.0187	−0.0147	−0.0218	−0.0146
7-Hydroxyquinoline, cis	0.0343	0.0325	0.0379	0.0273	−0.0094	−0.0158	−0.0153	−0.0143
7-Hydroxyquinoline, trans	0.0339	0.0326	0.0382	0.0269	−0.0132	−0.0146	−0.0146	−0.0141
2-Hydroxyquinoline, enol	0.0396	0.0356	0.0412	0.0326	−0.0132	−0.0118	−0.0127	−0.0122
2-Hydroxyquinoline, keto	0.0293	0.0282	0.0301	0.0240	−0.0095	−0.0181	−0.0158	−0.0197
o-Dimethoxybenzene, trans	0.0170	0.0148	0.0163	0.0155	0.0193	−0.0325	−0.0296	−0.0276
7-Azaindole	0.0080	0.0095	0.0130	0.0026	−0.0129	−0.0204	−0.0110	−0.0259
Phenylacetylene	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Benzonitrile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Aniline	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2-Hydroxypyridine	0.0140	0.0074	0.0084	0.0131	−0.0140	0.0035	−0.0003	−0.0205
2-Methylpyrimidine	0.0029	0.0002	0.0002	0.0025	0.0051	0.0087	0.0012	0.0034
5-Methylpyrimidine	0.0026	0.0000	0.0012	0.0000	0.0038	0.0022	−0.0011	0.0016
Pyrazine	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2-Pyridone	0.0519	0.0451	0.0470	0.0403	−0.0588	−0.0580	−0.0469	−0.0921
Catechol	0.0078	0.0104	0.0122	0.0073	−0.0045	−0.0228	−0.0243	−0.0178
Resorcinol, Isomer 1	0.0033	0.0006	0.0005	0.0029	−0.0025	−0.0032	−0.0031	−0.0060
Resorcinol, Isomer 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Resorcinol, Isomer 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hydroquinone, cis	0.0055	0.0030	0.0027	0.0043	−0.0056	−0.0047	−0.0045	−0.0100
Hydroquinone, trans	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tab. A.1.: Bond length alternation in the ground and excited state structures, continued.

	ground state				excited state			
molecule	B3LYP	DFTB	LC-DFTB	CC2	B3LYP	DFTB	LCDFTB	CC2
o-Methoxyphenol	0.0121	0.0122	0.0139	0.0111	−0.0151	−0.0259	−0.0254	−0.0186
m-Methoxyphenol, Isomer 2	0.0051	0.0026	0.0022	0.0047	−0.0010	−0.0069	−0.0014	−0.0034
m-Methoxyphenol, Isomer 4	0.0083	0.0032	0.0026	0.0076	−0.0029	−0.0063	−0.0040	−0.0086
p-Methoxyphenol, cis	0.0098	0.0048	0.0042	0.0084	−0.0023	−0.0017	−0.0028	−0.0088
p-Methoxyphenol, trans	0.0048	0.0020	0.0017	0.0045	0.0038	0.0021	0.0014	0.0009
o-Cresol, cis	0.0004	0.0018	0.0014	0.0004	−0.0016	−0.0006	−0.0035	0.0014
o-Cresol, trans	0.0044	0.0045	0.0049	0.0036	−0.0141	−0.0141	−0.0126	−0.0115
m-Cresol, cis	0.0034	0.0014	0.0010	0.0029	−0.0054	−0.0044	−0.0031	−0.0072
m-Cresol, trans	0.0005	0.0014	0.0011	0.0019	0.0032	−0.0041	−0.0038	−0.0036
p-Cresol	0.0034	0.0019	0.0020	0.0023	−0.0035	−0.0031	−0.0028	−0.0053

Tab. A.2.: Bond length alternation in the ground and excited state structures. The B3LYP optimized structures were taken from Ref. [90]. (TD-)DFTB/3OB is abbreviated by DFTB and (TD-)LC-DFTB2/OB2$^{0.3}$(base) by LC-DFTB.

	ground state			excited state		
molecule	B3LYP	DFTB	LC-DFTB	B3LYP	DFTB	LC-DFTB
Aminobenzonitrile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Aniline	0.0000	0.0000	0.0000	0.0006	0.0000	0.0000
Anisol	0.0067	0.0031	0.0026	−0.0004	0.0000	−0.0020
Anthracene	0.0092	0.0077	0.0064	0.0077	0.0067	0.0071
Azulene	0.0168	0.0149	0.0146	−0.0042	−0.0052	−0.0036
Benzaldehyde	0.0038	0.0007	0.0000	−0.0020	0.0014	−0.0002
Benzene	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Benzonitrile	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
β-Dinaphthyleneoxide	0.0230	0.0195	0.0250	−0.0021	−0.0059	−0.0069
Biphenyl	0.0128	0.0125	0.0126	0.0005	0.0027	−0.0001
Cinnoline	0.0307	0.0282	0.0287	0.0058	−0.0002	0.0007
DCS	0.0194	0.0194	0.0225	0.0116	0.0182	0.0036
Dimethylaminobenzonitrile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DMPD	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000
1-6-Epoxy-10-annulene	0.0000	0.0000	0.0000	0.0000	0.0033	−0.0056
Hydroquinone	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Indole	0.0105	0.0078	0.0112	−0.0249	−0.0235	−0.0372
Methyl-4-hydroxycinnamate	0.0249	0.0253	0.0288	0.0093	−0.0076	0.0017
Naphthalene	0.0327	0.0317	0.0369	−0.0301	−0.0235	−0.0316
Octatetraene	0.0970	0.0937	0.1115	0.0194	0.0272	0.0278
p-Benzoquinone	0.1462	0.1448	0.1372	0.1094	0.0963	0.0949
p-Diethynylbenzene	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p-Phenylenediamine	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Phenol	0.0019	0.0010	0.0009	−0.0046	−0.0043	−0.0035
Pyrene	0.0245	0.0232	0.0277	0.0029	0.0044	0.0029
Pyridine	0.0000	0.0000	0.0000	−0.0088	0.0001	−0.0046
Pyridone-lactam	0.0504	0.0451	0.0470	−0.0345	−0.0580	−0.0469
Pyridone-lactim	0.0137	0.0074	0.0084	−0.0034	0.0035	−0.0003
Pyrimidine	0.0000	0.0000	0.0000	−0.0002	0.0081	0.0000
Quinoline	0.0356	0.0344	0.0399	−0.0117	−0.0129	−0.0138
Quinoxaline	0.0392	0.0375	0.0431	−0.0140	−0.0157	−0.0138
Terylene	0.0313	0.0269	0.0358	0.0090	0.0095	0.0089
Tetrazine	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Toluene	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
t-Stilbene	0.0222	0.0223	0.0245	0.0027	0.0038	0.0003

Tab. A.3.: Adiabatic excitation energies obtained with (TD-)DFTB/3OB, (TD-)LC-DFTB2/OB2$^{0.3}$ (base), i. e. with repulsive parameters [92], and a combination of (TD-)LC-DFTB2/OB2$^{0.3}$ (base) used for optimization [92] and particularly optimized parameters for the excitation energies [104]. The values are in eV.

molecule	DFTB	LC-DFTB rep.	LC-DFTB combined
Tetraphenylporphine	1.75	1.93	1.72
Porphycene	1.80	1.95	1.79
Porphine	1.93	2.09	1.85
Chlorin	2.00	2.17	1.92
Tetracene	2.10	2.78	2.56
Perylene	2.52	3.12	2.85
Phenanthrene	3.58	4.46	3.96
Fluorene	4.02	4.74	4.22
Carbazole	3.71	4.60	4.12
Dibenzofuran	3.92	4.86	4.34
Pyrrolo[3.2-h]quinoline	3.46	3.88	3.88
Tryptamine-A-ph	-	4.72	4.23
Tryptamine-A-py	-	4.71	4.22
Tryptamine-A-up	3.74	4.71	4.22
Tryptamine-Ph-out	3.67	4.68	4.21
Tryptamine-Ph-up	3.69	4.61	4.17
Tryptamine-Py-out	3.83	4.70	4.21
Tryptamine-Py-up	3.74	4.72	4.23
2.4.6.8-Decatetraene	3.37	3.88	3.55
5-MO-salicylic acid	3.13	3.99	3.62
Salicylic acid	3.55	4.12	4.15
3P-1-propionic acid, gauche	4.57	5.59	5.24
3P-1-propionic acid, anti	4.65	5.79	5.17
1-Naphthol, cis	3.42	4.21	3.77
1-Naphthol, trans	3.57	4.32	3.88
2-Naphthol, cis	3.56	4.27	3.83
2-Naphthol, trans	3.51	4.23	3.80
7-Hydroxyquinoline, cis	3.01	3.43	3.44
7-Hydroxyquinoline, trans	2.98	3.40	3.41
2-Hydroxyquinoline, enol	3.20	3.56	3.58
2-Hydroxyquinoline, keto	2.39	3.24	3.28
o-Dimethoxybenzene, trans	4.33	5.07	4.47
7-Azaindole	3.93	4.28	4.21

Tab. A.3.: Adiabatic excitation energies, continued.

molecule	DFTB	LC-DFTB rep.	LC-DFTB combined
Phenylacetylene	4.70	5.37	4.98
Benzonitrile	4.76	5.41	5.04
Aniline	4.49	5.21	4.65
2-Hydroxypyridine	4.32	4.64	4.52
2-Methylpyrimidine	3.72	3.97	4.00
5-Methylpyrimidine	3.68	3.97	4.01
Pyrazine	3.40	3.79	3.80
2-Pyridone	2.55	3.36	3.36
Catechol	4.56	5.25	4.62
Resorcinol, Isomer 2	4.77	5.51	4.85
Resorcinol, Isomer 1	4.73	5.47	4.82
Resorcinol, Isomer 3	4.69	5.44	4.78
Hydroquinone, cis	4.19	4.81	4.26
Hydroquinone, trans	4.19	4.80	4.26
o-Methoxyphenol	4.49	5.21	4.58
m-Methoxyohenol, Isomer 4	4.63	5.42	4.78
m-Methoxyphenol, Isomer 2	4.59	5.38	4.74
p-Methoxyphenol, trans	4.06	4.74	4.21
p-Methoxyphenol, cis	4.07	4.74	4.21
o-Cresol, cis	4.65	5.34	4.72
o-Cresol, trans	4.64	5.35	4.73
m-Cresol, cis	4.69	5.43	4.79
m-Cresol, trans	4.70	5.44	4.79
p-Cresol	4.57	5.27	4.67
Acetaldehyde	3.83	4.24	4.01
Acetone	3.92	4.34	4.17
Acrolein	2.67	3.13	3.00
HCOOH	-	5.15	4.92
Aminobenzonitrile	4.35	5.06	4.52
Anisol	4.48	5.32	4.73
Anthracene	2.76	3.45	3.14
Azulene	1.90	2.26	1.99
Benzaldehyde	2.83	3.34	3.21
Benzene	5.14	5.92	5.24
β-Dinaphthyleneoxide	3.01	3.81	3.46

Tab. A.3.: Adiabatic excitation energies, continued.

molecule	DFTB	LC-DFTB rep.	LC-DFTB combined
Biphenyl	4.16	4.81	4.32
C2H2	4.76	6.16	5.60
CH2O	3.72	4.07	3.86
Cinnoline	2.12	2.49	2.59
2-Cyclopenten-1-one	2.74	3.30	3.28
DBH	3.77	3.75	3.90
DCS	2.47	3.59	3.28
Dimethylaminobenzonitrile	3.09	4.60	4.14
DMPD	3.32	4.30	3.85
1-6-Epoxy-10-annulene	3.06	3.57	3.23
Fluorene	4.02	4.74	4.22
Glyoxal	2.09	2.37	2.35
Hydroquinone	4.19	4.80	4.26
Indole	4.15	4.94	4.41
Methyl-4-hydroxycinnamate	2.61	3.43	3.45
Naphthalene	3.75	4.46	4.06
Octatetraene	3.57	4.06	3.73
p-Benzoquinone	1.51	1.83	2.01
p-Diethynylbenzene	4.27	4.96	4.48
p-Phenylenediamine	3.95	4.63	4.14
Phenol	4.72	5.43	4.80
Propynal	3.12	3.65	3.52
Pyrene	3.13	3.77	3.39
Pyridine	3.96	4.42	4.25
Pyridone-lactam	2.55	3.36	3.36
Pyridone-lactim	4.32	4.64	4.52
Pyrimidine	3.71	4.02	4.05
Quinoline	3.01	3.39	3.41
Quinoxaline	2.46	2.78	2.87
Terylene	1.98	2.50	2.31
Tetrazine	2.22	2.23	2.47
Toluene	5.02	5.81	5.15
t-Stilbene	3.50	4.09	3.72

Tab. A.4.: Excited state reorganization energies obtained with TD-LC-DFTB2 and TD-DFTB compared to the B3LYP results from Ref. [90] and the CAM-B3LYP results of this work.

molecule	B3LYP from Ref. [90]	DFTB	LC-DFTB	CAM-B3LYP
7-Azaindole	0.39	0.35	0.35	0.37
Acetaldehyde	0.45	0.74	0.64	0.41
Acetone	0.52	0.67	0.64	0.48
Acrolein	0.34	0.54	0.43	0.27
HCOOH	0.98	-	1.12	0.90
Aminobenzonitrile	0.18	0.25	0.30	0.17
Aniline	0.22	0.28	0.33	0.20
Anisol	0.15	0.27	0.25	0.14
Anthracene	0.22	0.19	0.28	0.25
Azulene	0.38	0.40	0.38	0.35
Benzaldehyde	0.32	0.43	0.35	0.25
Benzene	0.13	0.20	0.23	0.12
Benzonitrile	0.15	0.24	0.39	0.13
β-Dinaphthyleneoxide	0.19	0.16	0.28	-
Biphenyl	0.54	0.38	0.42	-
C2H2	1.99	3.62	3.31	1.90
CH2O	0.31	0.83	0.66	0.26
Cinnoline	0.41	0.40	0.30	0.37
2-Cyclopenten-1-one	0.37	0.54	0.45	0.27
DBH	0.07	0.17	0.16	0.07
DCS	0.09	0.33	0.27	-
Dimethylaminobenzonitrile	0.25	1.04	0.32	0.18
DMPD	0.22	0.52	0.34	0.25
1-6-Epoxy-10-annulene	0.14	0.28	0.38	0.18
Fluorene	0.15	0.28	0.40	0.14

Tab. A.4.: Reorganization energies obtained with TD-LC-DFTB2 and TD-DFTB compared to the B3LYP results from Ref. [90], continued.

molecule	B3LYP from Ref. [90]	DFTB	LC-DFTB	CAM-B3LYP
Glyoxal	0.09	0.20	0.17	0.08
Hydroquinone	0.18	0.20	0.23	0.16
Indole	0.29	0.31	0.34	0.13
Methyl-4-hydroxycinnamate	0.20	0.59	0.56	0.25
Naphthalene	0.29	0.27	0.39	0.13
Octatetraene	0.22	0.19	0.40	0.24
p-Benzoquinone	0.10	0.11	0.11	0.09
p-Diethynylbenzene	0.23	0.28	0.40	0.13
p-Phenylenediamine	0.26	0.28	0.35	0.25
Phenol	0.16	0.21	0.24	0.14
Porphyrin	0.02	0.02	0.02	0.01
Propynal	0.41	0.80	0.52	0.36
Pyrene	0.14	0.13	0.23	0.12
Pyridine	0.46	0.38	0.38	0.49
Pyridone-lactam	0.51	0.77	0.77	0.36
Pyridone-lactim	0.32	0.44	0.43	0.34
Pyrimidine	0.43	0.43	0.42	0.45
Quinoline	0.56	0.43	0.43	0.60
Quinoxaline	0.32	0.30	0.29	0.31
Terylene	0.12	0.08	0.20	-
Tetrazine	0.14	0.13	0.12	0.13
Toluene	0.14	0.20	0.22	-
t-Stilbene	0.28	0.24	0.44	-

Tab. A.5.: Stokes shifts calculated with different methods for CC2 optimized geometries from Ref. [89] and (TD-)LC-DFTB2/OB2$^{0.3}$ (base) optimized geometries. CAM-B3LYP is abbreviated by CAM. The values are in eV.

	CC2 geometries				LC-DFTB geometries			
molecule	CAM	B3LYP	DFTB	LC-DFTB	CAM	B3LYP	DFTB	LC-DFTB
Tetraphenylporphine	0.07	0.08	0.09	0.11	0.06	0.06	0.07	0.09
Porphycene	0.07	0.08	0.09	0.11	0.06	0.08	0.10	0.12
Porphine	0.04	0.04	0.05	0.06	0.03	0.03	0.04	0.05
Chlorin	0.02	0.02	0.03	0.03	0.03	0.02	0.02	0.04
Tetracene	0.33	0.26	0.29	0.38	0.35	0.27	0.31	0.42
Perylene	0.37	0.27	0.30	0.44	0.37	0.26	0.31	0.46
Phenanthrene	0.30	0.26	0.29	0.36	0.29	0.25	0.28	0.35
Fluorene	0.27	0.26	0.31	0.41	0.39	0.25	0.59	0.78
Carbazole	0.25	0.24	0.26	0.31	0.22	0.20	0.25	0.29
Dibenzofuran	0.24	0.23	0.26	0.31	0.27	0.23	0.29	0.42
Pyrrolo[3.2-h]quinoline	0.36	0.35	0.38	0.19	0.30	0.21	0.61	0.81
Tryptamine-A-ph	0.63	0.65	0.72	0.79	0.54	0.54	0.66	0.75
Tryptamine-A-py	0.65	0.67	0.73	0.80	0.55	0.55	0.66	0.75
Tryptamine-A-up	0.65	0.67	0.72	0.80	0.55	0.55	0.66	0.75
Tryptamine-Ph-out	0.69	0.71	0.74	0.81	0.64	0.64	0.72	0.82
Tryptamine-Ph-up	0.74	0.75	0.78	0.87	0.67	0.66	0.74	0.85
Tryptamine-Py-out	0.66	0.68	0.75	0.82	0.55	0.55	0.66	0.75
Tryptamine-Py-up	0.62	0.65	0.71	0.79	0.52	0.53	0.68	0.76
2.4.6.8-Decatetraene	0.48	0.26	0.43	0.69	0.52	0.29	0.46	0.74
5-MO-salicylic acid	0.59	0.51	0.54	0.67	0.38	0.32	0.37	0.48
Salicylic acid	1.26	1.20	1.66	1.86	0.26	0.21	0.87	1.07
3P-1-propionic acid, gauche	0.31	0.32	0.41	0.46	0.65	0.98	1.08	1.41
3P-1-propionic acid, anti	0.30	0.30	0.37	0.45	0.41	0.42	0.52	0.67
1-Naphthol, cis	0.67	0.62	0.60	0.73	0.54	0.47	0.53	0.67
1-Naphthol, trans	0.29	0.25	0.26	0.34	0.45	0.47	0.53	0.67
2-Naphthol, cis	0.34	0.31	0.34	0.43	0.36	0.34	0.44	0.55
2-Naphthol, trans	0.34	0.31	0.34	0.43	0.35	0.33	0.43	0.54
7-Hydroxyquinoline, cis	0.37	0.35	0.22	0.29	0.84	0.90	0.88	0.96
7-Hydroxyquinoline, trans	0.37	0.36	0.22	0.29	0.85	0.91	0.87	0.96
2-Hydroxyquinoline, enol	0.36	0.34	0.21	0.27	0.50	0.66	0.85	0.94
2-Hydroxyquinoline, keto	0.56	0.51	0.52	0.60	0.54	0.91	1.05	1.15
o-Dimethoxybenzene, trans	0.72	0.71	0.69	0.83	0.35	0.33	0.47	0.56
7-Azaindole	0.80	0.78	0.72	0.41	0.61	0.50	0.70	0.75
Phenylacetylene	0.31	0.31	0.40	0.45	0.92	0.77	1.23	1.73
Benzonitrile	0.31	0.31	0.40	0.45	0.69	0.69	0.91	1.21
Aniline	0.43	0.42	0.50	0.60	0.40	0.40	0.53	0.63
2-Hydroxypyridine	0.46	0.44	0.27	0.13	0.61	0.74	0.84	0.93
2-Methylpyrimidine	1.02	1.00	0.95	1.02	0.85	0.83	0.84	0.91
5-Methylpyrimidine	1.03	1.00	0.94	0.99	0.82	0.80	0.80	0.86

Tab. A.5.: Stokes shifts, continued.

	CC2 geometries				LC-DFTB geometries			
molecule	CAM	B3LYP	DFTB	LC-DFTB	CAM	B3LYP	DFTB	LC-DFTB
Pyrazine	0.70	0.71	0.81	0.68	0.20	0.19	0.26	0.29
2-Pyridone	1.19	1.53	1.58	1.70	0.87	1.31	1.48	1.60
Catechol	0.78	0.81	0.68	0.80	0.31	0.31	0.44	0.52
Resorcinol, Isomer 2	0.34	0.33	0.41	0.48	0.28	0.27	0.39	0.46
Resorcinol, Isomer 1	0.44	0.44	0.49	0.57	0.28	0.28	0.39	0.47
Resorcinol, Isomer 3	0.65	0.65	0.62	0.70	0.29	0.28	0.40	0.47
Hydroquinone, cis	0.38	0.36	0.39	0.46	0.32	0.30	0.38	0.46
Hydroquinone, trans	0.38	0.36	0.39	0.46	0.32	0.30	0.38	0.45
o-Methoxyphenol	0.67	0.67	0.64	0.76	0.33	0.32	0.45	0.54
m-Methoxyphenol, Isomer 4	0.39	0.38	0.44	0.52	0.28	0.27	0.41	0.47
m-Methoxyphenol, Isomer 2	0.59	0.58	0.57	0.66	0.28	0.28	0.40	0.47
p-Methoxyphenol, trans	0.36	0.33	0.36	0.45	0.31	0.29	0.37	0.45
p-Methoxyphenol, cis	0.37	0.35	0.37	0.45	0.31	0.29	0.37	0.45
o-Cresol, cis	0.37	0.36	0.40	0.46	0.30	0.29	0.40	0.47
o-Cresol, trans	0.44	0.44	0.45	0.52	0.36	0.36	0.43	0.50
m-Cresol, cis	0.35	0.34	0.40	0.46	0.28	0.28	0.38	0.45
m-Cresol, trans	0.33	0.32	0.39	0.44	0.29	0.29	0.39	0.45
p-Cresol	0.35	0.34	0.40	0.46	0.31	0.30	0.39	0.46

Tab. A.6.: Stokes shifts calculated with different methods for B3LYP optimized geometries from Ref. [90] and (TD-)LC-DFTB2/OB2$^{0.3}$ (base) optimized geometries. The Stokes shifts already given in Ref. [90] were recalculated to be consistent with the further calculations here. CAM-B3LYP is abbreviated by CAM. The values are in eV.

	B3LYP geometries				LC-DFTB geometries			
molecule	CAM	B3LYP	DFTB	LC-DFTB	CAM	B3LYP	DFTB	LC-DFTB
Acetaldehyde	1.16	1.19	1.30	1.38	1.19	1.22	1.41	1.49
Acetone	1.33	1.36	1.34	1.43	1.11	1.15	1.27	1.36
Acrolein	0.67	0.74	0.91	0.92	0.60	0.67	0.85	0.85
HCOOH	2.56	2.60	2.52	2.71	2.33	2.37	2.49	2.65
Aminobenzonitrile	0.33	0.33	0.41	0.49	0.36	0.35	0.48	0.57
Anisol	0.29	0.28	0.34	0.41	0.34	0.34	0.48	0.53
Anthracene	0.46	0.39	0.44	0.56	0.44	0.36	0.41	0.54
Azulene	0.78	0.80	0.88	0.93	0.65	0.68	0.75	0.80
Benzaldehyde	0.63	0.70	0.80	0.82	0.54	0.60	0.70	0.72
Benzene	0.26	0.26	0.36	0.40	0.39	0.39	0.50	0.59
Benzonitrile	0.28	0.29	0.37	0.41	0.69	0.69	0.91	1.21
β-Dinaphthyleneoxide	0.42	0.34	0.37	0.54	0.43	0.34	0.37	0.56
Biphenyl	0.85	0.68	0.81	1.12	0.51	0.31	0.59	0.81
C2H2	3.97	3.89	5.19	5.69	3.98	3.90	5.14	5.63
CH2O	0.91	0.94	1.25	1.31	1.18	1.22	1.55	1.62
Cinnoline	0.88	0.94	1.03	0.95	0.65	0.69	0.73	0.69
2-Cyclopenten-1-one	0.74	0.83	0.95	1.97	0.60	0.69	0.88	0.89
DBH	0.22	0.22	0.25	0.25	0.14	0.14	0.30	0.31
DCS	0.26	0.13	0.10	0.32	0.39	0.14	0.10	0.52
Dimethylaminobenzonitrile	0.53	0.66	0.86	0.79	0.41	0.47	0.67	0.68
DMPD	0.44	0.41	0.48	0.59	0.45	0.43	0.53	0.62
1-6-Epoxy-10-annulene	0.29	0.27	0.28	0.36	0.46	0.41	0.65	0.80

Tab. A.6.: Stokes shifts, continued.

	B3LYP geometries				LC-DFTB geometries			
molecule	CAM	B3LYP	DFTB	LC-DFTB	CAM	B3LYP	DFTB	LC-DFTB
Glyoxal	0.19	0.20	0.26	0.27	0.22	0.23	0.31	0.34
Hydroquinone	0.35	0.33	0.36	0.43	0.32	0.30	0.38	0.45
Indole	0.46	0.57	0.65	0.71	0.41	0.47	0.58	0.69
Methyl-4-hydroxycinnamate	0.47	0.33	0.45	0.45	0.72	1.08	1.06	1.15
Naphthalene	0.36	0.43	0.60	0.75	0.42	0.48	0.68	0.90
Octatetraene	0.45	0.24	0.43	0.67	0.56	0.33	0.54	0.85
p-Benzoquinone	0.21	0.21	0.21	0.23	0.18	0.18	0.19	0.22
p-Diethynylbenzene	0.47	0.31	0.53	0.70	0.47	0.30	0.59	0.78
p-Phenylenediamine	0.50	0.47	0.52	0.63	0.48	0.45	0.57	0.69
Phenol	0.35	0.35	0.41	0.47	0.29	0.29	0.40	0.46
Propynal	0.79	0.82	1.01	1.03	0.77	0.80	0.98	1.01
Pyrene	0.16	0.12	0.29	0.43	0.18	0.14	0.31	0.46
Pyridine	1.27	1.25	1.29	1.28	0.89	0.87	0.90	0.92
Pyridone-lactam	1.02	1.45	1.52	1.62	0.88	1.32	1.48	1.60
Pyridone-lactim	1.26	1.33	1.33	1.40	0.61	0.74	0.84	0.93
Pyrimidine	1.01	0.98	0.96	1.02	0.83	0.82	0.82	0.88
Quinoline	1.19	1.16	1.03	1.10	0.95	0.93	0.84	0.91
Quinoxaline	0.69	0.69	0.71	0.72	0.53	0.53	0.57	0.60
Terylene	0.27	0.16	0.20	0.34	0.29	0.16	0.21	0.37
Tetrazine	0.35	0.34	0.33	0.31	0.23	0.22	0.24	0.25
Toluene	0.27	0.27	0.36	0.40	0.38	0.38	0.49	0.59
t-Stilbene	0.63	0.48	0.55	0.82	0.63	0.47	0.56	0.84

Tab. A.7.: Calculated frequencies of anthanthrene in cm^{-1}. The first six frequencies correspond to translation and rotation, respectively.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
−1.77	0.00	−1.91	0.00	−1.23	−6.47	−1.94	−5.84
−1.40	0.00	−1.49	0.00	0.00	−1.37	−0.68	−4.65
−0.50	0.00	−0.45	0.00	0.00	−1.13	−0.37	−2.52
0.33	0.00	0.28	0.00	0.00	1.01	0.48	−2.35
0.60	0.00	0.49	0.00	1.14	3.85	0.83	3.14
1.05	0.00	1.10	0.00	1.40	6.34	1.57	4.04
65.66	66.93	66.50	74.09	60.83	60.89	62.54	63.61
97.70	96.53	100.64	103.71	91.09	87.98	95.99	91.13
162.08	160.38	163.65	168.02	150.36	150.16	156.49	159.66
188.62	178.31	192.65	183.11	175.39	167.12	188.94	179.19
201.28	190.97	206.23	197.28	186.95	177.19	197.13	185.47
267.66	251.98	270.71	262.99	262.58	255.94	285.54	269.97
289.41	275.63	294.27	282.41	270.22	257.73	287.51	282.92
292.76	284.47	298.80	290.25	270.98	264.42	292.47	283.57
317.76	306.86	322.29	316.49	295.14	289.15	313.26	310.00
345.21	343.21	349.31	347.72	337.71	334.69	372.58	370.14
389.40	378.83	394.11	388.07	367.79	357.09	388.36	384.82
400.00	383.41	406.34	392.72	382.73	376.69	420.28	414.95
455.70	451.74	460.53	456.43	447.63	421.39	488.79	444.52
473.96	470.45	479.09	475.55	463.50	443.06	509.31	485.18
497.72	472.15	509.12	486.86	468.63	458.48	513.38	503.98
504.49	487.90	510.14	504.28	473.69	465.59	517.49	510.53
515.14	499.03	525.65	513.87	486.55	470.20	537.37	518.22
533.83	518.34	541.07	529.45	498.42	489.90	548.21	532.36
535.99	530.37	546.50	538.40	516.30	490.56	566.62	546.14
557.63	532.76	564.19	547.66	531.32	527.03	587.62	580.98
560.69	552.95	570.48	561.95	537.98	528.44	593.01	583.56
565.50	564.53	570.65	571.81	553.89	551.27	612.77	616.36
580.58	566.74	591.91	579.39	557.90	557.48	621.02	622.04
608.80	599.53	615.70	606.53	600.70	590.45	659.67	649.21
658.96	637.84	672.23	654.17	618.03	590.54	686.73	660.28
663.70	653.35	673.74	661.35	657.87	645.75	727.35	710.19
713.37	703.62	728.83	711.59	666.79	650.59	748.10	720.93
719.59	707.71	734.57	714.09	670.72	652.37	749.64	727.54
753.61	734.03	762.79	742.09	728.36	693.62	810.26	772.63
763.45	735.35	774.03	745.18	738.15	699.04	824.38	787.96
765.61	749.21	780.56	764.75	748.56	724.21	827.13	821.39
773.33	757.91	783.33	766.30	752.14	729.61	840.62	822.91
774.56	762.93	791.66	776.75	757.01	746.79	845.01	833.11

Tab. A.7.: Calculated frequencies of anthanthrene in cm^{-1}, continued.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
808.66	780.71	824.35	793.62	767.42	748.35	860.25	849.87
834.31	815.50	843.27	829.05	767.54	748.62	883.76	857.81
838.37	823.41	856.93	835.66	782.99	759.20	888.18	864.51
846.58	837.93	869.72	851.08	790.25	759.64	906.91	886.09
873.65	852.26	898.31	871.57	817.94	771.59	915.85	904.37
896.95	857.11	920.99	875.10	819.25	778.16	943.54	909.09
900.49	872.70	924.95	885.81	826.38	802.18	952.40	910.36
905.39	901.23	926.40	915.12	843.32	803.89	974.56	934.86
911.50	901.84	935.48	923.72	844.46	820.37	977.53	938.75
916.91	920.70	945.04	951.95	893.77	858.55	996.46	991.23
956.00	921.93	988.88	952.42	893.89	860.97	1035.49	1010.64
956.25	930.82	989.75	959.38	904.76	878.33	1035.61	1011.81
966.32	930.87	990.27	961.14	906.38	878.72	1044.55	1020.14
967.38	954.62	1001.91	965.78	907.07	895.46	1044.62	1020.66
976.32	969.23	1002.52	989.33	975.63	973.47	1095.54	1102.13
1016.18	1012.90	1032.25	1028.99	1013.62	1003.22	1150.80	1141.42
1067.26	1061.12	1080.32	1076.72	1054.57	1049.35	1187.83	1185.79
1090.36	1080.75	1102.59	1096.44	1060.36	1053.76	1204.75	1200.70
1099.13	1104.71	1106.56	1120.83	1077.93	1068.98	1213.35	1209.91
1113.39	1107.67	1127.10	1124.32	1081.94	1086.37	1220.29	1225.79
1164.06	1161.07	1158.79	1164.93	1106.90	1097.08	1255.68	1250.80
1176.34	1168.18	1184.14	1170.71	1123.20	1110.04	1262.88	1258.08
1184.00	1187.53	1192.81	1192.64	1138.09	1138.80	1284.06	1282.75
1198.80	1194.45	1209.33	1201.84	1150.71	1139.11	1285.87	1283.67
1207.82	1201.77	1216.47	1211.08	1155.64	1147.95	1294.49	1292.56
1208.18	1203.24	1216.83	1212.38	1159.68	1156.61	1303.03	1297.80
1250.77	1247.22	1266.84	1259.33	1214.76	1205.38	1369.93	1358.10
1267.55	1255.39	1280.20	1259.38	1228.65	1210.23	1375.15	1363.16
1286.96	1288.37	1298.52	1287.78	1251.90	1246.10	1401.84	1405.86
1298.57	1304.82	1298.75	1302.87	1268.77	1258.45	1430.93	1406.08
1331.73	1313.59	1353.36	1332.79	1303.01	1289.75	1484.87	1457.01
1353.33	1323.03	1367.36	1338.55	1317.82	1310.64	1499.97	1498.43
1361.87	1374.06	1376.96	1375.54	1338.21	1324.89	1508.06	1512.71
1383.27	1374.91	1401.83	1390.08	1338.28	1343.63	1535.60	1529.34
1399.27	1401.23	1416.67	1404.75	1367.86	1351.65	1554.68	1537.32
1418.85	1402.04	1443.81	1425.07	1395.03	1374.81	1581.99	1565.28
1434.82	1417.53	1446.84	1431.31	1406.67	1382.12	1589.23	1576.28
1446.58	1433.77	1470.68	1447.52	1417.68	1392.66	1623.43	1586.91
1449.91	1445.10	1479.40	1462.59	1437.30	1409.93	1626.05	1603.07

Tab. A.7.: Calculated frequencies of anthanthrene in cm^{-1}, continued.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
1470.75	1447.46	1487.15	1462.75	1461.28	1418.04	1646.54	1613.01
1478.03	1466.66	1511.66	1484.96	1464.64	1430.65	1675.24	1618.76
1504.44	1474.99	1541.23	1509.99	1480.57	1469.52	1704.62	1667.51
1511.04	1488.12	1544.58	1511.78	1491.33	1472.89	1714.81	1673.40
1520.57	1517.06	1550.23	1544.97	1517.98	1502.53	1718.79	1708.72
1588.73	1545.62	1631.15	1578.41	1557.98	1522.95	1806.27	1741.74
1604.88	1562.09	1650.13	1593.97	1584.32	1530.32	1827.75	1762.45
1619.87	1573.19	1663.30	1604.88	1598.74	1545.91	1839.98	1768.53
1626.07	1592.96	1671.74	1631.28	1605.62	1563.68	1847.54	1794.19
1629.55	1596.79	1674.64	1635.30	1609.10	1565.27	1849.46	1799.70
1660.28	1626.69	1710.69	1663.65	1620.52	1585.43	1885.57	1820.29
1663.30	1630.60	1715.05	1667.60	1631.56	1592.74	1886.84	1845.78
3156.28	3158.30	3182.48	3185.90	2996.83	2999.25	3130.81	3134.93
3156.29	3158.39	3182.54	3186.18	2996.86	2999.80	3131.07	3135.34
3158.53	3161.14	3184.94	3187.72	3000.78	3000.60	3134.42	3135.94
3158.91	3161.58	3185.32	3188.47	3001.35	3001.04	3134.71	3136.28
3161.60	3164.37	3187.60	3190.46	3001.92	3001.12	3137.39	3136.57
3161.77	3164.75	3187.73	3190.97	3001.94	3002.08	3137.61	3136.96
3169.34	3171.79	3195.09	3199.22	3006.34	3006.65	3141.91	3144.65
3169.56	3172.36	3195.29	3199.60	3006.38	3006.76	3142.01	3144.90
3175.52	3178.90	3201.89	3205.69	3008.38	3008.59	3143.92	3145.91
3175.62	3179.16	3201.99	3206.02	3008.63	3009.58	3144.21	3146.38
3185.12	3188.35	3211.27	3214.04	3014.66	3014.18	3150.33	3151.13
3185.30	3188.67	3211.40	3214.37	3014.84	3014.44	3150.56	3151.31

Tab. A.8.: Calculated frequencies of TAPP in cm^{-1}. The first six frequencies correspond to translation and rotation, respectively. The excited state with the largest oscillator strength of the first three excitations was considered.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
−1.63	0.00	−1.25	0.00	−3.36	−8.59	−3.37	−10.00
−0.49	0.00	−1.17	0.00	−2.09	−6.58	−2.66	−7.26
−0.20	0.00	−0.65	0.00	−1.09	−3.27	−2.37	−6.43
0.56	0.00	−0.39	0.00	−0.90	−2.86	0.91	−2.72
2.07	0.00	2.70	0.00	3.13	−2.38	2.59	4.96
4.97	0.00	5.45	0.00	3.53	4.41	3.12	10.64
45.74	37.87	46.19	39.60	41.48	42.40	45.00	46.69
52.14	46.82	55.66	51.89	51.73	42.76	68.87	58.58
101.46	95.64	103.51	99.70	95.48	93.84	108.18	108.75
173.87	170.07	177.50	178.13	162.57	155.97	171.10	166.95
177.61	175.92	181.55	182.00	163.01	160.54	174.81	167.61
185.31	183.41	187.41	188.39	182.92	174.44	199.64	195.50
203.66	185.50	209.66	192.56	187.50	181.03	210.73	197.71
216.55	201.84	222.11	203.84	206.15	189.06	234.72	210.02
240.12	242.34	244.36	245.52	220.67	220.08	237.98	239.46
280.51	281.96	284.63	282.70	275.10	273.18	306.31	303.90
324.57	308.60	329.31	326.71	293.88	288.03	309.16	305.80
332.62	318.74	338.46	331.95	307.10	308.56	332.81	338.91
351.42	331.53	356.39	338.79	343.31	321.68	382.28	368.98
399.90	383.13	407.12	394.39	364.87	356.12	394.70	389.42
465.71	466.68	470.64	474.93	453.69	434.64	498.86	480.27
476.15	467.76	481.16	479.54	460.79	442.78	505.40	480.62
486.70	485.49	493.81	482.52	462.89	455.21	506.95	503.48
493.66	485.54	498.80	494.67	463.65	459.32	511.39	512.76
505.47	495.15	513.99	496.61	479.44	460.44	525.90	529.43
508.98	498.27	516.13	500.19	482.24	478.81	529.15	529.45
512.60	503.13	519.41	507.42	482.53	480.65	538.34	537.74
533.52	529.09	539.66	521.33	497.52	483.13	554.13	554.65
533.87	534.93	544.79	525.69	504.92	493.04	562.11	565.08
534.69	573.82	548.18	537.40	521.63	516.63	569.25	567.39
574.13	577.60	581.69	581.04	527.76	521.11	581.75	586.91
586.03	589.15	597.26	584.68	530.02	534.90	590.88	611.63
589.23	590.58	599.11	592.70	569.17	568.62	632.83	634.78
590.31	654.01	600.27	596.13	582.12	581.19	655.01	649.01
599.80	681.65	605.43	599.88	594.73	582.92	657.39	657.28
692.44	686.25	699.58	697.55	651.67	622.01	726.46	692.45

Tab. A.8.: Calculated frequencies of TAPP in cm^{-1}, continued.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
704.35	695.52	712.99	698.43	658.93	656.09	729.67	740.23
707.55	702.64	723.28	718.22	686.63	657.46	758.59	750.63
714.97	717.73	725.79	733.54	687.76	681.57	780.64	753.37
740.74	736.63	750.79	736.01	694.84	686.42	782.06	773.09
742.80	752.62	752.54	740.76	698.30	692.17	782.99	782.39
755.82	790.17	777.59	743.77	720.81	711.89	789.31	782.79
807.01	798.81	817.09	815.65	755.94	736.00	855.00	836.47
829.70	807.05	844.87	824.69	776.88	751.51	881.65	849.64
836.08	821.00	864.42	834.10	789.08	762.10	884.38	882.76
851.08	822.76	864.68	848.93	789.22	776.25	909.61	892.36
855.00	839.48	866.62	865.15	795.96	780.22	914.67	904.43
857.72	853.21	870.36	866.93	797.90	783.81	915.87	917.40
880.29	860.39	898.66	868.31	799.27	794.17	921.91	923.11
887.36	885.60	899.78	898.56	840.44	836.32	928.15	926.98
902.49	897.23	920.39	902.53	851.84	845.26	937.93	938.89
908.13	912.82	928.77	913.93	871.85	859.31	958.00	958.41
932.80	939.25	950.62	943.00	899.20	864.03	1023.81	1007.36
948.68	944.89	953.35	972.73	900.87	864.54	1024.38	1015.84
977.64	955.14	1012.06	973.16	903.60	868.05	1041.46	1020.75
979.73	969.10	1015.39	980.38	905.00	869.15	1042.28	1024.05
993.28	975.24	1017.35	993.15	911.37	869.78	1051.05	1035.24
997.29	978.91	1022.49	1001.77	911.41	911.50	1051.47	1040.72
999.88	997.48	1030.20	1010.40	935.47	911.74	1052.12	1042.58
1001.22	1074.29	1033.29	1011.16	944.76	925.03	1073.24	1093.01
1078.72	1082.63	1093.19	1019.83	1032.90	956.92	1209.29	1204.25
1082.01	1129.38	1104.01	1095.01	1071.29	1058.01	1220.01	1207.49
1134.37	1144.93	1132.90	1102.23	1077.66	1066.22	1237.43	1220.26
1135.98	1156.17	1154.94	1147.78	1080.14	1082.42	1249.73	1238.49
1151.58	1167.14	1161.29	1162.82	1114.63	1103.21	1256.44	1255.71
1164.13	1183.06	1161.73	1167.68	1121.17	1104.41	1258.38	1258.77
1168.31	1216.71	1179.28	1187.84	1128.20	1132.77	1267.51	1279.97
1190.76	1228.02	1197.09	1192.05	1142.13	1140.00	1295.09	1300.84
1246.78	1239.30	1268.65	1232.67	1169.01	1163.14	1360.18	1330.64
1249.13	1240.92	1273.58	1261.22	1181.62	1176.26	1360.94	1352.33
1267.35	1270.44	1274.72	1264.14	1207.89	1193.07	1370.83	1353.23
1274.57	1284.84	1289.72	1282.33	1211.10	1204.39	1408.63	1389.12
1278.02	1303.87	1290.39	1304.58	1241.57	1236.79	1420.68	1418.81
1302.15	1305.01	1322.98	1309.95	1254.52	1239.36	1422.69	1422.61
1312.17	1322.36	1328.64	1324.72	1257.42	1255.89	1428.87	1423.08

Tab. A.8.: Calculated frequencies of TAPP in cm^{-1}, continued.

B3LYP		CAM-B3LYP		DFTB/3OB-freq		LC-DFTB	
ground st.	excited st.	ground st.	excited st.	ground st.	excited st.	ground st.	excited st.
1342.32	1332.00	1360.72	1329.16	1304.49	1291.33	1457.57	1456.13
1350.31	1357.05	1373.10	1342.46	1317.81	1294.34	1509.08	1471.05
1366.36	1358.12	1390.32	1359.39	1337.65	1306.23	1513.62	1503.66
1386.86	1380.35	1410.95	1362.05	1349.42	1330.79	1527.00	1525.40
1387.48	1385.88	1417.75	1390.30	1364.27	1338.07	1571.63	1532.28
1395.38	1409.87	1426.54	1405.21	1373.79	1350.97	1583.61	1565.75
1429.15	1419.42	1446.21	1436.05	1415.47	1365.24	1593.76	1578.88
1430.98	1434.18	1458.37	1441.34	1417.29	1390.86	1613.35	1582.77
1448.61	1442.32	1464.76	1445.02	1430.60	1400.39	1629.96	1590.47
1457.57	1444.13	1499.49	1462.40	1434.65	1423.80	1667.01	1604.40
1497.69	1482.57	1524.21	1471.96	1460.47	1435.36	1670.72	1630.17
1505.07	1485.25	1535.31	1508.72	1488.37	1452.40	1694.35	1666.17
1509.53	1498.91	1541.81	1516.32	1494.30	1462.57	1721.73	1683.53
1510.05	1503.01	1547.53	1537.32	1495.88	1481.51	1722.53	1688.19
1510.62	1536.16	1559.77	1549.90	1509.78	1483.34	1722.94	1691.43
1556.54	1536.57	1603.46	1577.27	1513.38	1518.04	1746.33	1727.93
1562.23	1537.02	1607.04	1578.97	1547.06	1518.22	1774.82	1738.51
1591.15	1564.76	1640.07	1597.40	1560.75	1523.05	1796.08	1748.29
1600.65	1565.71	1652.95	1601.15	1583.31	1527.58	1833.69	1782.93
1624.01	1566.91	1673.02	1611.20	1592.08	1546.85	1845.96	1792.17
1635.22	1612.58	1682.08	1614.14	1601.08	1549.00	1855.85	1797.95
1646.51	1633.68	1702.89	1668.43	1605.13	1576.28	1868.58	1821.18
1659.55	1636.19	1709.02	1670.98	1621.00	1580.96	1881.28	1843.41
3157.94	3155.29	3185.81	3179.82	3001.28	2998.06	3077.42	3070.76
3158.10	3155.68	3185.95	3180.17	3001.73	2998.64	3077.56	3070.89
3187.81	3189.95	3213.37	3216.48	3005.11	3005.29	3105.05	3096.59
3187.89	3190.19	3213.48	3216.58	3005.14	3005.51	3105.47	3097.11
3193.09	3195.32	3218.49	3221.80	3009.13	3008.08	3109.02	3104.64
3193.29	3195.39	3218.71	3222.08	3009.26	3008.18	3109.05	3104.65
3204.11	3205.48	3230.08	3231.62	3018.29	3017.76	3154.87	3157.63
3204.26	3205.49	3230.20	3231.71	3018.52	3018.01	3154.93	3157.72
3218.64	3218.22	3247.42	3245.84	3019.15	3018.35	3155.53	3158.19
3218.92	3218.35	3247.68	3245.89	3019.56	3018.87	3155.53	3158.22

A.1.2. Analysis of States

As observed for TAPP, in the later application of TD-LC-DFTB2 in QM/MM simulations of prodan and 4-AP, it was found that the first excited state of TD-LC-DFTB2 is of different character than the first excited state predicted by DFT. In the previous benchmark, blindly the first excited states were compared to the reference data (except for the TAPP molecule). The question arises if these agree regarding orbital transitions. To clarify this, a comparison of the orbitals involved in the first transitions was done. The B3LYP orbitals were computed with ORCA using the geometries of Refs. [89] and [90]. The plots were created with the orca_plot program and visualized with VMD [100]. LC-DFTB2 orbital plots were created with waveplot and also visualized with VMD. The isovalue was set to ± 0.05, respectively. The ordering of the orbitals is not considered, only that the transition between the same orbitals, i. e. orbitals which look equally, gives the main contribution to the first excited state. The decision, which orbitals are considered as equal, is not strict. Small differences are tolerated and it is mainly looked on the character and symmetry of the main parts.

In Tab. A.9, the results of the comparison of the B3LYP main orbital transitions describing the first excited state to the LC-DFTB main orbital transitions describing the first excited state are shown. The results are classified as follows: (i) there is one orbital transition dominating the first excitation and the involved orbitals obtained with B3LYP and LC-DFTB agree; (ii) the LC-DFTB orbitals differ but are still similar to the B3LYP orbitals; (iii) LC-DFTB describes the first excitation with a different orbital transition and a higher state corresponds to the B3LYP orbital transition describing the first excited state; (iv) there are two or more orbital transitions which contribute similarly to the first excitation acoording to B3LYP and LC-DFTB describes the first excitation with only one of them or the largest weight is assigned to another transition; (v) the B3LYP and LC-DFTB results are completely different.

Molecules of category (iii) were reoptimized in the excited state corresponding to the first excited state of B3LYP. The AEEs and Stokes shifts are presented in Tables A.10 and A.11. The AEEs are logically larger than the ones using S1. The values of the hydroxyquinolines, 2-hydroxypyridine and methyl-4-hydroxycinnamate, which were underestimated before, are now overestimated to a similar extent. Salicylic acid and 7-azaindol instead deviate significantly more than before. The Stokes shifts in contrast were reduced and fit better to the reference data except for salicylic acid. In detail, the Stokes shifts computed with TD-LC-DFTB2/OB2$^{0.3}$ (base) used the optimized geometries in S2 and the absorption and fluorescence energies in S2. They are compared to (a) the B3LYP Stokes shifts obtained on the same TD-LC-DFTB2/OB2$^{0.3}$ (base) geometries but vertical excitation energies in the corresponding state of B3LYP and (b) to the B3LYP Stokes shifts obtained on the reference geometries of Refs. [89, 90], i. e. on the B3LYP geometry of methyl-4-hydroxycinnamate and on the CC2 geometries of the other molecules. The difference between references (a) and (b) are only the geometries. The Stokes shifts are very similar except for salicylic acid, which is encouraging, since it confirms that the TD-LC-DFTB2/OB2$^{0.3}$ (base) geometries are reasonable. The differences arising from the different methods for excitation energy calculation are a bit larger, seen from coloumns two and three in Tab. A.11.

Tab. A.9.: Orbital analysis of first excitation. The results of a comparison between B3LYP and LC-DFTB are sorted in four groups, (i)-(iv), which are defined in the text.

molecule	(i)	(ii)	(iii)	(iv)	(v)
Tetraphenylporphine	●				
Porphycene	●				
Porphine	●				
Chlorin	●				
Tetracene	●				
Perylene	●				
Phenanthrene				●	
Fluorene					●
Carbazole	●				
Dibenzofuran	●				
Pyrrolo[3.2-h]quinoline					●
Tryptamine-A-ph	●				
Tryptamine-A-py	●				
Tryptamine-A-up	●				
Tryptamine-Ph-out	●				
Tryptamine-Ph-up	●				
Tryptamine-Py-out	●				
Tryptamine-Py-up	●				
2,4,6,8-Decatetraene	●				
5-MO-salicylic acid	●				
Salicylic acid			●		
3P-1-propionic acid, gauche				●	
3P-1-propionic acid, anti		●			
1-Naphthol, cis	●				
1-Naphthol, trans	●				
2-Naphthol, cis	●				
2-Naphthol, trans	●				
7-Hydroxyquinoline, cis			●		
7-Hydroxyquinoline, trans			●		
2-Hydroxyquinoline, enol			●		
2-Hydroxyquinoline, keto			●		
o-Dimethoxybenzene, trans	●				
7-Azaindole			●		
Phenylacetylene				●	
Benzonitrile	●				
Aniline	●				
2-Hydroxypyridine			●		
2-Methylpyrimidine	●				
5-Methylpyrimidine	●				
Pyrazine	●				
2-Pyridone	●				

Tab. A.9.: Orbital analysis of first excitation, continued.

molecule	(i)	(ii)	(iii)	(iv)	(v)
Resorcinol, Isomer 2	●				
Resorcinol, Isomer 1	●				
Resorcinol, Isomer 3	●				
Hydroquinone, cis	●				
Hydroquinone, trans	●				
o-Methoxyphenol	●				
m-Methoxyohenol, Isomer 4	●				
m-Methoxyphenol, Isomer 2	●				
p-Methoxyphenol, trans	●				
p-Methoxyphenol, cis	●				
o-Cresol, cis					
o-Cresol, trans	●				
m-Cresol, cis	●				
m-Cresol, trans	●				
p-Cresol					
Acetaldehyde		●			
Acetone		●			
Acrolein		●			
HCOOH					
Aminobenzonitrile	●				
Anisol	●				
Anthracene	●				
Azulene	●				
Benzaldehyde	●				
Benzene		●			
β-Dinaphthyleneoxide					
Biphenyl					
C2H2	●				
Cinnoline	●				
2-Cyclopenten-1-one					
DBH	●				
DCS	●				
Dimethylaminobenzonitrile	●				
DMPD	●				
1-6-Epoxy-10-annulene					
Glyoxal		●			
Hydroquinone	●				
Indole	●				
Methyl-4-hydroxycinnamate			●		
Naphthalene	●				

Tab. A.9.: Orbital analysis of first excitation, continued.

molecule	(i)	(ii)	(iii)	(iv)	(v)
Octatetraene	●				
p-Benzoquinone	●				
p-Diethynylbenzene					
p-Phenylenediamine					
Phenol	●				
Propynal		●			
Pyrene					●
Pyridine		●			
Pyridone-lactim		●			
Pyrimidine		●			
Quinoline		●			
Quinoxaline	●				
Terylene	●				
Tetrazine	●				
Toluene		●			
t-Stilbene	●				

Tab. A.10.: AEEs recomputed for molecules of category (iii) considering the excited state which corresponds to the first excited state of B3LYP. This state was in all cases S2 using the $OB2^{0.3}$(base) parameters. The values are in eV.

molecule	S1	S2	B3LYP reference [89, 90]
Salicylic acid	4.12	4.64	3.89
7-Hydroxyquinoline, cis	3.43	4.30	3.85
7-Hydroxyquinoline, trans	3.40	4.26	3.62
2-Hydroxyquinoline, enol	3.56	4.32	3.95
2-Hydroxyquinoline, keto	3.24	3.91	3.74
7-Azaindole	4.28	4.78	4.17
2-Hydroxypyridine	4.64	5.35	4.83
Methyl-4-hydroxycinnamate	3.43	4.44	4.00

Tab. A.11.: Stokes shifts recomputed for molecules of category (iii) considering the excited state which corresponds to the first excited state of B3LYP. This state was in all cases S2. The $OB2^{0.3}$(base) parameter were used in the LC-DFTB2 calculations. The reference Stokes shifts were computed (a) with B3LYP on the LC-DFTB2 geometries and (b) with B3LYP on the CC2 geometries (set from Ref. [89]) or B3LYP geometries (set from Ref. [90], here: methyl-4-hydroxycinnamate) from Refs. [89, 90]. The values are in eV.

molecule	S1	S2	Ref. (a)	Ref. (b) [89, 90]
Salicylic acid	1.07	0.58	0.40	1.20
7-Hydroxyquinoline, cis	0.96	0.52	0.35	0.35
7-Hydroxyquinoline, trans	0.96	0.53	0.35	0.36
2-Hydroxyquinoline, enol	0.94	0.54	0.36	0.34
2-Hydroxyquinoline, keto	1.15	0.80	0.47	0.51
7-Azaindole	0.75	0.81	0.60	0.78
2-Hydroxypyridine	0.93	0.53	0.34	0.44
Methyl-4-hydroxycinnamate	1.15	0.65	0.31	0.33

Tab. A.12.: Statistical evaluation of the errors for the AEEs in eV, using CC2 as reference for the set of Ref. [89] and B3LYP as reference for the set of Ref. [90]. LC-DFTB2/$OB2^{0.3}$ (base) is abbreviated by LC rep. and LC-DFTB2/comb is abbreviated by LC comb. The states of TD-LC-DFTB2 were selected such that they correspond to the same state as the references.

	Set [89],[a]		Set [89],[b]		Set [90],[b]	
	LC rep.	LC comb.	LC rep.	LC comb.	LC rep.	LC comb.
MAX	0.92	−0.67	0.77	−0.55	1.36	0.80
ME	0.39	−0.04	0.43	0.00	0.28	0.01
MAE	0.45	0.18	0.48	0.14	0.38	0.17

[a] Analysis with respect to CC2.
[b] Analysis with respect to B3LYP.

A.2. QM/MM Simulations: Results for the Ground State

The ground state results not discussed in the main text are reported here and compared to experimental or computational reference data from the literature.

Flugi-2

The experimental absorption spectra of Flugi-2 in different solvents is presented in Ref. [115]. The shown absorption bands in the energy range between 335 nm and 435 nm are slightly solvent dependent. The absorption maximum in DMSO, located at 387 nm (3.20 eV), is a little red shifted compared to the maximum in decane (at 3.26 eV [114]). The band in DMSO shows a shoulder at the blue side of the maximum. In decane, at the corresponding place a small second peak is observed and an additional shoulder further on the blue side. In general, the absorption spectrum in decane is much broader than the spectra in the other solvents and the maximum has a little lower intensity than the others.

In Ref. [114], theoretical calculations of the absorption are shortly discussed. Flugi-2 was optimized in ground state at different levels of theory (CC2, LC-DFT/ωB97X, DFT/B3LYP) and vertical excitation energies were computed on the respective optimized structures with various methods. The calculations were done in vacuum, which is expected to be comparable to the nonpolar decane. CC2 for both optimisation and excitation energy calculation, provided an absorption energy of 3.22 eV and thus could almost reproduce the experimental value. Also B3LYP came very close to the experimental value with a predicted absorption energy of 3.29. The range-separated ωB97X functional in contrast overestimated it. On the structure optimized with ωB97X the excitation energy amounted to 4.17 eV, on the CC2 geometry it was 3.88 eV. The excitation energy of OM2 computed on the geometry optimised with CC2 amounted to 3.57 eV.

Here, with TD-LC-DFTB2/OB2 (base), a value of 3.54 eV is predicted. If the excitation energy is computed with the pure electronical parameter set on the geometry optimized with the OB2 (base) parameters, then the same value as with CC2 (3.22 eV) is obtained. The oscillator strength from TD-LC-DFTB2, however, is a little lower (0.252 compared to 0.478 [114]). The excitation energy from OM2 on the TD-LC-DFTB2/OB2 (base) geometry is 3.59 eV and the corresponding oscillator strength amounts to 0.294.
Excitation energies were computed with OM2 and TD-LC-DFTB2 with both parameter sets on the geometries sampled in the ground state simulation in vacuum. Absorption spectra were modelled by Gaussian fits to the resulting histograms of the excitaiton energies. Only the first excitation energies were taken into account since they had a significantly larger oscillator strength than the second excitation energies. Higher excited states have energies definitely outside of the investigated range of the absorption spectra. The obtained maxima from TD-LC-DFTB are the same as the excitation energies on the optimized geometry and the OM2 result is very close to it, too (Tab. A.13).

There is no more theoretical reference data for the absorption energies in DMSO. According to the experiment, the values should slightly shift to the red. This is not visible in the results here. Instead, the values shifted in the wrong direction (Tab. A.13). However, the differences are small.

Tab. A.13.: Absorption energies of Flugi-2. Reported are the maxima of Gaussian fits to histograms of the excitation energies computed with different methods on the geometries sampled with TD-LC-DFTB2 in vacuum or in DMSO. The values are in eV.

method for exc. energy	vacuum	DMSO
TD-LC-DFTB2/OB2 (base)	3.54	3.56
TD-LC-DFTB2/el. params	3.22	3.27
OM2	3.56	3.62

Prodan

The absorption spectra of prodan in different solvents are reported in Ref. [117]. Since the absorption band is very broad and does not show a clear peak, the authors provide λ'_{max} values, which show the average energy at 90 % of the intensity. The solvent shifts determined from these values are summarized in Tab. A.14 together with the results from the ground state QM/MM simulations. The excitation energies with the largest oscillator strengths from the first two excitations were used for the histograms, respectively. The resulting solvent shifts are of the right magnitude, but the shift from water is larger predicted than the shift from DMSO which different to the experiment. In particular, the DMSO shift is underestimated. Since the experimental spectra are broad, the results are acceptable.

Tab. A.14.: Absorption energies of prodan in different solvents. Reported are the maxima of Gaussian fits to histograms of the excitation energies computed with TD-LC-DFTB2 on the geometries sampled with the same method and OB2 (base) parameters. Solvent shifts are given with respect to vacuum (cyclohexane in exp.). The values are in eV.

method for exc. energy	vac/cyhex	DMSO	water	Δ DMSO	Δ water
TD-LC-DFTB2/OB2 (base)	3.89	3.75	3.67	-0.14	-0.22
TD-LC-DFTB2/el. params	3.50	3.38	3.32	-0.12	-0.18
experiment [117]	3.63	3.43	3.46	-0.20	-0.17

It was further checked if there are planar and non-planar species in the simulated ensemble. The dimethylamino moiety fluctuates around the planar geometry. The carbonyl moiety can flip by 180° in the alternative planar conformation but twisted states were not observed, neither in vacuum nor DMSO or water.

Finally, the vertical excitation energies are shown which were computed with different methods using the geometries optimized with CAM-B3LYP and TD-LC-DFTB2 in ground state (Tables A.15 and A.16).

4-Aminophthalimide

Experimental absoprtion maxima are reported in Ref. [180] and compared to the theoretical results of this work in Tab. A.17. The excitation energies with the largest oscillator

Tab. A.15.: Excitation energies of prodan optimized with TD-LC-DFTB2/OB2 (base) in vacuum computed with different methods. The values are in eV, oscillator strengths are shown in parentheses.

	TD-LC-DFTB2			
	OB2 (base)	el. params	B3LYP	CAM-B3LYP
S1	3.69 (4.6×10^{-5})	3.54 (0.17)	3.59 (0.29)	4.00 (0.11)
S2	3.93 (0.22)	3.68 (7.2×10^{-5})	3.72 (0.0063)	4.02 (0.026)
S3	4.54 (0.076)	4.06 (0.062)	3.76 (0.14)	4.19 (0.35)

Tab. A.16.: Excitation energies of prodan optimized with CAM-B3LYP in vacuum computed with different methods. The values are in eV.

	TD-LC-DFTB2			
	OB2 (base)	el. params	B3LYP	CAM-B3LYP
S1	3.83 (5.1×10^{-5})	3.69 (0.17)	3.70 (0.32)	4.13 (0.049)
S2	4.10 (0.23)	3.80 (4.5×10^{-5})	3.84 (0.0024)	4.15 (0.13)
S3	4.72 (0.080)	4.23 (0.069)	3.90 (0.077)	4.32 (0.28)

strenghts from the first three transitions were considered. As for prodan, the experiment shows a smaller colour shift in the absorption spectrum for water than for DMSO. This is not reproduced by TD-LC-DFTB2, independently of the paramter set. The predicted shift from DMSO is close to the experimental shift but the shift due to water solvent is overestimated by more than 0.2 eV.

Tab. A.17.: Absorption energies of 4-AP in different solvents. Reported are the maxima of Gaussian fits to histograms of the excitation energies computed with TD-LC-DFTB2 on the geometries sampled with the same method and OB2 (base) parameters. Solvent shifts are given with respect to vacuum (hexane in exp.). The values are in eV.

method for exc. energy	vac/hex	DMSO	water	Δ DMSO	Δ water
TD-LC-DFTB2/OB2 (base)	4.48	4.18	3.90	-0.30	-0.58
TD-LC-DFTB2/el. params	4.12	3.83	3.59	-0.29	-0.53
experiment [180]	3.67	3.32	3.36	-0.35	-0.31

Vertical excitation energies computed with different methods using the geometries optimized with CAM-B3LYP and TD-LC-DFTB2 in ground state are reported in Tables A.18 and A.19).

Tab. A.18.: Excitation energies of 4-AP optimized with TD-LC-DFTB2/OB2 (base) in vacuum computed with different methods. The values are in eV, oscillator strengths are shown in parentheses.

	TD-LC-DFTB2			
	OB2 (base)	el. params	B3LYP	CAM-B3LYP
S1	3.69 (1.6×10^{-5})	3.78 (3.1×10^{-5})	3.74 (0.066)	4.16 (0.076)
S2	4.11 (1.5×10^{-6})	4.10 (1.7×10^{-4})	3.77 (0.0074)	4.20 (0.0018)
S3	4.50 (0.079)	4.14 (0.068)	4.38 (2.0×10^{-5})	4.72 (0.021)

Tab. A.19.: Excitation energies of prodan optimized with CAM-B3LYP in vacuum computed with different methods. The values are in eV.

	TD-LC-DFTB2			
	OB2 (base)	el. params	B3LYP	CAM-B3LYP
S1	3.92 (1.6×10^{-5})	3.99 (5.5×10^{-5})	3.87 (0.076)	4.30 (0.081)
S2	4.37 (1.2×10^{-5})	4.34 (0.013)	3.97 (5.0×10^{-4})	4.40 (5.0×10^{-4})
S3	4.73 (0.086)	4.35 (0.060)	4.58 (0.019)	4.86 (0.037)

A.3. Optical Glucose Sensor: Partial Charges

The atomic partial charges derived for the glucose molecule in the binding pocket of the wildtype are plotted in Fig. A.1 for different cut-off radii for the environmental point charges and alternatively the point charges of complete residues in the proximity of glucose. Obviously, the amount of environment considered in the calculation of the ESP plays a minor role in the range between 4 to 8 Å. The charges of the same atoms are very similar and do not seem to converge with an increase of the cut-off radius. The detailed values are reported in Tab. A.20. A picture of the glucose with the atom labels is shown in the main text in Fig. 6.6.

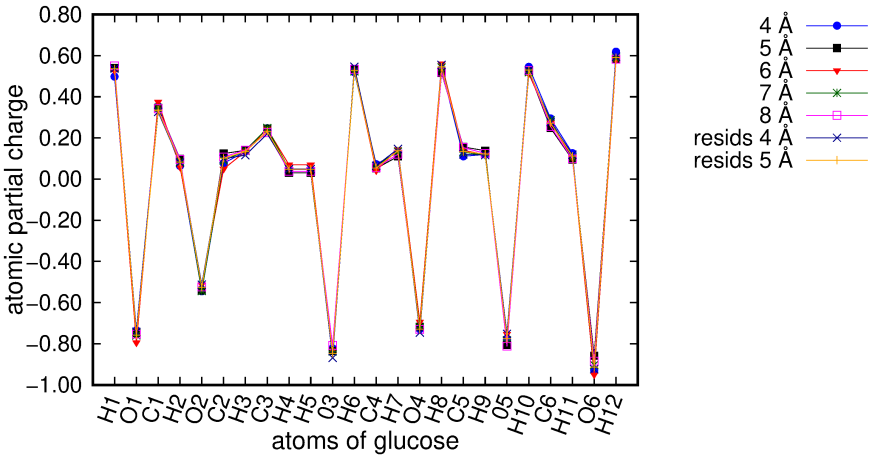


Fig. A.1.: Partial charges of glucose atoms polarized in the binding pocket of the wildtype by a different amount of environment.

Tab. A.20.: Atomic partial charges computed for the glucose molecule in the binding pocket of the wild-type. The ESP was calulated considering the environment as point charges within different distances from the glucose/ belonging to the closest residues as explained in the computational details part of chapter 6.

atom	4 Å	5 Å	6 Å	7 Å	8 Å	resids 4 Å	resids 5 Å
H1	0.498	0.540	0.527	0.537	0.550	0.537	0.536
O1	−0.735	−0.747	−0.794	−0.759	−0.754	−0.742	−0.758
C1	0.353	0.347	0.375	0.346	0.342	0.327	0.335
H2	0.064	0.098	0.059	0.085	0.095	0.080	0.083
O2	−0.545	−0.542	−0.512	−0.544	−0.524	−0.513	−0.525
C2	0.077	0.124	0.050	0.090	0.108	0.099	0.102
H3	0.134	0.140	0.134	0.137	0.141	0.116	0.134
C3	0.247	0.245	0.244	0.248	0.226	0.220	0.232
H4	0.046	0.031	0.070	0.048	0.038	0.049	0.048
H5	0.046	0.031	0.070	0.048	0.038	0.049	0.048
O3	−0.822	−0.835	−0.832	−0.837	−0.808	−0.869	−0.847
H6	0.519	0.533	0.529	0.524	0.534	0.546	0.529
C4	0.072	0.052	0.042	0.064	0.055	0.065	0.054
H7	0.131	0.110	0.145	0.133	0.121	0.147	0.138
O4	−0.729	−0.717	−0.694	−0.720	−0.731	−0.746	−0.729
H8	0.536	0.516	0.561	0.542	0.520	0.555	0.545
C5	0.111	0.156	0.140	0.131	0.153	0.120	0.134
H9	0.120	0.138	0.119	0.116	0.125	0.117	0.122
O5	−0.779	−0.806	−0.753	−0.783	−0.812	−0.752	−0.775
H10	0.545	0.522	0.512	0.519	0.528	0.528	0.530
C6	0.295	0.248	0.273	0.286	0.256	0.277	0.272
H11	0.126	0.095	0.105	0.105	0.099	0.119	0.112
O6	−0.930	−0.858	−0.950	−0.903	−0.884	−0.924	−0.913
H12	0.619	0.581	0.580	0.586	0.583	0.595	0.591

A.4. Exciton Transfer in the FMO Complex: Couplings

The couplings computed with TD-LC-DFTB2 along the trajectory of the simulation with constraints on bonds involving hydrogen are shown in Figs. A.2 and A.3. The fluctuations during the simulation shown in Fig. A.3 are important to understand the scatter plots shown in the main text (Fig. 7.16). The phytyl tail and the environment were negelcted in the calculations.

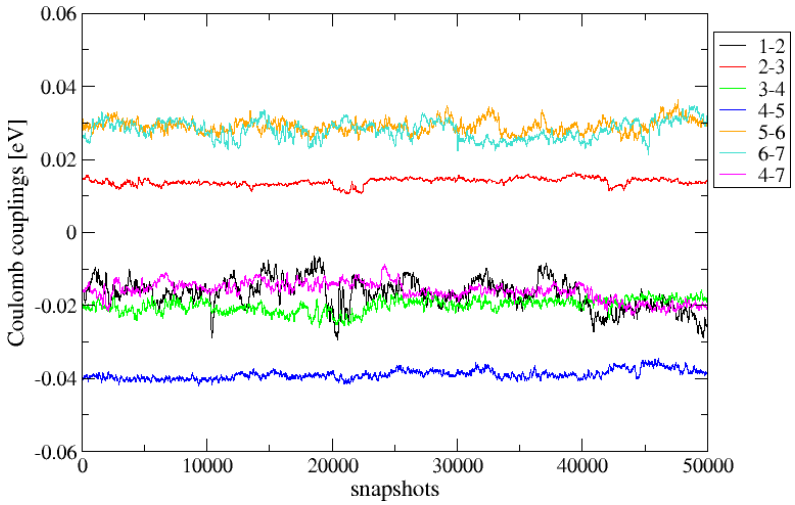


Fig. A.2.: Coulomb couplings in vacuum obtained with TD-LC-DFTB2 along the first 50 ns of the trajectory for all important pigment pairs. The running averages over 100 data points are shown.

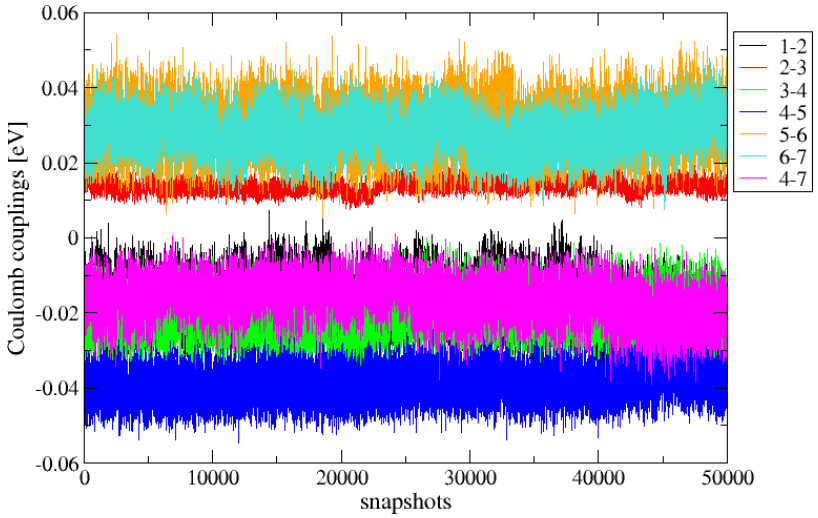


Fig. A.3.: Coulomb couplings in vacuum obtained with TD-LC-DFTB2 along the first 50 ns of the trajectory for all important pigment pairs. The values were not averaged to make the fluctuations visible.

Bibliography

[1] L. Tähkämö, T. Partonen, and A.-K. Pesonen. Systematic review of light exposure impact on human circadian rhythm. *Chronobiol. Int.*, 36(2):151–170, 2019.

[2] M. G. Neumann, C. C. Schmitt, G. C. Ferreira, and I. C. Corrêa. The initiating radical yields and the efficiency of polymerization for various dental photoinitiators excited by different light curing units. *Dent. Mater.*, 22:576–584, 2005.

[3] J. J. Kranz, M. Elstner, B. Aradi, T. Frauenheim, V. Lutsker, A. D. Garcia, and T. A. Niehaus. Time-Dependent Extension of the Long-Range Corrected Density Functional Based Tight-Binding Method. *J. Chem. Theory Comput.*, 13(4):1737–1747, 2017.

[4] M. Sokolov, B. M. Bold, J. J. Kranz, S. Höfener, T. A. Niehaus, and M. Elstner. 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. *J. Chem. Theory Comput.*, 17(4):2266–2282, 2021.

[5] F. Khan, T. E. Saxl, and J. C. Pickup. Fluorescence intensity- and lifetime-based glucose sensing using an engineered high-Kd mutant of glucose/galactose-binding protein. *Anal. Biochem.*, 399:39–43, 2010.

[6] L. L. E. Salins, R. A. Ware, C. M. Ensor, and S. Daunert. A Novel Reagentless Sensing System for Measuring Glucose Based on the Galactose/Glucose-Binding Protein. *Anal. Biochem.*, 294:19–26, 2001.

[7] R. M. De Lorimier, J. J. Smith, M. A. Dwyer, L. L. Looger, K. M. Sali, C. D. Paavola, S. S. Rizk, S. Sadigov, D. W. Conrad, L. Loew, and H. W. Hellinga. Construction of a fluorescent biosensor family. *Protein Sci.*, 11:2655–2675, 2002.

[8] V. Scognamiglio, V. Aurilia, N. Cennamo, P. Ringhieri, L. Iozzino, M. Tartaglia, M. Staiano, G. Ruggiero, P. Orlando, T. Labella, L. Zeni, A. Vitale, and S. D'Auria. D-galactose/D-glucose-binding Protein from *Escherichia coli* as Probe for a Non-consuming Glucose Implantable Fluorescence Biosensor. *Sensors*, 7:2484–2491, 2007.

[9] F. Khan, L. Gnudi, and J. C. Pickup. Fluorescence-based sensing of glucose using engineered glucose/galactose-binding protein: A comparison of fluorescence resonance energy transfer and environmentally sensitive dye labelling strategies. *Biochem. Bioph. Res. Co.*, 365:102–106, 2008.

[10] J.-H. Chen, H. Wu, C. Xu, X.-C. Liu, Z. Huang, S. Chang, W. Wang, G. Han, T. Kuang, J.-R. Shen, and X. Zhang. Architecture of the photosynthetic complex froma green sulfur bacterium. *Science*, 370:931–938, 2020.

[11] D. Wöhrle, M. W. Tausch, and W.-D. Stohrer. *Photochemie: Konzepte, Methoden, Experimente*. WILEY-VCH Verlag Gmbh, D-69469 Weinheim, 1st edition, 1998.

[12] P. Atkins and R. Friedman. *Molecular Quantum Mechanics*. Wiley-VCH, Weinheim, 5th edition, 2013.

[13] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer Science+Business Media, LCC, 3rd edition, 2006.

[14] G. Saroja, T. Soujanya, B. Ramachandram, and A. Samanta. 4-Aminophthalimide Derivatives as Environment-Sensitive Probes. *J. Fluoresc.*, 8:405–410, 1998.

[15] C. C. Vequi-Suplicy, Y. Orozco-Gonzales, M. T. Lamy, S. Canuto, and K. Coutinho. A new interpretation of the absorption and the dual fluorescence of Prodan in solution. *J. Chem. Phys.*, 153:244104–1–244104–13, 2020.

[16] A. Nakajima. Solvent Effect on the Vibrational Structures of the Fluorescence and Absorption Spectra of Pyrene. *Bull. Chem. Soc. Japan*, 44:3272–3277, 1971.

[17] M. A. J. Rodgers. Picosecond Fluorescence Studies of Xanthene Dyes in Anionic Micelles in Water and Reverse Micelles in Heptane. *J. Phys. Chem.*, 85:3372–3374, 1981.

[18] E. Krystkowiak, K. Dobek, and A. Maciejewski. Origin of the strong effect of protic solvents on the emission spectra, quantum yield of fluorescence and fluorescence lifetime of 4-aminophthalimide Role of hydrogen bonds in deactivation of S1-4-aminophthalimide. *J. Photochem. Photobiol. A*, 184:250–264, 2006.

[19] A. V. Fonin, S. A. Silonov, I. A. Antifeeva, O. V. Stepanenko, O. V. Stepanenko, A. S. Fefilova, O. I. Povarova, A. A. Gavrilova, I. M. Kuznetsova, and K. K. Turoverov. Photophysical Properties of BADAN Revealed in the Study of GGBP Structural Transitions. *Int. J. Mol. Sci.*, 22:11113, 2021.

[20] R. Crespo-Otero, Q. Li, and L. Blancafort. Exploring Potential Energy Surfaces for Aggregation-Induced Emission-From Solution to Crystal. *Chem. Asian J.*, 14:700–714, 2019.

[21] W. Rettig. Charge Separation in Excited States of Decoupled Systems-TICT Compounds and Implications Regarding the Development of New Laser Dyes and the Primary Processes of Vision and Photosynthesis. *Angew. Chem. Int. Ed. Engl.*, 25:971–988, 1986.

[22] A. V. Fonin, I. M. Kuznetsova, and K. K. Turoverov. Spectral properties of BADAN in solutions with different polarities. *J. Mol. Struct.*, 1090:107–111, 2015.

[23] B. Mennucci and C. Curutchet. The role of the environment in electronic energy transfer: a molecular modeling perspective. *Phys. Chem. Chem. Phys.*, 13:11538–11550, 2011.

[24] T. Renger and F. Müh. Understanding photosynthetic light-harvesting:a bottom up theoretical approach. *Phys. Chem. Chem. Phys.*, 15:3348–3371, 2013.

[25] J. Adolphs, F. Müh snd M. Madjet, and T. Renger. Calculation of pigment transition energies in the FMO protein. *Photosynth. Res.*, 95:197–209, 2008.

[26] D. Hayes and G. S. Engel. Extracting the Excitonic Hamiltonian of the Fenna-Matthews-Olson Complex Using Three-Dimensional Third-Order ELectronic Spectroscopy. *Biophys. J.*, 100:2043–2052, 2011.

[27] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory.* Wiley-VCH, Weinheim, 2nd edition, 2001.

[28] J. Reinhold. *Quantentheorie der Moleküle.* Springer Fachmedien Wiesbaden, 5th edition, 2015.

[29] F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon, and M. Rapacioli. Density-functional tight-binding: basic concepts and applications to molecules and clusters. *Advances in Physics: X*, 5(1):1710252, 2020.

[30] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864–B871, 1964.

[31] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[32] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, 1992.

[33] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.

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

[35] J. P. Perdew. Density-functional approximation for the correlation-energy of the inhomogenous electron gas. *Phys. Rev. B*, 33:8822–8824, 1986.

[36] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785, 1988.

[37] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[38] P. J. Stephens, F. J. Delvin, C. F. Chabalowski, and M. J. Frisch. Ab initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, 98:11623, 1994.

[39] T. Yanai, D. P. Tew, and N. C. Handy. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393:51–57, 2004.

[40] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.*, 115:3540–3544, 2001.

[41] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao. A long-range-corrected time-dependent density functional theory. *J. Chem. Phys.*, 120:8425–8433, 2004.

[42] E. Runge and E. K. U. Gross. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.*, 52(12):997–1000, 1984.

[43] R. van Leeuwen. Mapping from Densities to Potentials in Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.*, 82(19):3863–3866, 1999.

[44] C. A. Ullrich. *Time-Dependent Density Functional Theory.* Oxford University Press, New York, 1st edition, 2012.

[45] M. A. L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, and E. K. U. Gross. *Time-Dependent Density Functional Theory.* Springer-Verlag, Berlin Heidelberg, 1st edition, 2006.

[46] M. Elstner. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.*, 116:316–325, 2006.

[47] M. Gaus, A. Goez, and M. Elstner. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.*, 9(1):338–354, 2013.

[48] M. Gaus, Q. Cui, and M. Elstner. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, 2011.

[49] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haug, T. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

[50] W. M. C. Foulkes and R. Haydock. Tight-binding models and density-functional theory. *Phys. Rev. B*, 39(17):12520–12536, 1989.

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

[52] J. C. Slater and G. F. Koster. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.*, 94(6):1498–1524, 1954.

[53] M. Gaus, C.-P. Chou, H. Witek, and M. Elstner. Automatized Parametrization of SCC-DFTB Repulsive Potentials: Application to Hydrocarbons. *J. Phys. Chem. A*, 113:11866–11881, 2009.

[54] V. Lutsker, B. Aradi, and T. A. Niehaus. Implementation and benchmark of a long-range corrected functional in the density functional based tight-binding method. *J. Chem. Phys.*, 143:184107–1–184107–14, 2015.

[55] R. Baer and D. Neuhauser. Density Functional Theory with Correct Long-Range Asymptotic Behavior. *Phys. Rev. Lett.*, 94:043002–1–043002–4, 2005.

[56] E. Livshits and R. Baer. A well-tempered density functional theory of electrons in molecules. *Phys. Chem. Chem. Phys.*, 9:2932–2941, 2007.

[57] T. A. Niehaus, S. Suhai, F. Della Sala, P. Lugli, M. Elstner, G. Seifert, and Th. Frauenheim. Tight-binding approach to time-dependent density-functional response theory. *Phys. Rev. B*, 63:085108, Feb 2001.

[58] Fabio Trani, Giovanni Scalmani, Guishan Zheng, Ivan Carnimeo, Michael J. Frisch, and Vincenzo Barone. Time-Dependent Density Functional Tight Binding: New Formulation and Benchmark of Excited States. *J. Chem. Theory Comput.*, 7(10):3304–3313, 2011.

[59] A. Domínguez, B. Aradi, T. Frauenheim, V. Lutsker, and T. A. Niehaus. Extensions of the Time-Dependent Density Functional Based Tight-Binding Approach. *J. Chem. Theory Comput.*, 9(11):4901–4914, 2013.

[60] Y. Nishimoto. Time-dependent density-functional tight-binding method with the third-order expansion of electron density. *J. Chem. Phys.*, 143:094108–1–094108–11, 2015.

[61] F. Jensen. *Introduction to Computational Chemistry.* WILEY, England, 3rd edition, 2017.

[62] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Models.* John Wiley & Sons Ltd, England, 2nd edition, 2004.

[63] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[64] H. M. Senn and Q. Thiel. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.*, 48:1198–1229, 2009.

[65] X. Wang, S. Ramirez-Hinestrosa, J. Dobnikar, and D. Frenkel. The Lennard-Jones potential: when (not) to use it. *Phys. Chem. Chem. Phys.*, 22:10624–10633, 2020.

[66] E. G. Lewars. *Computational Chemistry.* Springer International Publishing AG Switzerland, 3rd edition, 2016.

[67] U. H. E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.*, 281:140–150, 1997.

[68] H. Fukunishi, O. Watanabe, and S. Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.*, 116(20):9058–9067, 2002.

[69] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109:7737–7744, 1998.

[70] P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman. Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science*, 235:574–576, 1987.

[71] G. M. Torrie and J. P. Valleau. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.*, 23:187–199, 1977.

[72] J. R. Gullingsrud, R. Braun, and K. Schulten. Reconstructing Potentials of Mean Force through Time Series Analysis of Steered Molecular Dynamics Simulations. *J. Comput. Phys.*, 151:190–211, 1999.

[73] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. *WIREs Comput. Mol. Sci.*, 1:826–843, 2011.

[74] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci.*, 99(20):12562–12566, 2002.

[75] L. Sutto, S. Marsili, and F. L. Gervasio. New advances in metadynamics. *WIREs Comput. Mol. Sci.*, 2:771–779, 2012.

[76] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.

[77] A. Barducci, G. Bussi, and M. Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.*, 100:020603–1–020603–4, 2008.

[78] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J. Comput. Chem.*, 30:1615–1621, 2009.

[79] T. Negami, K. Shimizu, and T. Terada. Coarse-grained molecular dynamics simulation of protein conformational change coupled to ligand binding. *Chem. Phys. Lett.*, 742:137144, 2020.

[80] M. Kim and A. E. Cho. The role of water molecules in stereoselectivity of glucose/galactose-binding protein. *Sci. Rep.*, 6:36807–1–36807–10, 2016.

[81] B. M. Bold, M. Sokolov, S. Maity, M. Wanko, P. M. Dohmen, J. J. Kranz, U. Kleinekathöfer, S. Höfener, and M. Elstner. Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes. *Phys. Chem. Chem. Phys.*, 22:10500–10518, 2020.

[82] E. Brunk and U. Rothlisberger. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.*, 115(12):6217–6263, 2015.

[83] R. S. Mulliken. Electronic Population Analysis on LCAO-MO Molecular Wave Functions.I. *J. Chem. Phys.*, 23(10):1833–1840, 1955.

[84] A. Géron and K. Rother. *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow : Konzepte, Tools und Techniken für intelligente Systeme.* Animals. O'Reilly, 1st edition, 2018.

[85] A. Y. Wang, R. J. Murdock, S. K. Kauwe, A. O. Olinyk, A. Gurlo, J. Brgoch, K. A. Persson, and T. D. Sparks. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chem. Mater.*, 32:4954–4965, 2020.

[86] R.-X. Chen, A. J. A. Aquino, A. C.-H. Sue, T. Niehaus, and H. Lischka. Characterization of Charge Transfer in Excited States of Extended Clusters of pi-Stacked Donor and Acceptor Complexes in Lock-Arm-Supramolecular Ordering. *J. Phys. Chem. A*, 123:4532–4542, 2019.

[87] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery Jr. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.*, 14:1347–1363, 1993.

[88] B. Hourahine, B. Aradi, V. Blum, et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.*, 152(12):124101, 2020.

[89] N. O. C. Winter, N. K. Graf, S. Leutwyler, and C. Hättig. Benchmarks for 0-0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data. *Phys. Chem. Chem. Phys.*, 15:6623–6630, 2013.

[90] R. Send, M. Kühn, and F. Furche. Assessing Excited State Methods by Adiabatic Excitation Energies. *J. Chem. Theory Comput.*, 7:2376–2386, 2011.

[91] M. T. do N. Varella, L. Stojanović, V. Q. Vuong, S. Irle, T. A. Niehaus, and M. Barbatti. How the Size and Density of Charge-Transfer Excitons Depend on Heterojunction's Architecture. *J. Phys. Chem. C*, 125(10):5458–5474, 2021.

[92] V. Q. Vuong, J. Akkarapattiakal Kuriappan, M. Kubillus, J. J. Kranz, T. Mast, T. A. Niehaus, S. Irle, and M. Elstner. Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules. *J. Chem. Theory Comput.*, 14:115–125, 2018.

[93] J.-D. Chai and M. Head-Gordon. Systematic optimization of long-rangecorrected hybrid density functionals. *J. Chem. Phys.*, 128:084106–1–084106–15, 2008.

[94] 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. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005.

[95] F. Neese. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2:73–78, 2012.

[96] S. Hirata and M. Head-Gordon. Time-dependent density functional theory within the tamm–dancoff approximation. *Chem. Phys. Lett.*, 314:291–299, 1999.

[97] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials (Chem. Phys. Letters 240 (1995) 283) (PII:0009-2614(95)00621-4). *Chem. Phys. Lett.*, 242(6):652–660, 1995.

[98] F. Weigend. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.*, 8(9):1057–1065, 2006.

[99] F. Neese, F. Wennmohs, A. Hansen, and U. Becker. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.*, 356(1-3):98–109, 2009.

[100] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[101] M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, and M. Elstner. Calculating Absorption Shifts for Retinal Proteins: Computational Challenges. *J. Phys. Chem. B*, 109:3606–3615, 2005.

[102] M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner. Computational photochemistry of retinal proteins. *J. Comput. Aided Mol. Des.*, 20:511–518, 2006.

[103] S. Knecht, C. M. Marian, J. Kongsted, and B. Mennucci. On the Photophysics of Carotenoids: A Multireference DFT Study of Peridinin. *J. Phys. Chem. B*, 117:13808–13815, 2013.

[104] N. Schieschke, B. M. Bold, P. M. Dohmen, D. Wehl, M. Hoffmann, A. Dreuw, M. Elstner, and S. Höfener. Geometry dependence of excitonic couplings and the consequences for configuration-space sampling. *J. Comput. Chem.*, 42(20):1402–1418, 2021.

[105] Y. Nishimoto. Time-dependent long-range-corrected density-functional tight-binding method combined with the polarizable continuum model. *J. Phys. Chem. A*, 123(26):5649–5659, 2019.

[106] V. Stehr, R. F. Fink, B. Engels, J. Pflaum, and C. Deibel. Singlet Exciton Diffusion in Organic Crystals Based on Marcus Transfer Rates. *J. Chem. Theory Comput.*, 10:1242–1255, 2014.

[107] KOALA, an ab-initio electronic structure program, written by S. Höfener, with contributions from A.-S. Hehn, J. Heuser, and N. Schieschke.

[108] S. Höfener. Coupled-cluster frozen-density embedding using resolution of the identity methods. *J. Comput. Chem.*, 35(23):1716–1724, 2014.

[109] S. Höfener. The KOALA program: Wavefunction frozen-density embedding. *Int. J. Quantum Chem.*, n/a(n/a):e26351, 2020.

[110] R. Berger, C. Fischer, and M. Klessinger. Calculation of the Vibronic Fine Structure in Electronic Spectra at Higher Temperatures. 1. Benzene and Pyrazine. *J. Phys. Chem. A*, 102:7157–7167, 1998.

[111] B. Moore, A. Charaf-Eddin, A. Planchat, C. Adamo, J. Autschbach, and D. Jacquemin. Electronic Band Shapes Calculated with Optimally Tuned Range-Separated Hybrid Functionals. *J. Chem. Theory Comput.*, 10:4599–4608, 2014.

[112] S. Höfener, B. A. R. Günther, M. E. Harding, and L. H. Gade. Understanding UV-Vis Spectra of Halogenated Tetraazaperopyrenes (TAPPs): A Computational Study. *J. Phys. Chem. A*, 123:3160–3169, 2019.

[113] R. Rüger, T. Niehaus, E. van Lenthe, T. Heine, and L. Visscher. Vibrationally resolved uv/vis spectroscopy with time-dependent density functional based tight binding. *J. Chem. Phys.*, 145:184102–1–184102–13, 2016.

[114] F. E. Wolff, S. Höfener, M. Elstner, and T. A. Wesolowski. Origin of the Solvatochromism in Organic Fluorophores with Flexible Side Chains: A Case Study of Flugi-2. *J. Phys. Chem. A*, 123:4581–4587, 2019.

[115] O. N. Burchak, L. Mugherli, M. Ostuni, J. J. Lacapère, and M. Y. Balakirev. Combinatorial Discovery of Fluorescent Pharmacophores by Multicomponent Reactions in Droplet Arrays. *J. Am. Chem. Soc.*, 133:10058–10061, 2011.

[116] A. B. J. Parusel, W. Nowak, S. Grimme, and G. Köhler. Comparative Theoretical Study on Charge-Transfer Fluorescence Probes: 6-Propanoyl-2-(N,N-dimethylamino)naphthalene and Derivatives. *J. Phys. Chem. A*, 102:7149–7156, 1998.

[117] J. Catalan, P. Perez, J. Laynez, and F. G. Blanco. Analysis of the Solvent Effect on the Photophysics Properties of 6-Propionyl-2-(dimethylamino)naphthalene (PRODAN). *J. Fluoresc.*, 1(4):215–223, 1991.

[118] G. Weber and F. J. Farris. Synthesis and Spectral Properties of a Hydrophobic Fluorescent Probe: 6-Propionyl-2-(dimethylamino)naphthalene. *Biochemistry*, 18(14):3075–3078, 1979.

[119] J. B. Massey, H. S. She, and H. J. Pownall. Interfacial Properties of Model Membranes and Plasma Lipoproteins Containing Ether Lipids. *Biochemistry*, 24:6973–6978, 1985.

[120] J. Zeng and P. L.-G. Chong. Interactions between Pressure and Ethanol on the Formation of Interdigitated DPPC Liposomes: A Study with Prodan Fluorescence. *Biochemistry*, 30:9485–9491, 1991.

[121] H. Rottenberg. Probing the Interactions of Alcohols with Biological Membranes with the Fluorescent Probe Prodan. *Biochemistry*, 31:9473–9481, 1992.

[122] H. A. Wilson-Ashworth, Q. Bahm, J. Erickson, A. Shinkle, M. P. Vu, D. Woodbury, and J. D. Bell. Differential Detection of Phospholipid Fluidity, Order, and Spacingby Fluorescence Spectroscopy of Bis-pyrene, Prodan, Nystatin,and Merocyanine 540. *Biophys. J.*, 91:4091–4101, 2006.

[123] M. Stawinoga, M. Sleiman, J. Chastain, and C. Richard. Usefulness of Fluorescent Probe Prodan To Gain Insight into thePolarity of Plant Cuticles. *J. Agric. Food Chem.*, 63:6932–6938, 2015.

[124] T. Saxl, F. Khan, D. R. Matthews, Z.-L. Zhi, O. Rolinski, S. Ameer-Beg, and J. Pickup. Fluorescence lifetime spectroscopy and imaging of nano-engineered glucose sensor microcapsules based on glucose/galactose-binding protein. *Biosens. Bioelectron.*, 24:3229–3234, 2009.

[125] T. Saxl, F. Khan, M. Ferla, D. Birch, and J. Pickup. A fluorescence lifetime-based fibre-optic glucose sensor using glucose/galactose-binding protein. *Analyst*, 136:968–972, 2011.

[126] N. Helassa, J. P. Garnett, M. Farrant, F. Khan, J. C. Pickup, K. M. Hahn, C. J. MacNevin, R. Tarran, and D. L. Baines. A novel fluorescent sensor protein for detecting changes in airway surface liquid glucose concentration. *Biochem. J.*, 464:213–220, 2014.

[127] C. Tiangco, D. Fon, N. Sardesai, Y. Kostov, F. Sevilla-III, G. Rao, and L. Tolosa. Fiber optic biosensor for transdermal glucose based on the glucose binding protein. *Sensors Actuat. B-Chem.*, 242:569–576, 2017.

[128] J. S. Marvin and H. W. Hellinga. Engineering Biosensors by Introducing Fluorescent Allosteric Signal Transducers: Construction of a Novel Glucose Sensor. *J. Am. Chem. Soc.*, 120:7–11, 1998.

[129] X. Ge, G. Rao, and L. Tolosa. On the Possibility of Real-Time Monitoring of Glucose in Cell Culture by Microdialysis Using a Fluorescent Glucose Binding Protein Sensor. *Biotechnol. Progr.*, 24:691–697, 2008.

[130] X. Ge, G. Rao, Y. Kostov, S. Kanjananimmanont, R. M. Viscardi, H. Woo, and L. Tolosa. Detection of Trace Glucose on the Surface of a Semipermeable Membrane Using a Fluorescently Labeled Glucose-Binding Protein: A Promising Approach to Noninvasive Glucose Monitoring. *J. Diabetes Sci. Technol.*, 7:4–12, 2013.

[131] S. Kanjananimmanont, X. Ge, K.S. Mupparapu, G. Rao, R. Potts, and L. Tolosa. Passive diffusion of transdermal glucose: Noninvasive glucose sensing using a fluorescent glucose binding protein. *J. Diabetes Sci. Technol.*, 8:291–298, 2014.

[132] S. Brown, P. N. Zambrana, X. Ge, D. Bagdure, A. L. Stinchcomb, G. Rao, and L. Tolosa. Minimally invasive technique for measuring transdermal glucose with a fluorescent biosensor. *Anal. Bioanal. Chem.*, 410:7249–7260, 2018.

[133] A. Balter, W. Nowak, W. Pawelkiewicz, and A. Kowalczyk. Some Remarks on the Interpretation of the Spectral Properties of Prodan. *Chem. Phys. Lett.*, 143(6):565–570, 1988.

[134] A. Kawski. Ground- and Excited-State Dipole Moments of 6-Propionyl-2-(dimethylamino)naphthalene Determined from Solvatochromic Shifts. *Z. Naturforsch.*, 54a:379–381, 1999.

[135] W. Nowak, P. Adamczak, and A. Balter. On the possibility of fluorescence from twisted intramolecular charge transfer states of 2-Dimethylamino-6-acylnaphthalenes. A quantum-chemical study. *J. Mol. Struct. (Theochem)*, 139:13–23, 1986.

[136] P. Ilich and F. G. Prendergast. Singlet Adiabatic States of Solvated PRODAN: A Semiempirical Molecular Orbital Study. *J. Phys. Chem.*, 93(11):4441–4447, 1989.

[137] A. Samanta and R. W. Fessenden. Excited State Dipole Moment of PRODAN as Determined from Transient Dielectric Loss Measurements. *J. Phys. Chem. A*, 104:8972–8975, 2000.

[138] B. C. Lobo and C. J. Abelt. Does PRODAN Possess a Planar or Twisted Charge-Transfer Excited State? Photophysical Properties of Two PRODAN Derivatives. *J. Phys. Chem. A*, 107:10938–10943, 2003.

[139] B. N. Davis and C. J. Abelt. Synthesis and Photophysical Properties of Models for Twisted PRODAN and Dimethylaminonaphthonitrile. *J. Phys. Chem. A*, 109:1295–1298, 2005.

[140] R. K. Everett, H. A. A. Nguyen, and C. J. Abelt. Does PRODAN Possess an O-TICT Excited State? Synthesis and Properties of Two Constrained Derivatives. *J. Phys. Chem. A*, 114:4946–4950, 2010.

[141] B. Mennucci, M. Caricato, F. Ingrosso, C. Cappelli, R. Cammi, J. Tomasi, G. Scalmani, and M. J. Frisch. How the Environment Controls Absorption and Fluorescence Spectra of PRODAN: A Quantum-Mechanical Study in Homogeneous and Heterogeneous Media. *J. Phys. Chem. B*, 112:414–423, 2008.

[142] A. Parusel. Semiempirical studies of solvent effects on the intramolecular charge transfer of the Fluorescence probe PRODAN. *J. Chem. Soc., Faraday Trans.*, 94:2923–2927, 1998.

[143] M. Raguž and J. Brnjas-Kraljević. Resolved Fluorescence Emission Spectra of PRODAN in Ethanol/Buffer Solvents. *J. Chem. Inf. Model.*, 45:1636–1640, 2005.

[144] B. A. Rowe, C. A. Roach, J. Lin, V. Asiago, O. Dmitrenko, and S. L. Neal. Spectral Heterogeneity of PRODAN Fluorescence in Isotropic Solvents Revealed by Multivariate Photokinetic Analysis. *J. Phys. Chem. A*, 112:13402–13412, 2008.

[145] A. Marini, A.Muñoz-Losa, A. Biancardi, and B. Mennucci. What is Solvatochromism? *J. Phys. Chem. B*, 114:17128—17135, 2010.

[146] V. I. Tomin and K. Hubisz. Inhomogeneous Spectral Broadening and the Decay Kinetics of the Luminescence Spectra of Prodan. *Opt. Spectrosc.*, 101(1):98–104, 2006.

[147] M. Novaira, M. A. Biasutti, J. J. Silber, and N. M. Correa. New Insights on the Photophysical Behavior of PRODAN in Anionic and Cationic Reverse Micelles: From Which State or States Does It Emit? *J. Phys. Chem. B*, 111:748–759, 2007.

[148] R. Adhikary, C. A. Barnes, and J. W. Petrich. Solvation Dynamics of the Fluorescent Probe PRODAN in Heterogeneous Environments: Contributions from the Locally Excited and Charge-Transferred States. *J. Phys. Chem. B*, 113:11999–12004, 2009.

[149] C. C. Vequi-Suplicy, K. Coutinho, and M. T. Lamy. New Insights on the Fluorescent Emission Spectra of Prodan and Laurdan. *J. Fluoresc.*, 25:621–629, 2015.

[150] O. M. Zharkova, S. I. Rakhimov, and Y. P. Morozova. Quantum-Chemical Investigation of the Spectral Properties of Fluorescent Probes Based on Naphthalene Derivatives (Prodan, Promen). *Russ. Phys. J.*, 56(4):411–419, 2013.

[151] S. Baral, M. Phillips, H. Yan, J. Avenso, L. Gundlach, B. Baumeier, and E. Lyman. Ultrafast Formation of the Charge Transfer State of Prodan Reveals Unique Aspects of the Chromophore Environment. *J. Phys. Chem. B*, 124:2643–2651, 2020.

[152] T. Soujanya, T. S. R. Krishna, and A. Samanta. The Nature of 4-Aminophthallmlde-Cyclodextrln Inclusion Complexes. *J. Phys. Chem.*, 96:8544–8548, 1992.

[153] G. Saroja and A. Samanta. Polarity of the micelle-water interface as seen by 4-aminophthalimide, a solvent sensitive fluorescence probe. *Chem. Phys. Lett.*, 246:506–512, 1995.

[154] K. Sahu, S. K. Mondal, D. Roy, R. Karmakar, and K. Bhattacharyya. Slow solvation dynamics of 4-AP and DCM in binary mixtures. *J. Photochem. Photobiol. A*, 172:180–184, 2005.

[155] L. Bucsiová and P. Hrdlovič. Medium Effect of Polymer Matrices on Spectral Properties of 4-Aminophthalimide and 4-Dimethylaminophthalimide. *J. Macromol. Sci. A*, 44(9):1047–1053, 2007.

[156] D. Singha, D. K. Sahu, and K. Sahu. Anomalous Spectral Modulation of 4-Aminophthalimide inside Acetonitrile/AOT/n-Heptane Microemulsion: New Insights on Reverse Micelle to Bicontinuous Microemulsion Transition. *J. Phys. Chem. B*, 122:6966–6974, 2018.

[157] T. Soujanya, R. W. Fessenden, and A. Samanta. Role of Nonfluorescent Twisted Intramolecular Charge Transfer State on the Photophysical Behavior of Aminophthalimide Dyes. *J. Phys. Chem.*, 100:3507–3512, 1996.

[158] A. M. Durantini, R. D. Falcone, J. D. Anunziata, J. J. Silber, E. B. Abuin, E. A. Lissi, and N. M. Correa. An Interesting Case Where Water Behaves as a Unique Solvent. 4-Aminophthalimide Emission Profile to Monitor Aqueous Environment. *J. Phys. Chem. B*, 117:2160–2168, 2013.

[159] D. C. Khara and A. Samanta. Comment on "An Interesting Case Where Water Behaves as a Unique Solvent. 4-Aminophthalimide Emission Profile to Monitor Aqueous Environment". *J. Phys. Chem. B*, 117:5387–5388, 2013.

[160] A. M. Durantini, R. D. Falcone, J. D. Anunziata, J. J. Silber, E. B. Abuin, E. A. Lissi, and N. M. Correa. Reply to "Comment on 'An Interesting Case Where Water Behaves as a Unique Solvent. 4-Aminophthalimide Emission Profile to Monitor Aqueous Environment'". *J. Phys. Chem. B*, 117:5389–5391, 2013.

[161] D. C. Khara, S. Banerjee, and A. Samanta. Does Excited-State Proton-Transfer Reaction Contribute to the Emission Behaviour of 4-Aminophthalimide in Aqueous Media? *ChemPhysChem*, 15:1793–1798, 2014.

[162] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52(2):255–268, 1984.

[163] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[164] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[165] S. Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 50(5):1055–1076, 1983.

[166] F.-E. C. Wolff. *Computer Simulations of Light-Triggered Processes of Chromophores in Complex Environments.* dissertation, Karlsruhe Institute of Technology, Germany, 2018.

[167] Origin Professional, version 2020b. OriginLab Corporation, Northhampton, MA, USA, 2020.

[168] W. Thiel. Program MNDO2005. Version 7.0, Max-Planck-Institut fuer Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Muelheim, Germany, 2005.

[169] O. Christiansen, H. Koch, and P. Jørgensen. The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.*, 243(5):409–418, 1995.

[170] R. Ahlrichs, M. Baer, M. Haeser, H. Horn, and C. Koelmel. Electronic structure calculations on workstation computers: the program system TURBOMOLE. *Chem. Phys. Lett.*, 162:165, 1989.

[171] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.

[172] J. S. García, M. Boggio-Pasqua, I. Ciofini, and M. Campetella. Excited State Tracking during the Relaxation of Coordination Compounds. *J. Comput. Chem.*, 40:1420–1428, 2019.

[173] T. A. Niehaus. Personal communication, 2021.

[174] T. Kubař. Personal communication, 2021.

[175] J. Li, P. Reiser, B. R. Boswell, A. Eberhard, N. Z. Burns, P. Friederich, and S. A. Lopez. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.*, 12:5302–5324, 2021.

[176] J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González, and P. Marquetand. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.*, 10:8100–8107, 2019.

[177] J. Westermayr, M. Gastegger, and P. Marquetand. Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J. Phys. Chem. Lett.*, 11:3828–3834, 2020.

[178] J.-K. Ha, K. Kim, and S. K. Min. Machine Learning-Assisted Excited State Molecular Dynamics with the State-Interaction State-Averaged Spin-Restricted Ensemble-Referenced Kohn-Sham Approach. *J. Chem. Theory Comput.*, 17:694–702, 2021.

[179] D. Noukakis and P. Suppan. Photophysics of aminophthalimides in solution I. Steady-state spectroscopy. *J. Lumin.*, 47:285–295, 1991.

[180] S. Mukherjee, K. Sahu, D. Roy, S. K. Mondal, and K. Bhattacharyya. Solvation dynamics of 4-aminophthalimide in dioxane–water mixture. *Chem. Phys. Lett.*, 384:128–133, 2004.

[181] S. C. Wörner. *Fluoreszente Aminosäuren als Elektronentransfer- und Transmembransonden.* dissertation, Karlsruhe Institute of Technology, Germany, 2019.

[182] John C. Pickup, Faeiza Hussain, Nicholas D. Evans, and Nabihah Sachedina. In vivo glucose monitoring: the clinical reality and the promise. *Biosens. Bioelectron.*, 20(10):1897–1902, 2005.

[183] M. J. Borrok, L. L. Kiesling, and K. T. Forest. Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures. *Protein Sci.*, 16:1032–1041, 2007.

[184] P. Herman, J. Vecer, I. Barvik Jr., V. Scognamiglio, M. Staiano, M. de Champdoré, A. Varriale, M. Rossi, and S. D'Auria. The role of calcium in the conformational dynamics and thermal stability of the D-galactose/D-glucose-binding protein from Escherichia coli. *Proteins Struct. Funct. Bioinformat.*, 61:184–195, 2005.

[185] N. K. Vyas, M. N. Vyas, and F. A. Quiocho. Sugar and Signal-Transducer Binding Sites of the *Escherichia coli* Galactose Chemoreceptor Protein. *Science*, 242:1290–1295, 1988.

[186] S. L. Mowbray, R. D. Smith, and L. B. Cole. Structure of the periplasmic glucose/galactose receptor of salmonella typhimurium. *Receptor*, 1(1-2):41—53, 1990.

[187] Jin yu Zou, Maria M. Flocco, and Sherry L. Mowbray. The 1·7 Å refined x-ray structure of the periplasmic glucose/galactose receptor from salmonella typhimurium. *J. Mol. Biol.*, 233(4):739–752, 1993.

[188] M. M. Flocco and S. L. Mowbray. The 1.9 A X-ray Structure of a Closed Unliganded Form of the Periplasmic Glucose/Galactose Receptor from Salmonella typhimurium. *J. Biol. Chem.*, 269(12):8931–8936, 1994.

[189] M. N. Vyas, N. K. Vyas, and F. A. Quiocho. Crystallographic Analysis of the Epimeric and Anomeric Specificity of the Periplasmic Transport/Chemosensory Protein Receptor for D-Glucose and D-Galactose. *Biochemistry*, 33:4762–4768, 1994.

[190] G. Ortega, D. Castaño, T. Diercks, and O. Millet. Carbohydrate Affinity for the Glucose–Galactose Binding Protein Is Regulated by Allosteric Domain Motions. *J. Am. Chem. Soc.*, 134:19869–19876, 2012.

[191] L. Unione, G. Ortega, A. Mallagaray, F. Corzana, J. Pérez-Castells, A. Canales, J. Jiménez-Barbero, and O. Millet. Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations. *Chem. Biol.*, 11:2149–2157, 2016.

[192] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr.and J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross,

V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09, revision a.02, 2009.

[193] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, 25:247–260, 2006.

[194] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[195] B. Aradi, B. Hourahine, and Th Frauenheim. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A*, 111(26):5678–5684, 2007.

[196] M. Cossi, N. Rega, G. Scalmani, and V. Barone. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.*, 24(6):669–681, 2003.

[197] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78:1950–1958, 2010.

[198] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general AMBER force field. *J. Comput. Chem.*, 25:1157–1174, 2004.

[199] I. S. Joung and T. E. Cheatham III. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.

[200] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101–1–014101–7, 2007.

[201] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[202] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *Proc. of EASC 2015 LNCS*, 8759:3–27, 2015.

[203] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. A tolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–854, 2013.

[204] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient , load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[205] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.*, 26:1701–1719, 2005.

[206] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7:306–317, 2001.

[207] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

[208] Z. Pang, M. Sokolov, T. Kubař, and M. Elstner. Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model. *Phys. Chem. Chem. Phys.*, 24:2441–2453, 2022.

[209] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. Plumed 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185:604–613, 2014.

[210] P. Pospišil, K. E. Luxem, M. Ener, J. Sýkora, J. Kocábová, H. B. Gray, A. Vlček, and M. Hof. Fluorescence Quenching of (Dimethylamino)naphthalene Dyes Badan and Prodan by Tryptophan in Cytochromes P450 and Micelles. *J. Phys. Chem. B*, 118:10085–10091, 2014.

[211] J. L. Asensio, A. Ardá, F. J. Cañada, and J. Jiménez-Barbero. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.*, 46(4):946–954, 2013.

[212] A. E. Enß. Simulation von Farbstoffbindestellen am Glucosebindeprotein (unpublished Bachelor thesis), 2019. Karlsruhe Institute of Technology, Germany.

[213] J. M. Berg, J. L. Tymoczko, G. J. Gatto jr., and L. Stryer. *Stryer Biochemie*. Springer-Verlag GmbH Deutschland, Heidelberger Platz 3, 14197 Berlin, Germany, 8 edition, 2018.

[214] G. He, H. Zhang, J. D. King, and R. E. Blankenship. Structural Analysis of the Homodimeric Reaction Center Complex from the Photosynthetic Green Sulfur Bacterium Chlorobaculum tepidum. *Biochemistry*, 53:4914–4930, 2014.

[215] Y.-F. Li, W. Zhou, R. E. Blankenship, and J. P. Allen. Crystal Structure of the Bacteriochlorophyll a Protein from Chlorobium tepidum. *J. Mol. Biol.*, 271:456–471, 1997.

[216] R. E. Fenna and B. W. Matthews. Chlorophyll arrangement in a bacteriochlorophyll protein from chlorobium limicola. *Nature*, 258:573–577, 1975.

[217] A. Ben-Shem, F. Frolow, and N. Nelson. Evolution of photosystem I - from symmetry through pseudosymmetry to asymmetry. *FEBS Letters*, 564:274–280, 2004.

[218] J. Wen, H. Zhang, M. L. Gross, and R. Blankenship. Native electrospray mass spectrometry reveals the nature and stoichiometry of pigments in the FMO photosynthetic antenna protein. *Biochemistry*, 50:3502–3511, 2011.

[219] A. N. Melkozernov, J. M. Olson, Y.-F. Li, J. P. Allen, and R. E. Blankenship. Orientation and excitonic interactions of the Fenna-Matthews-Olson Protein in membranes of the green sulfur bacterium Chlorobium tepidum. *Photosynth. Res.*, 56:315–328, 1998.

[220] J. Wen, H. Zhang, M. L. Gross, and R. E. Blankenship. Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *Proc. Natl. Acad. Sci.*, 106(15):6134–6139, 2009.

[221] D. Bína, Z. Gardian, F. Vácha, and R. Litvín. Native FMO-reaction center supercomplex in green sulfur bacteria: an electron microscopy study. *Photosynth. Res.*, 128:93–102, 2016.

[222] S. Maity, B. M. Bold, J. D. Prajapati, M. Sokolov, T. Kubař, M. Elstner, and U. Kleinekathöfer. DFTB/MM Molecular Dynamics Simulations of the FMO Light-Harvesting Complex. *J. Phys. Chem. Lett.*, 11:8660–8667, 2020.

[223] D. E. Tronrud, J. Wen, L. Gay, and R. E. Blankenship. The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynth. Res.*, 100:79–87, 2009.

[224] P. Kjellberg, Z. He, and T. Pullerits. Bacteriochlorophyll in Electric Field. *J. Phys. Chem. B*, 107:13737–13742, 2003.

[225] E. Thyrhaug, K. Zidek, J. Dostal, D. Bina, and D. Zigmantas. Exciton Structure and Energy Transfer in the Fenna-Matthews-Olson Complex. *J. Phys. Chem. Lett.*, 7:1653–1660, 2016.

[226] M. Rätsep, Z.-L. Cai, J. R. Reimers, and A. Freiberg. Demonstration and interpretation of significant asymmetry in the low-resolution and high-resolution Qy fluorescence and absorption spectra of bacteriochlorophyll a. *J. Chem. Phys.*, 134:024506–1–024506–15, 2011.

[227] T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming. Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature*, 434:625–628, 2005.

[228] R. M. Pearlstein. Theory of optical spectra of the bacteriochlorophyll a antenna protein trimer from *Prosthecochloris aestuarii*. *Photosynth. Res.*, 31:213–226, 1992.

[229] X. Lu and R. M. Pearlstein. Simulations of *Prosthecochloris* Bacteriochlorophyll a-protein optical spectra improved by parametric computer search. *Photochem. Photobiol.*, 57:86–91, 1993.

[230] D. Gülen. Interpretation of the Excited-State Structure of the Fenna-Matthews-Olson Pigment Protein Complex of *Prosthecochloris aestuarii* Based on the Simultaneous Simulation of the 4 K Absorption, Linear Dichroism, and Singlet-Triplet Absorption Difference Spectra: A Possible Excitonic Explanation? *J. Phys. Chem.*, 100:17683–17689, 1996.

[231] R. J. W. Louwe. Toward an Integral Interpretation of the Optical Steady-State Spectra of the FMO-Complex of *Prosthecochloris aestuarii*. 2. Exciton Simulations. *J. Phys. Chem.*, 101:11280–11287, 1997.

[232] S. I. E. Vulto, M. A. de Baat, R. J. W. Louwe, H. P. Permentier, T. Neef, M. Miller, H. van Amerongen, and T. J. Aartsma. Exciton Simulations of Optical Spectra of the FMO Complex from the Green Sulfur Bacterium *Chlorobium tepidum* at 6 K. *J. Phys. Chem. B*, 102:9577–9582, 1998.

[233] M. Wendling, M. A. Przyjalgowski, D. Gülen, S. I . E. Vulto, T. J. Aartsma, R. van Grondelle, and H. van Amerogen. The quantitative relationship between structure and polarized spectroscopy in the FMO complex of *Prosthecochloris aestuarii*: refining experiments and simulations. *Photosynth. Res.*, 71:99–123, 2002.

[234] J. Adolphs and T. Renger. How Proteins Trigger Excitation Energy Transfer in the FMO Complex of Green Sulfur Bacteria. *Biophys. J.*, 91:2778–2797, 2006.

[235] F. Müh, M. E.-A. Madjet, J. Adolphs, A. Abdurahman, B. Rabenstein, H. Ishikita, E.-W. Knapp, and T. Renger. α-Helices direct excitation energy flow in the Fenna-Matthews-Olson protein. *Proc. Natl. Acad. Sci.*, 104:16862–16967, 2007.

[236] M. S. am Busch, F. Müh, M. E. Madjet, and T. Renger. The Eighth Bacteriochlorophyll Completes the Excitation Energy Funnel in the FMO Protein. *J. Phys. Chem. Lett.*, 2:93–98, 2011.

[237] E. Gudowska-Nowak, M. D. Newton, and J. Fajer. Conformational and Environmental Effects on Bacteriochlorophyll Optical Spectra: Correlations of Calculated Spectra with Structural Results. *J. Phys. Chem.*, 94(15):5795–5801, 1990.

[238] S. Shim, P. Rebentrost, S. Valleau, and A. Aspuru-Guzik. Atomistic Study of the Long-Lived Quantum Coherences in the Fenna-Matthews-Olson Complex. *Biophys. J.*, 102:649–660, 2012.

[239] J. Gao, W. Shi, J. Ye, X. Wang, H. Hirao, and Y. Zhao. QM/MM modeling of environmental effects on electronic transitions of the FMO complex. *J. Phys. Chem. B*, 117:3488–3495, 2013.

[240] S. Jurinovich, C. Curutchet, and B. Mennucci. The Fenna-Matthews-Olson Protein Revisited: A Fully Polarizable (TD)DFT/MM Description. *ChemPhysChem*, 15:3194–3204, 2014.

[241] C. Olbrich, T. L. C. Jansen, J. Liebers, M. Aghtar, J. Strümpfer, K. Schulten, J. Knoester, and U. Kleinekathöfer. From atomistic modeling to excitation transfer and two-dimensional spectra of the FMO light-harvesting complex. *J. Phys. Chem. B*, 115:8609–8621, 2011.

[242] X. Jia, J. Mei, J. Z. H. Zhang, and Y. Mo. Hybrid QM/MM study of FMO complex with polarized protein-specific charge. *Sci. Rep.*, 5, 2015.

[243] G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mančal, Y.-C. Cheng, R. E. Blankenship, and G. R. Fleming. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, 446:782–786, 2007.

[244] G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship, and G. S. Engel. Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proc. Natl. Acad. Sci.*, 107(29):12766–12770, 2010.

[245] E. Thyrhaug, R. Tempelaar, M. J. P. Alcocer, K. Žídek, D. Bína, J. Knoester, T. L. C. Jansen, and D. Zigmantas. Identification and characterization of diverse coherences in the Fenna–Matthews–Olson complex. *Nat. Chem.*, 10:780–786, 2018.

[246] S. A. Oh, D. F. Coker, and D. A. W. Hutchinson. Variety, the spice of life and essential for robustness in excitation energy transfer in light-harvesting complexes. *Faraday Discuss.*, 221:59–76, 2020.

[247] M. Sokolov. QM/MM Molecular Dynamics Simulations for the Calculation of FMO UV/Vis Absorption Spectra (unpublished Master thesis), 2018. Karlsruhe Institute of Technology, Germany.

[248] C. Olbrich, J. Stümpfer, K. Schulten, and U. Kleinekathöfer. Quest for Spatially Correlated Fluctuations in the FMO Light-Harvesting Complex. *J. Phys. Chem. B*, 115(4):758–764, 2011.

[249] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Bess, and E. Lindahl. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.*, 6:459–466, 2010.

[250] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *J. Mach. Learn. Res.*, 18:6765–6816, 2018.

List of Abbreviations

Acrylodan	6-acryloyl-2-dimethylaminonaphthalene
AEE	adiabatic excitation energy
AO	atomic orbital
4-AP	4-aminophthalimide
Badan	6-bromoacetyl-2-dimethylaminonaphthalene
BLA	bond length alternation
BPC	binding pocket polarized charges
CC2	second-order approximate coupled cluster
CT	charge transfer
CV	collective variable
DFT	density functional theory
DFTB	density functional tight binding
ESP	electrostatic potential
GAFF	general Amber force field
GGBP	Glucose/galcatose binding protein
HF	Hartree Fock
HOMO	highest occupied molecular orbital
1,5-IAEDANS	5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid
IANBD amide	N,N'-dimethyl-N-(iodoacetyl)-N'-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)ethylenediamine
IANBD ester	4-(N-(iodoacetoxy)ethyl-N-methyl)amino-7-nitrobenz-2-oxa-1,3-diazole
IC	internal conversion
ICT	internal/intramolecular charge transfer
ISC	intersystem crossing
LC-	long-range corrected
LCAO	linear combination of atomic orbitals
LE	locally excited
LUMO	lowest unoccupied molecular orbital
MAE	mean absolute error
MD	molecular dynamics
MDCC	N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide

ME	mean error
MM	molecular mechanics
MO	molecular orbital
NBD	nitrobenzoxadiazole
NTO	natural transition orbital
PBC	periodic boundary conditions
PBP	periplasmic binding protein
PDA	point-dipole approximation
PES	potential energy surface
PICT	planar intramolecular charge transfer
PME	particle–mesh Ewald
QM	quantum mechanics / quantum mechanical
RBP	ribose binding protein
RESP	restrained electrostatic potential
RI	resolution of identity
RMSD	root mean square deviation
S0	ground state
SCC	self-consistent charge
SCF	self-consistent field
S_n	nth excited singlet state
TAPP	tetraazaperopyrene
TD-	time-dependent
TDA	Tamm-Dancoff approximation
TICT	twisted intramolecular charge transfer
T_n	nth excited triplet state
VR	vibrational relaxation
WPC	water polarized charges

List of Figures

2.1. Jablonski diagram 4
2.2. Potential energy surfaces in ground and excited state 6

3.1. Working principle of metadynamics 26
3.2. Scheme of neural network 30
3.3. Comparison of optimizers 32

4.0. Bond lenght alternations of optimized structures 37
4.1. Optimized geometries of aniline 39
4.2. Optimized geometries of aminobenzonitrile 39
4.3. Optimized geometries of dimethylaminobenzonitrile 40
4.4. Optimized geometries of DMPD 40
4.5. Optimized geometries of para-phenylenediamine 41
4.6. Optimized geometries of phenol 41
4.7. Optimized geometries of *trans*-stilbene 42
4.8. Optimized geometries of pyridine 42
4.9. Adiabatic excitation energies, molecules set 1 46
4.10. Adiabatic excitation energies, molecules set 2 47
4.11. Adiabatic excitation energies from different parameters sets, molecules set 1 48
4.12. Adiabatic excitation energies from different parameters sets, molecules set 2 49
4.13. Stokes shifts without vibrational progression, molecules set 1 50
4.14. Stokes shifts without vibrational progression, molecules set 2 51
4.15. Optimized geometries of salicylic acid 52
4.16. Optimized geometries of 3P-1-propionic acid 53
4.17. Optimized geometries of phenylacetylene and benzonitrile 53
4.18. Optimized geometries of methyl-4-hydroxycinnamate 54
4.19. Optimized geometries of hydroxyquinolines, pyrroloquinoline and 2-hydroxy-pyridine 54
4.20. Stokes shifts without vibrational progression from different parameter sets, molecules set 1 56
4.21. Stokes shifts without vibrational progression from different parameter sets, molecules set 2 57
4.22. Excited state reorganization energies 58
4.23. Chemical structures of anthanthrene and TAPP 59
4.24. Ground state frequencies of anthanthrene and TAPP 60
4.25. Excited state frequencies of anthanthrene and TAPP 61
4.26. Simulated vibronic progression 62

5.1. Chemical structure of Flugi-2 66
5.2. Chemical structure of prodan 66
5.3. Chemical structure of 4-aminophthalimide 68

5.4. Bond length alternation of Flugi-2 . . . 71
5.5. Clustered dihedral angles of Flugi-2 . . . 72
5.6. Dihedral angles of Flugi-2 versus simulation time . . . 73
5.7. Hydrogen bond between Flugi-2 and DMSO . . . 76
5.8. Fluorescence spectra of Flugi-2 computed with OM2 . . . 77
5.9. Fluorescence spectra of Flugi-2 computed with TD-LC-DFTB2 . . . 78
5.10. Orbitals of prodan . . . 83
5.11. Simulated states of prodan in vacuum . . . 85
5.13. Maximum force components during optimizations of prodan with and without state-tracking . . . 86
5.12. Excited state dipole moments during optimizations of prodan with and without state-tracking . . . 87
5.14. Orbitals of prodan in polar environment as obtained with TD-LC-DFTB2 . 88
5.15. Excited state dipole moments and fluorescence energies of prodan in first excited state in polar environment . . . 88
5.16. Orbitals of 4-AP obtained with TD-LC-DFTB2 in vacuum . . . 91
5.17. Bond length changes in 4-AP upon excitation . . . 92
5.18. Sampled state of 4-AP in vacuum . . . 92
5.19. Dipole moments and oscillator strengths of 4-AP in S3 in vacuum . . . 93
5.20. Dipole moments and oscillator strengths of 4-AP in S3 in DMSO . . . 93
5.21. Dipole moments and oscillator strengths of 4-AP in S1 in water . . . 94
5.22. Hydrogen bonds between 4-AP and water . . . 94
5.23. Computed fluorescence spectra of 4-AP . . . 95

6.1. Working principle of optical glucose sensor . . . 98
6.2. Crystal structures of GGBP . . . 99
6.3. Binding pocket of GGBP . . . 100
6.4. Chemical structure of badan . . . 103
6.5. Triple mutant H152C-badan/A213R/L238S . . . 104
6.6. Labels for atoms of glucose . . . 108
6.7. Partial charges, different methods . . . 110
6.8. Comparison of new charge sets . . . 111
6.9. Root mean square deviation along trajectory, wildtype . . . 112
6.10. Root mean square deviation along trajectory, triple mutant . . . 113
6.11. Hydrogen bonds between glucose and triple mutant . . . 113
6.12. Hinge and twist angle describing the conformational changes of GGBP . . . 114
6.13. Labels for atoms in badan . . . 116
6.14. Representative badan orientations . . . 117
6.15. Sampled conformations of the protein and badan . . . 118
6.16. Possible interaction between badan and Trp183 . . . 119
6.17. Distance between badan and Trp183 . . . 119
6.18. 2D histograms of protein conformation and badan orientation with glucose in the binding pocket . . . 121
6.19. 2D histograms of protein conformation and badan orientation without glucose in the binding pocket . . . 122
6.20. Arrangement of Trp183, glucose and badan . . . 123
6.21. Start structures of QM/MM simulations . . . 126

7.1. Photosynthetic complex of *Chlorobaculum tepidum* 130
7.2. FMO complex and embedded Bacteriochlorophylls 131
7.3. Chemical structure of Bacteriochlorophyll a 132
7.4. Truncated Bacteriochlorophyll a 136
7.5. Root mean square deviation of FMO along trajectory 137
7.6. Site energies from different trajectory pieces 138
7.7. Comparison of different ways to include the environment 139
7.8. Environmental effect on site energies 140
7.9. Environmental effect on Coulomb couplings 141
7.10. Coulomb couplings along the trajectory 143
7.11. Computed Coulomb couplings compared to literature 143
7.12. Learning curves for site energies 146
7.13. Excitation energy distributions from simulations with and without constraints on bonds involving hydrogen 146
7.14. Scatter plots for site energies 147
7.15. Learning curves for couplings 148
7.16. Scatter plots for couplings 149

A.1. Partial charges, different cut-offs 182
A.2. Running averages of Coulomb couplings 184
A.3. Fluctuations of Coulomb couplings 184

List of Tables

2.1. Colour of light depends on its energy . . . 3

4.1. Statistical evaluation of adiabatic excitation energies . . . 45
4.2. Statistical evaluation of excited state reorganization energies . . . 55
4.3. Reorganization energies of polycyclic aromatic hydrocarbons . . . 55
4.4. Relative peak positions in absorption spectra of anthanthrene and TAPP . . 63
4.5. Relative peak positions in emission spectra of anthanthrene and TAPP . . . 63
4.6. Comparison of computational times . . . 63

5.1. Cluster centers from k-means algorithm . . . 71
5.2. Atomic partial charges of excited Flugi-2 . . . 75
5.3. Calculated fluorescence energies of Flugi-2 . . . 79
5.4. Geometrical parameters of prodan in ground state . . . 81
5.5. Vacuum excitation energies for prodan on optimized geometries . . . 82
5.6. Geometrical parameters of prodan in first excited state . . . 83
5.7. Properties of prodan in first excited state . . . 84
5.8. Geometrical parameters of prodan in the third excited state . . . 85
5.9. Properties of 4-AP in vacuum . . . 91

6.1. Experimentally tested glucose sensors based on GGBP . . . 101
6.2. Comparison of new charge sets . . . 109
6.3. Hinge and twist angles of crystal structures . . . 114
6.4. Average number of water molecules surrounding badan . . . 124
6.5. Hydrogen bonds with badan . . . 125
6.6. Computed absorption maxima of badan attached to GGBP . . . 127

7.1. Experimental exciton energies . . . 133
7.2. Computed Coulomb couplings . . . 142
7.3. Excitonic energies . . . 142
7.4. Model performance with and without hydrogens, ntrain=10 000 . . . 145
7.5. Model performance with and without hydrogens, ntrain=300 000 . . . 145
7.6. Model performance under different conditions, couplings . . . 149

A.1. Detailed values of bond length alternation, molecules set 1 . . . 156
A.2. Detailed values of bond length alternation, molecules set 2 . . . 158
A.3. Detailed values of adiabatic excitation energies . . . 159
A.4. Detailed values of reorganization energies . . . 162
A.5. Detailed values of Stokes shifts, molecules set 1 . . . 164
A.6. Detailed values of Stokes shift, molecules set 2 . . . 165
A.7. Detailed values of vibrational frequencies, anthanthrene . . . 167
A.8. Detailed values of vibrational frequencies, TAPP . . . 170
A.9. Orbital analysis for molecules in benchmark set . . . 174

A.10.Recomputed adiabatic excitation energies 176
A.11.Recomputed Stokes shifts 177
A.12.New statistical evaluation 177
A.13.Computed absorption energies of Flugi-2 179
A.14.Computed absorption energies of prodan 179
A.15.Vacuum excitation energies for prodan optimized with TD-LC-DFTB2 . . . 180
A.16.Vacuum excitation energies for prodan optimized with CAM-B3LYP 180
A.17.Computed absorption energies of 4-AP 180
A.18.Vacuum excitation energies for 4-AP optimized with TD-LC-DFTB2 181
A.19.Vacuum excitation energies for 4-AP optimized with CAM-B3LYP 181
A.20.Detailed values of partial charges 183

Contributions to publications

- **chapter 4:**
  Monja Sokolov, Beatrix M. Bold, Julian J. Kranz, Sebastian Höfener, Thomas A. Niehaus and Marcus Elstner. 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. *J. Chem. Theory Comput.*, 17(4):2266–2282, 2021.
  My contribution was the application of the new implementation for TD-LC-DFTB2 forces. I performed all the geometry optimizations, frequency analyses and calculations of vertical excitation energies with (TD-)LC-DFTB, (TD-)DFTB and DFT, computed the investigated quantities (AEEs, Stokes shift, reorganization energy) and evaluated the results. I simulated vibronic progression by means of the program provided by S. Höfener.
  Text and figures of this publication are reproduced with changes.

- **chapter 6:**
  Ziwei Pang, Monja Sokolov, Tomáš Kubař and Marcus Elstner. Unravelling the mechanism of glucose binding in aprotein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model. *Phys. Chem. Chem. Phys.*, 24: 2441-2453, 2022.
  I derived some of the presented charge models for glucose (BPC models, models from QM/MM simulations) and contributed by running parts of the free MD simulations and discussing the results. The analysis of the Badan conformations was also my part. I further contributed with text in the introductory part.
  Parts of the text and figures are reproduced with changes.

- **chapter 7:**
  Beatrix M. Bold, Monja Sokolov, Sayan Maity, Marius Wanko, Philipp M. Dohmen, Julian J. Kranz, Ulrich Kleinekathöfer, Sebastian Höfener and Marcus Elstner. Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes. *Phys. Chem. Chem. Phys.*, 22:10500–10518, 2020.
  My contribution were all the results to the FMO complex from *C. tepidum* except the TrESP calculations.
  Text and figures are not reproduced. However, recalculated values for the site energies, couplings and excitonic energies presented in the publication are included in subchapter 7.4.1. These values were obtained with a different parameter set for TD-LC-DFTB2.

Sayan Maity, Beatrix M. Bold, Jigneshkumar D. Prajapati, Monja Sokolov, Tomáš Kubař, Marcus Elstner and Ulrich Kleinekathöfer. DFTB/MM Molecular Dynamics Simulations of the FMO Light-Harvesting Complex. *J. Phys. Chem. Lett.*, 11:8660–8667, 2020.
I contributed with QM/MM simulations.
The results from this publication are not reproduced.