
Efficient Response Methods for Light Harvesting Complexes

By

OLIVER JAMES HENRY FEIGHAN



School of Chemistry
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

JUNE, 2022

Word count: 0

ABSTRACT

Here goes the abstract

DEDICATION AND ACKNOWLEDGEMENTS

Here goes the dedication.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

TABLE OF CONTENTS

	Page
List of Tables	xi
List of Figures	xiii
1 Outline	1
2 Introduction	3
2.1 Efficient Response Methods	6
2.1.1 sTDA-xTB	6
2.2 Statistical Methods	8
2.2.1 Machine-Learning For Exciton Models	9
2.3 GPU Acceleration	12
2.4 Combining Approaches	13
3 Background Theory	15
3.1 Electronic structure	15
3.1.1 Density Functional Theory	15
3.1.2 Density Functional Tight Binding	18
3.1.3 Extended Tight Binding	21
3.2 Light-Matter Response Methods	26
3.2.1 Linear Response TD-DFT	26
3.2.2 Δ -SCF	29
3.2.3 Eigenvalue Difference	31
3.2.4 Transition Density and Dipole Moments	31
3.3 Large System Models	32
3.3.1 Excitation Energy Transfer and Frenkel Exciton Hamiltonians	32
4 Mean-Field excited states	37
4.1 Combining Δ -SCF With Lower Level Electronic Structure Theories	38
4.2 Benchmarking Δ -SCF	38
4.2.1 Reference Data and test set	38

TABLE OF CONTENTS

4.2.2	Small Systems	39
4.2.3	Non-orthogonality	42
4.2.4	LH2 Chlorophyll	43
4.2.5	xTB methods	48
4.3	Conclusions	53
5	Chlorophyll specific methods	57
5.1	Approximations in chl-xTB	58
5.1.1	Full TD-DFT Solutions	58
5.1.2	Approximations to Solutions	58
5.1.3	Integrals approximations	59
5.1.4	Q_y Transition	61
5.1.5	Changes to xTB parameters	65
5.1.6	Transition and Excited State Density from the Ground State	66
5.2	Parameterization	67
5.2.1	Objective Function	67
5.2.2	Minimisation Algorithms	68
5.2.3	Reference Data	73
5.2.4	Results	76
5.3	Cross-validation	79
5.3.1	Vibrational Mode Coupling	79
5.3.2	Absorption Spectra	87
5.4	Conclusions	88
6	Beyond Monomer Chlorophyll	91
6.1	Frenkel Exciton Hamiltonian	92
6.1.1	Exciton States	92
6.2	Same-functional Benchmarking	93
6.2.1	LH2 Dimers	94
6.2.2	Rotation	95
6.3	chl-xTB Excitons	99
6.3.1	Full Dimer comparison	99
6.4	Concentration Quenching	107
6.5	Conclusions	110
7	Light Harvesting Complexes	115
7.1	LH2	116
7.1.1	Spectral Density Method	116
7.2	Calculating LH2 Excitons	117

TABLE OF CONTENTS

7.2.1	Molecular Dynamics	117
7.2.2	Scaling	117
7.2.3	Coupling Values	118
7.2.4	Screening and Embedding	118
7.3	Spectra	123
7.3.1	Sites	123
7.3.2	Exciton states	126
7.3.3	Coupling	129
7.4	Assigning Specific Motions	132
7.4.1	Ether system	132
7.4.2	Huang Rhys Factors	134
7.4.3	N Axes Deformation	135
7.5	conclusions	139
8	Conclusions	141
8.1	Efficient Response Methods	142
8.2	Further Investigations into LHCs	144
A	Appendix A	147
A.1	Electronic Structure Codes	147
A.2	Computational Hardware	147
B	Appendix B	149
B.1	LH2 Data	149
B.1.1	Spectral Density Features	149
	Bibliography	159

LIST OF TABLES

TABLE	Page
4.1 Mean and standard deviations of the errors, in eV, against SCS-CC2 reference data. The <code>xtb4stda</code> entry represents the eigenvalue difference method that uses the eigenvalues output from this program.	51
5.1 Summary of the errors and correlations of transition properties for Q_y for a set of LH2 Bchla geometries from a range of established respond methods. The errors were calculated against PBE0/Def2-SVP TD-DFT data, for transition energies (ΔE) and transition dipole moments ($ \mu $).	74
5.2 chl-xTB parameters, optimized by the SLSQP procedure. Reference values for GFN1-xTB and sTDA-xTB are included, and served as initial guesses where available. Novel parameters all started from initial values of 1.0.	78

LIST OF FIGURES

FIGURE	Page
2.1 The trimer unit found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800).	4
2.2 Aggregate units found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800). This particular structure has a 9-fold symmetry, giving the circular ring structures.	5
2.3 Simplified schematic of a neural network showing how input values, such as the (flattened) Coloumb matrix, can be used to generate outputs by several matrix multiplications and activation functions (represented by the arrows). The values in the output layer for models discussed in the text would be Q_y transition energies or atom centered transition charges.	10
4.1 Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.	40
4.2 Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.	41
4.3 The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown in red, whilst those with the symmetric orthogonalisation correction are shown in black.	44
4.4 Transition energies from Δ -SCF for 26 chlorophyll geometries from the LH2 protein of purple bacteria, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.	45
4.5 Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 4.2.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.	46
4.6 A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.	50

LIST OF FIGURES

4.7	The distributions of errors compared to SCS-CC2 transition energies for the methods included in the Δ -xTB benchmarking.	52
5.1	Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.	62
5.2	The HOMO orbital of Bchla from PBE0/Def2-SVP DFT.	63
5.3	The LUMO orbital of BChla from PBE0/Def2-SVP DFT.	64
5.4	An example of the search through solution space using the Nelder-Mead method. The scatter points are positions of the Nelder-Mead simplex vertices, attempting to find the coordinates of the minimum in the HimmelBlau function.	71
5.5	A similar example of the search through solution space, but using the SQSLP method. The scatter points are positions of the iterative solution vector \mathbf{x} , again attempting to find the coordinates of the minimum in the HimmelBlau function.	72
5.6	Comparison of the Q_y transition energies predicted from the reference methods as well as chl-xTB (red) against PBE0/Def2-SVP values.	76
5.7	Comparison of the Q_y transition dipole moments predicted from the reference methods as well as chl-xTB (red) against PBE0/Def2-SVP values.	77
5.8	Normal modes of the nitrogen-magnesium centre of chlorophyll that have D_{4h} - D_{2h} (left) and D_{4h} - C_s (right) symmetry breaking components.	79
5.9	Change in the N_A - N_C displacements along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail.	80
5.10	Change in the N_A - Mg - N_C angle along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail. The small smaller variance than in figure 5.3.1 led to this metric to not be used in normal mode choices.	81
5.11	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	81
5.12	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	82
5.13	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	82
5.14	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	83
5.15	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	83

5.16 Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	84
5.17 Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	84
5.18 Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	85
5.19 Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.	85
5.20 Predicted and experimental absorption spectrum of chlorophyll in diethyl ether, only in the Q band region. Predicted spectra are shown without an energy shift on the left, and with an energy shift to match the experimental absorption maximum on the right. All spectra are normalised to have equal areas.	88
6.1 The correlation of lowest excitations in LH ₂ chlorophyll dimer systems predicted by the exciton model constructed from monomer TD-DFT and TD-DFT of the full dimer system. Scatter points are coloured by the coupling value between exciton states (left) and distance between magnesium centres (right).	94
6.2 Diagram of the artificial dimer system, showing the axes of rotations in green and the chosen sites for calculating distance between functional groups in orange.	96
6.3 The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_z axis.	97
6.4 The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_x axis.	98
6.5 The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_y axis.	98
6.6 Distributions of the lowest 5 transition energies of chlorophyll dimers predicted by TD-DFT with the CAM-B3LYP functional (grey) and PBE0 functional (red).	99
6.7 Scatter of the two lowest transitions from PBE0 TD-DFT against CAM-B3LYP TD-DFT, coloured by the coupling value from the exciton model using PBE0 monomer data.	101

LIST OF FIGURES

6.8	Transition energies of the exciton states calculated using PBE0 TD-DFT monomer data (left) and chl-xTB monomer data (right) against CAM-B3LYP full dimer TD-DFT transition energies, coloured by the coupling values.	104
6.9	Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_z axis.	105
6.10	Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_y axis.	105
6.11	Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_x axis.	106
6.12	Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_z axis.	107
6.13	Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_y axis.	108
6.14	Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_x axis.	109
6.15	Potential energy surfaces of the charge separated and vertical excitation states of chlorophyll dimer systems, using the difference in state energies as the coordinate, for a series of separations. Free energy changes (ΔA) and reorganisation energies (λ) are shown for each separation.	111
6.16	Potential energy surfaces of the charge separated and vertical excitation states of LH2 chlorophyll dimer systems, using the same coordinate as figure 6.4. The large free energy change and movement of charge separation state PES minimum to the left-hand side of the vertical excitation PES imply the LH2 protein environment inhibits transfer to the charge separated state.	112
7.1	a) a diagrammatic representation of coupling strengths and interactions between sites specified by ring labels, with thicker lines indicating stronger exciton coupling interactions. b) Exciton coupling values (in cm^{-1}) plotted as a function of distance between chlorophyll sites, illustrating the distance-dependent cutoff. c) Probability distributions of B850 coupling values, showing the difference in strengths between intra and inter dimer pairs. d) Probability distribution of B800-B850 interactions. . .	119

LIST OF FIGURES

7.2 Simulated absorption spectra of LH2, with (dashed line) and without (solid line) a screening factor for point charge interaction. An experimental line reconstructed from <i>et al.</i> is plotted in black. The simulated spectra are shifted to match the position of the 850 nm peak.	121
7.3 Density contributions of site transitions to exciton states for the vacuum (left) and screened (middle) Hamiltonian, with the difference shown in on the right.	122
7.4 Spectral density of the Q_y transition for a chlorophyll in the B800 ring in LH2. Axis scaling and units are chosen to best reproduce the spectrum reported by Mallus <i>et al.</i> [1].	124
7.5 Average spectral densities for Q_y transitions at sites in the B800, B850a and B850b rings.	125
7.6 Violin plots of absorption probabilities for exciton states, positioned by the average transition energy from the ground state. The width of each distribution indicates the density of values at absorption probabilities, with crosses marking the mean value. .	126
7.7 Spectral densities of exciton transition energies of LH2, weighted by the averaged absorption probability. The highest energy state is indicated in red. This state has a higher fluctuation in energy, causing excess noise in the spectral density.	127
7.8 Spectral density of exciton state coupling values (i.e. off diagonal elements of the exciton Hamiltonian) of LH2 coloured by average distance between chlorophyll sites.	130
7.9 Spectral density of the separation between nearest neighbour sites in LH2.	131
7.10 Spectral density of the Q_y transition for a chlorophyll in diethyl-ether.	133
7.11 A simulated spectral density of the Q_y transition for chlorophyll constructed from Huang-Rhys factors and the frequency of normal modes. Labeled peaks are chosen as modes that may correspond to features in other chlorophyll spectral densities.	136
7.12 Motions of the four central nitrogen atoms in bacterial chlorophyll for the vibrational modes labelled in figure 7.4.2.	137
7.13 Spectral density of the ratio $\frac{ N_A-N_C }{ N_B-N_D }$, averaged over chlorophyll sites in LH2.	138

OUTLINE

This thesis focuses on the problem of having efficient yet extendable methods for photochemical problems on large systems. The main developments is on how semi-empirical tight binding methods can be used to predict properties for the LH2 protein. It is argued in the introduction and literature review that many methods often compromise between accuracy of high level methods and efficiency of low level or statistical methods. This compromise is exacerbated when the size of systems gets bigger, for example with the LH2 protein where the chlorophyll system is made up of ~ 4000 atoms. The new work presented here includes a new benchmarking of previous methods and design of novel methods which could offer new insight to these problems. The novel method is used for systems ranging in size from a single chlorophyll to dimers and finally a full LH2 chlorophyll system. At appropriate stages it is benchmarked against relevant data, and is used to make novel arguments that explain chlorophyll system phenomena.

The work is structured into 7 chapters. **Chapter 2** introduces the ideas and theory referred to throughout, with **Chapter 3** being a literature review of how these ideas have been explored in recent work and where the gaps in this lead to the work reported here. The following chapters report on the results of the work done.

Chapter 4 reports on work done on investigating mean-field methods, such as Δ -SCF and eigenvalue difference, as a more approximate method for transition properties and whether it would be a good candidate for large bulk chlorophyll response properties. It also discusses how the underlying level of electronic structure theory affects the accuracy of transition properties. The novel work here is the benchmarking of chlorophyll transition properties with Δ -SCF and TD-DFT functionals, as well as the comparison of transition properties predicted with xTB based methods to higher level theories.

Chapter 5 then uses the findings of the previous chapter to inform design choices of a novel response method. This method uses previously used approximations in a new conjunction to

CHAPTER 1. OUTLINE

predict response properties. While it is not a general purpose method, its application to the Q_y transition of chlorophyll shows excellent performance at reproducing values from high level theories with great efficiency. It fulfills the criteria set out in the introduction and literature review, obtaining response properties with great efficiency whilst also being accurate to high level data as well as being extendable.

Chapter 6 focuses on dimer chlorophyll systems, using an exciton framework that can extend to multiple chlorophyll systems such as light harvesting proteins. Comparison to high level theories show that the new workflow can be expected to give good properties beyond monomer systems. With this workflow it starts to be possible to make novel arguments about light harvesting system phenomena. Specifically, the new workflow is used to explain how the transition from vertical excitation to a charge separation in chlorophyll dimers is suppressed by the LH2 protein scaffold, more than would be expected for the inter-chromophore separation observed in LH2. These are argued by calculating the rate constants of this transition for a series of chlorophyll dimer systems.

Chapter 7, the final results chapter, then reports on applying the novel method to the whole LH2 chlorophyll system. Using the high level of detail and efficiency new properties of this chlorophyll system are calculated, and further claims on the role of the protein scaffold are argued. These are mostly based on spectral densities of LH2 properties, partially characterised by comparison to other approaches of investigating the effect of environment on chlorophyll transitions. Whilst the novel workflow is utilised well, it is argued that with further work more improvements could be made in these final results.

The last chapter, **Chapter 8**, discusses the conclusions from the previous results chapters. Investigations such as applying the novel response method to systems beyond chlorophyll are discussed, as well as alterations to the exciton benchmarking and framework. More applications to LH2 and other light harvesting systems are proposed.

INTRODUCTION

Photosynthesis is the bedrock of life on this planet. It is often the first step in the food chain, establishing ecosystems from a near unlimited source of sunlight. The earliest evidence of photosynthetic organisms was formed around 3.5 billion years ago [2]. One of the most studied organisms are purple bacteria, [3]. These organisms use light harvesting complexes (LHCs) to absorb and stabilise energy from light which is eventually transferred to reaction centres [4]. The mechanism for creating adenosine triphosphate occurs at these reaction centres using the photonic energy for charge transfer. These complexes are extremely efficient, with the LHC found in purple bacteria being around 95% efficient at transferring the electronic energy to charge separation [5].

Many types of LHCs exist, differing by the type of chlorophyll pigments as well as structural features . Most often these complexes are formed of repeated units. For example the LH2 complex found in *acidophilus* is formed of a trimer unit with two bacterial chlorophyll *a* (BChla) chromophores in close proximity, aligned parallel along the porphyrin plane, with a third chromophore perpendicular and further apart (shown in figure 2) [6]. This unit is then repeated to form a circular structure with anywhere between an 8-10 fold symmetry depending on the species of bacteria (see figure 2) [1, 7]. The efficiency of these LHCs is due to these structures and the conformations of the chromophores and their effect on nuanced electronic energy transfer mechanisms [8].

Detailed computational study of these structures has been possible since the high level resolution of crystal structures [9, 10] and has produced a wealth of analysis from investigations of the properties of these complexes that could only be calculated from computational methods. Many of these studies explore the complex mechanisms of electronic energy transfer as well as environmental effects of the protein scaffold on spectroscopic properties of LHCs [11–14]. These

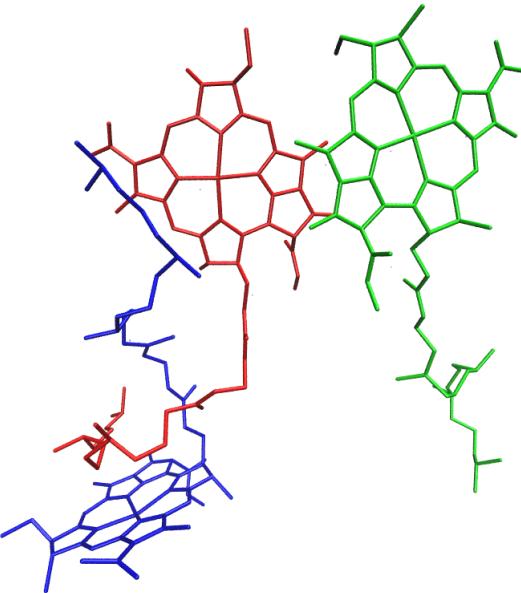


Figure 2.1: The trimer unit found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800).

studies show that the effects of the protein which occur on the atomistic level are an important aspect to capture in LHC models.

To investigate these properties it is necessary to use a model for the excited states of LHCs. These models use a Hamiltonian to describe the system state, similar to many other electronic structure problem. However due to the size of these complexes it is usually necessary to reduce the number of degrees of freedom in these Hamiltonians [1, 11]. For LH2, the chlorophyll molecules alone constituent 3780 atoms with an additional 6000 atoms for the entire scaffold [15, 16]. Including a membrane and explicit solvent can quickly lead to system sizes in the region of 300,000 atoms [17].

Usual electronic structure methods that describe molecular excited states (e.g. DFT) are not tenable for LHCs. Instead a Frenkel-Davydov model is more appropriate due to the weak coupling, and this model recaptures the delocalisation of excited states over pigment sites [18, 19]. This model constructs the Hamiltonian from intra-site energies and inter-site interactions, reducing the degrees of freedom of the Hamiltonian from the entire state to just single chromophores. A more formal description is given in the next chapter.

The Frenkel exciton Hamiltonian can either be constructed from static parameters, fit to experimental data or calculated theoretical values, or as functions of LHC geometry and/or time. A popular example of static parameters for LH2 was produced by Tretiak *et al.* [5]. To recover a truly atomistic treatment of LHCs it is necessary to take the second approach, which can be far more expensive to calculate as Hamiltonians are required for every frame in a time series of LHC geometries, explicitly treating the geometry variations. This also comes with the caveat

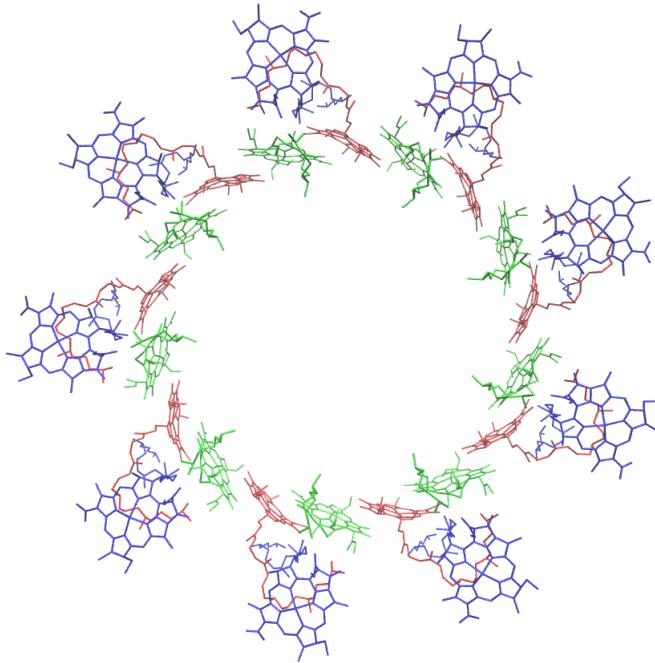


Figure 2.2: Aggregate units found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800). This particular structure has a 9-fold symmetry, giving the circular ring structures.

that methods used to construct the exciton Hamiltonian have to be quite accurate as geometry variations are usually very small, although these small variations are also used to justify the averaging approaches.

Often atomistic LHC models fall into a similar pattern (recently reviewed by Mennucci *et al.*)[20]. Electronic structure calculations are used to produce excited state properties for individual sites which in turn are used to construct Frenkel exciton Hamiltonians, ultimately giving the excited states of the whole LHC. The main design choices for LHC models are in choosing methods for these steps. The first choice is which electronic structure and response method to use to calculate intra-site properties, and then the second choice is how to use these properties to construct the exciton Hamiltonian. The decisions on these choices is highly dependent on the level of detail required for the exciton system (atomistic or coarse grain), and the volume of unique Hamiltonians required (how many frames of molecular dynamics (MD) are used). For a coarse-grain model, where the full geometry of the chlorophyll system is not important, then more approximate methods can be used.

For a truly atomistic approach often density functional theory (DFT) and linear response

methods (time-dependent DFT or TD-DFT) are used to calculate single site properties [20]. Due to the size of chlorophyll molecules, using high level methods such as coupled cluster and EOM-CCSD would not be reasonable. Even with TD-DFT methods only a limited number of explicit Hamiltonians could be constructed before the cost is too high. If a large number of LHC geometries need to be calculated then further approximations are necessary.

This trade-off between computational expense and fine-grain accuracy is the main issue in designing physical models for the excited state of LHCs. Often it is found that while coarse-grain models are good enough for large scale phenomena, smaller details are lost when not using high-level electronic structure and response calculations. Some solutions to this problem are offered in the literature which are discussed below, however each have limitations. These limitations sketch out the potential for a new kind of excited state LHC model, which is described at the end of this chapter.

2.1 Efficient Response Methods

The most obvious solution to this problem is to make the electronic structure and response calculations more efficient. This is often at the expense of accuracy. Many recent studies use tight binding methods such as TD-DFTB or ZINDO to calculate response properties [21–24]. These methods can include environmental effects (such as continuous solvent models or point charge embedding) and are implemented in a wide range of electronic structure packages. Comparison of low-level (TD-DFTB), mid-level (i.e. TD-DFT) and high-level (multireference configuration-DFT and complete active-space SCF) methods have demonstrated that often lower level methods are accurate enough but this is not given for every method and benchmarking is necessary [25–27].

Designing tight-binding models can be challenging due to the balance between accuracy and the parameters that need to be fit. This issue has been explored thoroughly by the xTB methods developed by Grimme *et al.* [28]. A more in-depth discussion of all of these methods is given in the next chapter, however as their excited state method (referred to as sTDA-xTB) is relevant to the discussion on LHCs models an outline and discussion of this method is given below.

2.1.1 sTDA-xTB

sTDA-xTB ("simplified Tann-Danoff Approximation - eXtended Tight Binding") is another method in the family of xTB methods developed by the Grimme group, and is parameterised for transition properties [29]. The accuracy in calculating transition energies with this method is very good with the error compared to high-level methods such as SCS-CC2 being around 0.3 - 0.5 eV.

Similar to other xTB methods, sTDA-xTB is based on tight-binding electronic structure that uses empirically fitted parameters and a minimal basis set. It was trained on a set of highly

accurate coupled cluster and density functional theory excitation energies, as well as atomic partial charges for inter-electronic interactions.

Unlike other xTB methods, coefficients in the basis set for sTDA-xTB are dependent on coordination number of the atom centres. This makes basis functions far more flexible, which would with fixed basis function could only be achieved by using diffuse or additional orbitals in the basis set. It also uses two sets of parameterized basis sets - a smaller valence basis set (VBS) and an extended basis set (XBS). Whilst this reduces the cost of having larger basis sets, it makes calculating the gradient of transition properties much more difficult.

The two basis sets are used to construct formally similar Fock matrix elements, although in practice they use different global parameters. The core Hamiltonian is similar to other DFTB methods that use a self-consistent charge (SCC) method to obtain molecular orbital (MO) coefficients. It is given by

$$(2.1) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left(k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left(h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

where μ, ν, l, l' are orbital and shell indices, k_μ^l are shell-wise Hückel parameters, h are effective atomic-orbital energy levels, $S_{\mu\nu}$ is the overlap of orbitals μ and ν , k_T is a global constant and \hat{T} is the kinetic energy operator. The charges used in the inter-electronic repulsion function are given by charge model 5 (CM5) [30] charges for the XBS Fock matrix. These are calculated using Mulliken charges obtained from diagonalising the Fock matrix with the VBS. The charges for the initial VBS Fock matrix are based on Gasteiger charges [31], modified by the parameterised electronegativities of atoms in the system.

The whole process for determining molecular orbitals can be summarized as:

1. Calculate modified Gasteiger charges for the first initial guess
2. Diagonalise Fock matrix in the VBS to get the first set of Mulliken charges
3. Compute CM5 charges
4. Diagonalise Fock matrix in the VBS again for final set of Mulliken charges.
5. Recalculate CM5 charges with this final set, and diagonalize the Fock matrix in the XBS.
6. The molecular orbital coefficients from this are then fed to the response theory.

The response theory for this method is based on previous work in the Grimme group on the simplified Tamm-Dancoff Approximation [32]. There are several approximations made between full linear response theory and the sTDA method. First is the Tamm-Dancoff approximation, where some transition characters are ignored. The second approximation is to use a Mataga-Nishimoto-Ohno-Klopman (MNOK) method to calculate 2 electron integrals instead of explicitly calculating them.

Transition charges are used to calculate these MNOK integrals. These charges q_{nm}^A (centred on atom A and associated with the transition $n \rightarrow m$) are computed using a Löwdin population analysis

$$(2.2) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m}$$

where the transformed coefficients $C'_{\mu n}$ are given by orthogonalising the original MO coefficients \mathbf{C}

$$(2.3) \quad \mathbf{C}' = \mathbf{S}^{\frac{1}{2}} \mathbf{C}$$

and μ is an index that runs over the atomic orbitals (AO). The MO coefficients are the solution of diagonalising the Fock matrix, similar to equation 3.62.

The approximations to full 2 electron integrals are given by charge-charge interaction damped by the MNOK[33–35] functions. For exchange and Coloumb type integrals, difference exponents are used, along with an additional free parameter to recover the amount of Fock exchange mixing in the original matrix element equation. These will be discussed in more detail in chapter 5, as they are a crucial part of designing a new response method for chlorophyll systems.

Lastly the single particle excited space used to construct the \mathbf{A} matrix is truncated more than that of normal TD-DFT. This reduces the number of elements that need to be calculated, reducing the time taken for diagonalisation whilst also capturing a broad enough spectrum of excitation energies.

The sTDA-xTB method is reported as having excellent accuracy against benchmarked data, and has been used to generate absorption spectra and other properties for large systems [29, 36–39]. At first glance, it would seem that this method would solve the issue of calculating many Frenkel exciton systems, however this is not given by the data so far. Much of the work reporting sTDA-xTB accuracy has been performed on a range of systems, and does not concern smaller geometry variations of a single system. The latter is more important for LHCs, as it is the variations in chlorophyll geometries that cause variations in the exciton system and these are relatively small [40]. Without any indication on how accurate sTDA-xTB is for a range of conformers it is difficult to say whether it would be better than previously used tight-binding methods. It may be better to start from methods that do have accurate correlations with system geometries, such as TD-DFT, and work from this to retain accuracy. This is explored in the next section on statistical method based on high-level data to generate new values for exciton models.

2.2 Statistical Methods

Making approximations in constructing the exciton framework is an alternative option to response method approximations. One of the simplest ways of doing this is by using static parameters fit

from experimental or calculated data [5, 41, 42]. These are referred to static in this work as the parameters do not vary between frames in a time series of LHC geometries. Using these static Hamiltonians negates any variation in intra-chromophore or protein scaffold geometry, but can still produce good predictions of physical phenomena.

If using a long timescale, where the full conformation space is well sampled, exciton Hamiltonians can be constructed from distributions of chromophore response properties (i.e. excitation energies and transition densities) [43]. These properties are distributed along a normal distribution when taken from a set of uncorrelated structures. The mean and standard deviations can then be used to define a distribution function which can be limitlessly sampled to construct Hamiltonians without the need for explicit calculations on structures. The Hamiltonians could utilise functions that take into account inter-chromophore geometries, for example by calculating coupling values from a distribution function of transition dipole magnitudes, or use distributions for all elements of the Hamiltonian matrix.

However even these methods are based on static parameters such as the mean and standard deviation and are not functions of time. This means they would be ill-suited for dynamic studies where structures at different times are correlated which is an important consideration for LH2 [44]. However recently machine-learning methods has been reported that would give time-dependent Hamiltonians still without the need for expensive calculations.

2.2.1 Machine-Learning For Exciton Models

Machine-learning models have been used in many areas of computational chemistry, especially in areas where both large amounts data and well-defined metrics make it easy to train methods [45–49]. At their heart, these methods are similar to the static statistical methods that have already been used for LH2 exciton systems, as they rely on parameters fit to high level data. However these models use machine-learning techniques to incorporate atomic geometry information so they are time dependent functions.

In 2016 Häse *et al.* reported on a multi-perceptron (neural network (NN)) model that predicts the Q_y transition for chlorophyll molecules [50]. Using this model, as well as fitted parameters for the exciton coupling, it was possible to calculate exciton population dynamics, as well as spectral densities for chlorophyll sites in the FMO light harvesting complex. Similar to other models a "Coloumb matrix" [51, 52] is used as descriptor of the chlorophyll systems defined as

$$(2.4) \quad M_{AB} = \begin{cases} \frac{1}{2}Z^{2.4} & \text{for } A = B \\ \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} & \text{for } A \neq B \end{cases}$$

where Z_A is some measure of the atomic charge on atom A , and \mathbf{R}_A is the position vector. It can be seen that the off-diagonal elements are simply the Coloumbic interactions, and diagonal elements are a polynomial of the atomic charges. This descriptor is popular due to the similarity

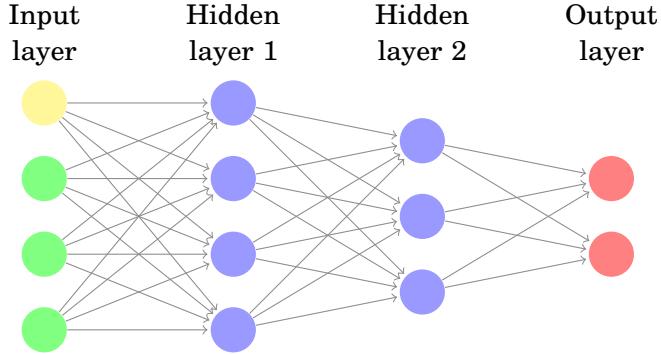


Figure 2.3: Simplified schematic of a neural network showing how input values, such as the (flattened) Coloumb matrix, can be used to generate outputs by several matrix multiplications and activation functions (represented by the arrows). The values in the output layer for models discussed in the text would be Q_y transition energies or atom centered transition charges.

in information that an electronic structure calculation would start from, namely the positions and nuclear charges of atoms [53].

This matrix is used as an input for a neural network model. A neural network is a series of matrix multiplications applied to input data that overall acts as a non-linear function. These multiplications are organised into layers with each element in the output of the multiplication referred to as a node. The first and last layers are called input and output layers and any steps in-between referred to as "hidden" layers [54].

A layer is a distinct vector of values. For example, flattening the Coloumb matrix into a 1D vector gives the input layer, and the Q_y transition energy is a the output vector (of 1 element). A vector at layer n , is multiplied with a coefficient matrix c to give the next layer

$$(2.5) \quad \mathbf{V}_{n+1} = \mathbf{c}_n \mathbf{V}_n$$

where the coefficient matrices \mathbf{c}_n are fitted to give the smallest deviations of the output layer against target data (i.e. error in predicted Q_y energies in this case). Often additional functions are used to modify these values, such as sinusoidal, sigmoid, linear ($\tilde{\mathbf{V}}_{n+1} = m\mathbf{V}_{n+1}$) or rectified linear unit ($\tilde{\mathbf{V}}_{n+1} = \max(0, m\mathbf{V}_{n+1})$) functions. In this way they are conceptually similar to biological neurons which take in electrical signals and through some mechanism are activated to send out a different signal. The coefficient matrices \mathbf{c}_n , and any other parameters in the activation functions, are fit by a back-propagation method which also uses parameters - these are referred to as hyper-parameters and can be optimised using a systematic grid-search to learn the best coefficient matrix. Other considerations such as over-fitting also need to be taken into account. The brief explanation of NNs here is fairly simplistic, and creating these models takes in-depth knowledge and experience to achieve good results.

This model predicted Q_y transition energies with around a 0.3 meV error for all of the 8 sites in the FMO complex [50]. This is exceptionally accurate, supporting the idea that atomic positions and nuclear charges contain all the information necessary to predict transition energies. It is noted in this work that the Nitrogen root mean squared deviations (RSMD) from the energy-minimised crystal structure correlates well with excited state properties. This implies that structural information is key to get transition properties correct and should be a large consideration when discussion models for chlorophyll transition properties. Exciton properties, such as the time series of exciton populations, were also well reproduced using these Q_y energies, although the coupling parameters were taken from other fits and not from a NN method.

Another method, developed by Farahvash *et al.*, utilised both neural networks and kernel ridge regression (KRR) to predict both site energies as well as exciton coupling parameters, giving a completely time dependent exciton Hamiltonian [55]. KRR is another machine-learning method that can be understood as two processes [56]. First is the ridge regression, which is similar to a linear regression model but with an additional factor to account for co-linear relationships between inputs. Regression models are multivariate linear models that follow the form

$$(2.6) \quad f'(\mathbf{X}) = \mathbf{X}\beta$$

where $f'(\mathbf{X})$ are the predicted values of some metrics $f(\mathbf{X})$ (i.e. Q_y energy or exciton coupling value), \mathbf{X} is the matrix of information used to predict the value f (i.e. flattened Coloumb matrix, referred to as the feature matrix) and β is a set of fitted coefficients that minimise the value $|f'(\mathbf{x}) - f(\mathbf{x})|$. The matrix β can be found by minimising the square of this value, a method known as the least-squares method, however this can lead to expensive terms when calculating the inner product of the feature matrix. Here the "kernel trick" is used to make these terms easier to calculate. This rearranges the minimisation of regression coefficients so that the inner products are not required, but requires a new function that compares the similarity of features. Glossing over some derivation details, the function f' becomes

$$(2.7) \quad f'_{\text{KRR}} = \sum_j^{N_x} \beta_j k(\mathbf{x}, \mathbf{x}_j)$$

where the N_x is the size of the feature vector \mathbf{x} (which is the linear model is stacked to form the matrix \mathbf{X}), and k is the kernel function. Often this is a gaussian function of the feature vector

$$(2.8) \quad k(\mathbf{x}, \mathbf{x}_j) = \exp\left(\frac{-(\mathbf{x} - \mathbf{x}_j)^2}{2\sigma^2}\right)$$

where σ is a fitted parameter [57]. Again these parameters are optimised by a systematic search through values, similar to the grid search referenced before.

In the Farahvash work a KRR model was developed for both the excitation energies and exciton coupling parameters. However it was found that a NN model predicted exciton couplings with greater accuracy, which was attributed to the more complex conformational space. This NN model calculated atomic centered transition charges which were then used to calculate coupling elements with small error against higher level methods. This corroborates another study on exciton coupling methods (discussed later) that argues that transition charge based methods are accurate enough for almost all needs.

These machine-learning models show that it is possible to generate time and geometry dependent functions that give either transition properties needed to construct exciton Hamiltonians or the full Hamiltonians themselves. However there are some issues with this approach. In contrast to the sTDA-xTB method, machine learning models do not use any formalism that treats the electronic structure explicitly. This makes it difficult to include any other affects such as continuous solvent models or point charge embedding that are often used in LHC quantum mechanics / molecular mechanics (QM/MM) models. Some models do use QM/MM methods to generate training data but this would only lead to pigeonholing the optimised model to the QM/MM system used. Additionally they would not able to calculate any other properties than those fitted - for example the models that return Q_y transition energies would not be able to return ground state energies. There are some practical considerations as well when designing these models such as the requirement of in-depth knowledge of machine-learning methods to retrain models. The cost of generating enough data to train these models is also very high level. For example the number of data point required to train the H  se and Farahvash models was on the order of 10^4 and 10^5 respectively. In summary, whilst these models are highly accurate enough to make time-dependent Hamiltonians they are limited by both their designs and their construction cost.

2.3 GPU Acceleration

An alternative approach to this problem would be to accelerate the TD-DFT calculations. This is possible using graphical processing units (GPUs) which have been used to accelerate many different types of molecular chemistry simulations. Programs written for GPUs can partition a limited number of basic operations over a massive number of parallel components. Often this is exploited to parallelise the bottleneck of expensive calculations which for TD-DFT is evaluating the electron integrals. Other examples of GPU usages include the large amount of operations required to calculate forces in MD simulations as well as the matrix multiplications of NNs discussed above.

This approach does not require any parameter optimisation or new formalism but does require appropriate hardware (GPU cards) and programs that can partition on GPUs correctly, however this is a common feature on high performance computers. GPU-acceleration has been used in

many other computational chemistry studies, and is a popular way of drastically increasing system sizes while keeping computing time down.

This approach has been used previously to construct Frenkel exciton Hamiltonians using ω PBEh/6-31G TD-DFT for every for the LH2 complex MD [58].

This model was benchmarked against TD-DFT calculations on a hexamer of bacterial chlorophyll *a* molecules from LH2, and had an average error of 0.008 eV. Comparisons to absorption spectra predicted by full TD-DFT and the exciton model were also favourable, as were predictions of dynamic properties of the exciton system.

Whilst computing the electron integral terms are far less time intensive, storing these values is a major issue. This is due to the lightweight memory restrictions on GPU cards, and even makes recalculation of some integrals more efficient than storage. This means that a larger basis sets may be too expensive to utilise. It could be argued that this lack of high level detail would make a niche method, one which uses more approximate response methods but is better at reproducing high level data, more suitable for some studies. This would lead back to using machine-learning methods, however this work proposes a novel approach outlined below.

2.4 Combining Approaches

The underlying issue of LHC models is clear - the systems are too large to explicitly calculate all transition properties required by high level methods. Making well chosen approximations has found success in solving parts of this issue, but often at with compromises. The sTDA-xTB method (and other tight-binding methods) are efficient but possibly not accurate enough to give meaningful results. Machine-learning methods are accurate and efficient but not extendable, use large amounts of expensive high level training data, and require a good understanding of machine-learning methods. Accelerating calculations with hardware can also be used but may still come up against memory issues, however inevitable progress in high performance machines will fix this eventually.

These shortcomings sketch out the need for a method that is efficient, accurate, extendable, memory light and easy to reproduce. Combining approaches would be expected to keep benefits whilst addressing shortcomings. The tight-binding approaches show that even a low level electronic structure is enough to capture the main features required for accurate transition properties. The machine-learning models show that is using a high degree of specificity can train methods to be extremely accurate. This forms the main thesis of this work - that there is merit in designing a method for light harvesting models that combines these benefits. The results presented in chapters 4 and 5 test this idea, and a discussion of the results propose reasons why these methods may or may not solve the shortcomings outlined above. Applications of these methods to LHC properties are also explored in chapters 6 and 7 to further explore whether the novel features of this combinatorial approach can elucidate new explanations of LHC phenomena.

BACKGROUND THEORY

There are three theory components that are frequently employed when modeling light harvesting complexes. First is the electronic structure theory, which is almost exclusively density functional theory (DFT) or density functional tight binding (DFTB). Second is the response method, for which the workhorse is often time-dependent DFT (TD-DFT) or TD-DFTB. In this work other response methods are discussed, and so these theories are outlined in this chapter as well. Third is the exciton method, again most frequently a Frenkel exciton Hamiltonian. In order to establish a basis for the reporting and discussion of work in later chapters, an outline of each of these three components are given here.

3.1 Electronic structure

3.1.1 Density Functional Theory

Density functional theory (DFT) is ubiquitous in electronic structure calculations. Its application to a wide range of system sizes as well as chemical systems make it an ideal choice for many situations, including chlorophyll and light harvesting systems. A brief overview of DFT is given here to contextualise its use in the results chapters.

At its heart DFT is based on the two Hohenberg-Kohn theorems. The first states that the ground state energy E_{GS} , is proven to have a one-to-one mapping to a functional of the electron density $\rho_{GS}(r)$

$$(3.1) \quad E_{GS} = E [\rho_{GS}(r)]$$

where $E[\rho_{GS}(r)]$ is the functional. The second theorem is closely related to the variation principle, stating that the *exact* ground state density also minimises the total energy. This minima corresponds to only one electron density. Whilst proven in principle, the exact functional of the electron density is unknown and so various approximations have been made. One popular method is the Kohn-Sham approach, where non-interacting electrons are used to generate the ground state density. The total energy in the Kohn-Sham approach is the sum of functionals

$$(3.2) \quad E_{\text{tot}}[\rho(r)] = E_{T_S}[\rho(r)] + E_V[\rho(r)] + E_J[\rho(r)] + E_X[\rho(r)] + E_C[\rho(r)]$$

where these terms correspond to the kinetic, (nuclear) potential, bare Coulombic, exchange and correlation interaction respectively. Due to the Coulombic approximation not including spin effects, it is necessary to include the exchange and correlation terms. The solution for the minimum energy would satisfy

$$(3.3) \quad \Delta \left[E_{\text{tot}}[\rho(r)] - \mu \left(\int \rho(r) dr - N \right) \right] = 0$$

with the constraint that the total number of electrons N is conserved. The value of μ is given by

$$(3.4) \quad \mu = \frac{\delta E_{\text{tot}}[\rho(r)]}{\delta \rho(r)}$$

which can be rewritten in terms of the kinetic energy and energy potentials

$$(3.5) \quad \begin{aligned} \mu &= \frac{\delta E_{T_S}[\rho]}{\delta \rho(r)} + v_V[\rho(r)] + v_J[\rho(r)] + v_X[\rho(r)] + v_C[\rho(r)] \\ &= \frac{\delta E_{T_S}[\rho]}{\delta \rho(r)} + v_{KS}[\rho(r)] \end{aligned}$$

where the energy potentials are combined into the Kohn-Sham potential $v_{KS}[\rho(r)]$ for convenience. As the electrons are non-interacting, this potential can be used to solve one-electron Schrödinger equations

$$(3.6) \quad \left[-\frac{1}{2} \nabla^2 + v_{KS}[\rho(r)] \right] \psi_i = \epsilon_i \psi_i$$

where ψ_i are one-electron wavefunctions, ϵ_i are Lagrange multipliers to ensure orthonormality, and the kinetic energy term $-\frac{1}{2} \nabla^2$ is given from the definition of the kinetic energy of non-interacting electrons

$$(3.7) \quad E_{T_S}[\rho(r)] = -\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle$$

The total electron density can be constructed from the one-electron wavefunction solutions

$$(3.8) \quad \rho = \sum_i^N |\psi_i(r)|^2$$

Again the issue is that the potential functional $v_{KS}[\rho(r)]$ is not known, and so approximations have to be made. Additionally, an initial guess of the electron density is needed to start the variational procedure, but this can readily be given from atomic densities or other methods.

In practice, a range of exchange-correlation functionals are employed, dependant on the problem at hand. The simplest functionals are employ the local density approximation (LDA), which assumes the electron density is the same as the uniform electron gas for all points. More complicated functionals use the density and density gradient, using the generalised gradient approximation (GGA). Meta-GGAs also use the second derivative. Hybrid functionals also use some fraction of the Hartree-Fock energy. Many of exchange-correlation functionals require parameterising against high level data.

The first DFT calculations were done on periodic systems, where infinite-domain plane wave functions where used as a basis set, however now most electronic structure packages use basis functions centred on atomic positions (although this is not always the case). The one-electron wavefunctions can be written as linear combinations of basis functions χ_j

$$(3.9) \quad \psi_i = \sum_{\mu}^n c_{\mu i} \chi_{\mu}(r)$$

where j is the index of the basis function, going up to the total number of functions n , and $c_{\mu i}$ is the molecular orbital (MO) coefficient of the basis function χ_{μ} (also referred to as an atomic orbital or AO) for orbital (one-electron wavefunction) ψ_i . Rewriting the electron density gives the density matrix \mathbf{D}

$$(3.10) \quad \rho = \sum_{\mu\nu} P_{\mu\nu} \chi_{\mu} \chi_{\nu}$$

where the elements of the density matrix are the product of MO coefficients

$$(3.11) \quad P_{\mu\nu} = \sum_i^N c_{i\mu} c_{i\nu}$$

Similarly, the Kohn-Sham potential can be written in matrix form

$$(3.12) \quad F_{\mu\nu} = \frac{\delta E_{KS}}{\delta P_{\mu\nu}}$$

where E_{KS} is the energy from the potential v_{KS} . Applying the variational principle gives the matrix equation

$$(3.13) \quad \mathbf{FC} = \mathbf{SC}\epsilon$$

where \mathbf{S} is the matrix of overlap matrix and ϵ is the diagonal matrix of one-electron wavefunction energies. This equation is similar to the Roothaan-Hall equations used to solve Hartree-Fock theory, and so the matrix \mathbf{F} is commonly referred to as the Fock matrix.

Having a convenient definition of the ground state electron structure is useful in calculating other properties. As explained in section 3.2.1, the ground state MO coefficients can be used to calculate vertical excitation energies as well as excited state and transition electron densities. These properties are necessary to understand the photochemical processes in light harvesting complexes.

3.1.2 Density Functional Tight Binding

In recent years there has been renewed interest in tight-binding methods, including tight binding methods derived from density functional theory (named density function tight binding or DFTB)[59]. These methods approximate the density functional energy by expanding into a Taylor series, based on the density fluctuations $\delta\rho$ [60]

$$(3.14) \quad \begin{aligned} E[\rho] &= E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}\left[\rho_0, (\delta\rho)^2\right] + E^{(3)}\left[\rho_0, (\delta\rho)^3\right] + \dots \\ &= E[\rho_0(r)] \\ &\quad + \int \frac{\delta E[\rho(r)]}{\delta\rho(r)} \Big|_{\rho_0} \delta\rho(r) \\ &\quad + \frac{1}{2} \int \int \frac{\delta^2 E[\rho(r)]}{\delta\rho(r)\delta\rho(r')} \Big|_{\rho_0} \delta\rho(r)\delta\rho(r') + \dots \\ &\quad + \frac{1}{p!} \int \int \dots \int \frac{\delta^p E[\rho(r)]}{\delta\rho(r)\delta\rho(r')\dots\delta\rho(r^{(p)})} \Big|_{\rho_0} \delta\rho(r)\delta\rho(r')\dots\delta\rho(r^{(p)}) + \dots \end{aligned}$$

with this series usually truncated between the first and third term[61]. These terms are analogous to some of the terms in the Kohn-Sham functional. The first term is called the band-structure energy - as it contains no density fluctuations is given by the sum of energies of single particle wavefunction

$$(3.15) \quad E[\rho_0(r)] = \sum_i \langle \psi_i | H[\rho_0] | \psi_i \rangle$$

where the Hamiltonian H contains the kinetic energy and electron-nuclear potential (often called the core Hamiltonian). The second order term corresponds to the Coulomb and exchange-correlation terms

$$(3.16) \quad \frac{1}{2} \int \int \frac{\delta^2 E [\rho(r)]}{\delta \rho(r) \delta \rho(r')} \left| \rho_0 \delta \rho(r) \delta \rho(r') \right| = \frac{1}{2} \int \int \frac{\delta^2 E_{XC} [\rho(r)]}{\delta \rho(r) \delta \rho(r')} + \frac{1}{|r - r'|} \left| \delta \rho(r) \delta \rho(r') \right|_{\rho_0}$$

with the other Taylor expansion terms collected into what is known as the repulsive energy term E^{Rep} . The common expression for DFTB (also known as DFTB2 or SCC-DFTB) energy is then given as the sum

$$(3.17) \quad E^{\text{DFTB}} = \sum_i \langle \psi_i | H [\rho_0] | \psi_i \rangle + E^{\text{XC}} + E^{\text{Rep}}$$

however it is possible to include other terms from this framework. It is also common to replace the continuous electron density with a point charge model. This is done by first approximating the charge fluctuations as a sum of fluctuations centred on atomic positions

$$(3.18) \quad \delta \rho(r) = \sum_A \delta \rho_A(r)$$

where $\rho_A(r)$ is the charge fluctuation on atom A . These atomic contributions are then expanded using a multipole expansion, truncated at the first term

$$(3.19) \quad \delta \rho_A(r) \approx \Delta q_A F_A^{00} Y^{00}$$

where F_A^{00} and Y^{00} are the multipole expansion coefficients - Δq_A is commonly referred to as the partial charge on atom A . As these charges are not an observable of the system, but an approximation of the electronic density, it is arbitrary which charge scheme is used. Often these are Mulliken charges but other methods can be used as will be seen in section 2.1.1. The benefit of using a point charge approximation to electron density is that integrals are far less expensive to calculate. The exchange-correlation term is then given by

$$(3.20) \quad E^{\text{XC}} = \frac{1}{2} \sum_{A,B} \Delta q_A \Delta q_B \gamma_{AB}$$

where the function γ_{AB} recovers the properties of electron-electron interactions. For example at large separation this function tends towards the Coulombic $\frac{1}{R_{AB}}$ interaction. At close separations, the γ function uses the chemical hardness of atoms to damping this interaction. For self-interaction (i.e. γ_{AA}), the energy term is equivalent to the second derivative of the energy with respect to the atomic partial charge, equal to the Hubbard parameter or twice the chemical hardness

$$\begin{aligned}
 E_{AA}^{\text{XC}} &= \frac{1}{2} \Delta q_A^2 \gamma_{AA} \\
 (3.21) \quad &= U_A \\
 &= 2\eta_A
 \end{aligned}$$

where U_A and η_A are the Hubbard parameter and chemical hardness respectively. For interatomic interactions the γ function is scaled by the Hubbard parameters of both atoms involved.

It is also assumed that density fluctuations will only occur in the valence space of atoms, and so core atomic orbitals do not need explicit treatment. As such, DFTB methods usually use a minimal valence basis set[28], although again this is not always the case. These valence atomic orbitals require orthogonalisation against the core orbitals of other atoms, usually achieved with a Schmit orthogonalisation. They are also often calculated from atomic DFT data, however with an additional harmonic constraint to prevent the orbitals from becoming too diffuse - the confinement potential is scaled to be within twice the covalent radius of each atom.

Solving for the ground state wavefunction follows employs a similar Roothaan-Hall matrix equation, although with the Fock matrix defined by equation 3.17. Whilst in normal DFT the convergence of the potential (or Fock matrix) between iterations indicates that the ground state solution has been found, many DFTB schemes use the difference in atomic charges. This scheme is called self-consistent charges (SCC), and is common in many tight binding models.

Whilst some parameters can be calculated from DFT or *ab initio* calculations, such as the Hubbard parameters and AO coefficients, many parameters required fitting against higher level or empirical data. For example the parameters for the repulsive energy term are often fit to bond length data. Additionally, many elements can be precomputed. For example the core Hamiltonian can be written in terms of the AO contributions

$$(3.22) \quad E [\rho_0(r)] = \sum_i \sum_{\mu\nu} c_{i\mu} c_{i\nu} H_{\mu\nu}$$

where the elements of $H_{\mu\nu}$ are precomputed due to not being dependent on the density fluctuations $\delta\rho$. Similarly elements of the AO overlap matrix $S_{\mu\nu}$ can be precomputed and tabulated. For many parameters a careful approach has to be taken to avoid systematic accuracy issues.

Much work has been done on benchmarking DFTB and expanding the formalism. Generally the accuracy of DFTB is found to be on par with DFT, and higher order derivative terms can be used to achieve better results. Most often it is found that the repulsive energy term is the trickiest to get correct, requiring the most parameterisation against reference data. Additional energy terms, such as Non-covalent interactions and spin-polarisation effects can also be included. The spin-polarisation is especially useful for excited states. Originally DFTB was formulated in a restricted ansatz, using doubly occupied orbitals instead of spin-orbitals. This restricted its application in a linear response framework as the excited states could not be properly treated.

However additional energy terms in the Fock matrix recover effects of using a spin-unrestricted ansatz, making predictions of excited states much more accurate.

Tight binding methods are usually used in investigations where the scale of the system of interest is too large for more usual methods, such as DFT or Hartree-Fock (HF) based methods, to be used. Previous methods of dealing with the size of these systems has been to turn to force-field methods, which do not use any quantum mechanical methods and only use classical methods to evaluate energies and gradients of systems. However, it has routinely been shown that these are inaccurate for many systems that involve proton transfer or metallic centers or the making and breaking of chemical bonds[62], which unfortunately covers many interesting biochemical systems, photosynthesis included. For these systems then, using tight-binding methods seems to be a good tradeoff between the expense of full DFT methods and the inaccuracies of classical methods. However, work on making DFT programs quicker, usually with efficient massively parallelized codes, is closing the gap where DFTB methods exist[63].

3.1.3 Extended Tight Binding

Recently the extended tight binding (xTB) family of methods, developed by the Grimme group, have been presented as another semi-empirical tight binding solution to investigating large chemical systems[28][64][65][66][29][67]. Many of these methods have been parameterized for geometry optimizations and frequencies of normal modes, and use novel approaches for non-covalent interactions. These are identified by the GFN prefix (Geometries, Frequencies and Non-covalent). These methods require far less pair-wise parameterisation whilst remaining efficient and accurate for large systems. These methods cover a range of detail in treating the electronic structure, often going in hand with the number of parameters required. For example, GFN2 uses a quadrupole expansion for electrostatics, that along with a detailed dispersion method makes GFN2 completely pairwise parameter free. GFN0 and GFN-FF on the other hand use very approximate methods and more parameters to increase computation efficiency.

The energy terms for xTB methods can be characterized by the order of density fluctuations they correspond to. For example, the zeroth order terms correspond to dispersion (either D3[68], D4[69] or a modified D4 method) and a halogen bonding correction. First order terms are calculated with an extended Huckel theory, and second and higher order terms are calculated by isotropic electrostatic and exchange-correlation terms. In the equations below, these energies are first labelled with a superscript (n) to identify the density fluctuation order, and a subscript to describe the interaction calculated. The second line in each equation gives the corresponding label found in the GFN-xTB publications. The GFN-xTB energies are then summarised as

$$(3.23) \quad \begin{aligned} E_{\text{GFN1-xTB}} &= E_{\text{disp}}^{(0)} + E_{\text{rep}}^{(0)} + E_{\text{XB}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\text{IES+IXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \\ &= E_{\text{disp}}^{\text{D3}} + E_{\text{rep}} + E_{\text{XB}}^{\text{GFN1}} + E_{\text{EHT}} + E_{\gamma} + E_{\Gamma}^{\text{GFN1}} \end{aligned}$$

$$(3.24) \quad \begin{aligned} E_{\text{GFN2-xTB}} &= E_{\text{disp}}^{(0,1,2)} + E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\text{IES+IXC}}^{(2)} + E_{\text{AES+AXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \\ &= E_{\text{disp}}^{\text{D4}'} + E_{\text{rep}} + E_{\text{EHT}} + E_{\gamma} + E_{\text{AEC}} + E_{\text{AXC}} + E_{\Gamma}^{\text{GFN2}} \end{aligned}$$

$$(3.25) \quad \begin{aligned} E_{\text{GFN0-xTB}} &= E_{\text{disp}}^{(0)} + E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + \Delta E^{(0)} \\ &= E_{\text{disp}}^{\text{D4}} + E_{\text{rep}} + E_{\text{EHT}} + E_{\text{EEQ}} + E_{\text{srB}} \end{aligned}$$

It can be seen that energy terms are treated with different methods even though they describe the same interaction. For example the dispersion interactions, which are all zeroth order with respect to density fluctuations, employ a D3, D4 and D4' (D4 modified) approach in GFN1-, GFN2- and GFN0-xTB respectively. The functional form of each of these energy terms is given below.

Common to all of these expressions is the extended Huckel theory energy term (E_{EHT}), derived from first order density fluctuations. This energy is given by tracing the Hamiltonian $H_{\mu\nu}^{\text{EHT}}$ with the valence one-electron density $P_{\mu\nu}$

$$(3.26) \quad E_{\text{EHT}} = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{\text{EHT}}$$

where μ, ν are the indices of atomic orbitals. The elements of the Hamiltonian are given by

$$(3.27) \quad H_{\mu\nu}^{\text{EHT}} = \frac{1}{2} K_{AB}^{ll'} S_{\mu\nu} (H_{\mu\mu} + H_{\nu\nu}) X(EN_A, EN_B) \Pi(R_{AB}, l, l') Y(\eta_l^A \eta_{l'}^B)$$

where A, B are atomic indices, and l, l' are the indices of atomic orbitals on atoms A, B respectively (ie $l \in A, l' \in B$), $K_{AB}^{ll'}$ are parameterised global scaling terms, S is the atomic orbital overlap, $H_{\mu\mu}$ are diagonal elements of the Hamiltonian, treating on-site energies, X is an environment-scaled electro-negativity EN function, Π is another distance-dependent function to correct for the distance-scaled interactions from the overlap matrix and Y corrects for kinetic energy integrals in GFN2- and GFN0-xTB but is discarded for GFN1-xTB.

Due to the lack of element pair-wise parameters (except for a few special cases in the global scaling constants), these terms are readily separable and so can either be considered or discarded for future parameterization work. For example, the Π , Y and X term deal with interactions that are more described by atomic environments rather than particular chemistry between atoms. Hence, for excited state theories they would not really have to be changed. This is corroborated by their absence in the Hamiltonian for the simplified Tamm-Dancoff (sTDA) xTB[29] method, which is discussed later.

In the GFN1 and GFN2 methods there is also the E_{γ} and E_{Γ}^{GFNx} terms, which are the energy terms from second and third order density fluctuations. The second order term, E_{γ} is common to GFN1 and GFN2, and is given by

$$(3.28) \quad E_{\gamma} = \frac{1}{2} \sum_{A,B} \sum_{l \in A} \sum_{l' \in B} q_l q_{l'} \gamma_{AB, ll'}$$

where q_l are shell-resolved Mulliken partial charges, A, B are atom indices and l, l' are shell indices. The γ operator describes short-range Coulombic interactions

$$(3.29) \quad \gamma_{AB,ll'} = \frac{1}{\sqrt{R_{AB}^2 + \eta_{AB,ll'}^{-2}}}$$

where R_{AB} is the internuclear distance between A and B, and η is a parameterized chemical hardness. The third order term is slightly different for GFN1 and GFN2, generally given as

$$(3.30) \quad E_\Gamma = \frac{1}{3} \sum_A^N q_A^3 \Gamma_A$$

where q_A is the atom partial charge (sum of the shell partial charges on that atom), and Γ_A is a different operator constructed from atom-wise parameters. These terms are analogous to the γ operators discussed in the general DFTB theory section. GFN2-xTB also includes higher order multipole interactions in the density-dependent terms, referred to as anisotropic electrostatic (AES) and anisotropic exchange-correlation (AXC) terms. The AES term is given by a sum of the monopole-dipole, monopole-quadrupole and dipole-dipole interactions

$$(3.31) \quad E_{AES} = E_{q\mu} + E_{q\Theta} + E_{\mu\mu}$$

The monopole-dipole interaction, monopole-quadrupole and dipole-dipole energy terms are given by

$$(3.32) \quad E_{q\mu} = \frac{1}{2} \sum_{AB} f_3(R_{AB}) \left[q_A \left(\mu_B^T \mathbf{R}_{BA} \right) + q_B \left(\mu_B^T \mathbf{R}_{AB} \right) \right]$$

$$(3.33) \quad E_{q\Theta} = \frac{1}{2} \sum_{AB} f_5(R_{AB}) \left[q_A \mathbf{R}_{AB}^T \Theta_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \Theta_A \mathbf{R}_{AB} \right]$$

$$(3.34) \quad E_{\mu\mu} = \frac{1}{2} \sum_{AB} f_5(R_{AB}) \left(\mu_A^T \mu_B \right) R_{AB}^2 - 3 \left(\mu_A^T \mathbf{R}_{AB} \right) \left(\mu_B^T \mathbf{R}_{AB} \right)$$

respectively, where μ_A is the dipole moment on atom A and where Θ_A is the quadrupole moment. The damping functions $f_n(R_{AB})$ follow a similar scheme to the dispersion models below, although with modified parameters.

The AXC term is given by

$$(3.35) \quad E_{AXC} = \sum_A \left(f_{XC}^{\mu_A} |\mu_A|^2 + f_{XC}^{\Theta_A} ||\Theta_A||^2 \right)$$

where $f_{XC}^{\mu_A}, f_{XC}^{\Theta_A}$ are element specific parameters. It can be seen that this exchange-correlation term only accounts for changes to the electron density around atom A.

Whilst the energy terms discussed so far have been derived from the Taylor expansion of electron density, other energy corrections are necessary to account for some of the shortcomings of the tight-binding approximations. The repulsion energy term common to all methods describes nuclear-nuclear interactions, different to the E_{rep} term in DFTB. This is defined as

$$(3.36) \quad E_{\text{rep}} = \frac{1}{2} \sum_{A,B} \frac{Z_A^{\text{eff}} Z_B^{\text{eff}}}{R_{AB}} e^{-\sqrt{\alpha_A \alpha_B} (R_{AB})^{k_f}}$$

where Z_A^{eff} is the effective nuclear charge on atom A, differing from the true nuclear charge Z_A by the core atomic electron density, k_f is a global parameter and α_A are atom-wise parameters.

The dispersion energy terms, whilst all correcting for the charge-average schemes introduced by the Kohn-Sham approach, vary in models. The D3 dispersion model gives this energy as

$$(3.37) \quad E_{\text{disp}}^{\text{D3}} = -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(CN_A, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB})$$

where C_n is the dispersion coefficients for dipole-dipole ($n = 6$) and dipole-quadrupole ($n=8$) interactions, which are functions of the coordination number CN_A , s_n is a scaling factor, and the Becke-Johnson damping function is given by

$$(3.38) \quad f_{\text{BJ-damping}}^{(n)}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + \left(a_1 \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} + a_2\right)^n}$$

where a_1, a_2 are also global parameters. The D4 model of dispersion differs by including a charge dependency in the dispersion coefficient function, giving

$$(3.39) \quad E_{\text{disp}}^{\text{D4}} = -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(q_A, CN_A, q_B, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB})$$

(here the charges calculated using the EEQ scheme, discussed below). For GFN2-xTB the D4 model is modified to include a 3-body term, giving

$$(3.40) \quad \begin{aligned} E_{\text{disp}}^{\text{D4}'} &= -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(CN_A, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB}) \\ &+ s_9 \sum_{A,B,C} \frac{(3 \cos(\theta_{ABC}) \cos(\theta_{BCA}) \cos(\theta_{CAB}) + 1) C_9(CN_A, CN_B, CN_C)}{(R_{AB} R_{AC} R_{BC})^3} \\ &\times f_{\text{damping}}^{(n)}(R_{AB}, R_{AC}, R_{BC}) \end{aligned}$$

where $f_{\text{damping}}^{(n)}(R_{AB}, R_{AC}, R_{BC})$ is a special damping function for the three-body term. It can be seen that again GFN2-xTB includes higher order terms to achieve greater accuracy against the test data.

Present also in GFN1- and GFN0-, but not GFN2-xTB, are correctional terms that are independent with respect to the electron density. For GFN1-xTB this is the halogen bonding energy term, which for a system of halogen bond acceptor A , donor B and halogen X is given by

(3.41)

$$E_{XB}^{\text{GFN1}} \sum_{AXB}^{N_{XB}} f_{ABX\text{-damping}} k_X \left[\left(\frac{k_{XR} R_{\text{cov},AX}}{R_{AX}} \right)^{12} - k_{X2} \left(\frac{k_{XR} R_{\text{cov},AX}}{R_{AX}} \right)^6 \right] \left[\left(\frac{k_{XR} R_{\text{cov},AX}}{R_{AX}} \right)^{12} + 1 \right]^{-1}$$

where k_{X2}, k_{XR} are global parameters but k_X is a halogen-specific parameter. The damping term is given as

$$(3.42) \quad f_{AXB\text{-damping}} = \frac{1}{2} \left(1 - \frac{1}{2} \frac{R_{XA}^2 + R_{XB}^2 - R_{AB}^2}{|R_{XA}| |R_{XB}|} \right)^6$$

Similar to DFT and DFTB, these energy terms (or more accurately their potentials) are used to construct the Fock matrix that is solved self-consistently in the Roos-Hall equations to give ground state MO coefficients and energies. However this is only true for GFN1- and GFN2-xTB, as these use Mulliken schemes for the partial charges. The GFN0-xTB uses a non-self consistent charge scheme referred to as electronegativity equilibration (EEQ), which means that only one diagonalisation of the Fock matrix is required for a ground state solution. These charges are given by solving the matrix equation

$$(3.43) \quad \begin{pmatrix} \mathbb{A} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ q_{\text{tot}} \end{pmatrix}$$

where q_{tot} is the total charge of the system, \mathbf{X} is a vector of electronegativities given by

$$(3.44) \quad X_A = \kappa_A \sqrt{mCN_A} - EN_A$$

where mCN_A is a modified coordination number, and \mathbb{A} is a charge-charge interaction matrix that damps interatomic interactions and returns a measure of the chemical hardness for intra-atomic elements

$$(3.45) \quad \mathbb{A}_{AA} = J_{AA} + \frac{2\gamma_{AA}}{\sqrt{\pi}}$$

$$(3.46) \quad \mathbb{A}_{AB} = \frac{\text{erf}(\sigma_{AB} R_{AB})}{R_{AB}}$$

where σ_{AB} is a geometric mean of the atomic radii α_A . Solving for these charges then gives the energy as

$$(3.47) \quad E_{EEQ} = \mathbf{q}^T \left(\frac{1}{2} \mathbb{A} \mathbf{q} - \mathbf{X} \right)$$

These charges are used in place of Mulliken charges that would be found in other energy terms, and as only a single diagonalisation of equation 3.43 is required to generate these charges it can be seen that GFN0-xTB does not need a self-consistent approach to get ground state solutions.

For the most part, the GFN-xTB methods use a minimal basis set, however with some exceptions for hydrogen atoms. These are constructed from Gaussian functions linearly combined into Slater type orbitals. The coefficients and number of Gaussian functions are intrinsic to each GFN-xTB method.

The parameters for the energy terms are either taken from DFT properties (such as the electronegativities or covalent radii) or by using a "top-down" approach to optimise the parameters to a set of target properties. This approach is very successful for all methods. GFN1-xTB has a standard relative deviation of around 1.1% against when compared to geometries from higher level methods, with GFN2-xTB performing similarly well. Generally the GFN-xTB methods predict non-covalent interaction energies with a mean absolute deviation of just over 1 kcal/mol, comparable to low level DFT methods and greatly outperforming other semi-empirical quantum methods. These successes against the target data shows that using partial charge interactions to replace full integrals and global/element-wise parameters to scale interactions works well in designing efficient but accurate methods.

Another benefit of the xTB approach to tight-binding is the lack of pair-wise parameters. Often using more approximate but faster energy terms leads to an increase in the number of fitted parameters, as in order to cover many edge cases it is necessary to include pair-wise parameters. The issue with these highly specific parameters is that they are rarely transferable to other chemical environments and often require reparameterisation for different target properties. They are often limited to the upper parts of the periodic table for this reason. However then GFN-xTB models were designed to emulate the ZDO type methods that only use element-wise parameters. This is best demonstrated in the GFN2-xTB method which uses no pairwise parameters.

3.2 Light-Matter Response Methods

3.2.1 Linear Response TD-DFT

Linear response time-dependant DFT (TD-DFT) is a well established method for calculating excitation energies and transition properties from only ground state information. It is formulated from the Runge-Gross theorem[70], which states that the time dependent density of a system can be mapped from the time-dependent external potential (for light-matter interactions, the external

potential is the light wave), and this mapping is unique. This is analogous to the Hohenberg-Kohn theorems used to derive ground states in section 3.1.1. Similar to the Taylor expansion approach used to derive energy terms for DFTB theory, the density response is usually expanded in terms of time derivatives. This is valid as long as the time scale for response is small, and the excited state density is not too different from the ground state. The name "linear response" is due to curtailing this expansion at first orders terms, which makes TD-DFT practical to calculate [71]. This is useful as only the ground state is needed to calculate perturbations to the first order, meaning that all transition properties can be calculated from the ground state[71].

Response theory is used to predict the changes in the electronic structure of a system over a time period, defined by the time-dependent Schrödiner equation

$$(3.48) \quad i \frac{\delta}{\delta t} \Psi(r, t) = \hat{H}(r, t) \Psi(r, t)$$

For excited state, this change is in response to an external light potential, defining the Hamiltonian as

$$(3.49) \quad \hat{H}(t) = \hat{H}^0 + v_{\text{ext}}(t)$$

where \hat{H}^0 is the unperturbed Hamiltonian, $V^{\text{ext}}(t)$ is the potential from the external field, which is usually taken to be an oscillating electric field and so is time t dependent. This Hamiltonian can then be used to describe a time-dependent set of Kohn-Sham equations[72]. The time-dependent Kohn-Sham Hamiltonian is given by

$$(3.50) \quad \hat{H}_{KS}[\rho](t) = \hat{H}_{KS}^0[\rho] + v_H[\rho](t) + v_{XC}[\rho](t) + v_{\text{ext}}(t)$$

where V_H and V_{XC} are the Kohn-Sham coloumb and exchange-correlation potentials respectively. The Runge-Gross theorem states that the density solution to this Hamiltonian can be uniquely mapped from the potential function, and so the time-dependent density can be written as a function of the potential function

$$(3.51) \quad \rho(t) = \rho[v_{\text{ext}}](t)$$

. The response of the density can be given by the integral of this external potential with what is referred to as the response function χ

$$(3.52) \quad \delta\rho(r, \omega) = \int d^3 r' \chi(r, r', \omega) \delta v_{\text{ext}}(r', \omega)$$

, which calculated the fourier transform of the time series $n(r,t) - n(r,t_0)$ which can be written explicitly for a Kohn-Sham system of electrons

$$(3.53) \quad \chi_{KS}(r, r', \omega) = \sum_{\mu\nu} (f_\mu - f_\nu) \frac{\psi_\mu(r)\psi_\mu^*(r')\psi_\nu(r')\psi_\nu^*(r)}{\omega - (\epsilon_\mu - \epsilon_\nu) + i\eta}$$

where f_μ is the occupation number of the orbital μ , and η is a positive infinitesimal number. Again similar to ground-state DFT, this response function is used in conjunction with correcting exchange-correlation terms to yield the response function of the real system. This would calculate the response, and therefore the excited state, explicitly. However this issue with the theory given so far is that it does not identify at which frequencies ω give any appreciable density response i.e. where the excited states are. However the response function can be written as a discretised set of states m

$$(3.54) \quad \chi(r, r', \omega) = \lim_{\eta \rightarrow 0^+} \sum_m \left[\frac{\langle 0 | \hat{n}(r) | m \rangle \langle m | \hat{n}(r') | 0 \rangle}{\omega - (E_m - E_0) + i\eta} - \frac{\langle 0 | \hat{n}(r') | m \rangle \langle m | \hat{n}(r) | 0 \rangle}{\omega + (E_m - E_0) + i\eta} \right]$$

, where $|0\rangle$ is the ground state with energy E_0 , $|m\rangle$ is an excited state with energy E_m , and \hat{n} is the density operator. It can be seen that the poles of the response function would be where $\omega = E_m - E_0$ or where the frequencies of light match the excitation energies. Eventually it can be shown that these poles can be identified by solving the eigenvalue equation

$$(3.55) \quad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

where ω are the excitation energies, and the vectors \mathbf{X} and \mathbf{Y} describe the electronic transitions in the basis of ground state molecular orbitals. The elements of matrices \mathbf{A} and \mathbf{B} are given by

$$(3.56) \quad A_{ia,jb}(\omega) = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \int dr_1 \int dr_2 \psi_i^*(r_1)\psi_a(r_1)f_{XC}(r_1, r_2, \omega)\psi_i(r_2)\psi_a^*(r_2)$$

$$(3.57) \quad B_{ia,jb}(\omega) = \int dr \int dr' \psi_i^*(r_1)\psi_a(r_1)f_{XC}(r_1, r_2, \omega)\psi_i(r_2)\psi_a^*(r_2)$$

i, j and a, b are occupied and virtual orbital indices respectively, ϵ_i are the orbital energies the ground state orbitals, and δ is the usual kronecker delta function. The kernel function f_{XC} is an exact frequency (excitation energy) dependent exchange-correlation functional, and as it is dependent on the excitation energy given by the solutions of this eigenvalue equation, can be seen to be self-consistent.

In some cases it is useful to make some approximations. First is that the coupling matrix elements are in fact zero. Hence the excitation energies are just the eigenvalue differences, giving

the eigenvalue difference method discussed below. Second is to say that the kernel function is actually frequency independent. This can then give the matrix elements in a more computable form. These could be calculated by any (DFT) method - for example if using a mix of (GGA) density functional and exact exchange, the matrix elements would be given by

$$(3.58) \quad A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) - \zeta(ij|ab) + (1 - \zeta)(ia|f_{XC}|jb)$$

$$(3.59) \quad B_{ia,jb} = (ia|jb) - \zeta(ij|ab) + (1 - \zeta)(ia|f_{XC}|jb)$$

where spatial notation has been used for brevity. ζ here is the amount of exact exchange mixing. The density functional term $(ia|f_{XC}|jb)$ is given by

$$(3.60) \quad (ia|f_{XC}|jb) = \int dr \int dr' \psi_i(r) \psi_a(r') \frac{\delta^2 E_{XC}^{\text{GGA}}}{\delta \rho(r) \delta \rho(r')} \psi_i(r') \psi_a(r)$$

It has been found that some density functionals are better than others for this approximation. Recently a benchmarking of different density functionals for the Q_y transitions in chlorophylls showed that the lowest error is around 0.1 eV [73]. TD-DFT has become the work-horse of excited state studies. It is on comparable accuracy to higher level methods, such as coupled cluster methods, but without the expense[74] and can be calculated from just ground state information with the linear response approximation.

3.2.2 Δ -SCF

Δ -SCF predicts the transition energy ΔE of a system as the difference of the single point energy E_n of two states:

$$(3.61) \quad \Delta E = E_2 - E_1$$

It is usually assumed that the excited state solution will be in a similar location to the ground state in the MO coefficient space. Therefore the ground-state MO coefficients can be used as an initial guess for the excited state. In its simplest form, the Δ -SCF method calculates the ground-state with normal DFT or other mean-field methods, and then calculates the excited state by rerunning the same method with the excited state occupation numbers. The second set of MO coefficients then give a full description of excited state.

The issue with finding the excited state solution is that the variation principle and SCF iterative procedure will try to find the global minimum, which is the ground state. The excited state is a local minimum, and so often is less reliable to find as a solution, especially from the standard SAD initial guess. For this reason it is often found that converging to the Δ -SCF

excited state will fail. Even when using the ground state as an initial guess with excited state occupations, normal SCF procedure may still collapse back to the ground state. Usually it is necessary to include additional changes to the SCF procedure, such as Fock damping, alternative DIIS methods and sometimes intermediate initial guess steps.

Initially, the excited state was calculated by relaxing the orbitals which contain the excited electron and hole in the ground state space, so that the excited state and ground state are orthogonal [75]. However, it was argued that this procedure would exacerbate the likelihood of collapsing to the ground state, and that the excited state was not a proper SCF solution [76]. Alternatively, an SCF like method was proposed, where instead of populating orbitals according to the Aufbau principle, orbitals which most resemble the previous iteration's orbitals should be occupied. This is known as the maximum overlap method (MOM). In the maximum overlap method, each iteration in an SCF procedure produces new molecular orbital coefficients by solving the Roothaan-Hall equations [77], generally given as an eigenvalue problem:

$$(3.62) \quad \mathbf{FC}^n = \mathbf{SC}^n \epsilon$$

where \mathbf{C}^n are the n^{th} orbital coefficient solutions, \mathbf{S} is the overlap of orbitals, and ϵ are the orbital energies. The Fock matrix \mathbf{F} is calculated from the previous set of orbital coefficients,

$$(3.63) \quad \mathbf{F} = f(\mathbf{C}^{n-1})$$

. The amount of similarity of orbitals can be estimated from their overlap,

$$(3.64) \quad \mathbf{O} = (\mathbf{C}^{\text{old}})^\dagger \mathbf{SC}^{\text{new}}$$

and for a single orbital can be evaluated as a projection,

$$(3.65) \quad p_j = \sum_i^n O_{ij} = \sum_v^N \left[\sum_\mu^N \left(\sum_i^n C_{i\mu}^{\text{old}} \right) S_{\mu v} \right] C_{vj}^{\text{new}}$$

where μ, v are orbital indices. the set of orbitals with the highest projection p_j are then populated with electrons. This method can be used for any excited state, with the caveat that the orbital solution will most likely be in the same region as the ground state solution. For a small number of low lying states this is generally true, and so Δ -SCF can be used to calculate a small spectrum of excited states [76].

Δ -SCF has been shown to be cheap alternative to TD-DFT and other higher level methods [78–80] without considerable losses of accuracy in certain cases especially for HOMO-LUMO transitions [81]. Additionally, as the excited state is given as solutions to SCF equations, the gradient of this solution can be given by normal mean-field theory. These gradients would be much cheaper than TD-DFT or coupled cluster methods, which is advantageous for simulating dynamics [79].

3.2.3 Eigenvalue Difference

Another approximation to full response theory is the eigenvalue difference method. Here there is assumed to be no response of the orbital energies and shapes when interacting with light. This would be recovered from the complete Cassida equation if the coupling elements in the **A** and **B** matrices were set to zero. Within this approximation, the transition energy is just the difference between the ground state energy of the orbital an electron has been excited to (ϵ_2) and the orbital has been excited from (ϵ_1),

$$(3.66) \quad \Delta E = \epsilon_2 - \epsilon_1$$

. Additionally, transition properties can be calculated by constructing transition density matrices from the ground state orbitals such that needing only a single SCF optimization is required. Generally, eigenvalue difference methods are not seen as accurate response methods, but can offer a quick and easy initial value [82].

3.2.4 Transition Density and Dipole Moments

Δ -SCF transition properties, such as the transition dipole moment, can be calculated from the SCF solutions for the ground and excited states. The reduced one-particle transition density matrix **D**²¹ can be written as

$$(3.67) \quad \mathbf{D}^{21} = |\Psi_1\rangle\langle\Psi_2|$$

where $|\Psi_n\rangle$ is the Slater determinant of state n , constructed from the set of spin orbitals $\{\phi_j^{(n)}\}$. Expressed in terms of the molecular orbitals coefficients **C**⁽ⁿ⁾, the transition density matrix is

$$(3.68) \quad \mathbf{D}^{21} = \mathbf{C}^{(2)} \text{adj}(\mathbf{S}^{21}) \mathbf{C}^{(1)\dagger}$$

where **S**²¹ is an overlap matrix with elements

$$(3.69) \quad S_{jk}^{21} = \langle\phi_j^2|\phi_k^1\rangle$$

. The dependence on the adjunct of the overlap is due to the use of Löwdin's rules for non-orthogonal determinants [83]. In the same way, the transition dipole moment is given by

$$(3.70) \quad \langle\Psi_2|\hat{\mu}|\Psi_1\rangle = \sum_{jk} \mu_{jk}^{21} \text{adj}(\mathbf{S}^{21})_{jk}$$

where $\hat{\mu}$ is the one-electron transition dipole operator, and μ_{jk} is the element of this operator corresponding to orbital indices j, k . The determinant of \mathbf{S}^{21} can be defined as the inner product of the two states involved in the transition

$$(3.71) \quad |\mathbf{S}^{21}| = \langle \Psi_2 | \Psi_1 \rangle$$

The general definition of the transition dipole

$$(3.72) \quad \mu^{1 \rightarrow 2} = \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle$$

can be expressed with this transition density matrix as

$$(3.73) \quad \begin{aligned} \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle &= \text{tr}(\hat{\mu} | \Psi_1 \rangle \langle \Psi_2 |) \\ &= \text{tr}(\hat{\mu} \mathbf{D}^{21}) \end{aligned}$$

3.3 Large System Models

There is nothing inherent in any of the theories above would exclude them from modeling an entire light harvesting system. Some reservations may be made about the accuracy of the results, but technically the size of the system is unlimited. However in practice system sizes are limited due to two factors - one is the computational power available, and the other is the scaling of calculations with respect to system size. DFT and DFTB for example both have a scaling of , but DFTB drastically decreases the overhead and scaling prefactors due to the many approximations used. This leads to DFT being used for systems of around 10^1 atoms at best, whereas DFTB can go up to 10^2 . This is all dependent on the programs, with some massively parallelised codes fixing these issues. It is also possible to use a linear scaling DFT formalism, but this only outperforms systems of huge sizes. A similar story is found in the response methods. TD-DFT, whilst far less expensive than coupled cluster or other higher level methods, is still limited to the 10^1 atom range, as is DFT based Δ -SCF and eigenvalue difference methods. Whilst it may be possible to use tight-binding schemes, there is a better approach to achieve the required scaling for light harvesting systems. This is done by splitting the system up into sites where excitations can occur (trivial for naturally occurring light harvesting complexes as these sites are just the chlorophyll molecules) and modeling the excitation energy transfer to generate excited states of the entire system. This is explained in the next section.

3.3.1 Excitation Energy Transfer and Frenkel Exciton Hamiltonians

Excitation energy transfer (EET) stabilizes the absorption of a photon of light, and it is thought that light harvesting systems have evolved to maximize this stability[7]. In regimes with strong

coupling (ie large transfer), excitation energy can flow back and forth between different sites. More explicitly, transfer can be described as the transition from an initial site to another site



where D and A are the donor and acceptor system respectively, and $*$ denotes the excited state. The superposition state can be given by the linear combination of the initial and final states:

$$(3.75) \quad |\phi\rangle = c_1 |D^* A\rangle + c_2 |DA^*\rangle$$

where c_n are the coefficients of each acceptor-donor state. In a Frenekl-Davydov model the total Hamiltonian for an aggregate of chromophore sites can be written as a sum of individual site Hamiltonians and inter-site interactions

$$(3.76) \quad \hat{H}_{\text{tot}} = \sum_m^N \hat{H}_m + \sum_m^N \sum_n^N (\hat{V}_{\text{el-el}} + \hat{V}_{\text{el-nuc}} + \hat{V}_{\text{nuc-nuc}})$$

where N is the number of sites, \hat{H}_m is on-site Hamiltonian (ie only treats electrons for the site m). $\hat{V}_{\text{el-el}}$, $\hat{V}_{\text{el-nuc}}$ and $\hat{V}_{\text{nuc-nuc}}$ are the electron-electron, electron-nuclear and nuclear-nuclear interactions between two sites respectively. It is assumed that the site Hamiltonians can give both the ground and excited states of the chromophore m , i.e. it satisfies the time-independent Schrödinger equation

$$(3.77) \quad \hat{H}_{a_m} |\phi_{a_m}\rangle = E_{a_m} |\phi_{a_m}\rangle$$

where a_m is the index of either the ground or an excited state for the chromophore m . There are then three approximations we can make to simplify the total Hamiltonian. First is to assume that there is only one site excitation, and this is to the first excited state. This reduces the complexity of any solution to the total Hamiltonian, and is known as the Heitler-London approximation[84]. Second is that the chromophore sites are well separated enough that the wavefunctions do not overlap[18]. We can then use a Hartree product to construct the aggregate wavefunctions from a basis set of individual chromophores

$$(3.78) \quad |\Psi_{\text{tot}}\rangle = \sum_{a_m} c_{a_m} |\Phi_{a_m}\rangle$$

$$(3.79) \quad |\Phi_a\rangle = \prod_m^N \phi_{a_m}$$

where c_{a_m} are the coefficients for each site, given from the eigenvectors of the time-independent Schrödiner equation. $|\Phi_{a_m}\rangle$ is the total chromophore state, and ϕ_{a_m} are the one-electron orbitals on the chromophore site m . The third approximation is to neglect the nuclear-electron and nuclear-nuclear interactions. This can be done if the effects of the surrounding nuclear environment are treated in the individual site Hamiltonians[85] - this includes the other chromophore sites as well as any other environments ie. the LH2 protein environment. Alternatively, this term can just be ignored.

The total Hamiltonian can then just be expressed as the site Hamiltonians and an electronic coupling element between two sites. The total Hamiltonian can then be constructed from the basis of single chromophores. Starting from the excitonic wavefunctions, the Hamiltonian can be written as

$$(3.80) \quad \begin{aligned} \langle \Phi_a | \hat{H}_{\text{tot}} | \Phi_b \rangle &= \sum_m^N \langle \Phi_a | \hat{H}_m | \Phi_b \rangle + \sum_m^N \sum_n^N \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle \\ &= E_{a_m}^{(m)} \prod_l^N \delta_{al} \delta_{bl} + \sum_m^N \sum_n^N \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle \end{aligned}$$

where the first term are the site energies, as the exciton states are orthogonal. The second term needs to be expanded into the basis set of individual chromophore sites, so we can calculate this term from individual response calculations. Each term in the double summation can be given as

$$(3.81) \quad \begin{aligned} \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle &= \sum_{i \in \mathbf{r}_m} \sum_{j \in \mathbf{r}_n} \langle \Phi_a | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \Phi_b \rangle \\ &= \left(\sum_{i \in m} \sum_{j \in n} \langle \phi_{a_m}^{(m)} | \phi_{a_n}^{(n)} | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \phi_{b_m}^{(m)} | \phi_{b_n}^{(n)} \rangle \right) \prod_{l \neq m, n}^N \delta_{a_l, b_l} \end{aligned}$$

where i, j are the indices of electrons on the sites m, n , and $\mathbf{r}_i, \mathbf{r}_j$ are their positions. With a few extra steps not included for brevity, this expression can then be written in terms of the transition densities of each site[85]

$$(3.82) \quad \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle = \int d\mathbf{r}_m \int d\mathbf{r}_n \frac{\rho_{a_m}(\mathbf{r}_m) \rho_{b_n}(\mathbf{r}_n)}{|\mathbf{r}_m - \mathbf{r}_n|}$$

where ρ_{a_m} is the transition density of transition a on site m . Often the Frenkel Exciton Hamiltonian is summarised as

$$(3.83) \quad \hat{H}_{\text{eff}} = \sum_{m=1}^N \epsilon_m + \sum_{m \neq n}^{N,N} V_{mn}$$

where the diagonal elements ϵ_m are site energies of the chromophores, and the off-diagonal elements J_{mn} are the coulombic coupling between the transition density of different sites. It is usual to take reduce the transition density with a multipole expansion[86]. This could either be done on a whole site scale or at a more detailed atom-in-site scale. Looking at the coarser-grained whole site scale first, the first term in the reduction, corresponding to monopoles, is zero as a local excitation will not produce an overall transition charge (as opposed to a non-local charge-transfer excitation, which can be included in the Frenkel Hamiltonian in some cases[87]). Using second term in this expansion gives a dipole-dipole interaction, referred to as the point-dipole method i.e.

$$(3.84) \quad V_{mn} = \frac{\mu_m \mu_n}{R_{mn}^3} - 3 \frac{(\mu_m \cdot \mathbf{R}_{mn})(\mu_n \cdot \mathbf{R}_{mn})}{R_{mn}^5}$$

and usually the expansion is stopped here. Alternatively reducing the transition density at the atomic level expansion would give an atom centred multipole expansion. The first term of which corresponds to transition charges, giving a Coulombic interaction

$$(3.85) \quad V_{mn} = \sum_{A>B, A \in m, B \in n} \frac{q_A q_B}{|R_A - R_B|}$$

. Other methods for coupling interactions also exist, such as the extended dipole, transition charges from electrostatic potential and transition density cube methods (i.e. a grid approach of discretising the continuous transition density). Intuitively, using a more detailed description of the site transition density results in more accurate couplings. However past the point dipole method, which often overestimates coupling energies, there are diminishing returns in using more expensive methods. At large separations (>20 Å), all of these methods converge to the same value.

Regardless of the interaction method used, all that is needed to calculate a rough Frenkel exciton Hamiltonian is the local excitations and transition properties for each site. This can be done with many response methods, such as time-dependent density functional theory (TDDFT) or mean-field methods like Δ -SCF . This is also often done with either DFT or DFTB based methods, or with statistical approaches.

MEAN-FIELD EXCITED STATES

Previous Published Work

Parts of the work presented in this chapter are also included in a paper published with Dr Susannah Bourne-Worster in March 2021[88]. These sections include parts of section 4.2 from 4.2.1 up to but excluding 4.2.5.

This chapter investigates the accuracy of Δ -SCF methods with both an ab initio DFT level of theory as well as with semi-empirical, tight-binding approximations. Δ -SCF is a well-known method for calculating transition energies, as well as ionisation potentials, electron affinities and other properties. Although slightly less accurate than other high-level response theories, it is much more efficient, requiring only two SCF solutions for the two states involved in a transition.

Whilst Δ -SCF has been benchmarked against other methods in previous studies, it has not been explicitly applied to chlorophyll systems in much detail before. It is possible that this response method could be a good replacement for the more expensive TD-DFT calculations used to obtain the response properties needed to make exciton models. This would work best when generating training data for the statistical methods discussed in section 2.2, as these require a smaller volume of calculation than a full time series of Hamiltonians.

For this second situation, the response method would still need to be as efficient as the lightweight machine-learning or tight-binding methods. Here it was investigated whether these Δ -SCF methods with tight-binding methods may solve this problem as well. These are novel response methods, and so required benchmarking against high-level data to assess their abilities.

This approach to finding efficient response methods for LHCs would then follow a similar pattern to that discussed in chapter 2. Using these more efficient response methods to calculate

transition properties, exciton Hamiltonians could be constructed either for a relatively smaller set of geometries using DFT based Δ -SCF or for a relatively larger set using tight-binding based Δ -SCF . This is all dependent on their performance.

Transition properties were calculated for a range of molecules, as well as for a small set of chlorophyll geometries, using variety of different basis sets, density functionals, response methods and electronic structure methods. Most of the work was compared to either high-standard EOM-CCSD or SCS-CC2 reference to draw conclusions on the accuracy of each method. Addressing the issue of non-orthogonality between the ground and excited states was also investigated for Δ -SCF . Generally it was found that the semi-empirical Δ -SCF methods may not be applicable to LHC models for reasons discussed in the following sections.

4.1 Combining Δ -SCF With Lower Level Electronic Structure Theories

It is clear that the efficiency of Δ -SCF methods could be increased by swapping out the usual DFT electronic structure methods with a tight-binding or semi-empirical approach. This would not cause any changes in the response method, only in the source of molecular orbital coefficients used to construct the excited states. There are several frameworks that could have been chosen for this purpose, but the recently published GFN-xTB method, parameterized by the Grimme group [65], was ultimately used. This method has been parameterized for geometries, frequencies and non-covalent interactions, and uses an extended version of Hückel theory. The name GFN-xTB is an acronym for "Geometries, Frequencies, Non-Covalent - eXtended Tight Binding". This method was chosen as a similar response method that calculates transition properties, the sTDA-xTB method, also exists and is the precursor to the GFN-xTB methods. Additionally the GFN-xTB method is implemented in many packages, including the QCORE package, and this significantly reduces the amount of effort required to implement and test. This became an important factor, necessary for the project discussed in the next chapter.

There are many forms of GFN-xTB, dependent on the level of theory of treated inter-electron interactions, named GFN1-, GFN2- and GFN0-xTB. Due to considerations about the gradients of transition properties (discussed in the conclusion chapter), GFN2-xTB was not included in this benchmarking. The method of using GFN1-xTB and GFN0-xTB electronic structure properties for Δ -SCF transitions is referred to as Δ -xTB .

4.2 Benchmarking Δ -SCF

4.2.1 Reference Data and test set

Full scale chlorophyll molecules are too large to be able to calculate a high-level benchmark. It would not be possible to test Δ -SCF data against coupled cluster data. Instead a stepwise

approach was used, covering a large chemical space by using a range of molecules before focusing on niche systems. This range of molecules could be smaller so high-level coupled cluster data could be generated.

A test set of small molecules, which would cover the same range of elements as found in organic chromophores and biological molecules, was chosen as to benchmark both the Δ -SCF methods as well as TD-DFT. The test set chosen was previously used by the Grimme group to parameterise and test the sTDA-xTB method [65], and as it was constructed to design new response methods it would contain a wide range of systems to appropriately benchmark the Δ -SCF methods.

The test set consisted of 109 small molecules. Each system was closed-shell, contained 12 atoms or less, and contained on H, C, N, O and F atoms. The three lowest energy singlet excited state transition energies were calculated using EOM-CCSD with an aug-cc-pVTZ basis set. These results were generated using the Gaussian 16 program [89].

4.2.2 Small Systems

Transition properties for this test set were calculated using TD-DFT and Δ -SCF , both using a CAM-B3LYP functional and aug-cc-pVTZ basis set. The transitions were assigned to the EOM-CCSD results by comparing transition dipoles, energies and the character of the MOs involved in the transitions. Symmetry assignments were also used where possible, but this was not the case for all systems either due to unsuccessful labelling or defaulting to a non-Abelain group. Symmetry labelling was also only available for TD-DFT calculations, as these were performed with Gaussian 16. Δ -SCF calculations were done with the QCORE program.

As the Δ -SCF singlet transition is not a correct representation of a true singlet excitation, which is a superposition of both spin-conserving $\alpha \rightarrow \alpha, \alpha$ and spin-flipping $\alpha \rightarrow \alpha, \beta$ excitations, the spin-purification formula:

$$(4.1) \quad \Delta E_S = 2\Delta E^{i,\alpha \rightarrow \alpha, \alpha} - \Delta E^{i,\alpha \rightarrow \alpha, \beta}$$

was used to correct for the true singlet excitation energy ΔE_S [90].

The results of comparing transition energies and transition dipole magnitudes are shown in figures 4.1 and 4.2.

Overall, the excitation energies calculated with Δ -SCF are as accurate at predicting EOM-CCSD energies as TD-DFT. The mean error is 0.35 eV, with a standard deviation of 0.25 eV. This is a marginal improvement on the TD-DFT results, which has a mean and standard deviation of 0.41 eV and 0.27 eV respectively. Transition dipoles were similarly accurate to the reference data, although Δ -SCF performs slightly worse in this respect. The mean and standard deviation in the absolute value of transition dipole moment, $|\mu|$, was 0.07 a.u. and 0.08 a.u. respectively. For TD-DFT, the mean and standard deviation were 0.03 a.u. and 0.06 a.u. (the atomic unit here

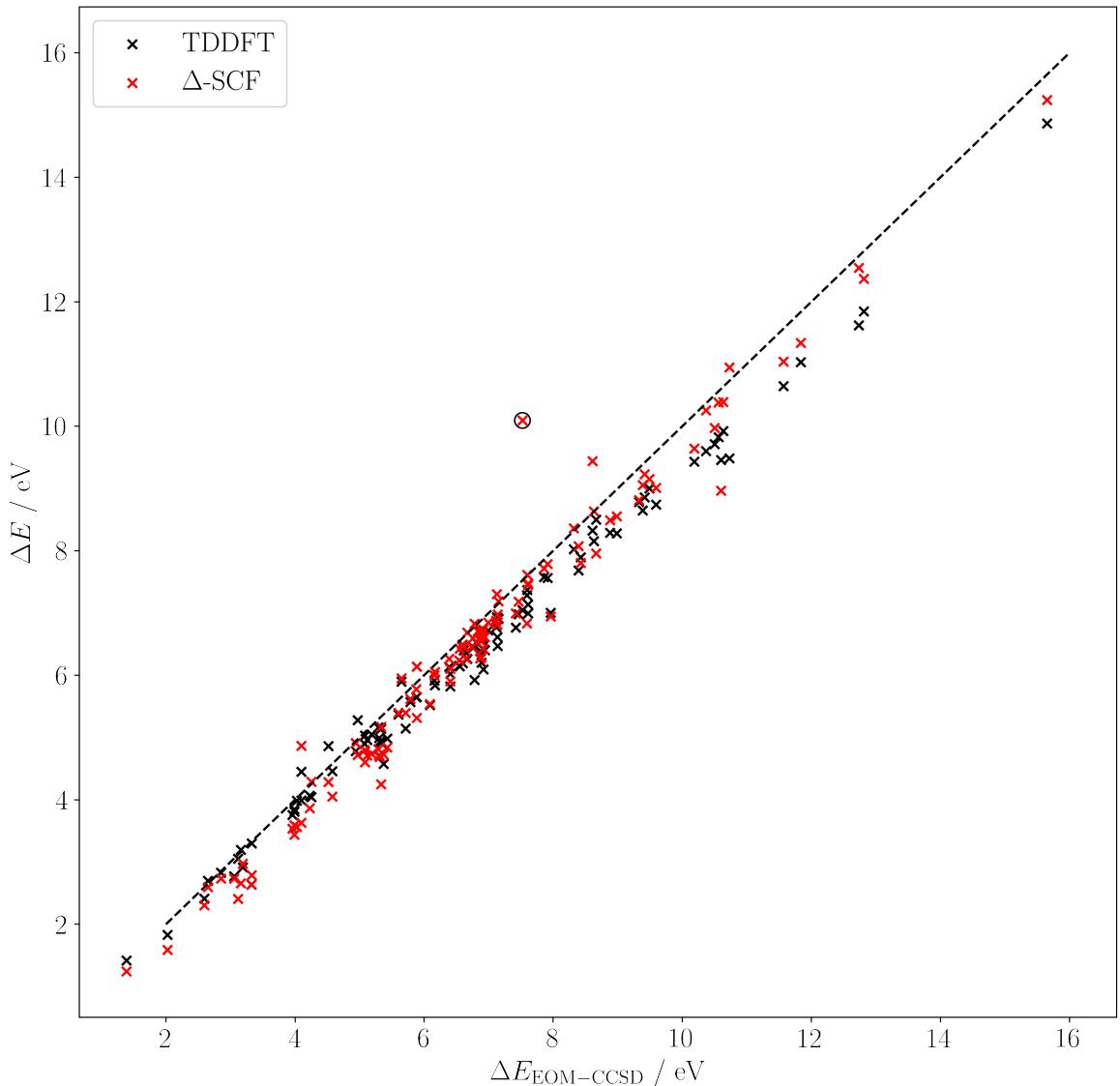


Figure 4.1: Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.

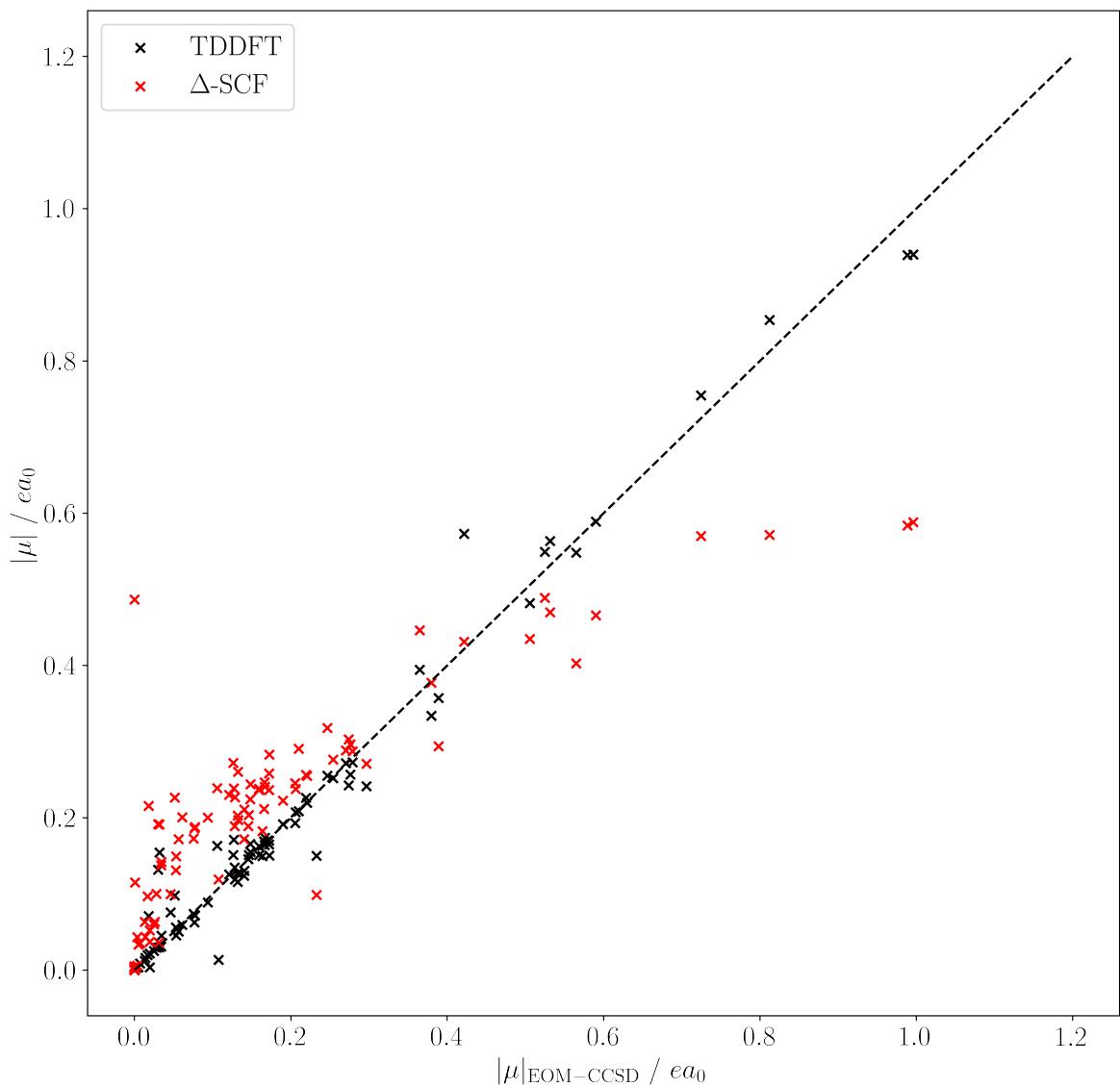


Figure 4.2: Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.

being equal to ea_0). The outlier circled in figure 4.1 is an ethene dimer system, and shows the inability of Δ -SCF to capture a mixed excited state. The two LUMO orbitals in this dimer system include in-phase and out-of-phase combinations of the π -antibonding orbitals, and are very close in energy. The HOMO orbitals are the same on both ethene molecules, being the π -bonding orbitals, which are degenerate in energy. The first excited state is predicted by TD-DFT and EOM-CCSD to be a mix of these two close HOMO-LUMO transitions. However Δ -SCF cannot include this mixed behaviour as it only considers single transitions. Δ -SCF predicts two transition energies of 7 eV and 10 eV, whilst TD-DFT and EOM-CCSD predict 7.5 eV. The outlier in figure 4.2 is due to Δ -SCF finding a different but still valid description of the transition dipole for a stretched benzene system, where the HOMO-1, HOMO, LUMO and LUMO+1 orbitals are all degenerate. The Δ -SCF transition dipole magnitude agrees with the value of an equally mixed HOMO - LUMO+1 and HOMO-1 - LUMO transition, which given the degeneracy is an equally valid description of the transition.

In summary, Δ -SCF can be seen to accurately predict transition properties to a EOM-CCSD level of accuracy with as much success as TD-DFT, except in cases of mixed transitions.

4.2.3 Non-orthogonality

Generally the ground and excited states calculated for Δ -SCF transition, being solutions from two separate SCF cycles, will not be completely orthogonal. The Slater determinants $|\Psi_n\rangle$, are constructed from a set of mutually orthogonal orbitals $\{|\phi_j^{(n)}\rangle\}$, such that orbitals will be orthogonal within the same state. However there is no orthogonality constraint on sets of orbitals derived from independent SCF cycles so it is possible for the overall states Ψ_1 and Ψ_2 to overlap such that the inner product,

$$(4.2) \quad S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle$$

will be non-zero. Similarly, there will be a non-zero transition charge,

$$(4.3) \quad q^{21} = \langle \Psi^2 | \Psi^1 \rangle$$

, which breaks the origin-independence property of the transition dipole moment. In this way, any transition dipoles that do not have their centre at the origin will have a systematic error based on this overlap and the distance from the origin. For vertical transitions, this transition charge should be zero, and so all transition dipole moments calculated with non-orthogonal Δ -SCF would always have this error.

In order to fix this issue, the standard transformation to symmetrically orthogonalise the two states was applied, which also would preserve as much character of the original states as possible. The transformation is given by

$$(4.4) \quad |\Psi_{\tilde{v}}\rangle = \sum_v |\Psi_v\rangle \left[\mathbf{S}^{-\frac{1}{2}} \right]_{v\tilde{v}}$$

where \mathbf{S} here is a block matrix

$$(4.5) \quad \mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

with S being the overlap value of the two states $\langle \Psi_2 | \Psi_1 \rangle$.

It was found that using this method for correcting the non-zero overlap of states, the origin-independence of the transition dipole moment was recovered (see figure 4.2.3). The transition dipole for each molecule in the test set systems was calculated for molecules translated by 100Å in the xyz direction. This would induce an error for the non-orthogonalised states, which was corrected by symmetric orthogonalisation. It should be noted that whilst this effect is dependent on how large the overlap may be, and it could be argued that with a small overlap this effect may not be large, having any large translation of the molecule (on the order of hundreds of angstroms) can be seen to lead to completely incorrect transition dipole magnitudes. In a large protein system, where chromophores can easily be tens or hundreds of angstrom from the overall system origin, this would obviously present a much larger problem than for a vacuum-phase small molecule.

4.2.4 LH2 Chlorophyll

Having demonstrated that TD-DFT is a good proxy for high-level methods, Δ -SCF was benchmarked against TD-DFT data for a series of chlorophyll molecules. This would show how well Δ -SCF methods could be used in simulating a whole LH2 system.

The PBE0 functional with Def2-SVP basis set [91, 92] was used to calculate both TD-DFT and Δ -SCF properties. This has been used previously for BChla structures, and has been shown to be a good balance between accuracy and cost [43].

The high degree of correlation between Δ -SCF and the TD-DFT results demonstrates that the variations in transition energies are due to differences in geometry, rather than any noise from random error. The systematic error is expected, and all methods used would have an associated error that is usually removed by a constant parameter.

The error in transition dipole magnitudes is larger than that of the small test set. This error is about 0.42 a.u. larger, but without EOM-CCSD or another high-level method, it's unclear whether this error might be from TD-DFT or Δ -SCF. Additionally, there is a clear correlation between the transition dipole magnitudes from TD-DFT and Δ -SCF, and so whilst quantitative results would not be possible, qualitative statements could be made from Δ -SCF data. For example, whilst the exact value of a transition dipole moment at a given geometry may not be accurate, the change

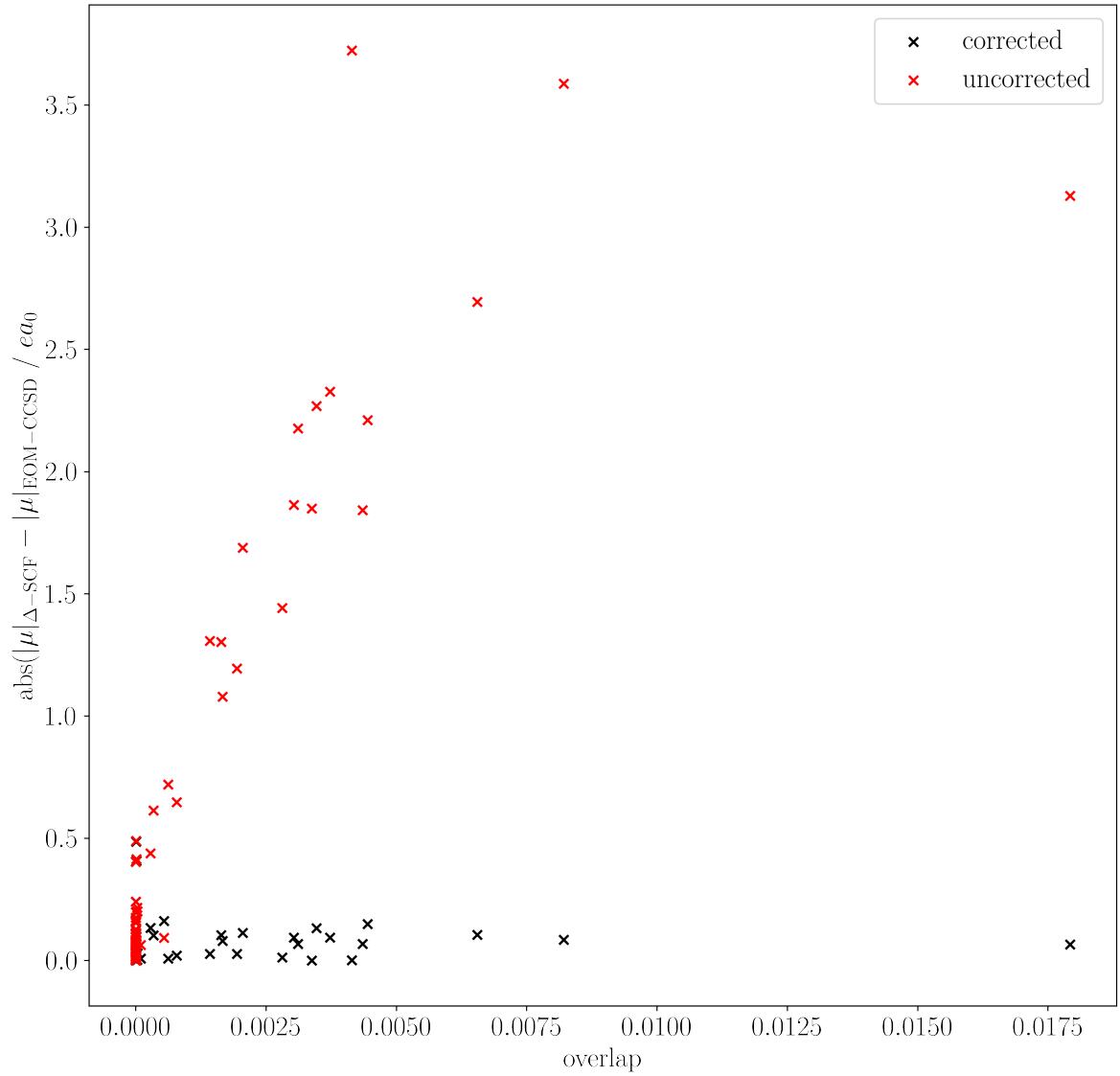


Figure 4.3: The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown in red, whilst those with the symmetric orthogonalisation correction are shown in black.

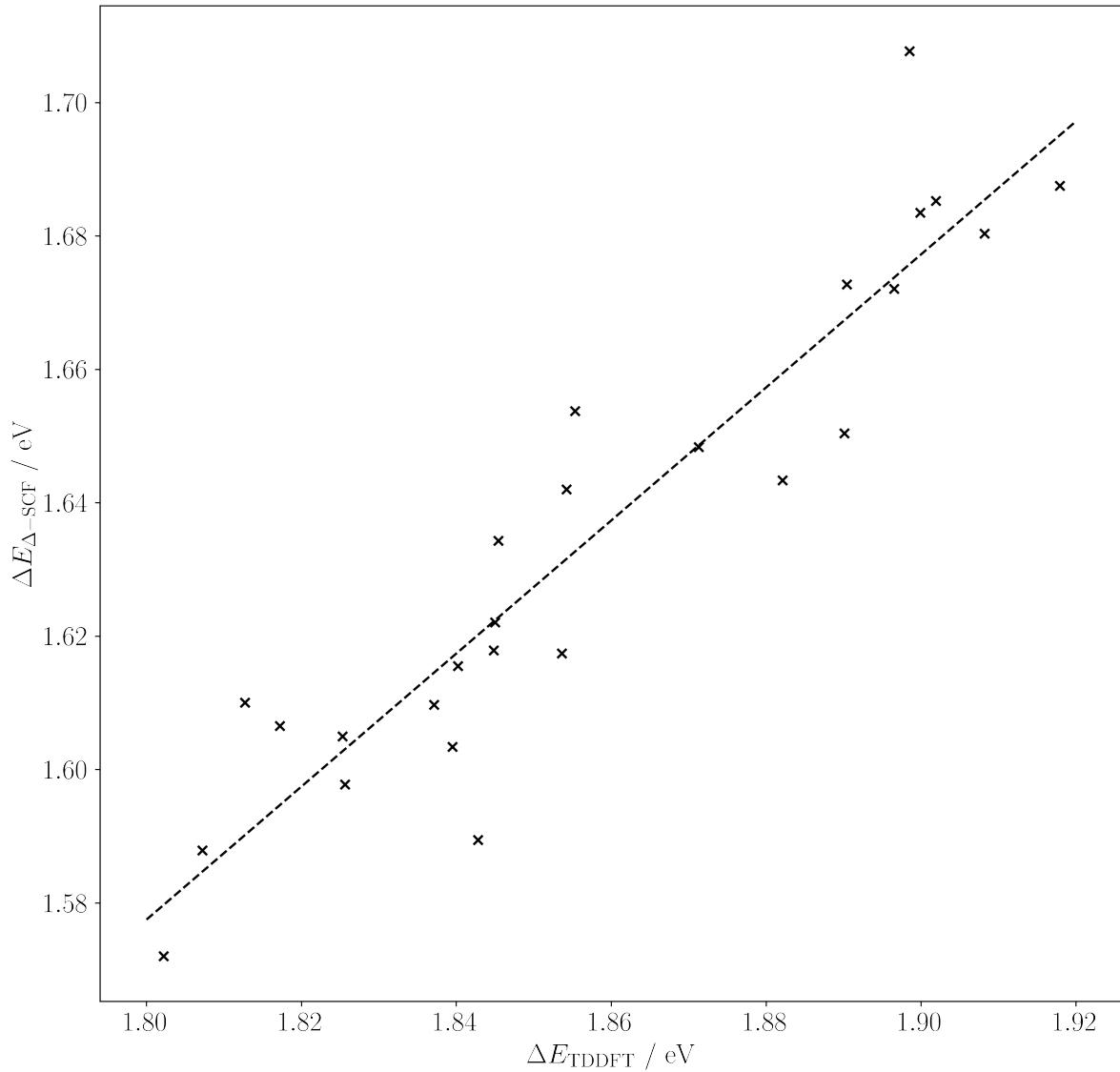


Figure 4.4: Transition energies from Δ -SCF for 26 chlorophyll geometries from the LH2 protein of purple bacteria, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

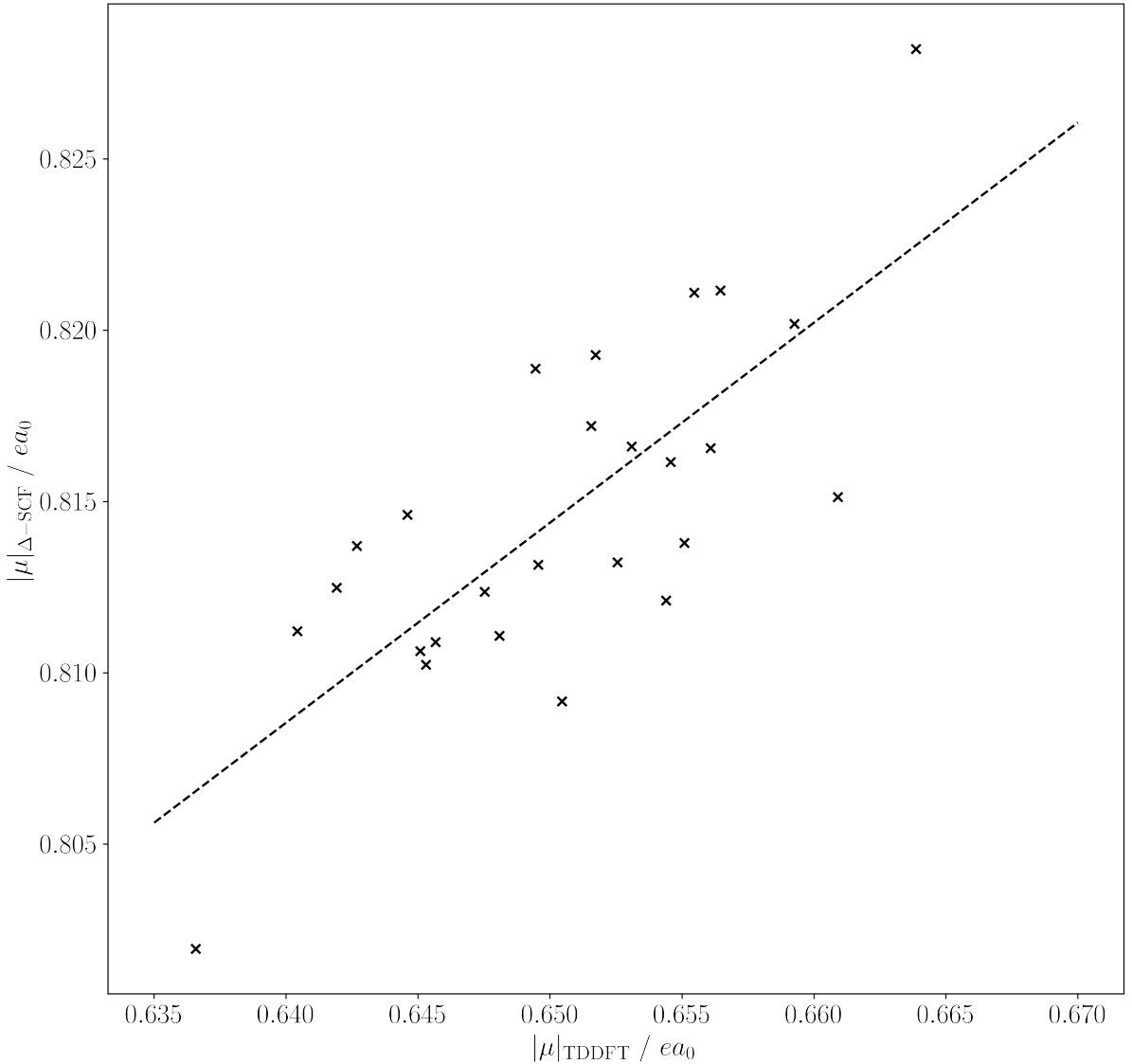


Figure 4.5: Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 4.2.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

in transition dipole moment from several geometries would be similar in variation to TD-DFT calculated properties.

An important point to take note of is that it was not possible to calculate the excited state for all 27 chlorophyll geometries, as one geometry repeatedly collapsed back down to the ground state. Methods such as including previous iterations' Fock matrices into the current Fock matrix (i.e. Fock damping), using intermediate initial guesses like half-electron promotions, and alternative DIIS procedures, had little effect on improving this collapse. Other methods could have been tried, such as the initial maximum overlap method (iMOM), where the projection in the MOM procedure is based on one static set of orbitals rather than the previous SCF cycle, but implementing and testing these methods is outside the scope of this work.

4.2.5 xTB methods

Δ -xTB was tested on the same small molecule benchmark set. Its performance was compared to other methods with a range of approximations for calculating transition properties, including a proxy for sTDA-xTB results. To include this in the comparison it was necessary to use spin-component-scaled second order coupled-cluster (SCS-CC2) [93, 94] reference data produced by Grimme et al. [95]. Two types of Δ -xTB were investigated, based on GFN1-xTB and GFN0-xTB properties. GFN0-xTB is similar to the GFN1-xTB method, but excludes any charge dependent terms in its Fock matrix so is not self-consistent. Therefore only a single diagonalisation is necessary for GFN0-xTB, and the same molecular orbital coefficients are used for ground and excited state. A transition energy from Δ -SCF with this method would functionally be the same as an eigenvalue difference. The eigenvalue differences from sTDA-xTB, given by the `xtb4stda` program [29], were also included in this benchmarking. This was done to observe whether the electronic structure or the response method is more important in achieving accuracy with the sTDA-xTB method. This was done with the `xtb4stda` program, the first of two programs that runs a version of xTB which provides molecular orbital coefficients and energies for the sTDA program to use to calculate the transition properties.

The full list of methods included in the comparison is:

- High level TD-DFT, with a range separated functional CAM-B3LYP and a aug-cc-pVTZ basis set.
- Lower level TD-DFT, with a PBE0 functional and smaller Def2-SVP basis set.
- Δ -SCF with CAM-B3LYP/aug-cc-pVTZ.
- Δ -SCF with PBE0/Def2-SVP.
- linear response with GFN1-xTB and GFN0-xTB.
- Δ -SCF with GFN1-xTB and GFN0-xTB, named Δ -xTB .
- Full sTDA-xTB.
- sTDA-xTB eigenvalue difference.

4.2.5.1 Post-SCF Assignment of Symmetry

A source of error that hasn't been discussed in much detail so far is the assignment of transitions between different methods. A known problem of Δ -SCF methods is that the excited state SCF cycle may not converge to the correct state, or it might collapse back to the ground state. This could be seen in the symmetry of the excitation - if Δ -SCF has converged to a different transition than TD-DFT and CC2, the symmetry label of transition would also be different. The benchmarking discussed previously used symmetry labels to assign TD-DFT transitions to EOM-CCSD but as

noted earlier this was not always possible, and assigning symmetry labels to Δ -SCF was not possible. Instead, transition dipole orientations and plots of MOs were used. Whilst a possible approach for Δ -xTB, the difference between Δ -xTB valence basis sets and full DFT basis sets caused issues. Instead the issue of assigning symmetry to Δ -SCF transitions was investigated in more detail.

Assigning symmetry to Δ -SCF transitions would require assigning symmetry to both molecular orbitals and the overall wave-function of a molecule. Most electronic structure codes have two choices in assigning symmetry to orbitals - either all of the SCF code will treat symmetry from the outset, or nothing is assigned in the SCF code and assignment will happen post-SCF. Both these approaches have benefits and drawbacks.

The first method means the symmetry to be known at any point in the SCF procedure, and allows the Hamiltonian to be organized into a block matrix which is more efficient to diagonalise. This can be especially useful when solving for a large basis set or large system, as the matrix diagonalisation can be partitioned and parallelised over several cores or nodes on a cluster computer. However this works best if the system is highly symmetric, which is often only the case when treating optimized systems. This discounts systems from molecular dynamics simulations as well as many biological systems.

The second approach of assigning symmetry after the SCF cycles does not fix these drawbacks but it does allow for codes which originally didn't have symmetry assignment to be used without rewriting SCF code. The obvious drawback of doing assignment post-SCF is that symmetry can't be utilized during the SCF procedure.

The symmetry label of an orbital is given by the character of that orbital in symmetry subspaces of the point group of the system. These characters are calculated by transforming the molecular orbital coefficients \mathbf{C} into each subspace A by using the transformation matrix \mathbf{T} , defined by

$$(4.6) \quad \tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{C}$$

, and then summing the coefficients to obtain the character

$$(4.7) \quad P_A = \sum_v |\tilde{\mathbf{C}}_{A,\mu v}|$$

, where μ, v are indices for the atomic and molecular orbitals respectively. The molecular orbital with character equal to 1 in a subspace would then have that symmetry label, and would be a well defined assignment. However, in practice this was not always clear cut and so the highest subspace character was taken as the assignment.

The steps for assigning orbital symmetry post-SCF is as follows:

1. Determine the point group of the molecule, from the atomic positions. This determines the symmetry subspaces.

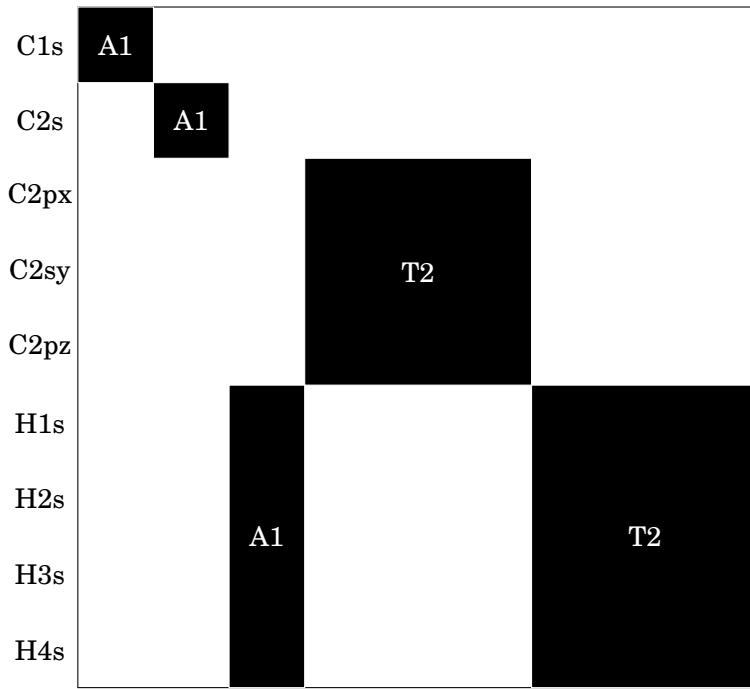


Figure 4.6: A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.

2. Get the symmetry adapted linear combinations (SALCs) of atomic orbitals for each subspace.
3. Use the SALCs to construct the transformation matrix T that will .
4. Assign the one electron molecular orbital (MO) for these subspace characters with the symmetry adapted linear combinations.
5. Multiply the one electron MO symmetries together to find the symmetry of the overall wavefunction.

This procedure was implemented and tested on methane with a minimal STO-3G basis set, using the open source library libmsym [96] for point group assignment functions and finding SALCs. The MOs for an optimised methane geometry with an STO-3G basis set were correctly assigned using this method (two occupied orbitals and one unoccupied orbital of A1 symmetry, and three occupied and unoccupied orbitals of T2 symmetry). A diagrammatic representation of the subspace characters of molecular orbitals is shown in figure 4.2.5.1. The overall wavefunction symmetry can then be expressed as the product of all MO symmetries, reduced with the reduction formula

$$(4.8) \quad n = \frac{1}{h} \sum_R \xi_r(R) \xi_i(R)$$

Method	Mean / eV	Standard Deviation / eV
TD-DFT CAM-B3LYP/aug-cc-pVTZ	-0.18	0.34
TD-DFT PBE0/Def2-SVP	-0.06	0.79
Δ -SCF CAM-B3LYP/aug-cc-pVTZ	-0.14	0.28
Δ -SCF PBE0/Def2-SVP	-0.62	0.50
TD-GFN1-xTB	0.27	1.47
TD-GFN0-xTB	-0.41	1.32
Δ -SCF GFN1-xTB	-0.12	2.11
Δ -SCF GFN0-xTB	-1.50	1.08
xtb4stda	4.39	1.26

Table 4.1: Mean and standard deviations of the errors, in eV, against SCS-CC2 reference data. The `xtb4stda` entry represents the eigenvalue difference method that uses the eigenvalues output from this program.

, where ξ_r, ξ_i are the reducible and irreducible representations respectively, h is the order of the group and R is the subspace. This correctly produced the overall symmetries of ground state systems for methane, as well as water.

Generally the subspace characters were binary for these easy test cases, however MOs of excited states were found to be mixed, making any assignment unclear for many MOs. This was also the case for non-abelian groups, where there would be degenerate subspaces. This is similar to the problem of assigning symmetry for the TD-DFT and EOM-CCSD transitions in the earlier benchmarking. Often this caused reduction of ground state wavefunctions to give non-physical answers.

Overall while improving some features of Δ -SCF methods is in the scope of this project, this type of assignment is flawed for LH2 system. Although it would be a useful feature for the benchmarking test-sets, chlorophyll molecules would be far from symmetric and so this type of assignment could not have been expected to have worked. Ultimately transitions could not be confidently assigned for Δ -SCF with this method. Hence while careful consideration was made for this problem, benchmarking of the Δ -xTB methods could be expected to still have some errors associated to this issue. In the end the assignment of symmetry was based on the previously used inspection of transition dipole orientations and transition density plots.

4.2.5.2 Δ -xTB Benchmarking Results

The distributions of errors to SCS-CC2 data for each of the benchmarking methods, as well as a generated distribution of sTDA-xTB results, are shown in figure 4.7. The means and standard deviations are reported in table 4.1.

Overall, both Δ -xTB methods are inaccurate - far too inaccurate to be used as a viable method for transition properties of chlorophyll, or any other system. The mean error GFN1- Δ -xTB was -0.12 eV, and has a significant standard deviation of 2.11 eV. GFN0- Δ -xTB had a larger mean error

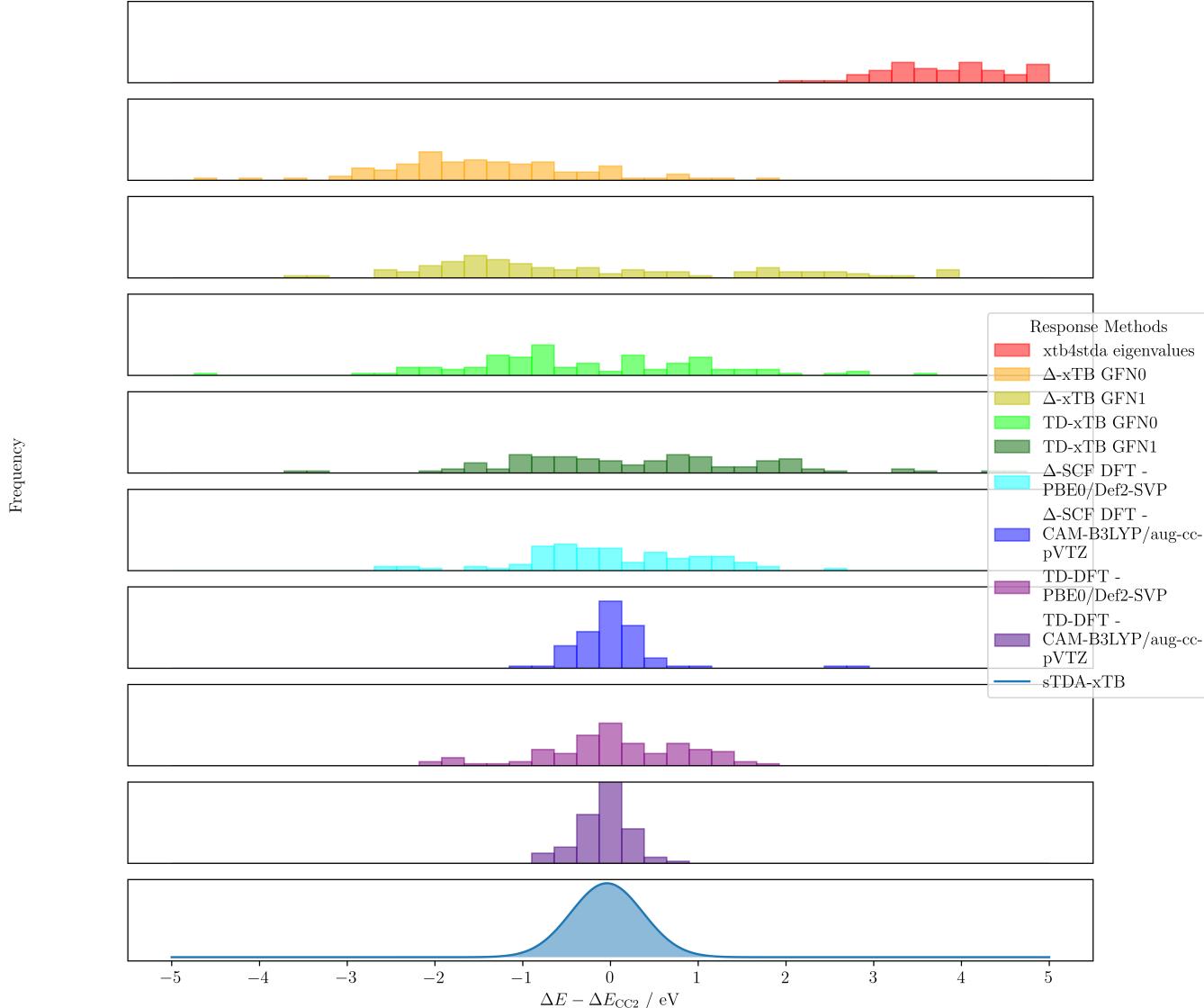


Figure 4.7: The distributions of errors compared to SCS-CC2 transition energies for the methods included in the $\Delta\text{-xTB}$ benchmarking.

of -1.50 eV, and whilst a slightly smaller standard deviation of 1.08 eV, this is still well beyond a usable accuracy.

The DFT methods, both the linear response and the Δ -SCF methods, are still shown to be accurate at predicting excitation energies, with means and standard deviations within ranges previously reported in the above sections.

Due to these results it is argued that decreasing the level of theory for the electronic structure dramatically decreases the accuracy of transition properties. The highest level electronic structure method has the best accuracy. CAM-B3LYP/aug-cc-pVTZ TD-DFT and Δ -SCF have mean errors of -0.18 eV and -0.14 eV and standard deviations of 0.34 and 0.28 eV respectively. Both methods are well within the accuracy needed to predict geometry-based variations. The outliers in the Δ -SCF results are known mixed transitions, as discussed earlier with the ethene dimer system. The lower level PBE0/Def2-SVP methods have a marked decrease in accuracy. On going from higher-level DFT to lower level, the standard deviation approximately doubles for both TD-DFT (0.34 eV to 0.79 eV) and Δ -SCF (0.28 eV to 0.50 eV). Again, the PBE0/Def2-SVP TD-DFT and Δ -SCF are comparable, with standard deviations of 0.79 eV and 0.59 eV, although the mean for Δ -SCF has a significant shift of -0.62 eV.

Comparing the DFT and GFN based methods, we can see the same trend that lowering the level of theory gives worse transition properties. Overall, the most inaccurate method is the eigenvalue difference methods based orbital energies (eigenvalues of the Hamiltonian diagonalisation) from the `xtb4stda` method. The means and standard deviation was 4.39 eV and 1.26 eV respectively, a huge difference to the values for values for full sTDA-xTB (-0.04 eV and 0.41 eV respectively)[29]. Arguably then the sTDA method, and not the underlying xTB method, can make up a large part of the accuracy for predicting transition properties. A similar response theory then might be expected to perform equally well, which is investigated in more detail in the next chapter.

The result that GFN-xTB based methods are not accurate is not unexpected. As opposed other DFT methods, which use *ab initio* or first principle parameters, the xTB methods were fit to target properties and so would not be expected to be suitable for other properties outside the training data [28]. Whilst the lack of pair-wise parameters and "top-down" parameterisation approach which gives GFN-xTB a better number and specificity of parameters compared to other methods, these parameters only extend the accuracy of predicted properties to different chemical systems and not to different properties altogether.

4.3 Conclusions

The transition properties of a test set of small molecules has been benchmarked with multiple Δ -SCF , TD-DFT and high-level methods. It has been shown that DFT based Δ -SCF and TD-DFT methods can reproduce the same transition energies and transition dipole magnitudes as EOM-

CCSD to within reasonable levels of accuracy. For the set of small molecules, transition energies were predicted with a mean of less than 0.5 eV, and 0.07 a.u. for transition dipole magnitudes. Additionally, the issue of breaking the origin independence property of transition dipoles has been shown to be fixed by using a symmetric orthogonalisation of the two originally non-orthogonal states.

For a small set of BChla geometries, it was found that the same level of accuracy for transition energy could be found between Δ -SCF and TD-DFT, where EOM-CCSD was too expensive to calculate. The error was well within the range of TD-DFT energy variation, shown in the high correlation coefficient, and so Δ -SCF could be reasonably expected to give correct geometry-dependent transition energies. Whilst the accuracy is slightly reduced for transition dipole moments, the appreciable degree of correlation implies that qualitative statements would be valid.

With all of the above benchmarking, reliably obtaining and assigning transitions predicted from Δ -SCF has proved to be an unsolved issue. Either Δ -SCF is formally unable to predict the correct character of transitions, as showcased in the ethene dimer mixed transition outlier, or it is unreliable in finding excited state solutions. This is best shown in the exclusion of a geometry of chlorophyll that could not be made to converge to the correct excited state.

To solve the inability of currently implemented Δ -SCF to assign symmetry labels, a post-SCF method of assigning MO and full wavefunction symmetry was investigated, but ultimately proved beyond the scope of this project. Whilst able to assign labels for small, trivial systems of STO-3G water and methane, non-trivial excited states and more complex systems did not work. Whilst there is more work that could be done in this area, it was decided that this should be moved to potential further work on Δ -SCF methods.

GFN-xTB based methods, named Δ -xTB were found to be inaccurate to the point where it they would not be a useful proxy to higher level methods. Due to the similarity in results for linear response and Δ -SCF methods over a range of electronic structure methods this drop in accuracy is attributed to the different electronic structure theory rather than the response method. This implies that altering the electronic structure method could lead to great improvements in the accuracy of a new response method.

The aim of this chapter was to determine whether Δ -SCF methods, which have a simple gradient theory and are less costly than TD-DFT, could provide a sufficiently accurate description of transition properties for use in an *ab initio* exciton framework. It has been shown that this is true with the condition that the underlying theory is sufficiently high. However there is still the outstanding issue for calculating properties for a large volume of chlorophyll structures. The efficiency of DFT calculations may be too low for a large exciton framework or for large monomer systems like Bchla. Semi-emperical Δ -SCF methods, which would be efficient enough, prove innaccurate in their current form. However, it is demonstrated that a correct electronic structure and a "top-down" parameterisation could make an accurate semi-emperical method, which is

4.3. CONCLUSIONS

investigated in the following chapter.

CHLOROPHYLL SPECIFIC METHODS

This chapter reports on designing and parameterising a novel method for response properties, referred to as chl-xTB. The framework and theory for the method is outlined in section 5.1. Parameterisation details are given in section 5.2, including the reference data used to create the training data, the objective function and optimisation algorithms. The accuracy of this new method is showcased in the final section 5.3.

From the previous chapter it was found that the tight-binding based Δ -xTB methods are not the solution to LHC model issues outlined in the introduction. They are too inaccurate, and the problem of comparing transitions to high-level methods makes analysis difficult. Additionally, the issues with convergence to excited states mean this method would be unreliable.

However, as stated in the previous chapter, the sTDA-xTB method does imply that a tight-binding based approach would work, the only issue is the response method. In the introduction it was argued that sTDA-xTB was not suitable for LHC exciton models as the minute variations in chlorophyll geometry may not be accurately treated. In this chapter it is posited that method that is specific to chlorophyll would solve this problem. This is similar to the machine-learning models discussed, where models predicted only Q_y transition energies. Here it is argued that electronic structure methods could still be used, and would be much easier to parameterise than NN or KRR methods, if the response method was as lightweight as possible.

Approximations to TD-DFT are proposed in order to make this lightweight response method. The aim of this process is to end up with a model that has as few parameters as possible to parameterise. This would reduce the chances of overfitting to data but would also increase the reusability of the approach reported here, an important feature of any LHC model. These are combined with the GFN1-xTB model to give a novel response method, which is referred to as chl-xTB.

This method fits into the wider picture of LHC models in a similar way as Δ -SCF . Exciton Hamiltonians would be constructed from the transition properties calculated by the chl-xTB method for chlorophyll geometries. As this method would be at least as efficient as other tight-binding methods, if not more due to the lightweight response method, calculating Hamiltonians for a large number of LHC geometries would be possible, an idea explored in section 6.4 and chapter 7.

5.1 Approximations in chl-xTB

5.1.1 Full TD-DFT Solutions

In linear-response TD-DFT, excitation energies and transition characters are given by the solutions of the non-Hermitian eigenvalue equation

$$(5.1) \quad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

where \mathbf{A} , \mathbf{B} are matrices whose elements describe the perturbation of the electron density in response to light interaction, the \mathbf{X} , \mathbf{Y} solutions are coefficients of excitations, similar to CIS, and eigenvalues ω are the excitation energies.

The elements of \mathbf{A} and \mathbf{B} correspond to descriptions of the virtual-occupied and occupied-virtual contributions respectively, and in TD-DFT, with a global hybrid density functional, are given by

$$(5.2) \quad A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2(i|a|jb) - a_x(i|j|ab) + (1 - a_x)(ia|f_{XC}|jb)$$

$$(5.3) \quad B_{ia,jb} = 2(i|a|bj) - a_x(ib|aj) + (1 - a_x)(ia|f_{XC}|bj)$$

where indices a, b and i, j refer to virtual and occupied orbitals respectively, ϵ_i is the orbital energy of orbital i , a_x is the value of non-local Fock exchange in the XC functional f_{XC} , and δ_{ij} is the Kronecker delta function. The intergrals (here in Mulliken notation) are Coloumb type for the \mathbf{A} matrix and exchange type for the \mathbf{B} matrix. The leading term in the \mathbf{A} matrix is the orbital energy difference, which only contributes to the diagonal elements due to the Kronecker deltas.

5.1.2 Approximations to Solutions

There are three approximations made in the chl-xTB method. First is the Tamm-Danoff approximation (TDA), which neglects "de-excitation" contributions to TD-DFT solutions. Second is a diagonal dominant approximation to eigenvalue solutions, where diagonal elements of a matrix

are taken as an approximation to the real eigenvalues. The third is to use a monopole approximation to calculate two electron integrals as point charge interactions, employing MNOK operators to accurately capture Coulomb and exchange functions. All these approximations reduce the amount of expensive electron integrals required to construct the full eigenvalue equation, whilst still capturing the important aspects of the electron density response.

5.1.2.1 Tamm-Dancoff approximation and Diagonal Dominant A matrices

One of the earliest approximations applied to full linear response was the Tamm-Dancoff approximation, where only virtual-occupied contributions are calculated. This sets all of the elements of **B** matrix to zero, and reduces the full eigenvalue equation to

$$(5.4) \quad \mathbf{AX} = \omega \mathbf{X}$$

where the definitions of elements are the same as above, but it should be noted that the solutions **X** will be different to full TD-DFT. This is formally the same as a CIS problem, and has been reported as a way to get back to CIS from full TD-DFT.

As an eigenvalue problem, solving for ω requires constructing and diagonalising the full **A**. The XC functional is dependent on ω values and so iterating through diagonalisation is needed to find stable, self-consistent solutions.

If the matrix is diagonal dominant, the diagonal elements of the **A** matrix can be used as an approximation to the eigenvalue solutions. This would be in the limit where the coupling elements (the off-diagonal elements) tend to zero, which would be the case for an excitation that is mostly made up of a single transition. Excitation energies would be given by

$$(5.5) \quad \omega_{ia} \approx A_{ia,ia} = (\epsilon_a - \epsilon_i) + 2(i|ia) - a_x(ii|aa) + (1 - a_x)(ia|f_{XC}|ia)$$

If the integrals are assumed to be small compared to the first term, the orbital energy difference, then the orbital energy difference can be taken as an approximation to the full excitation energy, recovering the eigenvalue difference method from the previous chapter.

This approximation also reduces computational cost by removing the need to calculate all of the off-diagonal elements as well as the diagonal elements for transitions that are not of interest. This approach would be expected to work best where the overall transitions are not mixed. If transitions are mixed, then the coupling elements would be non-zero, and depending on the proximity in energy of single transitions, the coupling values could have a large effect. This is considered in section 5.1.4.

5.1.3 Integrals approximations

Whilst using this diagonal dominant approximation removes a large portion of the integrals, those that are left would still require an iterative, self-consistent treatment. This can be simplified

by approximating these integrals as point charge interactions, as the charges would not be dependent on excited state solutions.

Atom centered transition charges for transition $m \rightarrow n$ can be given in the Mulliken scheme by summing the reduced one-electron transition density \mathbf{D}^{mn} multiplied the orbital overlap \mathbf{S} for all orbitals p, q centred on atom A

$$(5.6) \quad q_A^{mn} = \sum_{p \in A} D_{pq}^{mn} S_{pq}$$

. Partial charges can be given as the difference between the electronic charge and the nuclear charge Z_A

$$(5.7) \quad q_A = Z_A - \sum_{p \in A} D_{pq} S_{pq}$$

where \mathbf{D} is the reduced one-electron density. This applies to any state, so can be used to calculate ground state and excited state partial charges.

MNOK operators can recover the important behavior of full electron integrals that is lost in classical charge-charge interaction, especially for the small distance limit. For both coloumb and exchange type integrals, this can be done with a short-range damped MNOK operators. The integral is approximated by

$$(5.8) \quad (pq|rs) \approx \frac{1}{2} \sum_A^N \sum_B^N q_A^{mm} q_B^{mn} \Gamma_{AB}$$

where N is the total number of atoms in the system and p, q, r, s are electron indices for both occupied and virtual orbitals, and A and B are indices for the atoms. For Coloumb type integrals, the operator Γ_{AB} is given by

$$(5.9) \quad \Gamma_{AB}^J = \left(\frac{1}{(R_{AB})^{y_J} + (a_x \eta)^{-y_J}} \right)^{\frac{1}{y_J}}$$

where R_{AB} is the interatomic distance and η is the average of the chemical hardnesses. The chemical hardness of a given atom is defined as

$$(5.10) \quad \eta(A) = \frac{\delta^2 E(A)}{\delta^2 N^2}$$

but in practice is precomputed and used as a static parameter. y_J and a_x are global parameters, with the later used to recover the effects of Fock-exchange mixing in the short-distance limit. For exchange type, the operator is

$$(5.11) \quad \Gamma_{AB}^K = \left(\frac{1}{(R_{AB})^{y_K} + \eta^{-y_K}} \right)^{\frac{1}{y_K}}$$

where the y_K parameter replaces the y_J parameter in the Coulomb type operator.

As the a_x parameter can "mop up" many of the exchange effects, the density functional in equation 5.5 is also neglected to further reduce computational cost. The final form of the expression used to calculate excitation energies is then given by

$$(5.12) \quad \omega_{ia} = (\epsilon_a - \epsilon_i) + \sum_{A,B}^N \left(2q_{ia}^A \Gamma_{AB}^K q_{ia}^B - q^A \Gamma_{AB}^J q^B \right)$$

where the exchange term uses transition charges q_{ia} , and the Coulomb term uses partial charges from the ground state density.

It can be seen that the inclusion of MNOK operators introduces global parameters. These would require optimisation to a training set. Initially it was investigated whether parameterising to a training set with a broad range of systems and transitions would be possible. However it was found that parameter optimisation procedures, whilst improving upon the accuracy of the Δ -xTB methods of the previous chapter, could not break into the accuracy needed to investigate chlorophyll systems well enough (discussed in more detail in 5.2.3). This is due to many transitions being mixed, which would not be consistent with conditions needed for the diagonal dominant approximation. Additionally, issues with autonomous optimisation assigning transitions without symmetry made optimisation workflows difficult (see previous chapter section 4.2.5.1).

However a method that works for a wide set of systems is not required for LH2, as only the Q_y transition is of interest for many models. Parameterising to a single, well-defined transition would be a much better approach for this investigation. By reducing the scope of systems, the specificity of parameters can dramatically increase. This reduces the need to include more parameters to improve accuracy, as well as decreasing the amount of training data needed. Looking at other transitions and systems then is outside the scope of this work, but could be a large area for further work for similar investigations. An outline of the Q_y transition, and its applicability to these approximations, is given below.

5.1.4 Q_y Transition

The Q_y transition is a good candidate to test approximations to full linear response theory. It is a well-defined transition which makes assignment easy, and has been thoroughly analysed in the literature. It is also the most important transition in light harvesting systems, so an accurate treatment is necessary.

The Q_y transition is the one of the two transitions that make up the Q band in the absorption spectra of chlorophyll, the other being the Q_x transition. It is well known that the Q_y transition

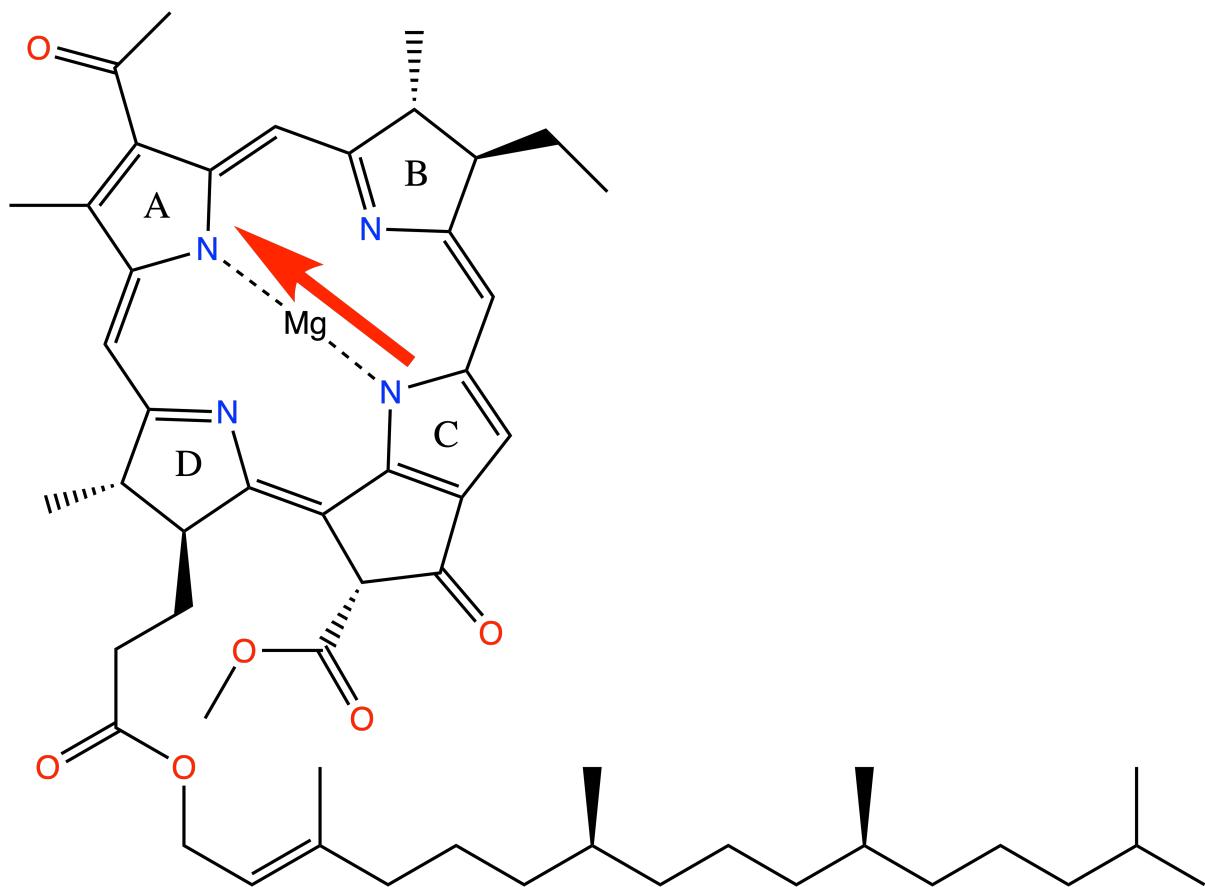


Figure 5.1: Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.

is important for electronic energy transfer, and predicting both transition energies and dipole moments correctly is important in constructing frameworks that model this transfer. The Q_y transition is mostly HOMO-LUMO in character (96%), with a small amount of HOMO-1 - LUMO+1 (remaining 4%). The analogous transition in the unsubstituted tetraphorphyrin ring has the transition dipole along the molecular axis defined by the N atoms, however due to the asymmetry introduced by substitutions and geometry deformations, this is usually not the case for BChla with a deviation from this axis of around 12 °. The Q_y transition in chlorophyll has its dipole component lying mostly the N_A - N_C axis, with Q_x lying orthogonally along the N_B - N_D axis. Q_x has the reverse character to the Q_y transition, being mostly HOMO-1 - LUMO+1.

Plots of the electron density of the HOMO and LUMO show how this transition is delocalised over large sections of the porphyrin ring, with approximate C_2 symmetry along the molecular axes. Notable contributions can be seen in the functional groups, giving the modified transition behaviour seen in different versions of chlorophyll.

It has recently also been shown that the high correlation between the eigenvalue difference of HOMO-LUMO orbitals and full TD-DFT excitation energies implies that the HOMO-1 - LUMO+1

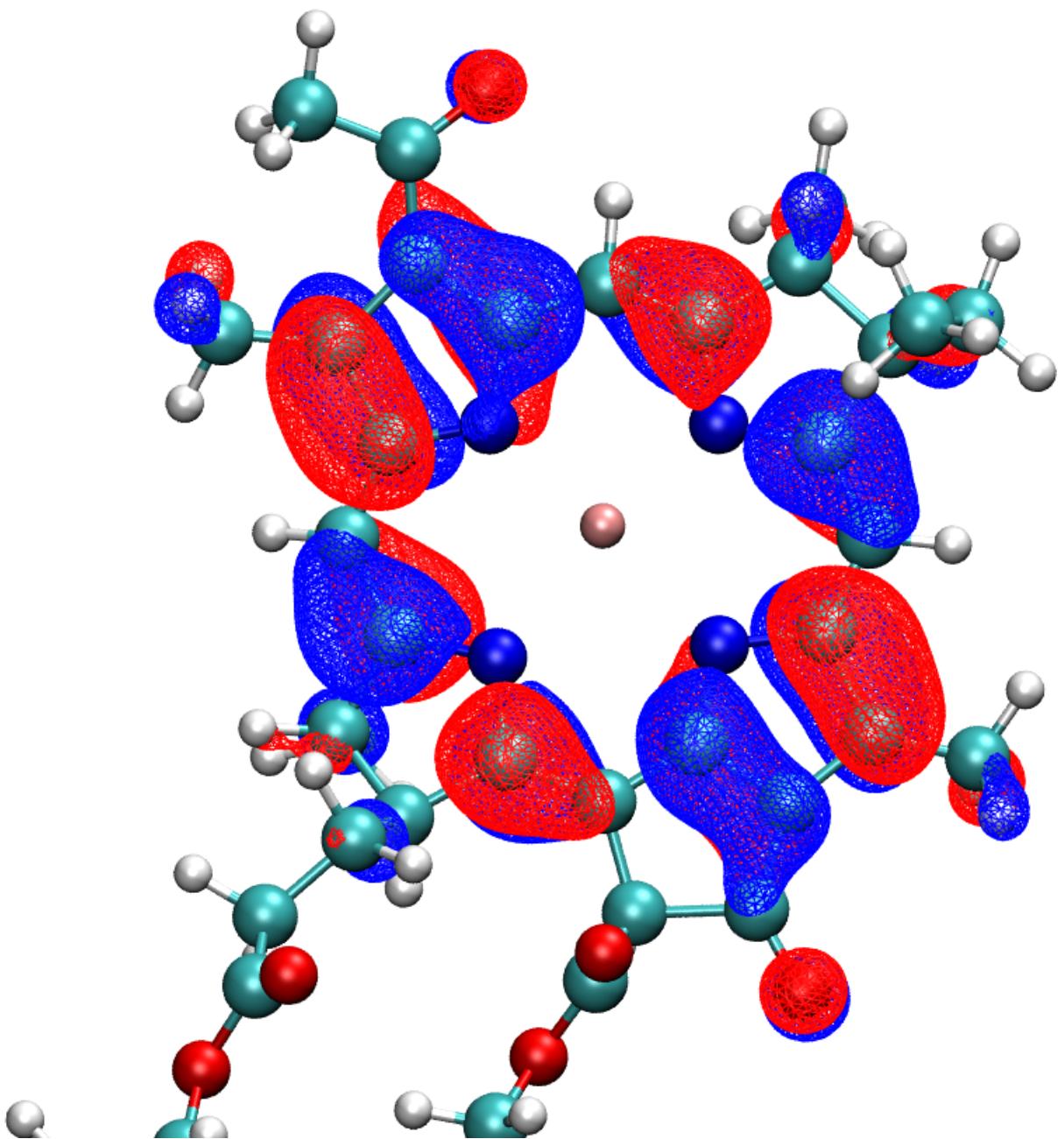


Figure 5.2: The HOMO orbital of Bchla from PBE0/Def2-SVP DFT.

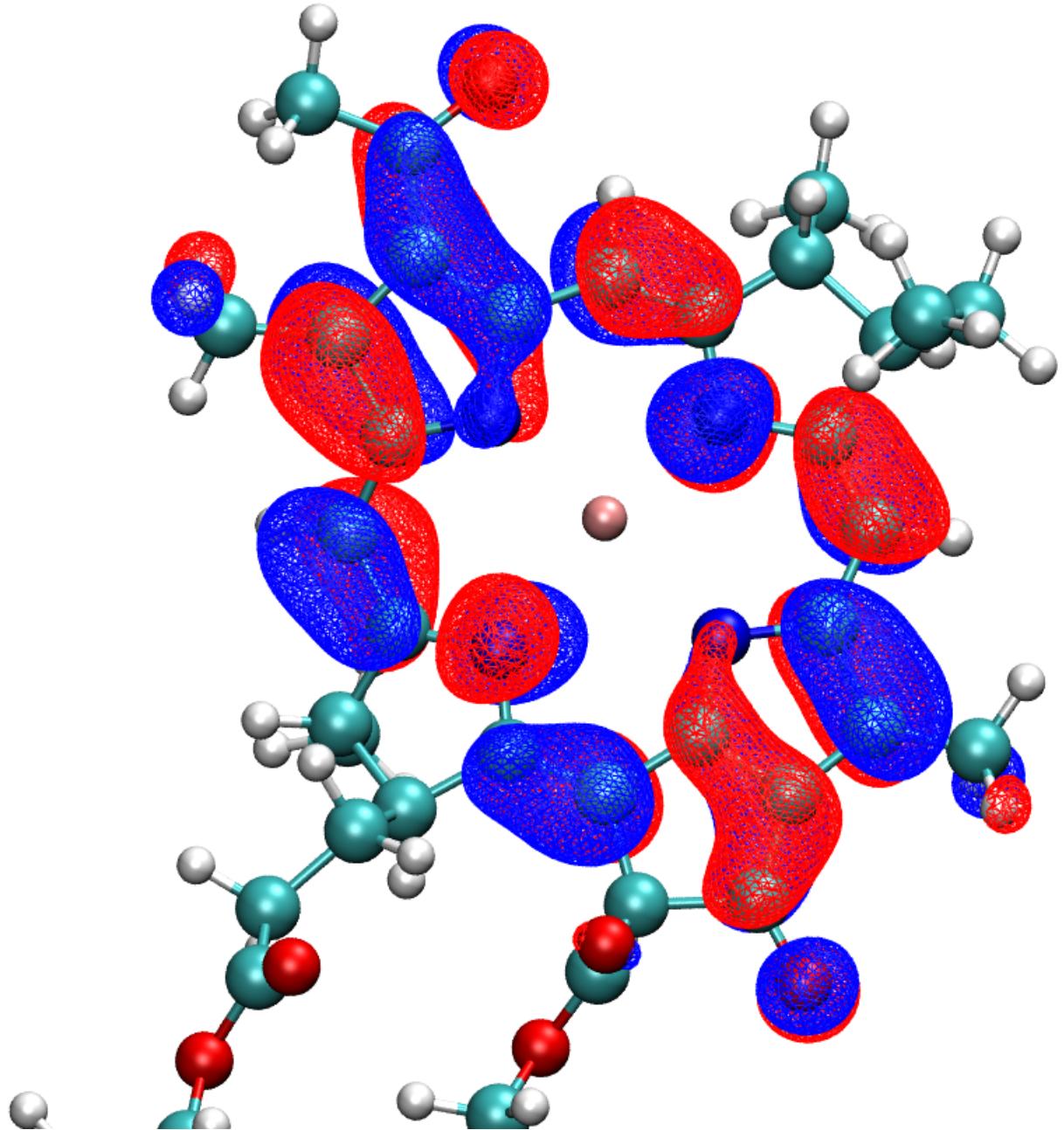


Figure 5.3: The LUMO orbital of BChla from PBE0/Def2-SVP DFT.

transition can be excluded from the transition character. This is also supported by the results for the chlorophyll test set in the previous chapter, as Δ -SCF with it's single transition treatment is able to capture Q_y transitions with good accuracy.

The high amount of single-transition character in the Q_y transition makes it ideal for the approximations to transition energies set out earlier. The lack of coupling elements that would appear in the \mathbf{A} matrix justifies the use of the diagonal dominant approximation.

Additionally, the well-defined transition make assignment trivial, and the transition dipole orientation to the N_A - N_C axis has been used as a metric for the accuracy of transition dipole moments. As a single value this is ideal for autonomous optimisation, both for use in an objective function as well as for discarding outlier transitions.

5.1.5 Changes to xTB parameters

As found in the last chapter, the underlying electronic structure can have a huge effect on the accuracy of transition properties, in some cases being more effective than a change in the response method. The xTB methods in particular, both linear response and Δ -xTB , were particularly ill-suited for transition properties. Whilst DFT methods could have been used as for Mulliken partial and transition charges, this would not be any more efficient than the Δ -SCF methods discussed in the previous chapter. A tight-binding, semi-empirical approach for the electron structure is still required. To improve the applicability of the xTB methods for transition properties, some of the parameters would need to be altered. A top-down approach can be used for these alterations, as it has been shown to work well for the GFN-xTB and sTDA-xTB methods.

As discussed in the introduction chapter, the GFN1-xTB Fock matrix is made of both charge-dependent and charge-independent terms. The form of these terms, and definitions of parameters, is also given in the introduction chapter. Only the charge dependent terms would have any effect on the partial and transition charges, thereby effecting the transition properties. The charge dependent terms are the first, second and third density fluctuation terms. The first order term is the leading term, and so altering parameters in this term would have the greatest effect. Only some of these parameters are "free", with others calculated from physical or *ab initio* values. The "free" parameters are the Hückel parameters k_l , where l is the angular momentum of the orbital, and the global scaling parameters. The global scaling parameters are used to adjust interactions for some pairs of elements where the global parameters led to erroneous properties, usually observed in the bond lengths. It was found that for the Q_y transition, only Mg and N interactions had to be scaled.

An obvious drawback of altering these parameters to fit transition properties is that they would lose their specificity to the GFN-xTB training set. However it is not required, or in the scope of this work, to find a semi-empirical method that would be able to calculate all properties for chlorophyll. chl-xTB is not used to calculate optimised geometries or hessians of chlorophyll, as other methods would be expected to much perform better. If it were necessary to include

more target properties in the parameter optimisation the accuracy to any one target would decrease, making the method worse overall. For example, the sTDA-xTB and GFN-xTB methods use different electronic structure parameters for this reason - one for transition properties, the other for geometries.

5.1.6 Transition and Excited State Density from the Ground State

Similar to Δ -SCF, the ground state orbital coefficients are taken to be a good approximation to the (Q_y) excited state coefficients. It was assumed that the excited state could be calculated without orbital relaxation, so that both the excited state and ground state could be constructed from the same set of molecular orbitals. The transition density is then calculated as

$$(5.13) \quad \mathbf{D}^{01} = |\Psi^0\rangle\langle\Psi^1|$$

with $|\Psi^0|$, $|\Psi^1|$ (the ground and excited state respectively) being constructed from the same set of molecular orbital coefficients \mathbf{C} but with different sets of occupation numbers for the ground and excited state (labelled $n^{(0)}$ and $n^{(1)}$ respectively). The ground state and excited state density can be similarly calculated as

$$(5.14) \quad \mathbf{D}^{(0)} = \sum_i n_i^{(0)} C_{ip}^{(0)} C_{iq}^{(0)}$$

$$(5.15) \quad \mathbf{D}^{(1)} = \sum_i n_i^{(1)} C_{ip}^{(0)} C_{iq}^{(0)}$$

. These density matrices are used to calculate the partial and transition charges with the Mulliken scheme, which is in turn used to calculate the MNOK integrals. The transition and molecular dipoles can be calculated as the trace of the dipole operator with the density.

It should be noted the ground and excited states would be orthogonal in this scheme, as they share the same set of MO coefficients. Additionally, the problematic excited state convergence is unnecessary as only the ground state cycle of the SCC procedure is needed. Excited state properties, such as the molecular dipole and partial charges can also be calculated from the excited state density, which will be important for the exciton framework of the next chapter.

It was found, however, that transition dipoles calculated using this method were much larger than TD-DFT. This was also observed for the Δ -SCF and eigenvalue difference methods, implying that the inclusion of the HOMO-1-LUMO+1 transition is key to accurately describing the transition density. To recover this, an additional parameter D_{sc1} was included to scale the transition density (i.e. $\tilde{D}^{01} = D_{sc1}D^{01}$ where \tilde{D} is the scaled transition density) which both yielded the correct transition dipole magnitudes as well as drastically increasing the accuracy of the method overall. This is discussed more in section 5.2.3.

5.2 Parameterization

An objective function and algorithm is required to find minima in parameter space when fitting models, where the minima correspond to an optimised set of parameters. Reference data is also required as a target. All of these elements are important for minimising the amount of error in the final method, as well as determining how well a method can perform when used for different problems.

5.2.1 Objective Function

The first metric that should be minimised is the root mean squared error (RMSE), where the error is to the reference transition energies, giving the objective function as

$$(5.16) \quad f_{\text{RMSE}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_i^N (\Delta E_i - \Delta E_{i,\text{ref.}})^2}$$

where \mathbf{x} is the set of parameters, ΔE_i , $\Delta E_{i,\text{ref.}}$ are the transition energies for system i from the chl-xTB method and reference method respectively, N is the number of systems used to calculate the objective function value (this can be different when looking at training and testing sets).

However just using the RMSE has two issues. First is that other transition properties are not included in the optimisation, and so no comments can be made on the accuracy of, for example, transition dipoles. This can be fixed by including a metric for the error in other properties of interest. The other issue is that a low RMSE does not guarantee a high correlation. A measure of the correlation can be given by the coefficient of determination

$$(5.17) \quad R^2 = 1 - \frac{\sum_i^N (\hat{y}_i - y_i)^2}{\sum_i^N (\hat{y}_i - \bar{y})^2}$$

where \hat{y} , y are the predicted and reference values respectively and \bar{y} is the average of the reference values. The correlation is a better metric for determining if chl-xTB has a small enough random error to predict transition properties, however it may not account for systematic errors. Both a low RMSE and high R^2 value are needed to optimise fully to the reference data, and the two metrics are not fully mutually inclusive.

By including both RMSE and R^2 values, for transition energies as well as dipole magnitudes, the full objective function becomes

$$(5.18) \quad f_{\text{full}}(\mathbf{x}) = \lambda_1 \text{RMSE}(\Delta E) + \lambda_2 \text{RMSE}(|\mu|) + \lambda_3 (1 - R^2(\Delta E)) + \lambda_4 (1 - R^2(|\mu|))$$

where λ_n are weights necessary to keep all of the terms to a similar range. This provides stability to the optimisation procedure, such that no one term dominates the solution space.

5.2.2 Minimisation Algorithms

Finding the optimal parameters for the chl-xTB method is a nonlinear problem. The parameters so far discussed can not be used to create a linear function that would reproduce the value of the objective function. Therefore it's necessary to use heuristics that can solve non-linear problems.

5.2.2.1 Nelder-Mead

The Nelder-Mead method, as implemented in SciPy, is a modified version of a simplex algorithm, that uses a n -dimensional shape to define a test region, and iteratively searches the n -dimension space by reflecting the vertices of the test region. The test region, or more specifically the shape described by its vertices, is the simplex. The simplex has $n + 1$ vertices - for example, a 2-dimensional problem would have a triangular simplex. The algorithm starts with an initial simplex guess. It is important that the initial guess covers enough area to avoid descending into any local minima, whilst not being too large as to not take into account finer details of the parameter space. The simplex is propagated by using a central value of the set of vertices, and using this to either expand, contract or shrink the simplex, or reflect on of the vertices. For example to find the parameters \mathbf{x} that correspond to a minimum of the function $f(\mathbf{x})$

$$(5.19) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

with initial simplex vertices $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$, the first step is to order the function values of the vertices

$$(5.20) \quad f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$$

and calculate the centroid of the set of vertices, excluding the worst vertex \mathbf{x}_{n+1} . The next steps then propagate the simplex, first by testing whether a reflection point \mathbf{x}_r is better than the worst vertex used to calculate the centroid

$$(5.21) \quad \mathbf{x}_r = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1})$$

where \mathbf{x}_0 is the centroid point. There are then a set of three possibilities for the value of $f(\mathbf{x}_r)$. First is that it the best value found so far, and so the simplex should be expanded along the centroid-reflected vertex axis

$$(5.22) \quad \mathbf{x}_e = \mathbf{x}_0 + \gamma(\mathbf{x}_r - \mathbf{x}_0).$$

- . The corresponding vertex of the two function values $f(\mathbf{x}_r), f(\mathbf{x}_e)$ then replaces the "worst" vertex \mathbf{x}_{n+1} .

A second possibility is that the function value for the reflected vertex is better than the worst vertex used to calculate the centroid, but worse than the best value, $f(\mathbf{x}_1) \leq f(\mathbf{x}_r) \leq f(\mathbf{x}_n)$. In this case the \mathbf{x}_{n+1} vertex is replaced by the reflected vertex.

The last possibility is that the reflected vertex has a greater function value than any vertex used to calculate the centroid. In this case a new point (contraction), or set of points (shrink) are used to propagate the simplex. Depending on whether this function value is greater or less than the worst vertex in the simplex (\mathbf{x}_{n+1}), the contracted point is either inside or outside of the simplex

$$(5.23) \quad f(\mathbf{x}) = \begin{cases} \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_r - \mathbf{x}_0) & \text{if } f(\mathbf{x}_r) < f(\mathbf{x}_{n+1}) \\ \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_{n+1} - \mathbf{x}_0) & \text{otherwise } f(\mathbf{x}_r) \geq f(\mathbf{x}_{n+1}) \end{cases}$$

if the contracted point \mathbf{x}_c is give a smaller function value than the reflected point for the first case, or the worst point for the second case, it then replaces the worst simplex vertex.

The final possibility is that both the contracted point function value is greater than either the reflected point or the worst point. In this case, the entire simplex is shrunk around axes to the best vertex

$$(5.24) \quad \mathbf{x}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$$

for $i \in \{1, \dots, n\}$.

Once either the worst vertex or all of the vertices are replaced, the new simplex is used as the start of a further iteration. Iterations are stopped once a termination criteria is met, such as a vertex function value being below a threshold.

Several versions of this method exist, that add additional constraints. This can include keeping the volume of the simplex constant, which can promote a steepest descent approach.

5.2.2.2 Sequential Least-Squares Quadratic Programming

The sequential least-squares quadratic programming (SLSQP) method is fundamentally different to the previous Nelder-Mead method, and follows a quasi-Newton procedure with additional factors to treat constraints.

The general problem is similar to Nelder-Mead, namely to solve

$$(5.25) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

however with an arbitrary amount of constraint functions c

$$(5.26) \quad c_i(\mathbf{x}) = 0$$

$$(5.27) \quad c_j(\mathbf{x}) \leq 0$$

where i, j are indices of the constraint functions. It is assumed that the space of f and c_n are one-to-one mappable on the space of x , and also is continuously differentiable. Starting from an initial value of \mathbf{x}_0 , a search direction d^k and step length α_k are used to propagate the set of parameters by

$$(5.28) \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

. The search direction, analogous to the ratio of function value to gradient in the Newton-Raphson method, is calculated by solving the Lagrange function

$$(5.29) \quad \mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_n^m \lambda_n g_n(\mathbf{x})$$

with a quadratic approximation, that reduces the problem to a quadratic programming subproblem

$$(5.30) \quad \min_d f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 \mathcal{L}(\mathbf{x}_k, \lambda_k) d$$

where the last term is often short-handed as the **B** matrix. This is the sequential quadratic programming method. A linear least squares subproblem could be used instead of quadratic programming, which would give the subproblem as

$$(5.31) \quad \min_d \|(\mathbf{D}_k)^{\frac{1}{2}} (\mathbf{L}_k)^T d + (\mathbf{D}_k)^{-\frac{1}{2}} (\mathbf{L}_k)^{-1} \nabla f(\mathbf{x}_k)\|$$

where the matrices **L**, **D** are from a diagonal decomposition of **B**

$$(5.32) \quad \mathbf{L}_k \mathbf{D}_k (\mathbf{L}_k)^T = \mathbf{B}_k$$

. With the solutions for \mathbf{d}_k solved by these subproblems, the parameter vector \mathbf{x} can be propagated until similar termination criteria as the Nelder-Mead method.

A visualisation finding a parameter set \mathbf{x} corresponding to a local minimum of the Himmelblau function (a standard benchmark for optimisation algorithms) is shown in figures 5.2.2.2 and 5.2.2.2. The former shows the Nelder-Mead simplex vertices, with the latter showing iterations in the \mathbf{x} space. It can be seen that the Nelder-Mead algorithm evaluates more points and covers a broader space, whereas the SLSQP algorithm follows a more direct gradient descent.

Both these methods were used to find optimal chl-xTB parameters, and it was found that the SLSQP method performed best, both in terms of the number of iterations, stability, and in the overall value of the objective function. This could be due to the addition of constraints, however it is hard to say as the wrapping of SciPy make both methods black-boxes that are hard to investigate further. The results of the optimisation is discussed in further detail in section 5.2.4.

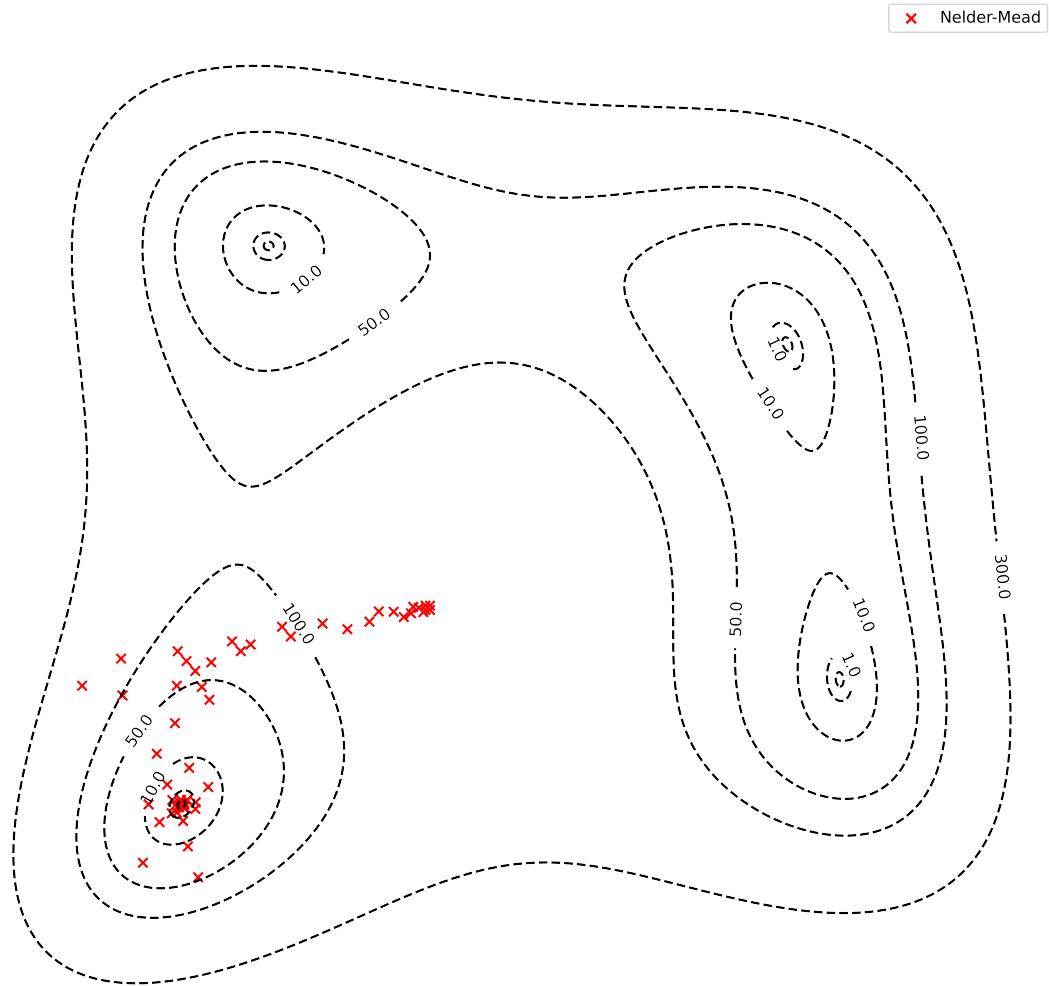


Figure 5.4: An example of the search through solution space using the Nelder-Mead method. The scatter points are positions of the Nelder-Mead simplex vertices, attempting to find the coordinates of the minimum in the HimmelBlau function.

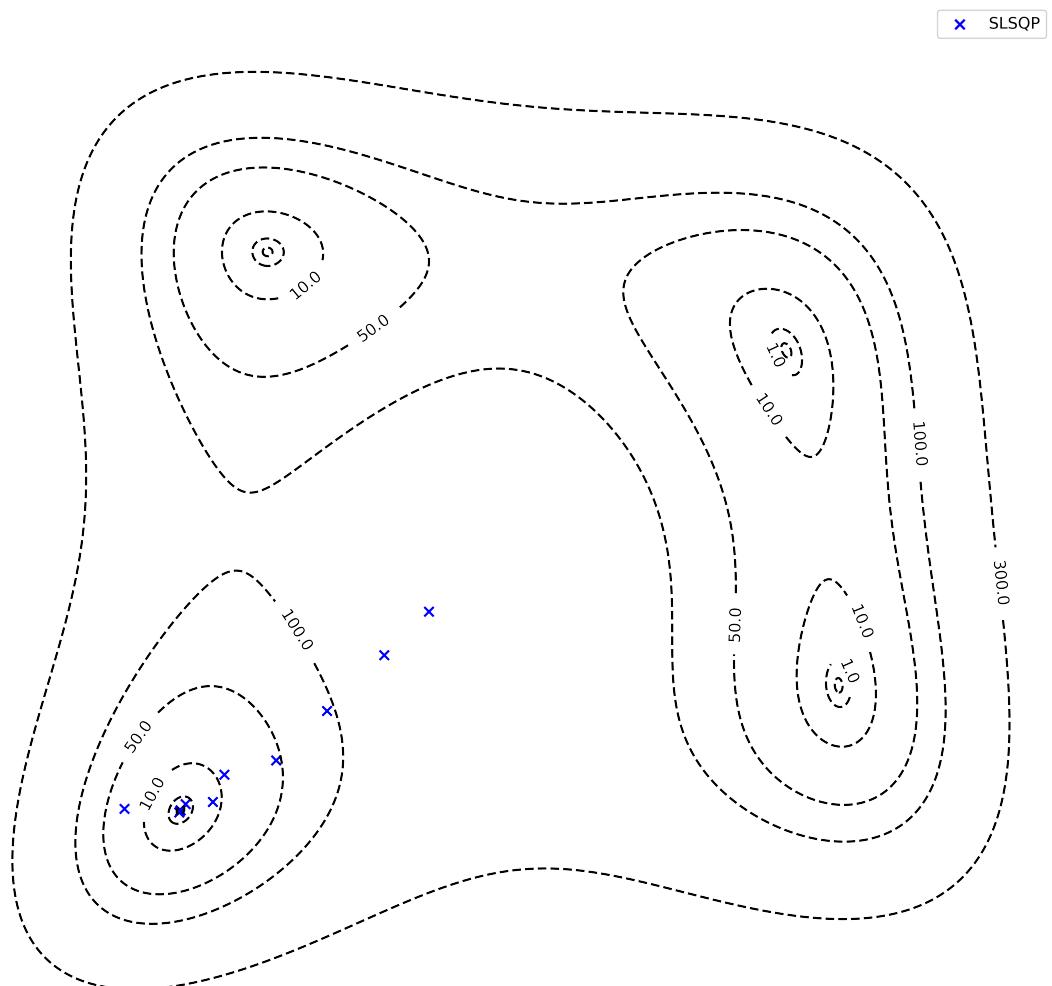


Figure 5.5: A similar example of the search through solution space, but using the SQSLP method. The scatter points are positions of the iterative solution vector \mathbf{x} , again attempting to find the coordinates of the minimum in the HimmelBlau function.

5.2.3 Reference Data

The geometries for the training set used to optimise the chl-xTB method were taken from molecular dynamics of the LH2 protein used in separate work[43]. The geometries of LH2 were chosen from uncorrelated snapshots, although each of the 27 chlorophylls from each snapshot were included in the training data to account for the differences in binding pockets. The stochastic collection of LH2 snapshots were chosen to cover a range of chlorophyll conformations to reduce the amount of artificial bias towards any particular correlation of geometries.

Training the chl-xTB parameters was done against PBE0 data. This data was chosen for the best accuracy-cost ratio, as well as having been previously used to investigate exciton properties for the LH2 system [43]. Additionally, from the outset it was unknown how much training data would be necessary, and so keeping potential future costs of expanding the training data down was another factor in choosing this functional.

	RMSE(ΔE)	$R^2(\Delta E)$	RMSE(μ)	$R^2(\mu)$
CAM-B3LYP	0.147	0.795	0.172	0.919
ω -B97XD	0.200	0.650	0.214	0.876
BLYP	0.053	0.871	0.384	0.129
Δ -SCF	0.225	0.847	1.455	0.566
$\Delta\epsilon$	0.218	0.875	1.511	0.500
ZINDO	0.596	0.339	1.526	0.396

Table 5.1: Summary of the errors and correlations of transition properties for Q_y for a set of LH2 Bchl_a geometries from a range of established respond methods. The errors were calculated against PBE0/Def2-SVP TD-DFT data, for transition energies (ΔE) and transition dipole moments ($|\mu|$).

Transition properties were calculated with a range of methods, covering levels of theory that would be useful to comparable to chl-xTB. These include an eigenvalue difference approach, Δ -SCF , and TD-DFT with different levels of theory. This was done so that the performance of any parameterisation results could be benchmarked with an idea of how accurate the method could be expected to be.

The errors and correlations of the reference data are shown in table 5.2.3. The methods included in the benchmarking were Δ -SCF , TD-DFT and eigenvalue difference all using the PBE0 functional and Def2-SVP basis set. Also included are CAM-B3LYP and BLYP functionals with Def2-SVP basis sets, as a higher and lower level reference respectively. The basis set was not changed as it has been found that the basis set has less importance on the accuracy than the functional [43].

There is a large variation in correlation and RMSE values for the reference methods to PBE0. This variation sets a reasonable expectation of how well a new method might perform. The BLYP functional performs best with an RMSE of 0.053 eV, and is most correlated (of the TD-DFT methods) with an R^2 value of 0.871. Second best is CAM-B3LYP, with an RMSE of 0.147 eV. ω -B97XD performs relatively well, with an RMSE value of 0.200 eV, but has the lowest R^2 value of 0.650. This illustrates how a low RMSE and high correlation are not mutually inclusive, and so must both be present in the objective function. The variance in these different TD-DFT methods, all of which have been used in studies on chlorophyll, show how it is difficult to assign a true value to transition energy for a set of geometries. Therefore as long as the accuracy of chl-xTB is within the range of methods shown here, it would also be valid.

The single transition methods also perform well, corroborating the earlier statement that only treating a HOMO-LUMO transition can give accurate results. The Δ -SCF and eigenvalue difference methods have slightly higher RMSE values, both around 0.22 eV. However the correlation is much higher at of 0.847 and 0.875 respectively. Hence, due to the dominance of HOMO-LUMO transition character, treating the transition as mixed is not necessary in order to achieve accuracy for transition energies. The story for transition dipoles, however, is different.

The agreement of transition dipole magnitudes is much lower than excitation energies. The

RMSE of Δ -SCF and eigenvalue difference methods is significantly higher (1.455 a.u. and 1.511 a.u. respectively) than the TD-DFT methods (0.172 a.u., 0.214 a.u. and 0.384 a.u. for CAM-B3LYP, ω -B97XD and BLYP respectively). The average magnitude of PBE0 transition dipoles is 2.751 a.u., with the average for Δ -SCF and eigenvalue difference being 4.287 a.u. and 4.342 a.u. respectively. This disparity is attributed to the lack of inclusion of Q_x transition character. As this direction of this dipole of this transition is orthogonal to Q_y it may reduce the transition dipole magnitude, similar to the effect seen in the outliers in the previous chapter.

Whilst there is a high degree of correlation between the higher level TD-DFT methods, with 0.919 and 0.876 for CAM-B3LYP and ω -B97XD respectively, the other methods have a much lower correlation to PBE0 transition dipoles. BLYP is the worst correlated, with a R^2 value of 0.129, which is fully uncorrelated. Δ -SCF and eigenvalue difference show a slight correlation at around 0.5.

Also included in the benchmarking was the semi-empirical method ZINDO. This had a poor accuracy across the board, with RMSE and R^2 values of 0.596 eV and 0.339 for transition energies, and 1.526 a.u. and 0.396 for transition dipoles. It was thought this might have good accuracy compared to TD-DFT and serve as a benchmark for how well a semi-empirical method might perform, but this turned out to not be the case.

Overall an RMSE to transition energies and dipoles of 0.15 eV and 0.2 a.u. is necessary to claim that transition properties can be calculated on a usable level of accuracy. Other cross-validations are necessary, and will be discussed in section 5.3, but for the optimisation this provides a reasonable accuracy benchmark. Whilst a high correlation of around 0.8 for both transition energies and transition dipoles is possible, it can be seen that the latter may not be possible for a single transition method.

5.2.3.1 Training and testing set

From the full set of PBE0 data, 100 random geometries were chosen for the training set and 507 geometries for the test set. The test set was used at the end of the optimisation procedure to validate how well the parameters perform on points outside of the training data. The sizes of each set was chosen to achieve a subset mean error (how far the mean of the subset is from the full set of data) in transition energies below 0.15 eV, whilst keeping a large number of geometries for the testing set.

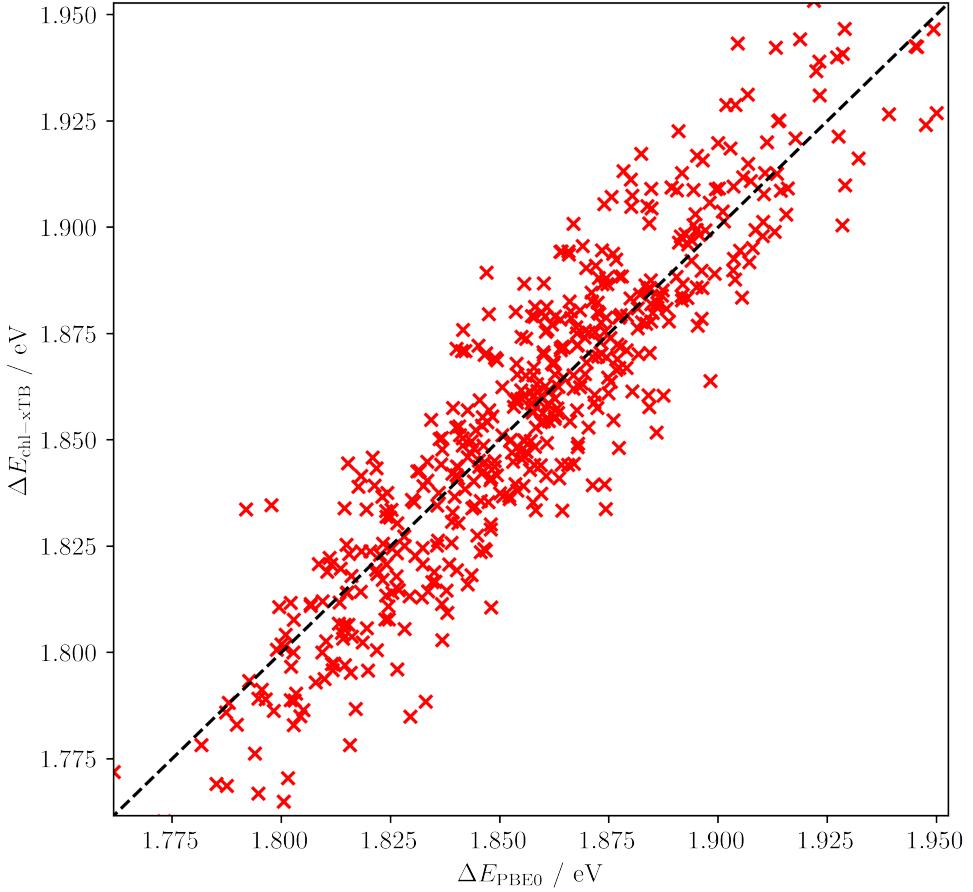


Figure 5.6: Comparison of the Q_y transition energies predicted from the reference methods as well as chl-xTB (red) against PBE0/Def2-SVP values.

5.2.4 Results

The chl-xTB method was parameterised to the PBE0/Def2-SVP training data, using the SLSQP method. Overall, chl-xTB performs extremely well considering the limitations discussed above. Transition energies and dipole magnitudes are predicted well within acceptable RMSE and correlation limits.

The final parameters for the chl-xTB method are given in table 5.2. The best performing set of parameters had an RMSE of excitation energy of 0.014 eV with an R^2 value of 0.88, and an RMSE of transition dipole magnitude of 0.057 a.u. with an R^2 value of 0.40. Repeated optimisation runs gave parameter and objective function minima to similar values, and the difference in these values can be attributed to the complex solution space. These values for RMSE are well within the values for TD-DFT with various functionals, and the R^2 of transition energy is equally good. While the correlation in transition dipole magnitude is low, it is near to the expected correlation from Δ -SCF and eigenvalue difference. It can also be seen in figure 5.2.4 that the variation in

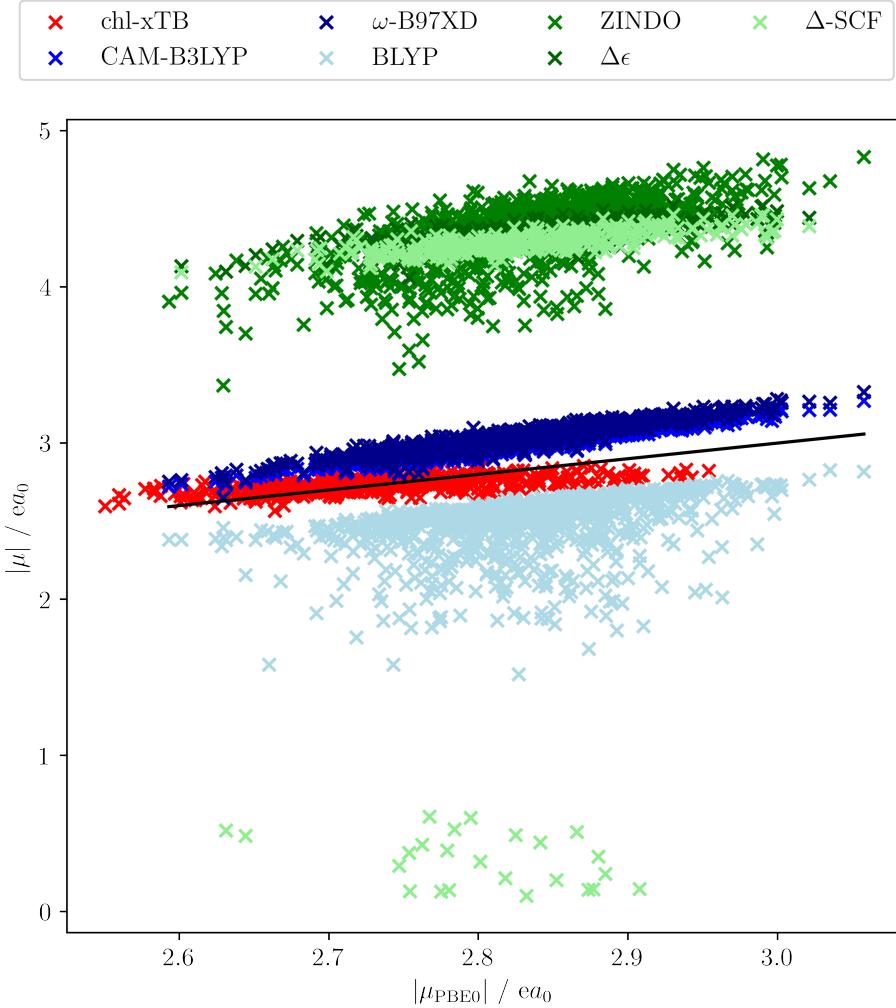


Figure 5.7: Comparison of the Q_y transition dipole moments predicted from the reference methods as well as chl-xTB (red) against PBE0/Def2-SVP values.

chl-xTB transition dipole magnitude is much smaller than Δ -SCF, eigenvalue difference and ZINDO, as well as being close to the mean from PBE0. This is a better behaviour, similar to the statistical method used before [43], than the other methods with low correlations.

It was also found that better minima of the objective function were found when using the SLSQP method for optimisation instead of the default Nelder-Mead method. Minima were found in a smaller number of iterations, reducing the overall CPU time required. This is in line with benchmarked SLSQP solutions in a non-linear multidimensional space. It was also investigated whether a reduction in the amount of parameters was possible, by only training the response parameters and not the Hamiltonian parameters, however this did not achieve the same levels of accuracy as using both sets of parameters.

The initial guess for parameters were the corresponding GFN1-xTB and sTDA-xTB parame-

Hamiltonian	chl-xTB	GFN1-xTB
k_s	1.462	1.850
k_p	2.694	2.250
Mg_p	0.902	-
Mg_s	1.053	-
N_p	1.044	-
N_s	1.281	-
$Mg_s\text{-}N_s$	1.468	-
$Mg_s\text{-}N_p$	1.023	-
$Mg_p\text{-}N_s$	1.067	-
$Mg_p\text{-}N_p$	1.402	-
Response		sTDA-xTB
y_K	2.147	2.000
y_J	4.012	4.000
a_x	0.067	0.500
D_{scale}	0.636	-

Table 5.2: chl-xTB parameters, optimized by the SLSQP procedure. Reference values for GFN1-xTB and sTDA-xTB are included, and served as initial guesses where available. Novel parameters all started from initial values of 1.0.

ters, or 1.0 for new parameters such as the Mg, N and transition density matrix scaling.

The optimised values do not differ much from the original GFN1-xTB and sTDA-xTB parameters (given for reference in table 5.2), with the exception of the a_x parameter. This parameter is far lower than the sTDA-xTB equivalent, which has a value of 0.500, but is in line with other methods that use similar MNOK approximations for coulomb-type integrals in response methods [97].

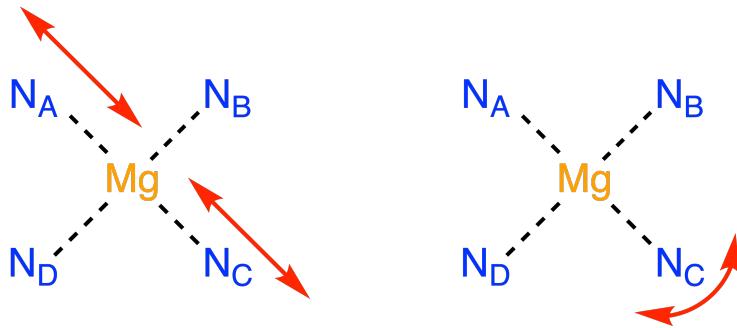


Figure 5.8: Normal modes of the nitrogen-magnesium centre of chlorophyll that have D_{4h} - D_{2h} (left) and D_{4h} - C_s (right) symmetry breaking components.

5.3 Cross-validation

5.3.1 Vibrational Mode Coupling

Whilst the stochastic selection of BChla geometries should represent a large section of the conformational space in LH2, it is not explicitly given that chl-xTB would perform equally well along important vibrational modes. Explicitly testing the values predicted by PBE0 and some of the reference methods as well as optimised chl-xTB would show how well the geometry dependence has been "learnt". These values would show how well chl-xTB predicts the coupling of vibrational modes and transition properties, as well as potentially reducing some error cancellation.

The geometries for this test were not taken from BChla for two reasons. There are 140 atoms in BChla, giving the number of normal modes is 414, and with 10 coordinates being calculated along each normal mode this represents a large number of geometries that would require reference data with expensive functionals and basis sets. Additionally, the normal modes would need to be calculated from an optimised geometry. The phytol tail in BChla (and chlorophyll in general) make geometry optimisations difficult due to the large degrees of freedom in rotations along the carbon chain. Without an accurately optimised geometry for the normal mode hessian, the predicted displacement vectors for the normal modes would be useless. Therefore the normal modes and transition properties were calculated for a truncated BChla with a hydrogen atom replacing the phytol tail, which made geometry optimisation possible, and also reduced the total number of vibrational modes.

Normal modes with the strongest coupling to the Q_y transition were chosen to most effectively scan the conformational space. These can be found by looking at modes which break the symmetry component of the Q_y transition. In an ideal model, the magnesium and nitrogen centre have D_{4h} symmetry with the Q_y transition lying along the N_A-N_C axis, and so vibrational modes with asymmetric components along this axis will couple to the transition. The movement of N_A-N_C atoms which induce D_{4h} - D_{2h} and D_{4h} - C_s symmetry breaking are shown in figure 5.3.1.

Normal modes that would also have this symmetry breaking component were indentified

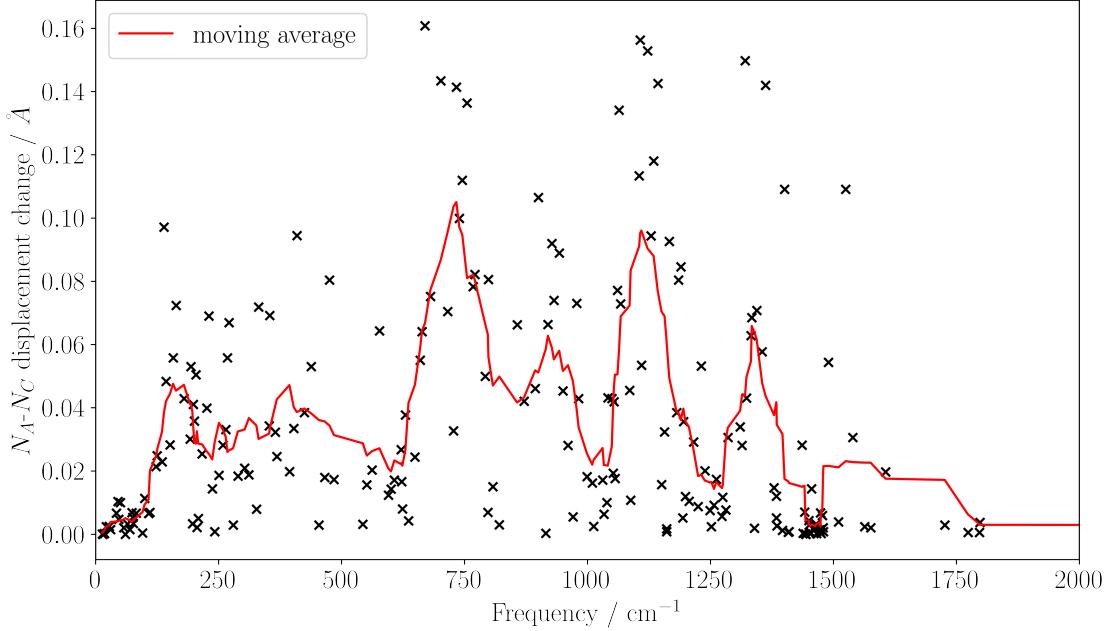


Figure 5.9: Change in the N_A - N_C displacements along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail.

by how much the N_A - N_C displacement would change along each normal mode. A plot of these values, as well as the moving average, is shown in figure ???. A similar scan was made of the N_A - Mg - N_C angle, however less variance in this value was found and the peak positions did not match previously reported frequencies for strong coupling. Normal modes with the largest N_A - N_C were then chosen for the transition properties scan. The normal modes that were chosen had frequencies at 669.6, 701.7, 733.3, 745.5, 755.1, 1105.0, 1107.0, 1122.2, 1142.9, 1320.3, 1361.8 cm^{-1} , which roughly correspond to previously identified normal modes with frequencies of around 728 and 1156 cm^{-1} [98].

The geometry was propagated for each selected normal mode such that the sum of all atomic displacements from the optimised geometry was in units of 1 \AA . This was done up to 3 \AA as it was found for most normal modes the energy difference between the optimised geometry and 3 \AA displaced geometry was greater than the thermal energy at 300 K (i.e. in an environment at 300 K, a chlorophyll system would not be expected to deform past these boundaries). At each increment, the Q_y transition energy and dipole was calculated using chl-xTB, as well as TD-DFT with PBE0/Def2-SVP and CAM-B3LYP/Def2-SVP levels of theory, Δ -SCF and eigenvalue differences (both using PBE0/Def2-SVP). A quadratic fit was made for each of the response methods and is shown in the plots. Some points for the Δ -SCF method are not shown due to issues with excited state convergence.

It can be seen that chl-xTB predicts PBE0 transition energies with a high degree of accuracy. Transition dipole magnitudes are predicted with worse accuracy, but still capture the general

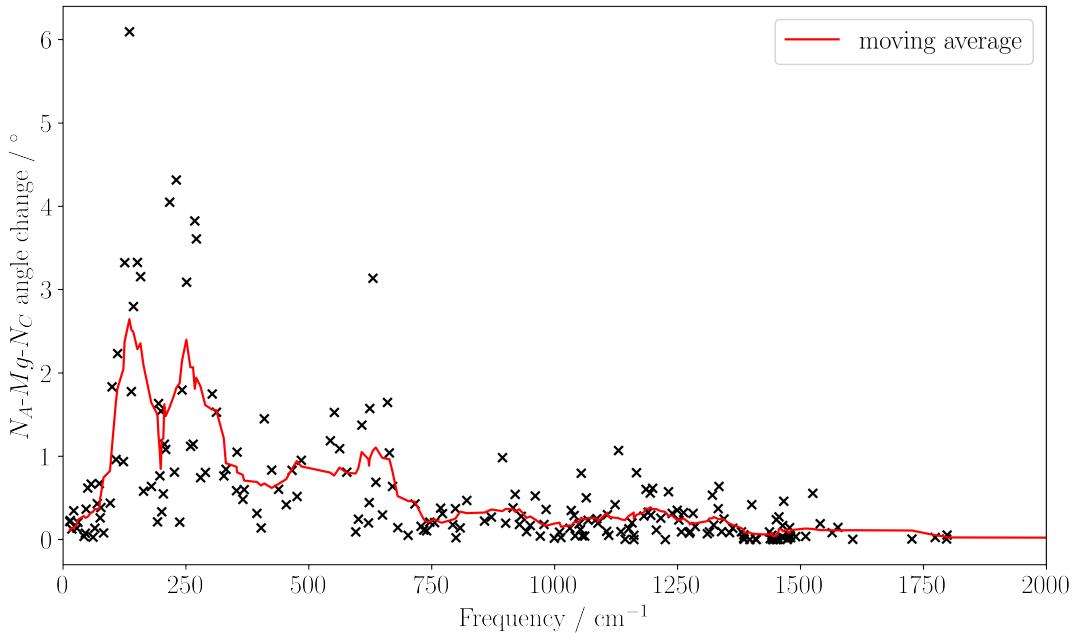


Figure 5.10: Change in the N_A -Mg- N_C angle along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail. The small smaller variance than in figure 5.3.1 led to this metric to not be used in normal mode choices.

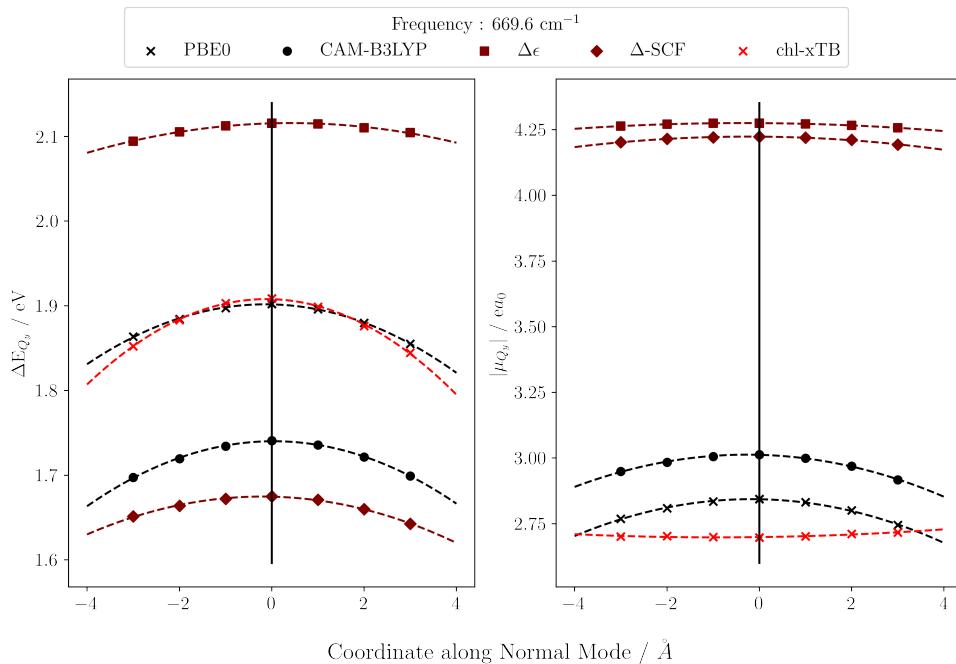


Figure 5.11: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

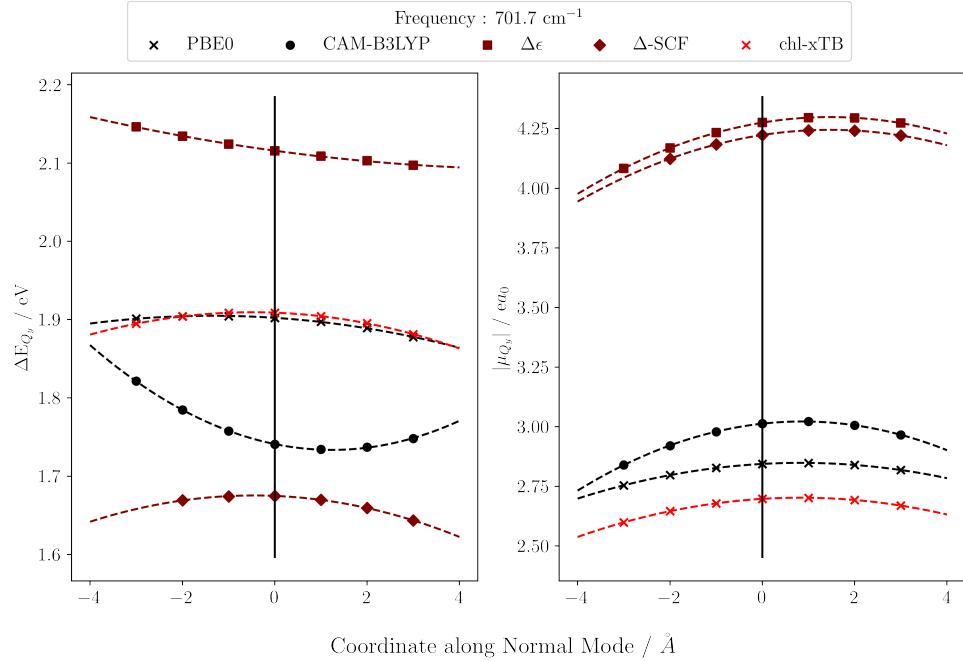


Figure 5.12: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

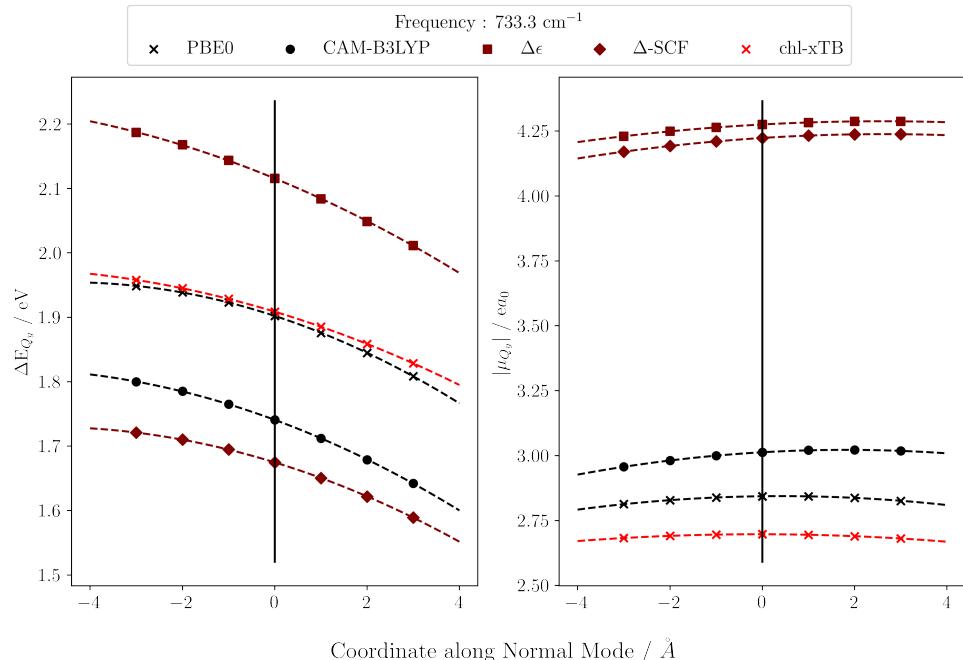


Figure 5.13: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

5.3. CROSS-VALIDATION

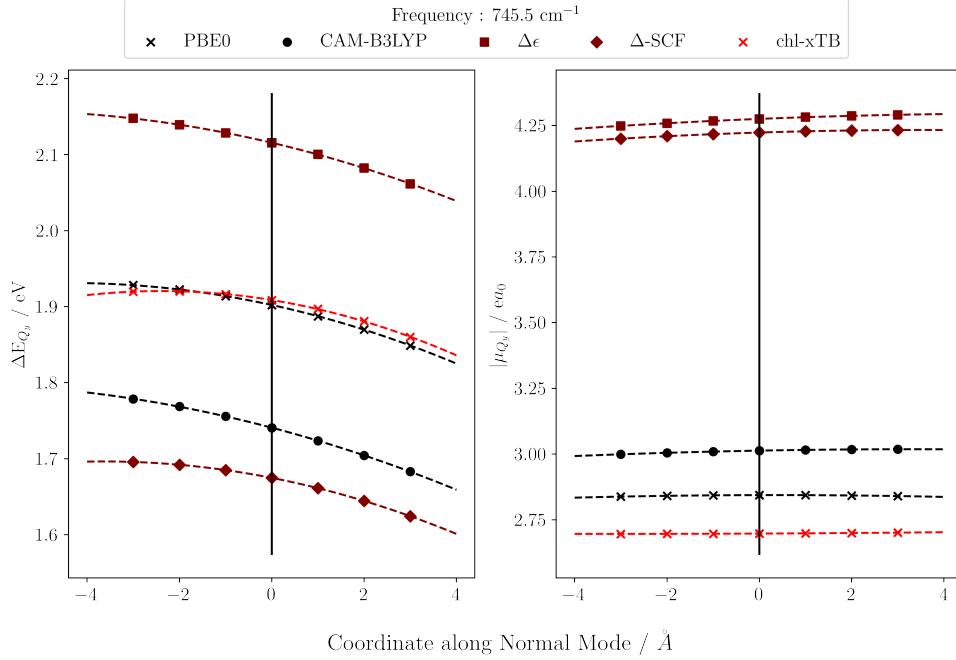


Figure 5.14: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

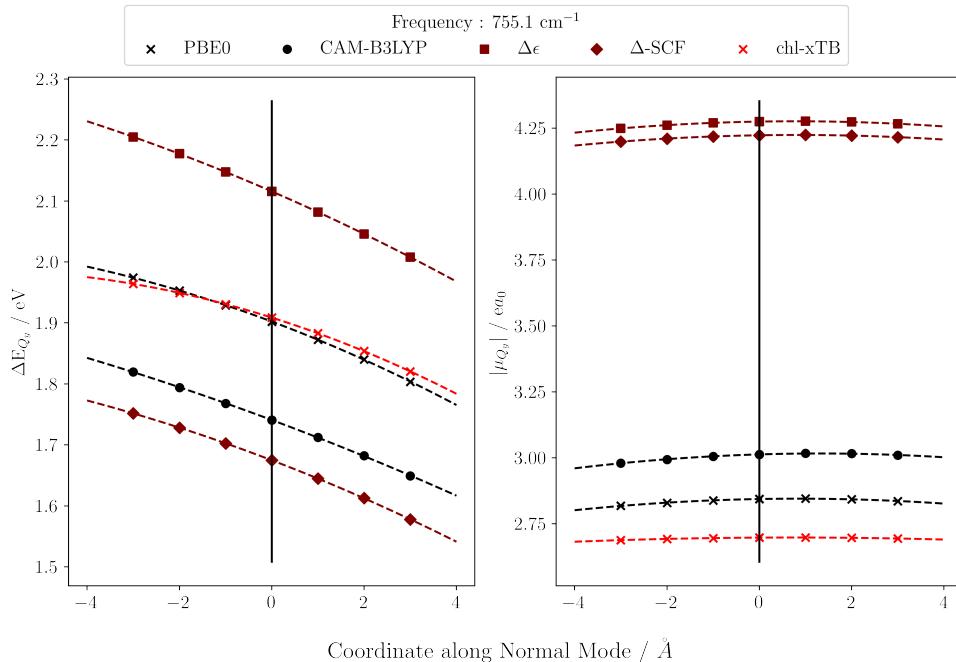


Figure 5.15: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

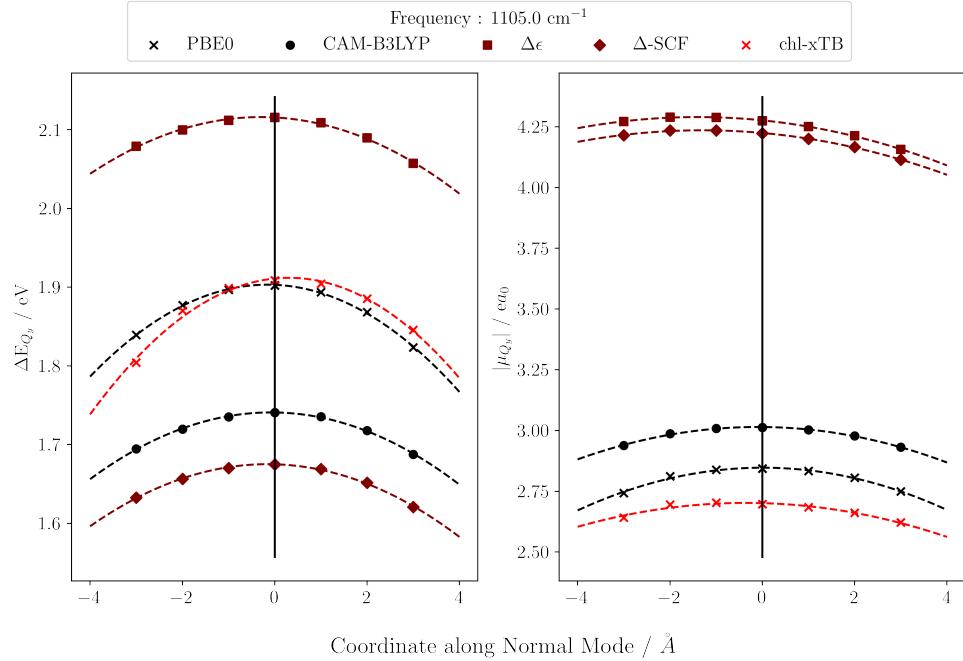


Figure 5.16: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

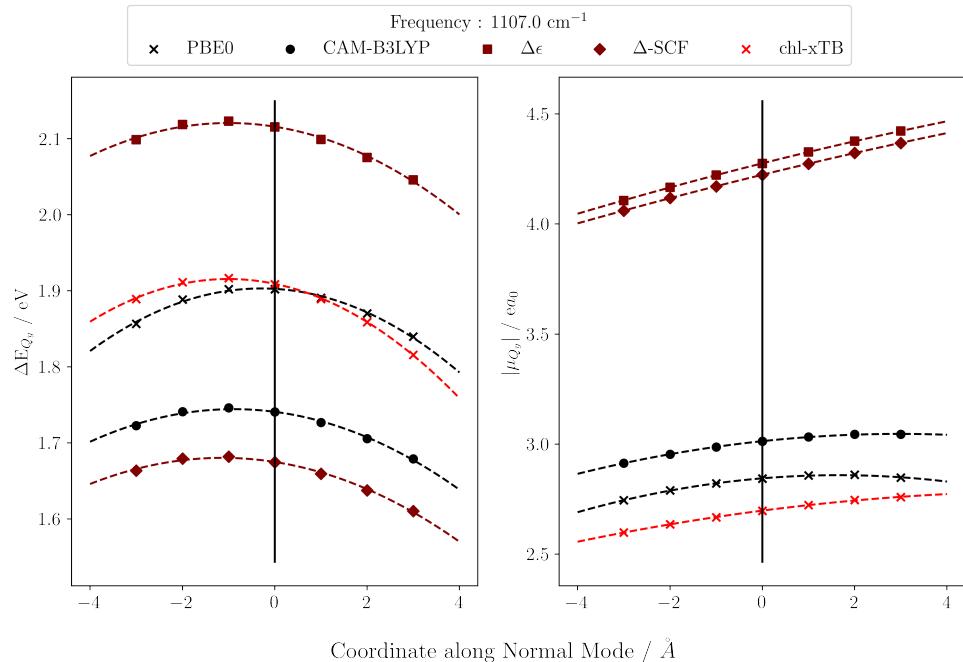


Figure 5.17: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

5.3. CROSS-VALIDATION

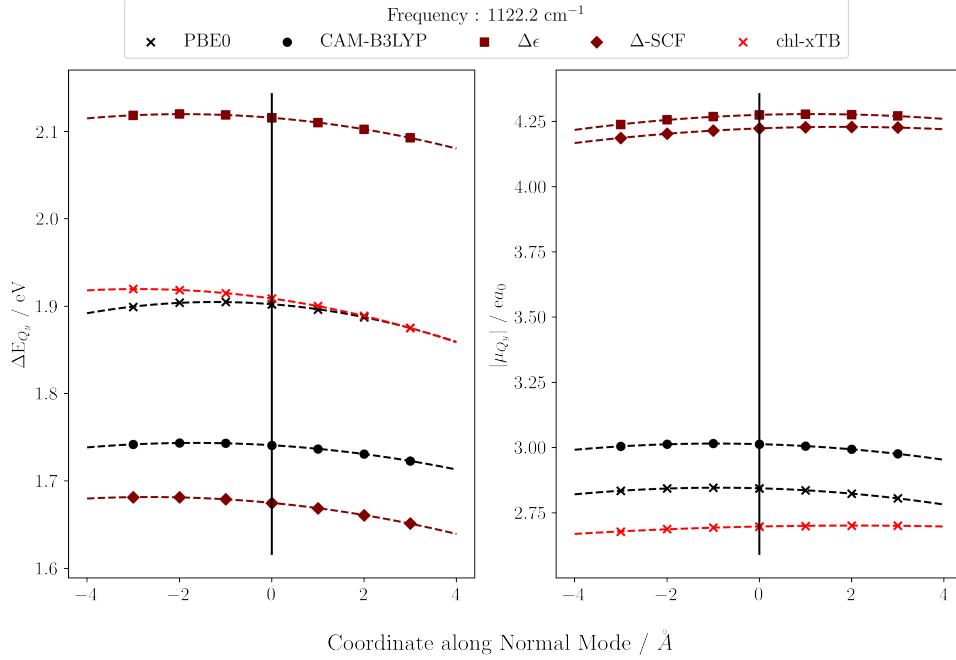


Figure 5.18: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

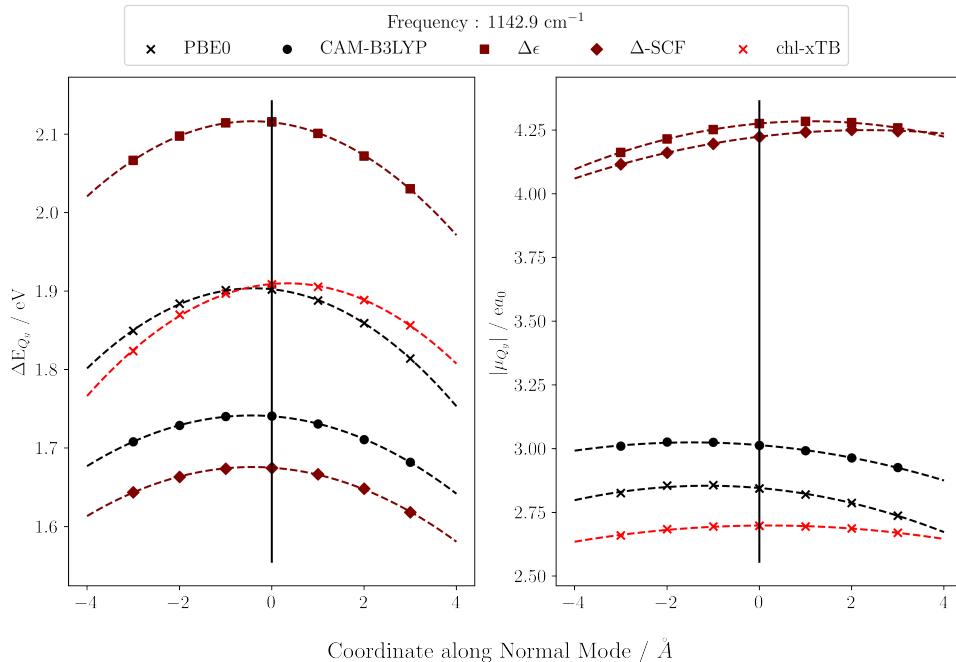


Figure 5.19: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with some of the reference response theories.

behaviour of PBE0 results well. chl-xTB consistently predicts PBE0 energies to within the same level of accuracy as achieved in the parameterisation test set. From the quadratic fits it can be seen that the gradients, turning points and curvature is well aligned between PBE0 and chl-xTB, especially compared to Δ -SCF and eigenvalue difference. It is also clear how important the transition density scaling factor is in achieving accuracy for the transition dipole magnitudes, with Δ -SCF and eigenvalue difference being well above the region where PBE0 and CAM-B3LYP values sit.

5.3.2 Absorption Spectra

So far all of the benchmark and tests on chl-xTB have been for gas phase systems and have no environmental effects. However in reality chlorophyll molecules would be embedded in many different environments, for example the LH2 protein. Although the training set took structures that have been perturbed by the LH2 protein, it is important to test the behaviour of properties predicted by chl-xTB when explicitly embedded. The obvious case for this would be predicting an absorption spectra for chlorophyll when embedded by an explicit solvent. Additionally, it should be tested whether the method could work for other chlorophyll systems other than Bchl_a. This would be important for future investigations into chlorophyll systems, but for the remaining work here it is not as important, and so is not investigated fully.

The absorption spectra was calculated using frames from an MD trajectory of chlorophyll A in an explicit diethyl ether solvent. An explicit solvent was used to account for inhomogeneous broadening in the spectrum. The MD was performed with the OpenMM toolkit. Forcefield parameters for the chlorophyll were taken from a bespoke parameterisation for photosystem II [99], with the rest of the system using the OpenForceField. The structure for chlorophyll was taken from the same source as the bespoke forcefield, and packed with explicit solvent using the tools in the `Mistral` package. Equilibration and production steps were done with a Langevin integrator set to 300 K and a timestep of 0.5 femtoseconds. The system energy was minimised before running a 10ps equilibration. Frames were then taken from a 2ns simulation time, with structures taken every picosecond.

Transition properties were calculated for chlorophyll structures from every frame. This was done with the chl-xTB method, as well as PBE0 and CAM-B3LYP TD-DFT, both using the Def2-SVP basis set. Experimental data for the absorption spectra was also taken from Katz *et. al* [100] [REF]. Embedding effects were included in the chl-xTB Fock matrix with a particle mesh Ewald method. The real space term was calculated using QCORE, whereas the more complicated spline reciprocal space term was calculated with the `HelpME` library [REF]. The absorption spectra for each method are shown as the absorption probability at excitation energies in nm, normalised such that the area under all spectra matched that of the experimental spectrum. The absorption spectra, both with and without a single-parameter shift of excitation energies, can be seen in figure 5.20.

chl-xTB performs equally well as TD-DFT methods at simulating absorption spectra, although constrained by the limits of the method and the training data. It can be seen that the chl-xTB lineshape is similar in position and width to the PBE0 method. This is highly encouraging, as the functional groups on chlorophyll A can have a large effect on the Q_y transition, and so the good agreement here provides evidence that important chemical features are captured in systems outside the training set. Although the chl-xTB lineshape is wider than the experimental spectrum, the CAM-B3LYP and PBE0 lines are also wider and so chl-xTB is still within a reasonable expectation of accuracy. All of the predicted spectra show lack of the lower intensity

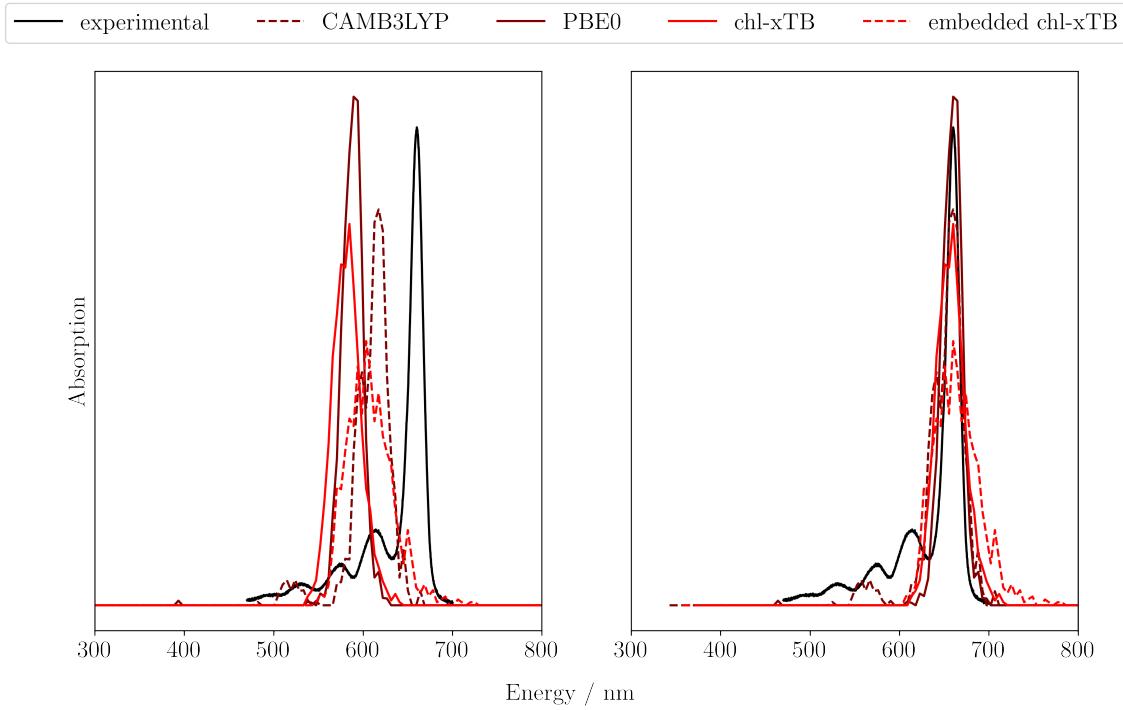


Figure 5.20: Predicted and experimental absorption spectrum of chlorophyll in diethyl ether, only in the Q band region. Predicted spectra are shown without an energy shift on the left, and with an energy shift to match the experimental absorption maximum on the right. All spectra are normalised to have equal areas.

peaks that are present in the experimental spectrum. Assignment of these peaks is shown in figure 5.20, and it can be seen that these are due to other Q transitions. The lack of Q_x peaks are expected, due to these transitions not being included in the predicted spectra, however the lack of the other Q_y peak can be seen in all methods bar CAM-B3LYP. There gives two explanations of why chl-xTB is missing this peak. First is that the PBE0 data does not include this feature, and so it would be unreasonable to expect it to be learnt from the training data. Second is that the training data include geometries from the LH2 complex, and so would not cover the conformations that give this lower intensity peak in a diethyl-ether system. Including Bchla geometries from different environments might fix this issue, but as this work only considers light harvesting systems this is outside of the current scope. The embedded chl-xTB spectra is both red-shifted and broadened slightly, which is also observed in simulated spectra from other methods.

5.4 Conclusions

It has been shown that novel approximations to the full linear-response eigenvalue equation can give accurate predictions of transition properties from high-level methods, which when combined with efficient electronic structure methods can give a very powerful tool for predicting transition

properties. However it is clear that limitations in the optimisation procedure create systemic issues with the final method.

For other studies, it may not have been necessary to include both the altered xTB framework and the novel response approximations. If the efficiency required was slightly less, a low level DFT calculation may have been good enough to achieve the required accuracy. This approach may have also been easier to parameterise, requiring less global parameters and more closely following the sTDA formalism. However as stated in the introduction, the efficiency of the desired methods does require semi-empirical efficiency, and so studying the response approximations outside of an xTB framework is not in the scope of this work.

Additionally the accuracy of the chl-xTB method against training data is most likely in large part due to the specificity. In general this is not an issue as many complex systems require highly specific treatment, especially for large scale systems such as chlorophyll. Applying the optimisation workflow to another system may be expected to work well on two conditions. First is how well a single transition character could approximate higher order transitions. This is easily sketched out by comparing other single transition methods, such as the Δ -SCF and eigenvalue difference methods, against any training data. Second is how well the electronic structure can be improved by altering the Hückel and scaling parameters. Altering parameters in other GFN-xTB terms might be required for more complex systems, which would require more training data to optimise.

Although any system outside of chlorophyll is outside the scope of this work, it would be beneficial to know how far the optimisation workflow can be pushed. Due to the relatively small size of the training data that may be required, a reoptimisation procedure could be fairly short. Additionally, depending on the system, the training data could be improved by either using a higher level method or by including a larger conformational range. The issue here would be knowing the limits of the workflow in greater detail, so it is known beforehand whether a candidate system and transition would be expected to work. However if these limits are found, many studies that require large numbers of calculations could be bootstrapped, requiring less data for a higher level of accuracy.

In terms of modelling light harvesting complexes, chl-xTB fulfills all the criteria required set out in the introduction. It would be efficient enough to calculate a huge volume of properties in a reasonable amount of time. Additionally the benefit of using the xTB framework is that the memory and CPU requirements are fairly low. This would allow for good scaling on high performance computers, using chl-xTB in a highly parallelised program (discussed in more detail in chapter 7). Additionally the accuracy of the method overcomes the statistical crutches that are used in other methods, making the models of light harvesting complexes far more detailed.

Overall, chl-xTB performs as well as can be expected from choices made in collecting the training data. Predictions of transition energies and transition dipoles are reliably close to the PBE0 values, and well within the error between different high level DFT functionals. It could

CHAPTER 5. CHLOROPHYLL SPECIFIC METHODS

be expected that improvements in the training data would yield better accuracy. This might be done by extending the training data in either the conformational space, or by investigating other systems. Applying chl-xTB to light harvest models forms the subject of the next chapters.

BEYOND MONOMER CHLOROPHYLL

Previous Published Work

Parts of the work presented in this chapter are also included in a paper published with Dr Susannah Bourne-Worster in tbd. These are reported in section 6.4.

This chapter investigates whether chl-xTB is still accurate for oligomer chlorophyll systems, beyond the monomers reported in the previous chapter. While monomer transition properties have been shown to be well predicted by chl-xTB, it is still necessary to show how this would perform in exciton frameworks that are used for LHCs.

The work here is in two parts. First is the work done on benchmarking a dimer exciton framework. Chlorophyll dimer systems were chosen as a test as opposed to the larger systems found in LHCs as their small size makes high-level methods tenable. Hence while this comparison may not be as relevant as a system closer to LHCs it would allow for more detailed comparisons.

The second part applies the dimer exciton system to predict the rate of transition to a charge separation state in chlorophyll dimers, and important mechanism in the quenching of excited states. This is done to illustrate how the novel response method designed in the previous chapter can solve the issues presented in the introduction. Aspects such as the efficiency, accuracy and extendability were all necessary in order to calculate these rates, and this would have been more difficult without the novel chl-xTB response method.

The methods discussed in the introduction are good generators of exciton Hamiltonian matrix elements. They are workable solutions to the problem of dealing with the large systems present in LHCs. However it could be argued that they are limited by some design choices, for example the machine-learning methods not being able to include embedding effects. It is explored here whether chl-xTB solves some of these issues or is similarly constrained by design choices (i.e. the density functional and conformer space used to construct the training data).

Applications to larger LHCs would be trivial after establishing that the dimer exciton system would work. This would just be a matter of using a system with more basis sites, for example the 27 chlorophyll sites in LH2. This exciton system is analysed in the next chapter as part of a investigation into LH2 properties that can be reliably calculated with the chl-xTB system.

6.1 Frenkel Exciton Hamiltonian

6.1.1 Exciton States

As stated in the introduction, a large majority of models of light harvesting systems utilise a Frenkel Hamiltonian model of exciton states, where the weak coupling between monomers mean that a model can be constructed from properties of the individual sites.

To recap the exciton theory from chapter 3, the exciton states $|\Psi\rangle$ can be constructed as a Hartree product of the states of individual sites (the monomers or chromophores in LHCs)

$$(6.1) \quad |\Psi\rangle = \Pi_m |\phi_m\rangle$$

where $|\phi_m\rangle$ is the monomer state on sites m . These monomer states form the basis function of the overall excitonic states. As the exciton is modelled to be localised to a specific site, an exciton state with an exciton at site i is given by

$$(6.2) \quad |\Psi\rangle^i = |\phi_i\rangle^* \Pi_{m \neq i} |\phi_m\rangle$$

where $|\phi_i\rangle^*$ is the excited state of monomer site i . States with more than one exciton are possible in this framework, however this is not usual for many light harvesting system models, as explained in the introduction. The Hamiltonian, including the "ground state" where there are no excitons in the system, is given by

$$(6.3) \quad H = \begin{bmatrix} E_0 & V_{0,(1,1)} & \cdots & V_{0,(N,1)} \\ V_{0,(1,1)} & E_{(1,1)} & \cdots & V_{(1,1)(N,1)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{0,(N,1)} & V_{(1,1)(N,1)} & \cdots & E_{(N,1)} \end{bmatrix}$$

where N is equal to the total number of individual sites. The diagonal terms E is the sum of the site energies and a point charge interaction

$$(6.4) \quad E_0 = \sum_m e_m + \sum_{m \neq n, A \in m, B \in n} \frac{q_m^A q_n^B}{r_{AB}}$$

where e_m is the energy of site m and q_m^A is the charge centered on atom A in site m . Hamiltonian elements corresponding to a single excitation are similarly given except with the excited state energy and charges

$$(6.5) \quad E_{(m,1)} = e_m + \delta e_m + \sum_{n \neq m} e_n + \sum_{n,A \in m, B \in n} \frac{q_m^{*A} q_n^B}{r_{AB}} + \sum_{n,p \neq m, A \in n, B \in p} \frac{q_n^A q_p^B}{r_{AB}}$$

where δe_m is the excitation energy of site m , and charges marked q^* are the excited state charges. The inter-chromophore term has been replaced to explicitly show that both interaction of excited state point charges with other sites, as well as ground-state/ground-state interactions, are included.

The off diagonal elements are the coupling elements between all exciton states. For coupling to the ground state, these are given by

$$(6.6) \quad V_{0,(m,1)} = \sum_{n,A \in m, B \in n} \frac{q_m^{\text{tr},A} q_n^B}{r_{AB}}$$

where charges marked q^{tr} are transition charges. It can be seen that this coupling element includes an electrostatic interaction between all sites and the excited site. This is different to coupling elements between two single exciton states, given as

$$(6.7) \quad V_{(m,1),(n,1)} = \sum_{A \in m, B \in n} \frac{q_m^{\text{tr},A} q_n^{\text{tr},B}}{r_{AB}}$$

which just involve sites m, n which have local excitations.

This Hamiltonian for exciton states is slightly different to those previously reported in the literature, as the ground state, where no excitons are present, is not usually included. Usually the local excitations are sufficiently high in energy and the couplings between ground and excited states are weak enough that the ground state eigensolution is usually unmixed with the local excitations. The excited state eigensolutions have very little ground state character, and so the block matrix of just excited state contributions would return the Hamiltonian used in many other studies.

6.2 Same-functional Benchmarking

It has been well established in many other studies that the Frenkel exciton Hamiltonian can predict dimer response properties well. This type of study was repeated here to set a benchmark for the later comparison with chl-xTB. Well known causes of error between the exciton model and full TD-DFT were explored, such as inter-chromophore distance and treatment of coupling of

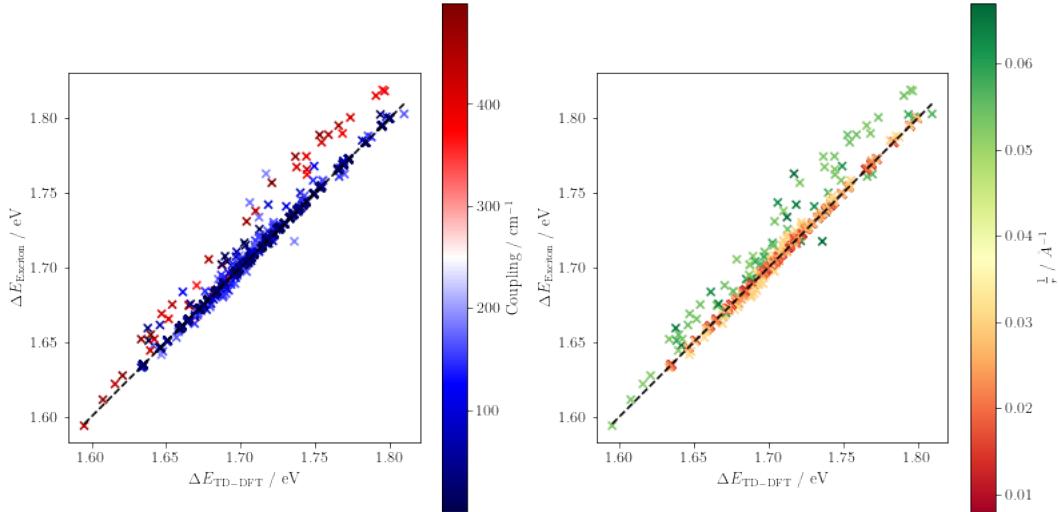


Figure 6.1: The correlation of lowest excitations in LH2 chlorophyll dimer systems predicted by the exciton model constructed from monomer TD-DFT and TD-DFT of the full dimer system. Scatter points are coloured by the coupling value between exciton states (left) and distance between magnesium centres (right).

excited states. This was done with two studies: first comparing the transition energies predicting by the exciton model against full TD-DFT for a set of chlorophylls from LH2; second with a conformational scan of chlorophyll dimers along rotational axes, to pinpoint if errors may be due to any specific interactions.

6.2.1 LH2 Dimers

Response properties were calculated for a series of chlorophyll dimers taken from the LH2 protein, using CAM-B3LYP/Def2-SVP TD-DFT. The chlorophyll dimer systems were taken from the previously used set of LH2 MD structures [43], with the phytol tails removed for ease of calculation. Transition energies were calculated for the full dimer systems, and compared with energies from the exciton model. The exciton model was constructed from monomer TD-DFT calculations on the two chlorophylls in each dimer. The transition energies for the exciton model were calculated as the difference between the excited and ground states. The scatter plot of these transition energies are shown in figure 6.2.1.

A clear relationship can be seen between the full dimer calculations and the exciton model. The systems in which there is any error can be seen to be correlated to the distance between the

monomers, with a lower distance giving a higher error. There is a smaller amount of correlation with the coupling value from the exciton Hamiltonian (taken as the coupling between the excited states, as generally the coupling value to the ground state is lower), implying that there are effects that are not included in the exciton model rather than a systematic error in the coupling. These effects could be from not including higher energy transitions, such as the Q_x transition, or charge transfer transitions. However for the large majority of cases the exciton model predicts full dimer TD-DFT with decent accuracy, within the range of previously reported exciton models.

6.2.2 Rotation

Scans of dimer conformations were done to further investigate breakdowns in the exciton model. As explored in the introduction, many previous investigations treated inter-chromophores coupling with a point dipole interaction:

$$(6.8) \quad v_{mn} = \frac{\vec{\mu}_m \cdot \vec{\mu}_n}{R_{mn}}$$

where $\vec{\mu}_m$ is the dipole on site m and R_{mn} is the inter-chromophore distance. From this expression it can be seen that the two inter-chromophore coordinates are the angle and distance between the monomers in the dimer pair. With an idea of how distance affects exciton model accuracy from the LH2 dimers above, the effect of angle on accuracy was investigated.

This was done by artificially constructing dimer systems with controlled angles between either the planes of the porphyrin rings or the N_A - N_C or N_B - N_D axis. All dimers were constructed from two truncated chlorophyll molecules with the porphyrin planes initially parallel and overlapping. The axis along the magnesium atoms was used to define the separation, and as this axis is approximately the cross product of the Q_y (N_A - N_C axis) and Q_x (N_B - N_D axis) transition dipoles, this axis is labelled Q_z . After separation, the angle between monomers was increased in increments of 3.6 degrees up to a full 360 degree rotation for all of the Q_y , Q_x , Q_z axes as axes of rotation. The magnesium atom was the centre of rotation. For the Q_z rotation the separation was 7 Å, a little less than the average separation of 9 Å found in LH2 - this was done to maximise the coupling in the exciton model to exaggerate any errors. For the Q_y and Q_x axes this separation was around 15 Å to be able to fit a full rotation. The initial system and the axes of rotation can be seen in figure 6.2.2.

In order to make assignment of exciton states easier, the geometries of the two truncated chlorophyll monomers were altered such that there would be a distinct gap in their transition energies. This gap was not wider than the variation in transition energies found in LH2 monomers. More detail on assigning transition energies to the correct excited state is discussed below.

Comparisons of the excited state energy can be seen in figures 6.2.2, 6.2.2, 6.2.2. The same trends in the profile of excited state energies can be seen both the full dimer system and exciton models. The minima and maxima are found at the same positions in both, and the qualitative

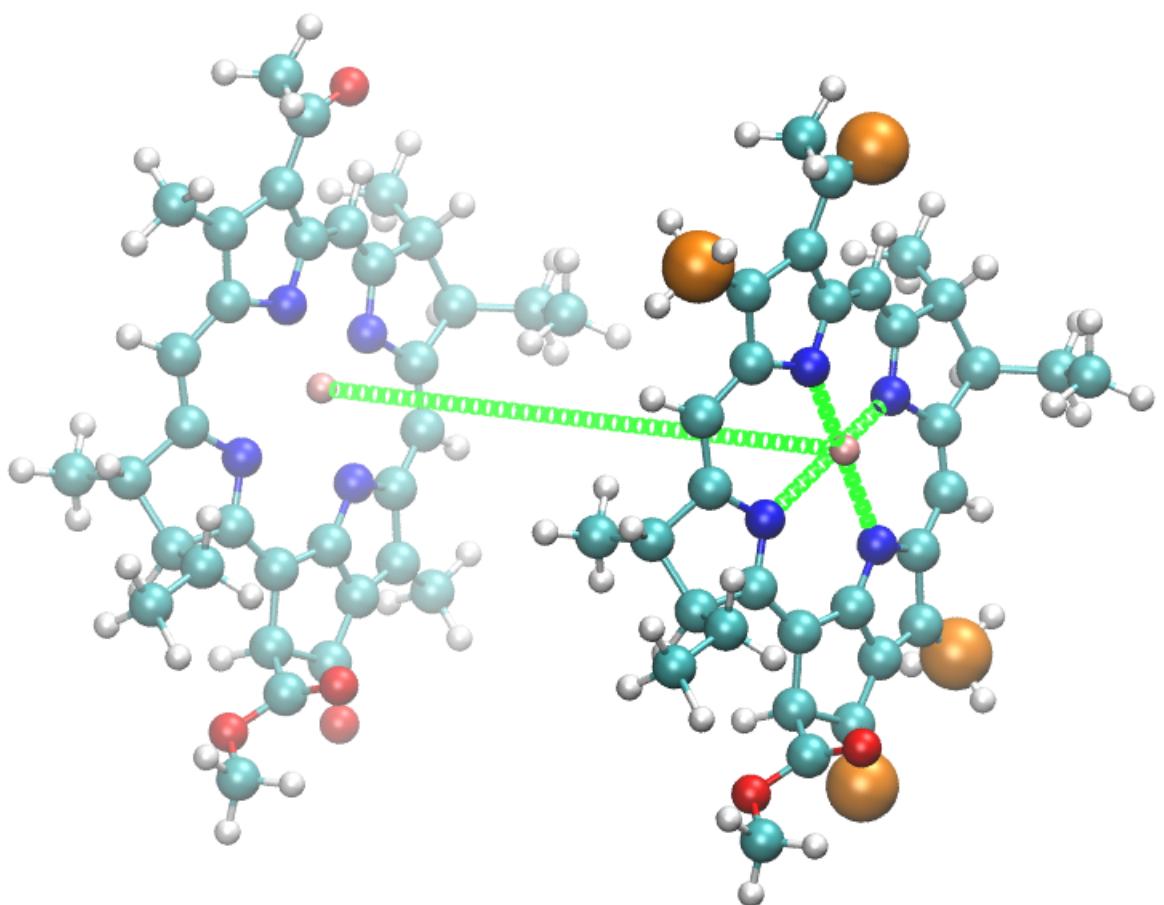


Figure 6.2: Diagram of the artificial dimer system, showing the axes of rotations in green and the chosen sites for calculating distance between functional groups in orange.

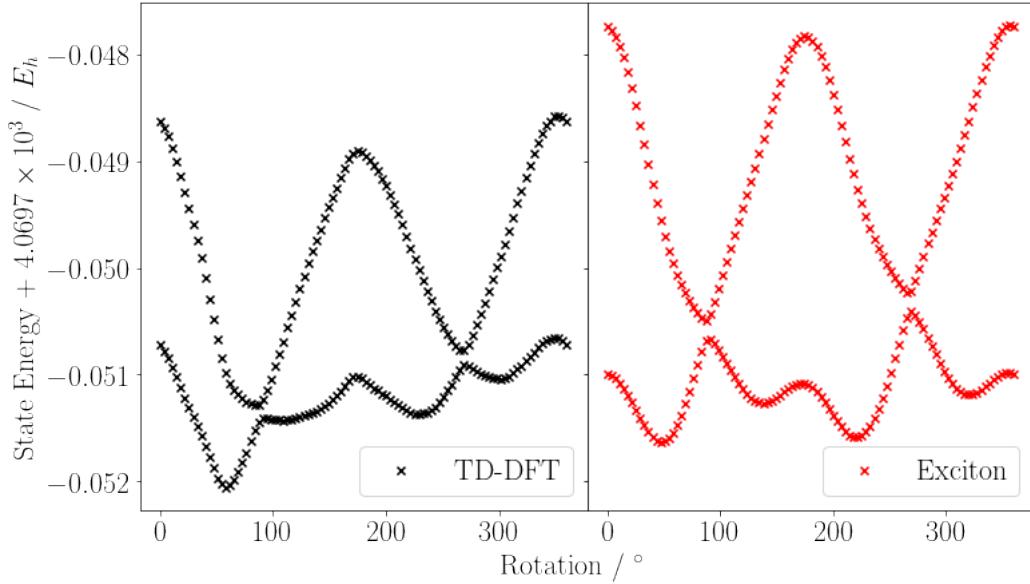


Figure 6.3: The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_z axis.

trends in curvature are also the similar. The only major discrepancy is seen in the 150-200 ° region in the Q_x profile, where the exciton model predicts the higher excited state at a little above the TD-DFT profile. A similar effect is seen in the LH2 dimers, where the errors in transition energies are almost all overestimates, and is attributed to a higher coupling value in the exciton model than is present in TD-DFT. This higher coupling could be due to the asymptotic behaviour of a bare point charge interaction. Using a short range damping operator might correct this, but this would be outside the scope of this work. It should be noted that the angle between nearest neighbour chlorophylls is either $\approx 0^\circ$ in the B850 rings, or $\approx 90^\circ$ for B800-B850 pairs, and so much of this space is not explored in the LH2 protein. For other light harvesting complexes there would be more variation in angle between chlorophyll monomers. Additionally, non-nearest neighbour pairs would also explore more of the angle space, however these dimers are further apart so would have much lower coupling values.

Overall it can be seen that the exciton model predicts TD-DFT excited states and transition energies well. Reasons for error are due to known issues with the exciton model. These could be addressed by using more detailed methods for calculating the coupling values between exciton states, or by including other transitions that may change the character of these states.

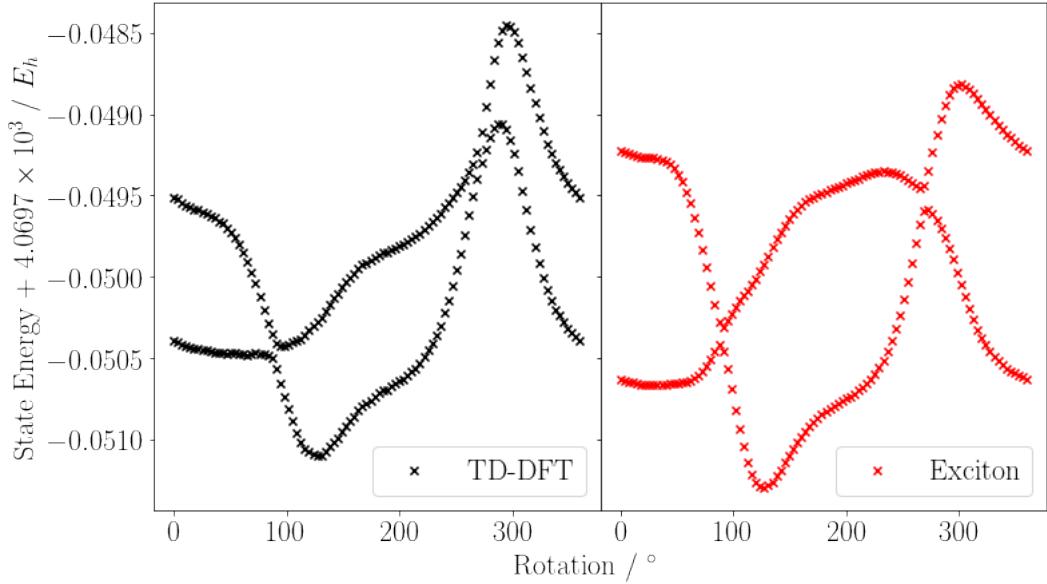


Figure 6.4: The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_x axis.

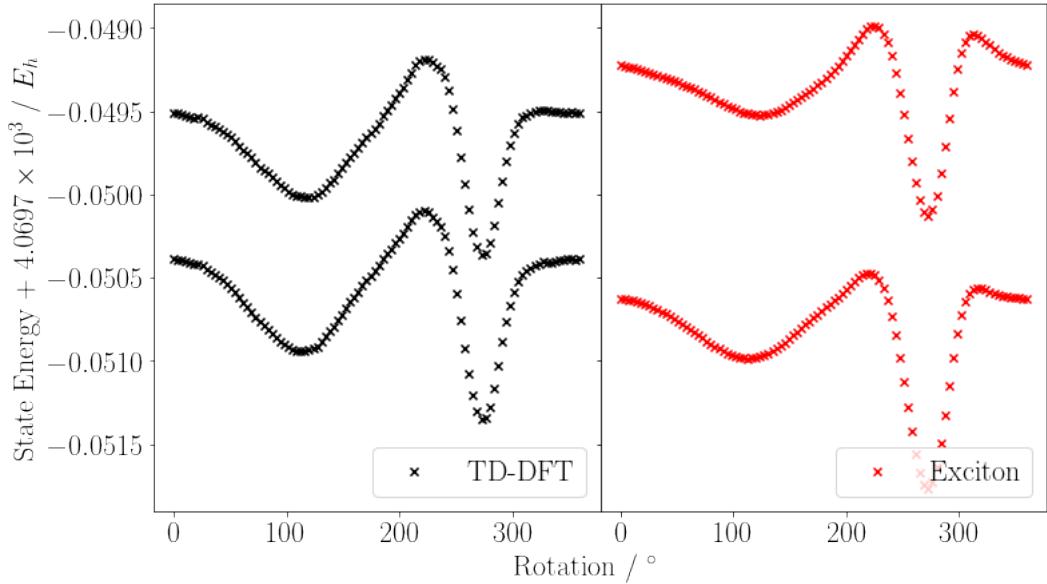


Figure 6.5: The dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the Q_y axis.

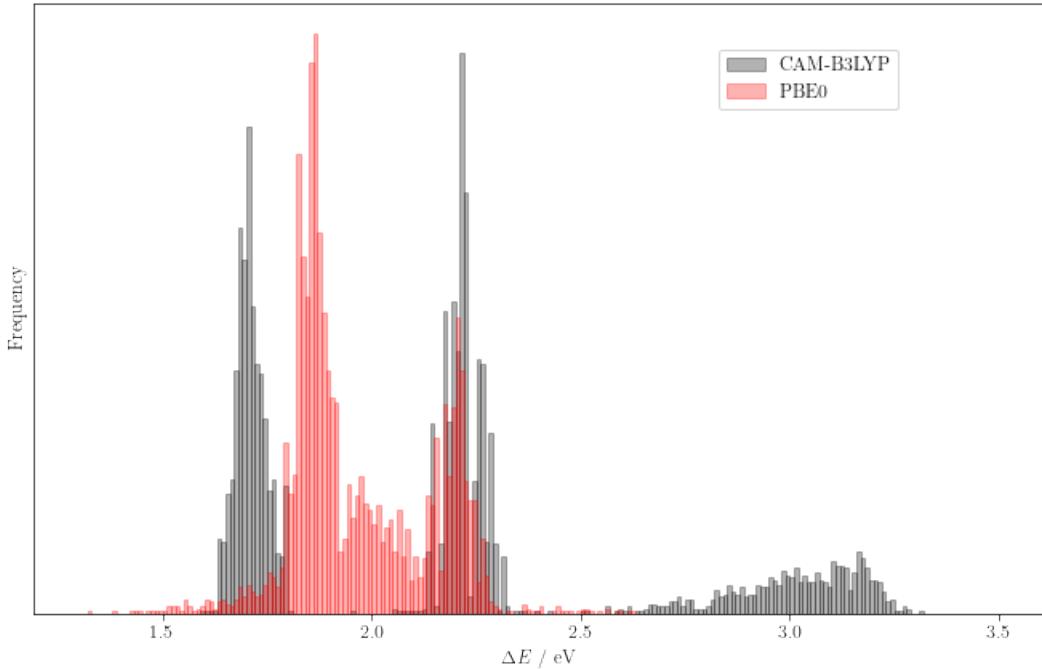


Figure 6.6: Distributions of the lowest 5 transition energies of chlorophyll dimers predicted by TD-DFT with the CAM-B3LYP functional (grey) and PBE0 functional (red).

6.3 chl-xTB Excitons

6.3.1 Full Dimer comparison

6.3.1.1 Discarding PBE0

While in the previous section it was possible to compare like-for-like with CAM-B3LYP dimer and monomer properties, it was found this was not the case for chl-xTB as well as the PBE0/Def2-SVP reference method. For chl-xTB, the obvious flaw is that the method was parameterised for monomer chlorophyll properties and not dimers. Additionally it was found that dimer response properties had systematic errors, which is attributed to the underlying xTB framework.

Ideally, the next-best comparison to replace dimer chl-xTB would have been dimer PBE0. However it was found that the transitions from PBE0/Def2-SVP were not well defined, having charge transfer character that is not present in the CAM-B3LYP dimer transitions. This effect can be seen in the distribution of transition energies, shown in figure 6.3.1.1.

Here the individual Q_y and higher energy transitions that are well resolved by CAM-B3LYP are not well resolved when using PBE0. This leads these transitions to be uncorrelated to CAM-

B3LYP transitions, as seen in figure 6.3.1.1. Hence the PBE0 dimer data would not make a good comparison for either the PBE0 exciton or the chl-xTB exciton models, and so was discarded from benchmarking. For the remaining sections, the exciton model is compared to CAM-B3LYP TD-DFT.

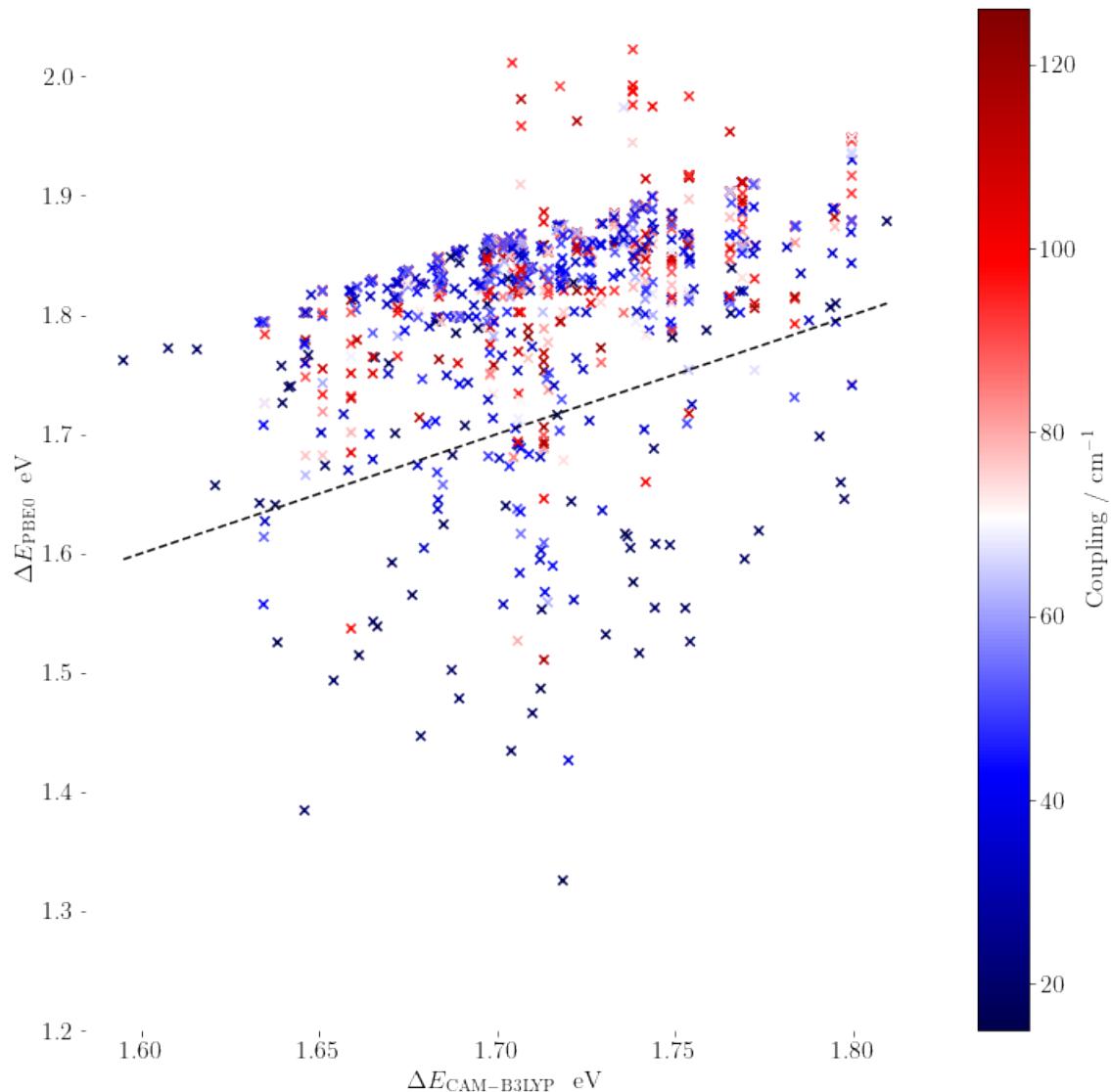


Figure 6.7: Scatter of the two lowest transitions from PBE0 TD-DFT against CAM-B3LYP TD-DFT, coloured by the coupling value from the exciton model using PBE0 monomer data.

6.3.1.2 Assignment of States

In the CAM-B3LYP comparisons, both TD-DFT and the exciton model gave excited states in the same energy ordering, making assignment straightforward. However it was found that for some dimer pairs, the exciton states constructed with PBE0 and chl-xTB would predict different energy ordering. For example, for an A-B dimer system, one model might predict the lower energy for a state with the exciton localised on monomer A, whereas CAM-B3LYP might predict the state with the exciton on B as the lower energy state. It should be noted that the CAM-B3LYP exciton model and CAM-B3LYP full dimer data were always consistent, and it was only when comparing to PBE0 or chl-xTB was this effect apparent.

It was therefore necessary to find the location of the excitons in both dimer and exciton models. For the TD-DFT dimer result, the exciton location was taken as the molecule where the transition charge distribution centered around. This centre \mathbf{c} was calculated by taking the average of the charge positions \mathbf{r} weighted by the absolute transition charge value $|q^{\text{tr}}|$

$$(6.9) \quad \mathbf{c} = \frac{\sum_i |q_i^{\text{tr}}| \mathbf{r}_i}{\sum_i |q_i^{\text{tr}}|}$$

where i is an index for all atoms in the dimer. Whilst the "centre of charge" is generally a poorly defined property for a system of point charges with zero total charge, these centres predicted with the above equation were found to be located close to the Mg centre of the truncated chlorophylls, and so is assumed to be a decent metric for the "transition centre".

For the Frenkel exciton result, the location was taken as the monomer corresponding to the dominant character in the eigenvector solution. For example, if the diagonal elements of the Hamiltonian were in order of ground state, transition on A, transition on B, and the eigenvector solution of the lower energy exciton state had the greatest value in its second element, the "transition on A" character, the exciton was assigned as localised on monomer A.

6.3.1.3 Comparison

Figure ?? shows the comparison of the exciton model, constructed with PBE0 and chl-xTB monomer data, against dimer TD-DFT data with CAM-B3LYP (i.e. the same domain as figure 6.2.1). Whilst the correlation is not as good as was found with CAM-B3LYP monomer data, there is still a clear relationship that supports the hypothesis that using the exciton model with lower level methods can still reproduce qualitative responses found from higher level data.

The systematic shift observed in both is equivalent to the shift found in monomer transitions in the previous chapter. The standard deviation of the errors was also found to be similar to the values reported in the last chapter. The two possible leading causes of error would be the exciton model or the theory used for monomer properties. From the CAM-B3LYP exciton comparison in figure 6.2.1, it was concluded that the correlation of close proximity of monomers, related to

high coupling values, to overestimation of transition energy implied that the exciton model was the leading cause of error. This implies that where there is large separation and low coupling between chromophores, the exciton model would have little effect and so any error here would be due to the monomer model used. This explains the behaviour seen in the PBE0 and chl-xTB exciton scatters - the error in the low coupling regions is due to the monomer theory rather than the exciton theory. When there is large coupling, the error in monomer properties is compounded with error in the exciton model. It can be seen that for high coupling values there is either a cancellation of errors, with some points being firmly in the middle of the pack, or addition of errors, giving the outliers.

The conclusion from these benchmarks is slightly different to the earlier same-functional benchmarking. Whereas in the earlier benchmark, improvements in the exciton model would be expected to give better transition energies, here improvements in the underlying monomer methods would give the greatest improvement. This is a similar conclusion to the choice of electronic theory against choice of response theory discussed in chapter 4. However as the chl-xTB method behaves similarly to PBE0, it would be hard to justify changes other than changing the functional for the training set data. The choice to do this would be almost arbitrary - a PBE0 exciton model has been shown to work well in past investigations, and so even though there is error to CAM-B3LYP data it is still good enough for investigations on light harvesting complexes.

6.3.1.4 chl-xTB Scans

Similar to the CAM-B3LYP scans above, the profile of excited dimer states was a useful test in detailing the accuracy of chl-xTB excitons. For the same reasons as the LH2 dimers, this could not be benchmarked against chl-xTB or PBE0 dimer properties, and so was compared to CAM-B3LYP again. Similar qualitative reproductions of the excited states were found, although interestingly with slightly better separations of the two excited states than before. Apart from differences in the magnitudes of the monomer charges, it is hard to attribute this improved accuracy to anything notable.

In this case where there is little random error from geometry variation, the exciton model rather than chl-xTB properties is most likely the leading cause of error. This does not contradict the LH2 dimer conclusion as these setups test two different things. For LH2 dimers, the exciton model had to predict correct variations in transition properties for a set of intra-chromophore geometries, similar to the study in the previous chapter. The range of inter-chromophore geometry is limited for the most, especially for nearest neighbours. In contrast, in this study the intra-chromophore geometry is constant, and the inter-chromophore geometry is changed. This removes the compound issue between chl-xTB and CAM-B3LYP monomer properties being different for intra-chromophore geometries, and just focuses on whether a non-varying chl-xTB transition can supply decent properties for a varying dimer system. One observation that showcased this difference was that the coupling values predicted by earlier version of chl-xTB were much larger

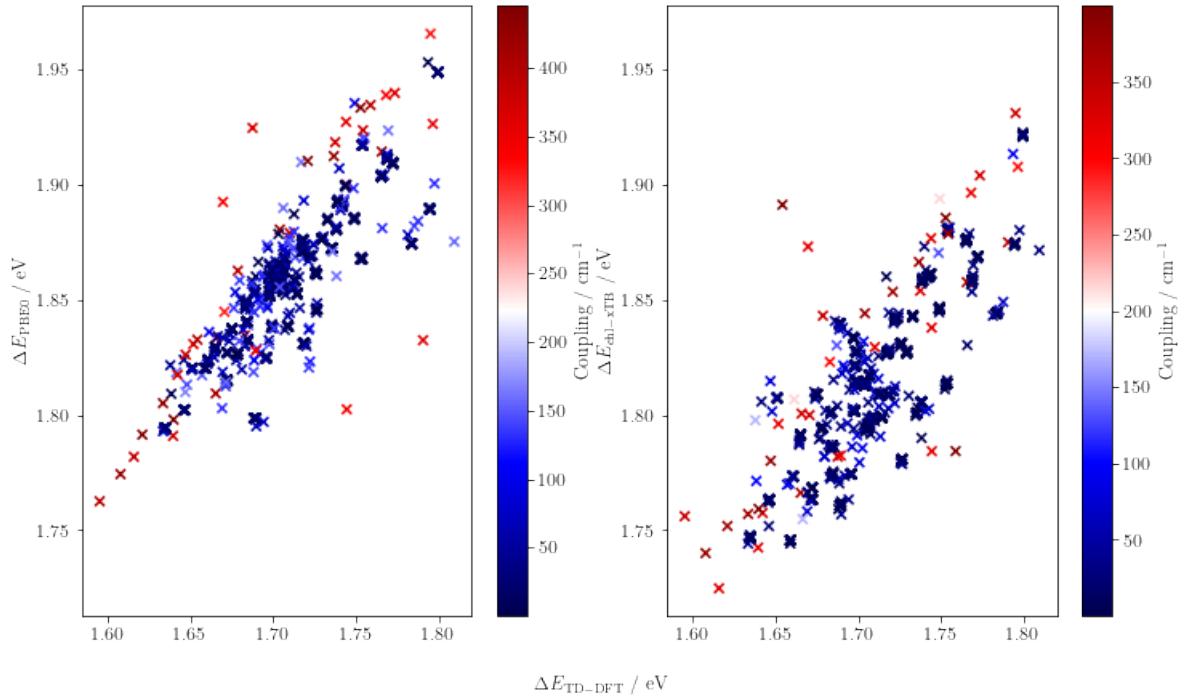


Figure 6.8: Transition energies of the exciton states calculated using PBE0 TD-DFT monomer data (left) and chl-xTB monomer data (right) against CAM-B3LYP full dimer TD-DFT transition energies, coloured by the coupling values.

than CAM-B3LYP. This was attributed to the much larger transition dipole magnitude that, observed in Δ -SCF and the eigenvalue difference methods of the previous chapter, and led to the inclusion of the transition density matrix scaling factor. Without removing the variation due to intra-chromophore geometry, this effect may not have been clear.

As the exciton framework has been shown to breakdown at small separations, it was postulated that there should be a good correlation between proximity of electron density and errors in predicting excited states. The places where the two chlorophylls would have a closest approach would be in the functional groups attached to the porphyrin ring. The centre of functional groups which have the most amount of transition density are shown in fig with orange spheres. These atoms were used to calculate the maximum reciprocal distance between the chlorophylls, which would correspond to the leading terms in the exciton coupling elements.

As the energy value of the excited states is very different between CAM-B3LYP and chl-xTB, a better comparison would be in the error of transition energies from the ground exciton state to the excited states, taken as the difference between the states. The profile of the sums of the absolute values of the error in these two transitions along the angle of rotation, alongside the reciprocal

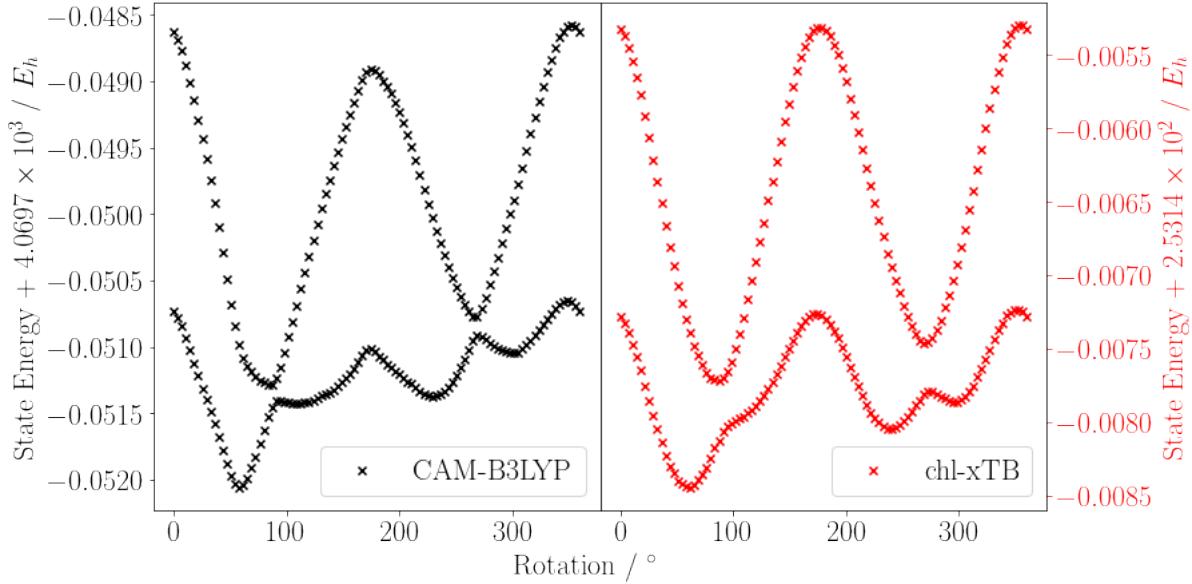


Figure 6.9: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_z axis.

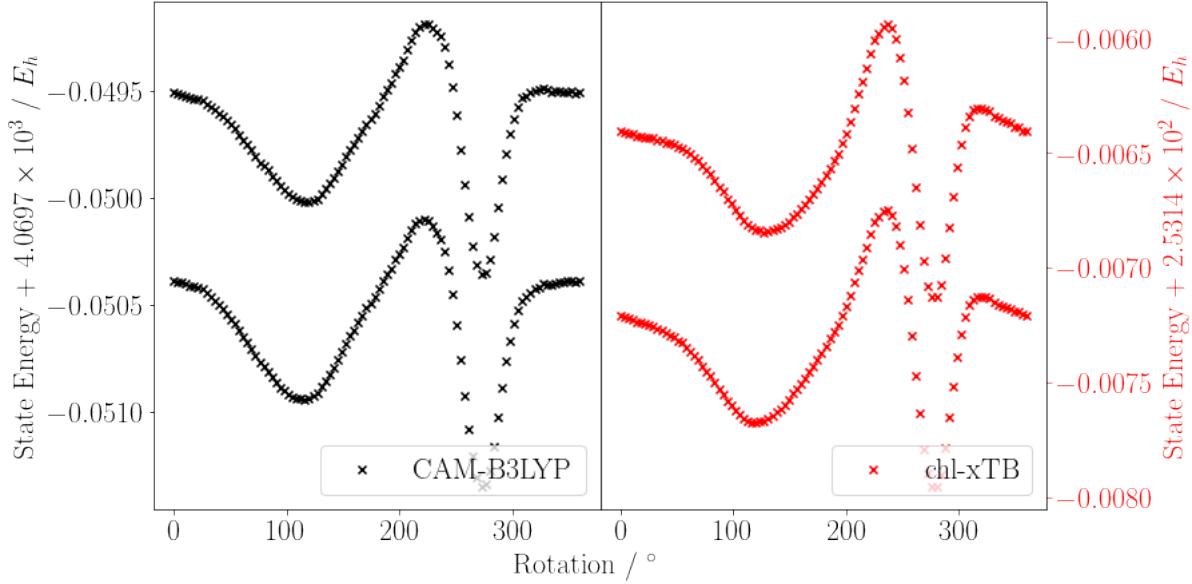


Figure 6.10: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_y axis.

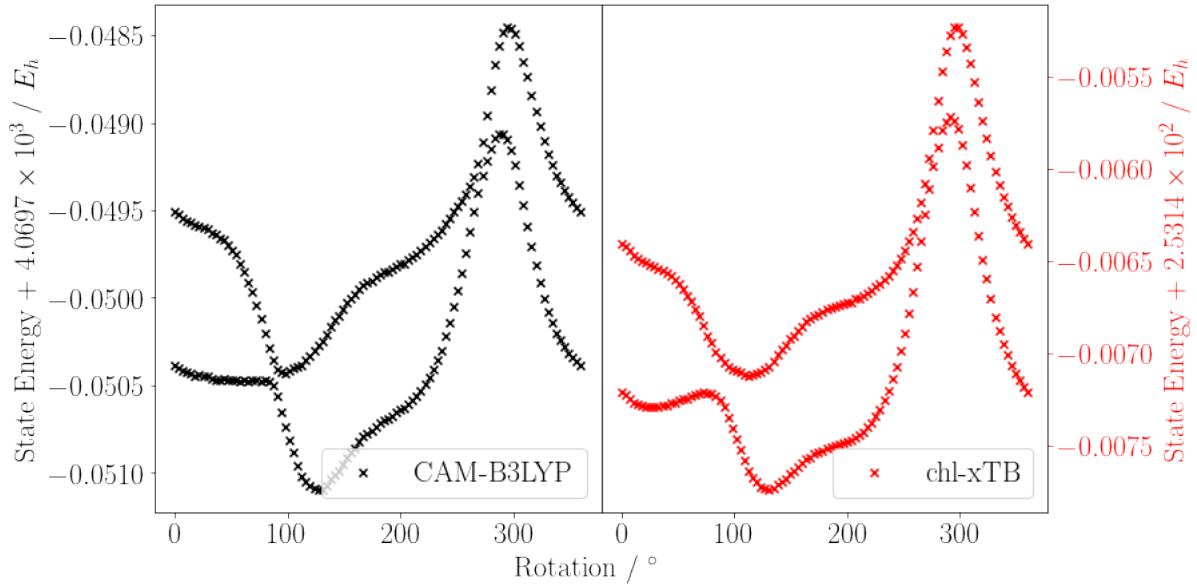


Figure 6.11: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the Q_x axis.

distance of the closest functional groups on the porphyrin ring is shown in figures 6.3.1.4, 6.3.1.4 and 6.3.1.4. It can be clearly seen that peaks in error in transition energy correspond with peaks in the proximity of the functional groups. It would be expected that a better approximation of electron interaction at close separations, such as the MNOK integrals used in the previous chapter, would improve the behaviour in these regions, but these issues are not in the scope of this work. As stated earlier, these separations are artificially small to exacerbate differences for clearer observation. Additionally the errors in transition energies are on the order of miliHartree, well within a reasonable range of other functionals.

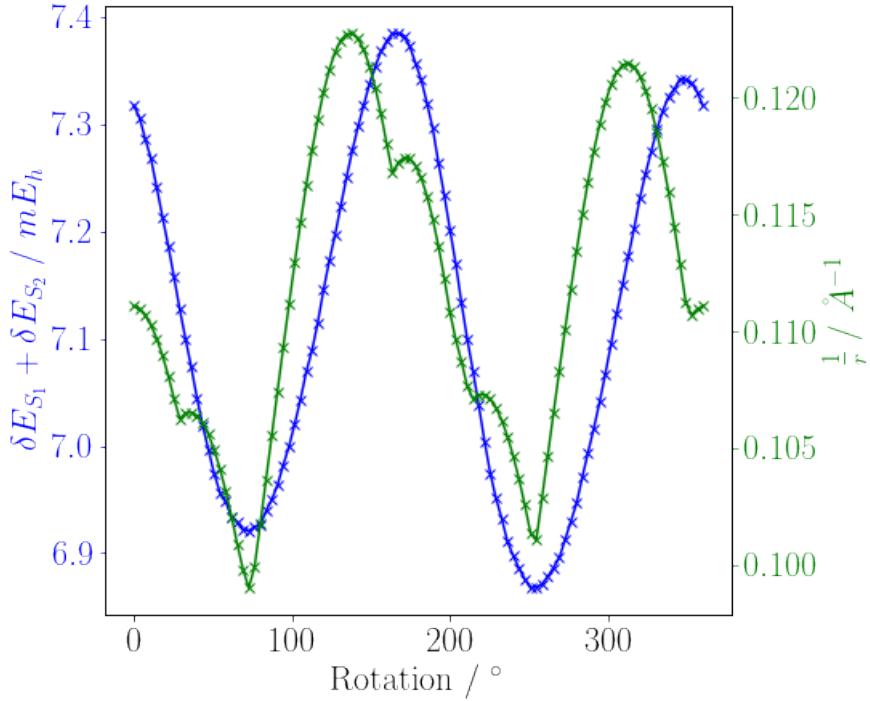


Figure 6.12: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_z axis.

6.4 Concentration Quenching

An ideal use case for the chl-xTB exciton dimer model was found to be reconstructing potential energy surfaces (PESs) of excited states, where a systematic scan of a reaction coordinate would not be possible. In these cases the PESs can be calculated from statistical methods, but these require a large volume of data to be accurate. This volume prohibits the use of full dimer calculations, as well as high level monomer TD-DFT that could be used to construct exciton models.

This is the case for calculating rate constants of conversion between the excited state (ES) and charge separation (CS) state of a chlorophyll dimer, which is important for understanding the quenching mechanism of excited chlorophyll dimer states. This rate constant is dependent on the free energy change ΔA and reorganisation energy λ of transitioning from the ES to CS state, and is given by

$$(6.10) \quad k_{\text{CS}} = \frac{2\pi}{\hbar} |H_{AB}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} \exp\left(\frac{-(\lambda + \Delta A)^2}{4\lambda k_B T}\right)$$

where H_{AB} is the coupling value between the ES and CS states, k_B is the usual Boltzmann factor and T is the temperature. ΔA and λ are calculated from the minima of the ES and CS PESs.

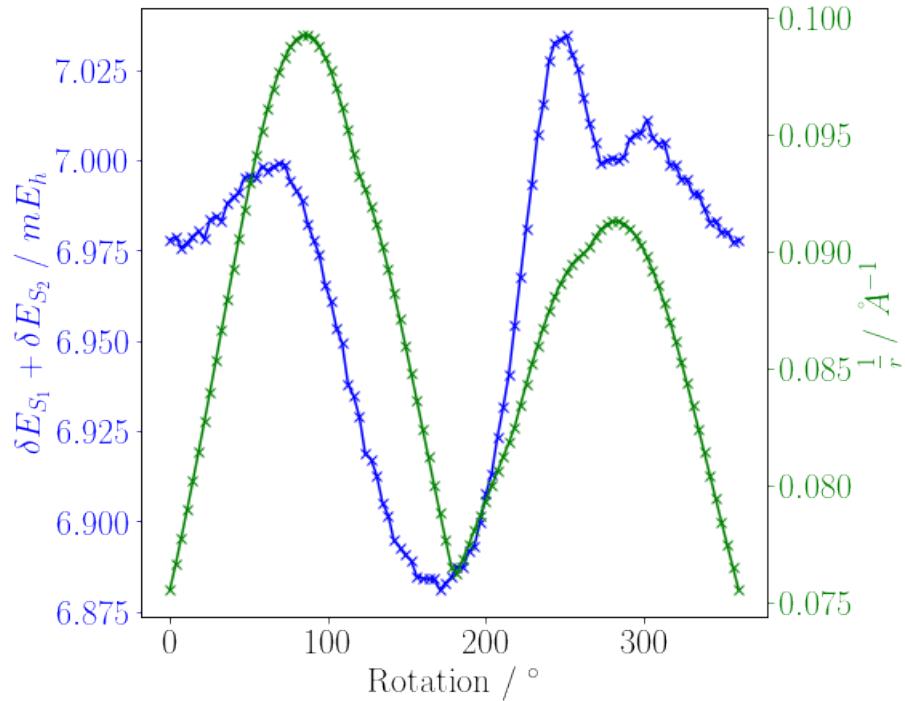


Figure 6.13: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_y axis.

These surfaces however are dependent the both the internal geometry of the dimers and the solvent system around it, and it is not possible to reduce this space to simple reaction coordinates to perform systematic scans. Instead, the PES can be reconstructed using the distribution of energy values that it defines, calculated from a large volume of chlorophyll dimer geometries.

This was done for a series of chlorophyll dimers, using the difference between the ES and CS energies as the reaction coordinate (and the space over which the ΔE are distributed), as this would include all of the internal chlorophyll coordinates and solvent coordinates.

The PESs was taken as a quadratic function of the excitation energy ΔE

$$(6.11) \quad V(\Delta E) = \frac{kT}{2} \left(\frac{\Delta E - \mu}{\sigma} \right)^2$$

where μ and σ are the mean and standard deviation of the excitation energies, the Boltzmann distribution from this PES is a normal distribution

$$(6.12) \quad P(\Delta E) = \exp \left(-\frac{1}{2} \left(\frac{\Delta E - \mu}{\sigma} \right)^2 \right)$$

and so values for μ and σ also define the PES.

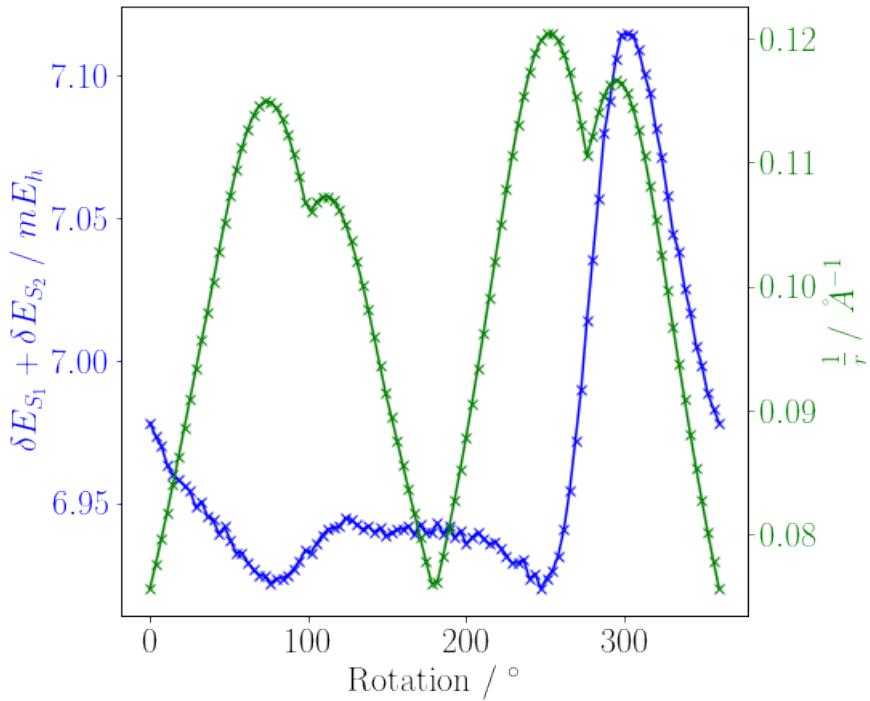


Figure 6.14: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the Q_x axis.

The rate $k_{ES \rightarrow CS}$ was calculated for a series of chlorophyll dimers at a constant separation, solvated in diethyl ether. The rate was also calculated for a pair of chlorophylls embedded in the LH2 protein. MD simulations were used to generate the thermally distributed geometries of the dimer systems. The diethyl ether systems were prepared using Packmol, solvating chlorophyll dimers into 100 Å diethyl ether boxes, with forcefield parameters taken from OpenForceField version 1.3.0. For the solvent systems, an additional constraint was used to keep the dimer at a constant separation in order to properly inspect the effect of chlorophyll distance on the charge separation rate. These constraints were set to 8, 10, 12, 14 and 20 Å. Each system was equilibrated for 10 ps at 300 K with a Langevin Integrator, with a production simulation of 500 ps performed afterwards. Geometries were taken every 1 ps for 500 total geometries. The LH2 geometries were calculated using the same MD method but using the LH2 structure and forcefield developed by Ramos *et al.*.

Dimer excited states were calculated for all of the geometries, using the Frenkel exciton Hamiltonian and chl-xTB. The lowest energy exciton states were taken for the distribution to reconstruct the PES. To account for environmental effects, the solvent and other chlorophyll(s) were included as point charges (with charge values taken from the forcefields used) to polarize the chl-xTB calculations. Fitting normal distribution values μ and σ was done with the

`scipy.stats` module. Good fits were achieved with this method, justifying the assumption that the PESs follow a quadratic function.

The coupling value $H_{AB} = \langle ES | \hat{H} | CS \rangle$ was calculated for a set of dimers using Fragment Orbital DFT with a PBE/3-21++G level of theory. FODFT will become less accurate as the distances between fragments get larger due to basis set truncation errors, so instead of calculating separate values for all separations the 10 Å case was used to better approximate further separations using an exponential decay. Additionally, truncated chlorophylls with the phytol tail removed were used to make the FODFT calculations feasible.

The rate constant calculated for the 10 Å angstrom dimer case was 0.0004 ns^{-1} . The fluorescence rate k_f is given as 0.2 ns^{-1} , implying that charge separation is not a competitive pathway for quenching the excited state. However this value assumes a full relaxation of the geometries to a new equilibrium. Taking values for ΔA and λ from the reaction coordinate at initial photoexcitation (i.e. where the ground state dimer distribution would sit) gives a more competitive rate constant of 0.01 ns^{-1} . Other considerations also suggest the rate constant calculated with this method will be an underestimate, making it even more competitive.

The rate constant also decreases with the distance between the chlorophyll dimers, as can be seen in the λ and ΔA values for 8 Å, 10 Å, 12 Å, 14 Å and 20 Å shown in figure 6.4. As the systems get more separated both ΔA and λ get greater, exponentially reducing the rate constant. This is in contrary to the previously held assumption that there is a critical distance where this rate drops to zero, although the exponential dropoff in rate will be sharp. Additionally, the coupling value H_{AB} will also decrease with distance. This is due to the dependence on overlap between the donor and acceptor orbitals in the electron transfer, which exponentially decreases with distance.

Comparing the constrained systems to the LH2 system reveals how the protein structure biases the rate constant to be even lower. The average separation of chlorophylls in the LH2 system is around 9 Å, which is in the region where transition to the CS state is most competitive. This is in contrast to the "better designed" FMO complex, where the average separation of 12 Å would drastically reduce the quenching pathway. However, as can be seen in figure 6.4, the free energy change is much higher for LH2 than for even the 20 Å case, attributed to the destabilisation of the CS state. Additionally, the crossing point for the electron transfer is to the right of the CS state minima, implying that any transition from ES to CS would quickly reverse. This illustrates how the LH2 protein actively reduces the rate constant for quenching which would otherwise be more present at such close separations.

6.5 Conclusions

It has been shown that a Frenkel exciton Hamiltonian constructed using chl-xTB can give reliable transition properties for aggregate chlorophyll systems. Whilst some issues were found in benchmarking against relevant PBE0 data, comparisons to CAM-B3LYP were still possible and

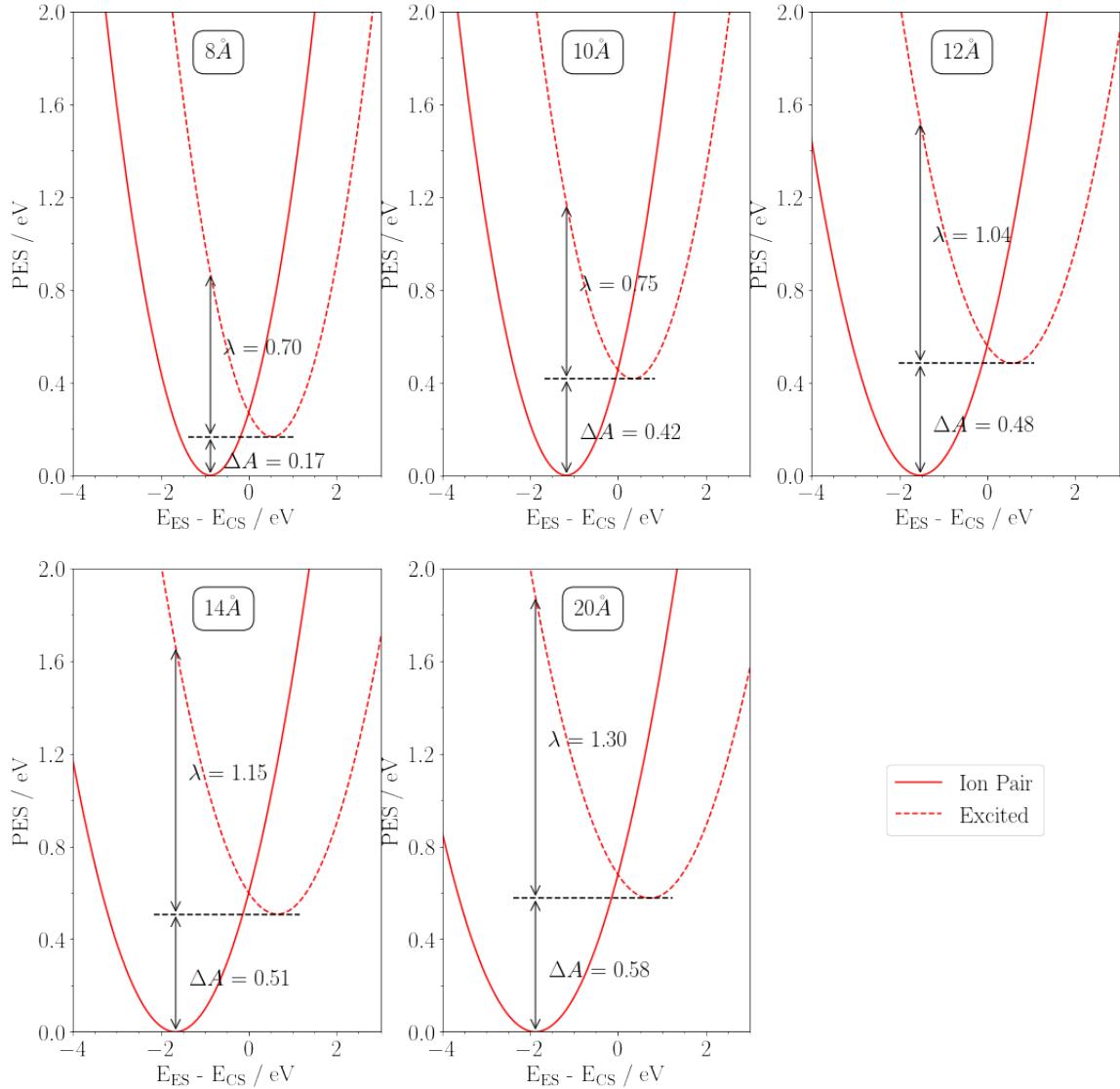


Figure 6.15: Potential energy surfaces of the charge separated and vertical excitation states of chlorophyll dimer systems, using the difference in state energies as the coordinate, for a series of separations. Free energy changes (ΔA) and reorganisation energies (λ) are shown for each separation.

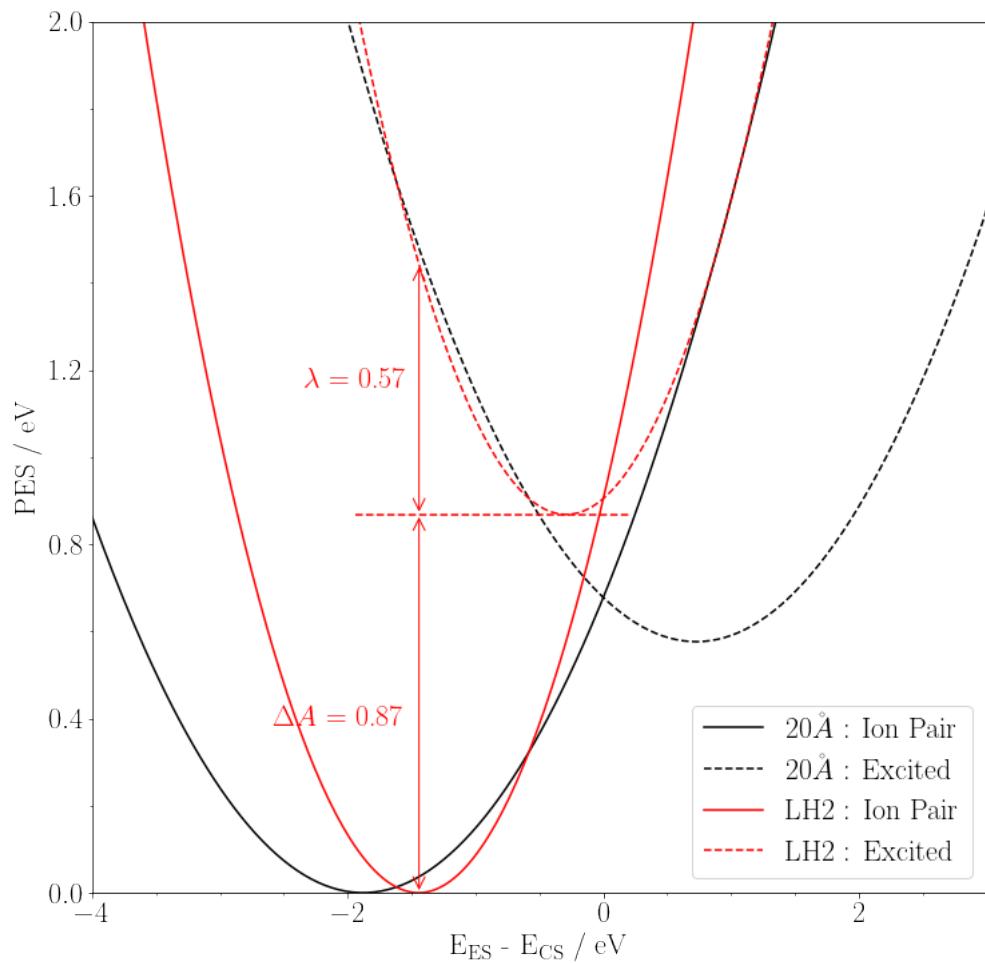


Figure 6.16: Potential energy surfaces of the charge separated and vertical excitation states of LH2 chlorophyll dimer systems, using the same coordinate as figure 6.4. The large free energy change and movement of charge separation state PES minimum to the left-hand side of the vertical excitation PES imply the LH2 protein environment inhibits transfer to the charge separated state.

showcase the accuracy of the chl-xTB exciton method. The quantitative accuracy of transition energies for LH2 dimers is on the order as expected from the training data. As chl-xTB and PBE0 based excitons predict CAM-B3LYP data with similar levels of accuracy, it is not expected that improvements to chl-xTB optimisation methods would have appreciable effect in this specific test. Instead the qualitative trends in excited state energies for artificial dimer geometries imply that chl-xTB excitons can be expected to predict variations due to geometry changes well, and this is the main focus of this work.

The benefits of the efficiently calculating chl-xTB excitons is illustrated in predicting the rate constants of transitions between excited states in chlorophyll dimers. This study is used to construct important arguments that explain quenching behavior in chlorophyll dimers and the effect of the LH2 protein scaffold. This work shows how the chl-xTB exciton method can be used for light harvesting studies that may not be possible with other response methods. Similar to the arguments in the section 5.4, it is the ability to bootstrap high level data that is the main benefit of this kind of method.

Some of the errors in the chl-xTB exciton model may be improved by using other interaction methods beyond the point charge interaction. This could be done using MNOK operators similar to the work done in the previous chapter, although with different parameters.

It can be seen how chl-xTB design choices have created a method that can be readily applied to light harvest systems. For this work, chl-xTB had to be shown to be able to predict transition properties for a range of LH2 dimer conformations, but no explicit consideration is given to other systems. The qualitative trends found in the dimer geometry scans imply that this method would work for other light harvesting complexes. However one outcome of this study could be that the training data range would need to be extended, similar to the conclusion of the previous chapter.

LIGHT HARVESTING COMPLEXES

This chapter reports on the application of the chl-xTB exciton framework in calculating the spectral density of the exciton states of LH2. By comparing this spectral density to spectra from other systems, as well as properties such as the Huang-Rhys factors for chlorophyll normal modes and chlorophyll separations, it is argued that the protein scaffold has little effect on exciton states outside of affecting individual chromophore geometries, and that intrinsic properties of chlorophyll play a much greater part in determining the spectral density.

This is the final step in establishing a new response model for light harvesting complexes. Having shown in previous chapters that chl-xTB can reliably calculate monomer and dimer properties, it remains to be tested how this model performs for a full light harvesting complex.

Similar to the previous chapter this falls into two components. First is how the chl-xTB based exciton framework is compared to previous models, which is discussed in subsection 7.2.3. This is an important step in establishing the reliability of any properties of LHCs that are calculated.

The second section applies this framework to calculate properties that similar to the charge transfer state transition rates would be difficult to do without chl-xTB. The chosen property was the spectral density of the exciton state of LH2. This would have been difficult to achieve with the methods discussed in the introduction due to the accuracy need with respect to small chlorophyll geometry variations, as well as the prediction of transition properties other than Q_y energies. This accuracy is a key requirement of obtaining reliable spectral densities, which is why a large emphasis has been given to achieving highly accurate predictions in the preceding chapters.

Once the performance of the chl-xTB framework is established, it would then be possible to discuss further applications to LHCs or other systems. Similar to the spectral densities discussed here, these would focus mainly on the features of chl-xTB that separate it from previously reported methods.

7.1 LH2

In the introduction it was stated that spectral densities of individual sites can be calculated with an explicit electronic treatment for each frame of a time series. For spectral densities of larger systems, such as the 27-site LH2 exciton system, the increased volume of calculation necessary mean that approximations have to be made that average over intra-site properties, reducing the level of detail. Due to the efficiency of the chl-xTB exciton method, it is possible to calculate the transition properties for the whole exciton system for each frame of a time series.

This chapter tries to answer two questions about environmental frequency coupling to the exciton states. The first is whether there is presence of low frequency modes in the spectral density, which would indicate a large scale motion of the protein scaffold. It can be seen on AFM images of the LH2 protein that there can be some deformations of the circular symmetry, which could be due to ring-breathing vibrational modes in the chlorophyll site positions. It is not fully understood whether this would have an effect on the exciton states. The second questions would be by how much is the exciton state spectral density different to that of the sites. This would be an indicator of whether the protein structure promotes some special interaction between sites, or if the protein only affects intra-chromophore geometry. Whilst these two questions only probe vague details of the protein structure, there are other conclusions possible that would be useful for further studies. The amount of intra- verses inter-site difference would inform how much detail is needed for future studies on LH2. It would also help in the design of artificial systems, which is discussed later.

7.1.1 Spectral Density Method

The spectral density describes how the oscillations in the surrounding environment can effect a property. Whilst it is not limited to transition energies, it is usually used when looking at light harvesting complexes to assign functions to the protein structure, as well as necessary for population dynamics.

There are several definitions of spectral density. This chapter uses the definitions laid out by Mallus *et al.* [1]. This describes the spectral density as the real part of the Fourier transform of the classical autocorrelation function. The autocorrelation $C(t_j)$ is the correlation of deviation of transition energies with itself

$$(7.1) \quad C(t_j) = \frac{1}{N-j} \sum_{k=1}^{N-j} \Delta E(t_j + t_k)(t_k)$$

where j, k are indices in the time series, N the total number of frames in the time series, and ΔE the deviation of transition energy to the mean. The spectral density is then given by

$$(7.2) \quad J(\omega) = \frac{\beta\omega}{\pi} \int_0^\infty dt C(t) \cos(\omega t)$$

where ω are the frequencies and β is the inverse temperature $\frac{1}{k_B T}$.

In practice, the spectral density is usually calculated from a discrete series of a property calculated from the frames of an MD simulation. Hence the autocorrelation and Fourier transforms are also discrete. This places a limit on the frequencies that the spectral density can cover, proportional to both the length of the MD for the low frequency limit and the interval between frames for the high frequency limit. The high frequency limit is usually around 1-2 fs in order to capture all atomic movements. The lower frequency limit is an open ended limit, dependent on the number of frames that are used in total.

7.2 Calculating LH2 Excitons

7.2.1 Molecular Dynamics

The time series of exciton states was calculated from a series of structures of LH2 generated by simulations run with the OpenMM package. The forcefield and geometry was the same as used by Ramos *et al.* [17]. For each simulation, the system was equilibrated for 60 ps. The production workflow length was 300 ps, with frames taken every 2 fs similar to the work done by Mallus *et al.* [1]. A Langevin integrator set at 300 K was used with a timestep of 2 fs. Non-bonded interactions were treated with a PME method.

7.2.2 Scaling

Exciton states were calculated with the same method as discussed in the previous chapter, extending beyond the dimer system to include all chlorophyll sites. The elements in exciton Hamiltonian were calculated using chl-xTB properties. It was necessary to implement a highly parallelised version of the code to calculate chl-xTB properties with good scaling. It was found that using a parent program to partition each chlorophyll site to run in serial on a single core had the best performance, with the data then being collected back in to calculate the exciton Hamiltonian. As the machines available had around 20 cores, the scaling of each exciton run was similar to a single chl-xTB calculation.

The main bottleneck in calculating these states was the volume of chl-xTB properties needed, as it was found construction and diagonalisation of the exciton Hamiltonian was negligible. A 300 ps MD simulation, saving frames every 2 fs, generates 150,000 individual frames requiring 4,050,000 chl-xTB runs to construct the exciton Hamiltonian. The time for each chl-xTB calculation is ≈ 1 second, so this presents about 46 days of CPU time. This was significantly reduced by using the highly parallelised program. Whilst have this level of detail is necessary for spectral density investigations, other properties of the LH2 system that can be obtained with more statistical methods may still be more efficient use of resources.

7.2.3 Coupling Values

A summary of the coupling values, broken down by ring-interaction type, are shown in figure 7.2.3. The coupling values obtained from the chl-xTB exciton framework match those of previous studies well. Coupling values are highest for B850-B850 interactions, distributed around a centre of 350 cm^{-1} , which corresponds well with previously used values between 238 and 771 cm^{-1} [6]. Intra-dimer (where the dimer is a subunit of LH2 chlorophylls consisting of a B850a, B850b and B800 chlorophyll) couplings are markedly stronger than inter-dimer couplings, which also corresponds well with previous estimated of inter-dimer couplings being 100 cm^{-1} less than intra-dimer couplings. Inter-ring couplings are much lower, with distributions around 35 cm^{-1} and 10 cm^{-1} , consistent with previous estimates of between 24 and 31 cm^{-1} .

7.2.4 Screening and Embedding

Recent studies that calculate exciton models for light harvesting systems employ screening factors as well as point charge interactions to compensate for some of the embedding effects of the protein scaffold. These were also tested, but ultimately not used for the spectral density in favour of a vacuum model.

The point charge embedding used in the previous chapter for LH2 dimers was implemented with the required PME periodic terms using the open source `hε1PME` library. The CPU time of the PME terms was profiled using the LH2 MD frames, and it was found that due to implementation problems with re-using splines and grid positions the time for each chl-xTB calculation increased from ≈ 1 second to ≈ 40 seconds. Additionally the PME functions required multiple cores in order to get the best performance, which competed with the parallelisation of chl-xTB runs. As improving PME implementations are out of the scope of this work (beyond what was already done to achieve the work in the last chapter), and considering that the embedding only marginally changes the Q_y transition, this embedding scheme was not applied to the exciton system used in this chapter.

A screening term was also implemented. This followed the same form as that reported by Mallus *et al.* [1], where any point-charge interaction has a prefactor screening term that is dependent on the distance between point charges. For example for a coupling term between two exciton states, a single element in the sum of chromophore-chromophore interactions, would be given by

$$(7.3) \quad V_{(m,1),(n,1)} = \sum_{A \in m, B \in n} \frac{f}{4\pi\epsilon_0} \frac{q_m^{\text{tr},A} q_n^{\text{tr},B}}{r_{AB}}$$

where the definition of variables are the same for equation 7.3. The scaling factor f is given by

$$(7.4) \quad f = A \exp(-BR_{ij}) + f_0$$

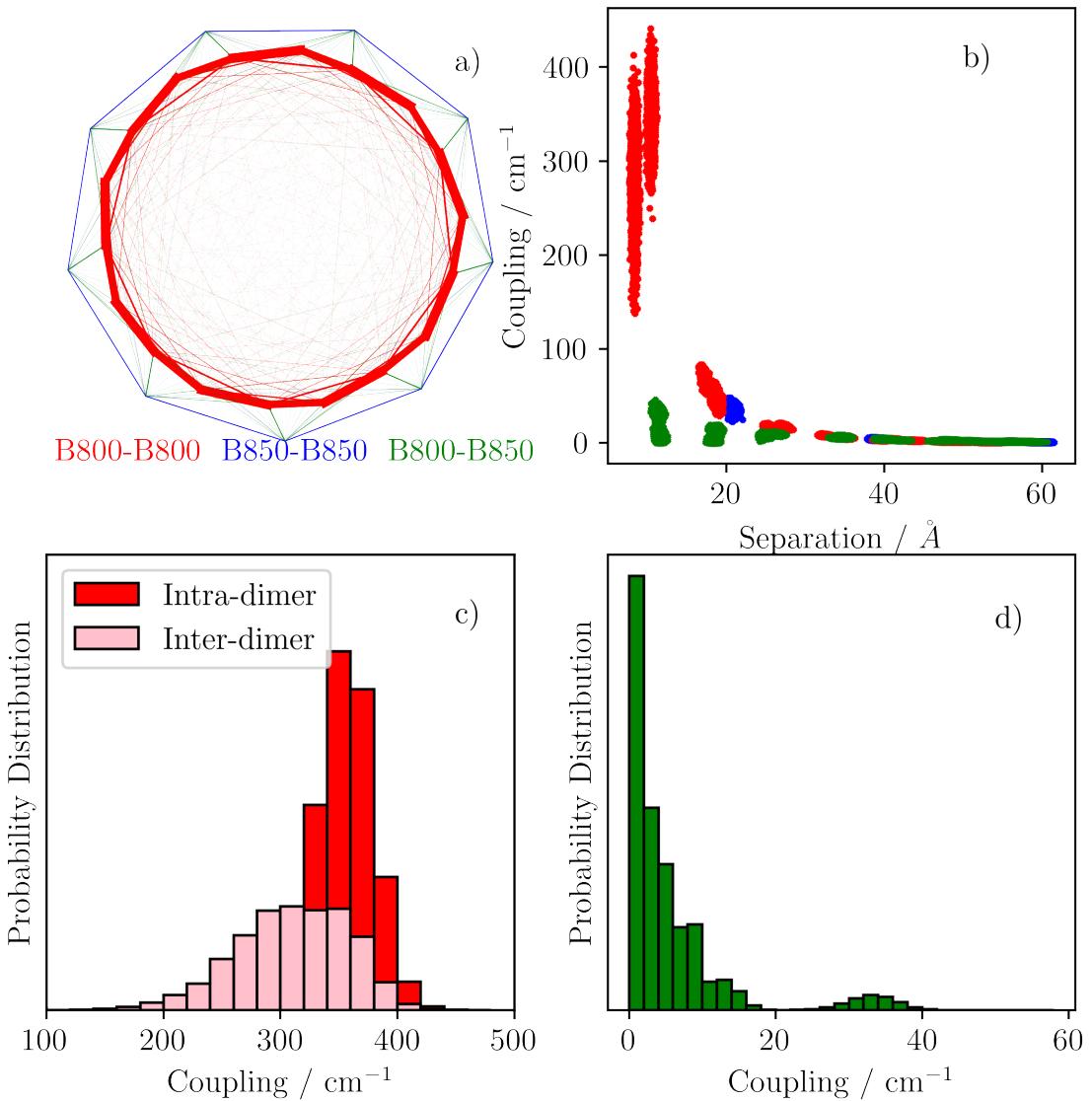


Figure 7.1: a) a diagrammatic representation of coupling strengths and interactions between sites specified by ring labels, with thicker lines indicating stronger exciton coupling interactions. b) Exciton coupling values (in cm^{-1}) plotted as a function of distance between chlorophyll sites, illustrating the distance-dependent cutoff. c) Probability distributions of B850 coupling values, showing the difference in strengths between intra and inter dimer pairs. d) Probability distribution of B800-B850 interactions.

where A , B and f_0 are constants.

When employing this screening factor, it was found the only effect was to decouple the B800 and B850 exciton states. This is best demonstrated in the simulated absorption spectra for LH2 with and without the screening factor, as well as the breakdown of exciton states into the density contributions from each site.

7.2.4.1 Absorption Spectra

Absorption spectra of LH2 is usually calculated by calculating the intensity of transitions for each exciton state and plotting these against the wavelengths of transitions to these states.

The intensity of transition I_k from the ground exciton state Ψ_0 to a (one) exciton state Ψ_k is given by

$$(7.5) \quad I_k \propto E_k |\langle \Psi_k | \hat{\epsilon} \cdot \mu | \Psi_0 \rangle|^2$$

where E_k is the energy of the state $\langle \Psi_k | H | \Psi_k \rangle$, $\hat{\epsilon}$ is a unit vector in the direction of the polarisation of the incident light, which to mimic sunlight should follow a random distribution. As the one-exciton states are mostly at the same energy the energy factor E_k can be neglected. The overlap term can then be expanded into the monomer basis giving

$$(7.6) \quad I_k \propto \left| \sum_{j=1}^N c_{kj} \epsilon \cdot \hat{\mu}_j \right|^2$$

where j, N are the index and total number of chlorophyll sites respectively, c_{kj} is the eigenvector coefficient of state k at site j , and μ_j is the transition dipole moment of chlorophyll j

$$(7.7) \quad \mu_j = \langle \phi_j^{(1)} | \mu | \phi_j^{(0)} \rangle$$

. As this intensity is dependent on the orientation of the unit vector ϵ , the average intensity can be found by either calculating the intensities for a large distribution of randomised unit vectors and taking the average, or by taking the analytic spherical average. In the large number limit, the statistical method converges to the analytic answer.

The simulated LH2 spectra with and without screening factors, alongside the experimental spectrum, are shown in fig 7.2.4.1. It can be seen that including the screening factor does produce better splitting of the B800 and B850 peaks. The poor fit of the B800 peak to the experimental spectrum has been well discussed in the literature. A lack of deformation of the ring structures is the most probable reason for the poor fit of the B800 peak in both screened and unscreened spectra.

The lack of different features in the absorption spectra imply that the only effect of using the screening terms is to lower the B800 energy states to better fit the experimental spectrum.

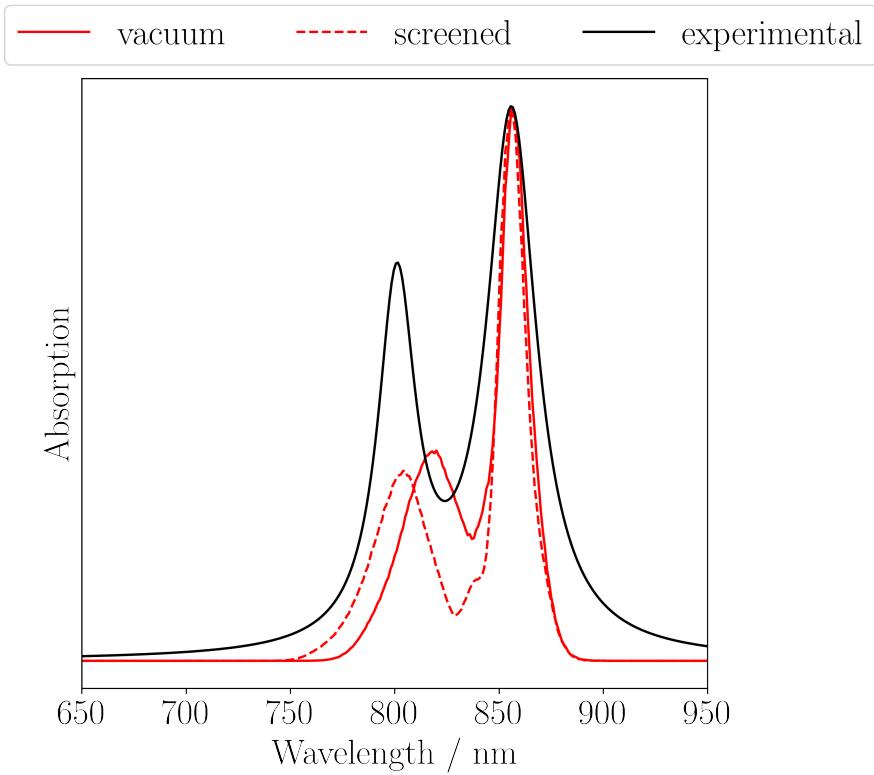


Figure 7.2: Simulated absorption spectra of LH2, with (dashed line) and without (solid line) a screening factor for point charge interaction. An experimental line reconstructed from *et al.* is plotted in black. The simulated spectra are shifted to match the position of the 850 nm peak.

This would be due to the inclusion of B850 site character into exciton states localised in the B800 ring, but the lack of significantly differing features implies this mixing may be minimal. The next section takes a more detailed look at the density contributions of each site to the exciton states.

7.2.4.2 Site Contributions to Exciton Density

The eigenvector solutions contain the amount of site character in each state, with the square of this being equal to the exciton density on a specific site. For example, the second element in each eigenvector is the corresponding amount of character from an excitation on the first chlorophyll site (the first element corresponding to the ground state, with no excitation on any chlorophylls). The average of density values is shown in figure 7.2.4.2 for both exciton states calculated with and without the screening factor, as well as the difference between them. Overall, the density contributions show that the ring structures are already significantly decoupled in the vacuum model, with the screening terms only adding slightly to this effect.

The main difference is in the amount of density shared between states localised on B800 and B850 sites. The plot of the density difference shows how density is localised more on B800

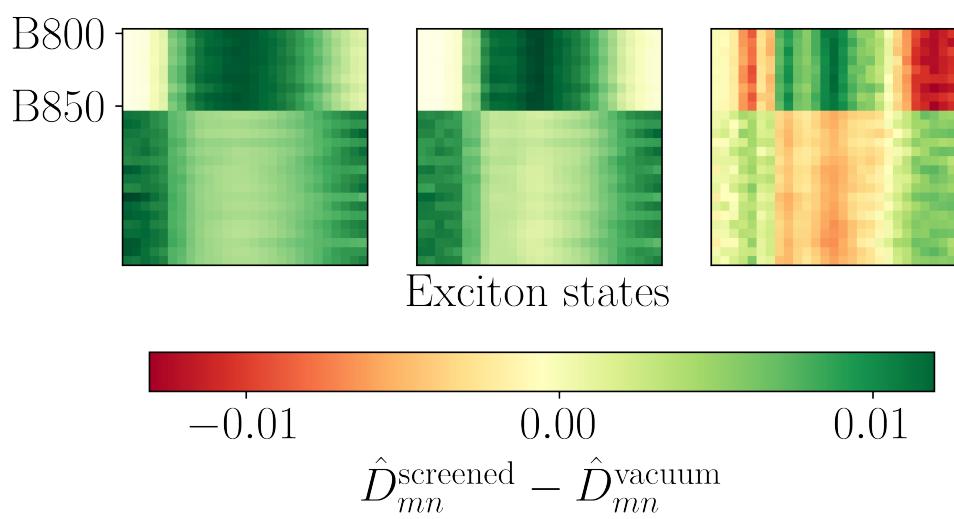


Figure 7.3: Density contributions of site transitions to exciton states for the vacuum (left) and screened (middle) Hamiltonian, with the difference shown on the right.

sites for exciton states calculated with the screening factor. This is also true for the B850 sites. However it does not reduce the delocalisation of intra-ring sites - the amount of density shared between B850 sites with B850 sites stays effectively the same. Even without the screening factor, most states are made of either only B800 or B850 character, with little mixing found.

Overall, whilst it was possible to include embedding effects into the chl-xTB exciton model, it may not change the overall qualitative behaviour of the exciton states. Calculating the chlorophyll system in a vacuum would still be a valid choice of model, and this model has shown to be effective in the previous investigations.

7.3 Spectra

The following section reports on the features of spectral densities calculated using the chl-xTB and exciton method. A more detailed breakdown of peak positions, heights and assignments can be found in appendix B.

7.3.1 Sites

In order to confirm the spectral densities calculated with chl-xTB excitons are reasonable, spectral densities of monomer chlorophyll Q_y transitions were benchmarked against literature data. This was done for a B800 site of LH2 chlorophyll, and the spectrum can be seen in figure 7.3.1.

The use of eV as units of the spectral density and well as frequency is consistent with previous literature reports. This first spectrum is compared to the spectral density reported by Mallus *et al.* [1], and uses the same axis scalings (max of 0.03 eV for $J(\omega)$ and 0.1 for $\hbar\omega$). Similar features such as the peaks at 0.09, 0.08 and 0.06 eV are found in both. The peak at 0.022 eV is not found in the literature comparison. The $J(\omega)$ values of the 0.08 and 0.06 peaks match well, although the 0.09 eV peaks is smaller in this work. These small discrepancies can be explained by the different forcefields and response methods used. It should be noted that this is in the low frequency region of the spectral density, and the major features are found at much higher frequencies. Features in this region are attributed to environmental forces [1], and so difference in the forcefield and MD methods used would be expected to cause small differences in the spectra. However the good match of peak positions and the generally important features indicate that using chl-xTB can reproduce spectral densities from higher level methods quite well.

The spectra shown in figure 7.3.1 shows an expanded range of frequencies up to 0.25 eV, as well as a breakdown by ring type. The three spectra here are calculated as the average for sites in the B800, B850a and B850b rings. Whilst many peak positions are similar in all three types, there are differences in $J(\omega)$ values that indicate differences in protein environment. All three show a major feature at 0.165 eV, with magnitudes of around 0.2 eV. The next set of largest features are around the 0.2 eV mark. Here differences in the coupling can be seen. For example the 0.21 eV peak present in B850a and B850b sites, the largest feature after the 0.165 eV at a

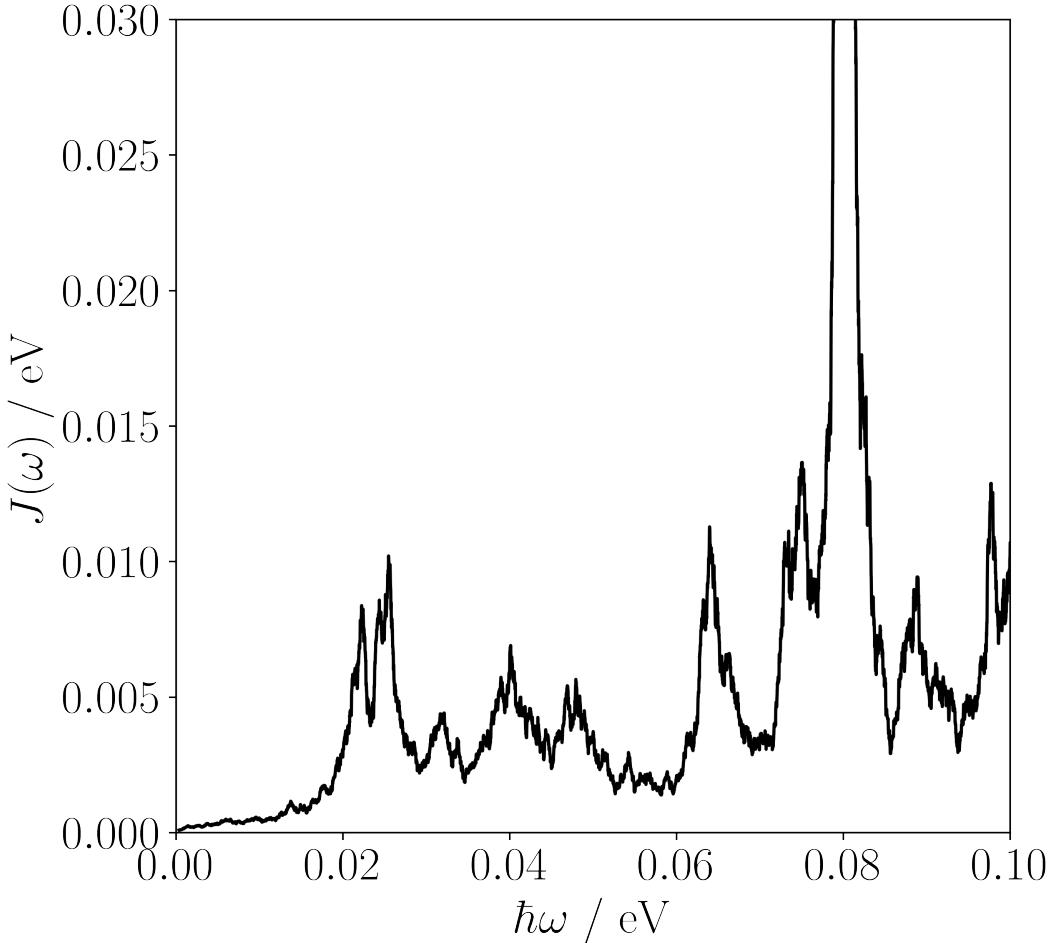


Figure 7.4: Spectral density of the Q_y transition for a chlorophyll in the B800 ring in LH2. Axis scaling and units are chosen to best reproduce the spectrum reported by Mallus *et al.*[1].

magnitude of 0.185 eV, is not present in the B800 sites. Similarly the 0.205 eV (height 0.117 eV) present in B800 is not found in the B850 spectra. A full comparison of peak positions and heights can be found in appendix B. For features found in both ring type sites, it is generally the case B800 values are higher than B850, which is in line with previous arguments that the more polar environment around B800 lead to greater variations in transition energy [22].

Generally it can be seen that spectral densities calculated with the chl-xTB method correspond well with previously reported observations. Major features appear at previously reported frequencies and magnitudes, and environmental effects are reproduced well. It is discussed later how the efficiency of the chl-xTB method would allow for better workflows in calculating spectral densities.

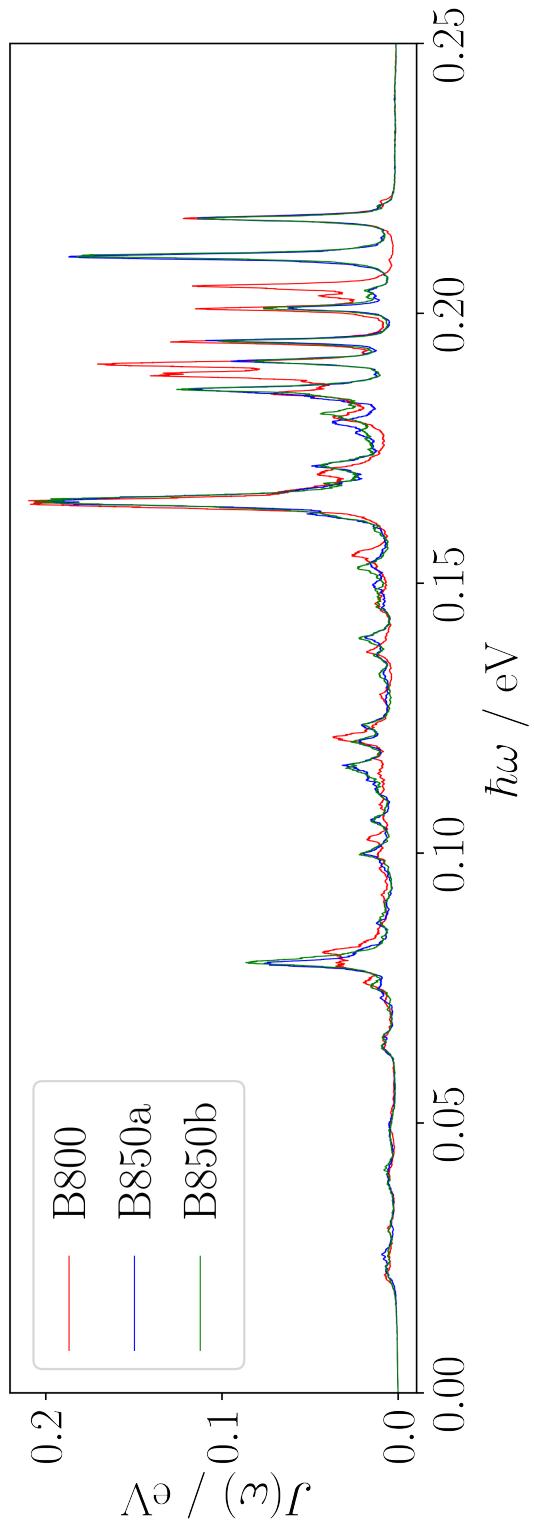


Figure 7.5: Average spectral densities for Q_y transitions at sites in the B800, B850a and B850b rings.

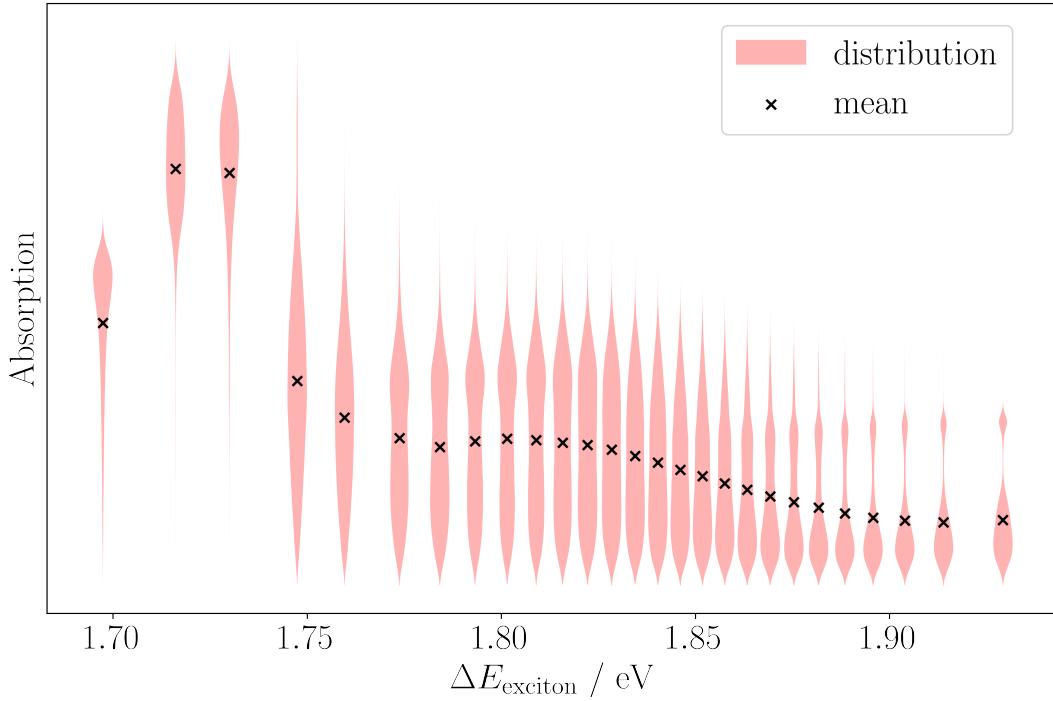


Figure 7.6: Violin plots of absorption probabilities for exciton states, positioned by the average transition energy from the ground state. The width of each distribution indicates the density of values at absorption probabilities, with crosses marking the mean value.

7.3.2 Exciton states

This section reports on the calculation the spectral density for exciton states. The work done here differs from similar work previously reported in the literature as it was possible to perform electronic structure calculations for every chlorophyll geometry in every frame of the MD simulation. Due to the strong correlation of geometry variations to Q_y transition property variations, it is expected that the features in the spectral density are representative of the environmental coupling to the exciton states, rather than any artificial features from a statistical method. It was found that the spectral densities of exciton transition energies closely match the monomer site spectra, implying that the environment coupling mainly affects intra-chlorophyll properties and not the exciton coupling values.

The exciton spectral density was calculated using the time series of the exciton transition energies (calculated as the difference between a given state and the exciton ground state) of each state (bar the ground state). The spectra shown in figure 7.3.2 have been weighted in color by their time-average absorption probabilities (shown in figure 7.3.2) to clearly present the features in the spectra. From figure 7.3.2 it can be seen that there are two "bright" states, which are the 2nd and 3rd lowest in transition energy. These absorption probabilities only consider ground

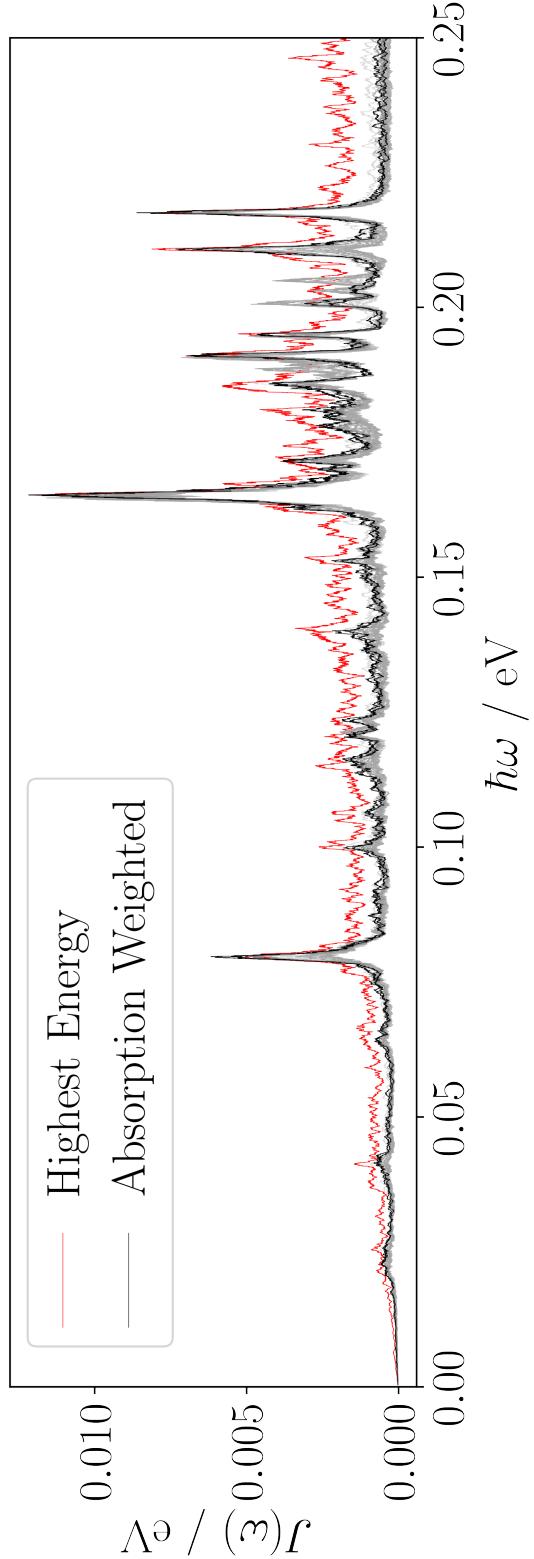


Figure 7.7: Spectral densities of exciton transition energies of LH₂, weighted by the averaged absorption probability. The highest energy state is indicated in red. This state has a higher fluctuation in energy, causing excess noise in the spectral density.

to excited transitions, and not the transition probability between exciton states. There is one outlier state in these spectra, marked in red - this corresponds to the highest energy state, and the increase in $J(\omega)$ values are attributed to increased variations in this state's energy, which would exacerbate the noise in the spectrum. Discussion of this state is limited due to this effect, as well as the fact that it may not be the most important state to consider as the absorption probability to this state is the lowest.

Looking at the 2nd lowest energy state, the main features appear at near identical frequency positions as the site spectra. The largest feature is at 0.165 eV, with a magnitude of 12.183 meV. Similarly, peaks at 0.211 eV and 0.218 eV have large amplitudes at 6.775 meV and 8.62 meV respectively. Full assignments can be found in appendix B.

From the change in scale on the y-axis, it is obvious that the environmental effect on exciton states is much smaller than for sites. This is explained by the lack of correlation between chlorophyll motions, generally cancelling out any variation in Q_y transition properties, mostly staying close to the mean. This would reduce the fluctuation of exciton transition energies, which in turn would decrease the magnitude of any peak in the spectral density.

The similarity of the exciton state and the site spectra implies that the protein environment effect on the transition at a single chlorophyll site is much greater than any effect on the exciton coupling between chlorophyll sites. What this indicates is the lack of any large scale environmental effects on multiple chlorophyll sites, and that all of the environment effects are localised at each site. This does not exclude the possibility that exciton coupling variations can be present, just that these would most likely be due to intra-chlorophyll variations, and not, for example, a large scale movement of the protein structure to bring chlorophylls closer together or change the angle between porphyrin planes. The next section looks at the variations in coupling values in greater detail.

7.3.3 Coupling

The spectral density of the coupling terms were calculated with the same method as the state and site transition energies, and are shown in figure 7.3.3.

It can be seen that there are far fewer features in the coupling spectral density than in the site and state transition spectra. Additionally there is a broad feature at the low frequency end of the spectrum, around 0.01 eV. Intuitively, the magnitude of the spectral density decreases as the separation of chlorophylls increase, with anything but nearest neighbours showing significant value against the strongest coupling spectrum.

Taking the largest valued spectrum, where the distance between the two chlorophylls was the smallest with an average 15.9 Å, the high frequency features correspond to features found in the exciton state and site spectra. The strongest features are found at 0.201, 0.203 and 0.205 eV, with strengths of 0.416, 0.457 and 0.372 meV respectively. These correspond with major features in the site spectra, implying that these are due to intra-chlorophyll variations rather than any change in the protein scaffold.

Previous arguments about coupling terms have said that the change in distance is the controlling factor. The effect of distance was also investigated by calculating the spectral density of the interchromophore distance (reported without units as these are not physically meaningful). These spectra can be seen in 7.3.3.

It can be seen that while there is some correspondence to the exciton coupling spectra in the low frequency region, there is little correspondence to features in the high frequency region. The strongest features at 0.169, 0.175, 0.180 and 0.182 eV do not correspond to frequencies present in other spectra. The largest feature in this distance spectral density is at 0.003 eV, which is at a similar frequency to a low frequency peak in the coupling spectra. However, these peaks are qualitatively different, with the exciton coupling peak being much broader. This low frequency peak in the exciton coupling spectra is relatively weak compared to the intra-chlorophyll motion coupling peaks.

The lack of peaks in the coupling spectra from 0.03 eV to 0.168 eV also supports the argument that intra-chromophore variations, and not the protein structure movements, are the determining factor in exciton state transition energy variations.

Overall, it can be seen that whilst some exciton coupling variations may be due to changes in the protein scaffold moving chlorophylls, generally this has little affect on the exciton state spectral densities. The similarity between major features in the coupling spectral density and site and state spectra imply again that almost all variation originates at chlorophyll sites.

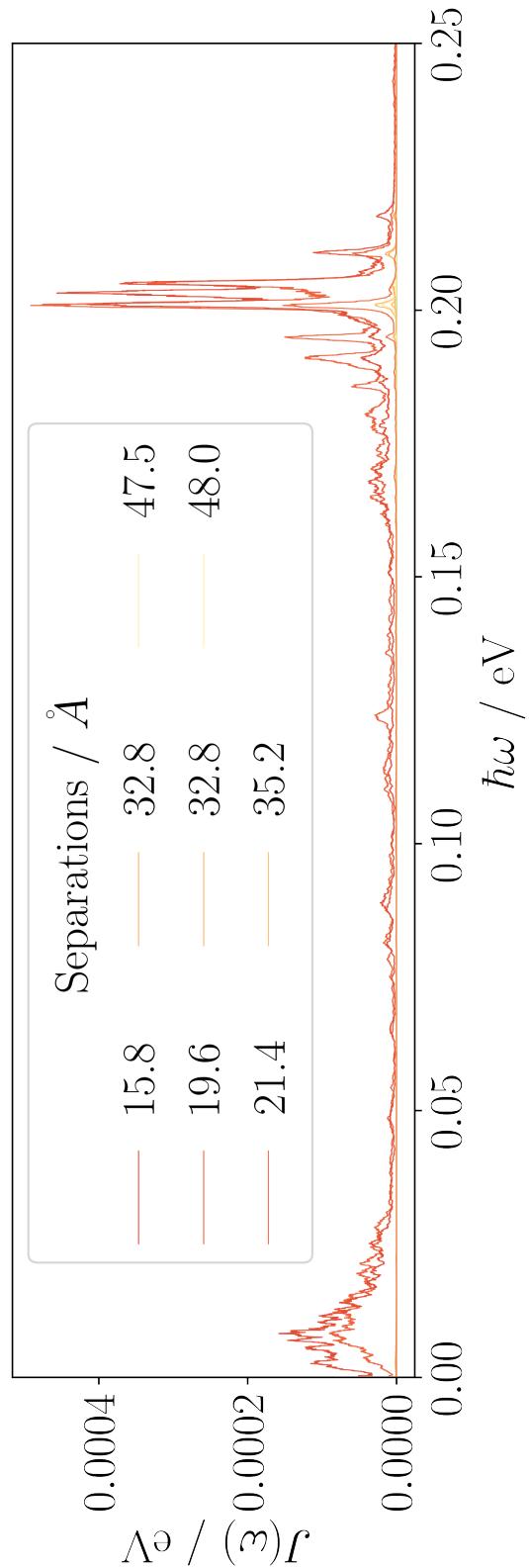


Figure 7.8: Spectral density of exciton state coupling values (i.e. off diagonal elements of the exciton Hamiltonian) of LH2 coloured by average distance between chlorophyll sites.

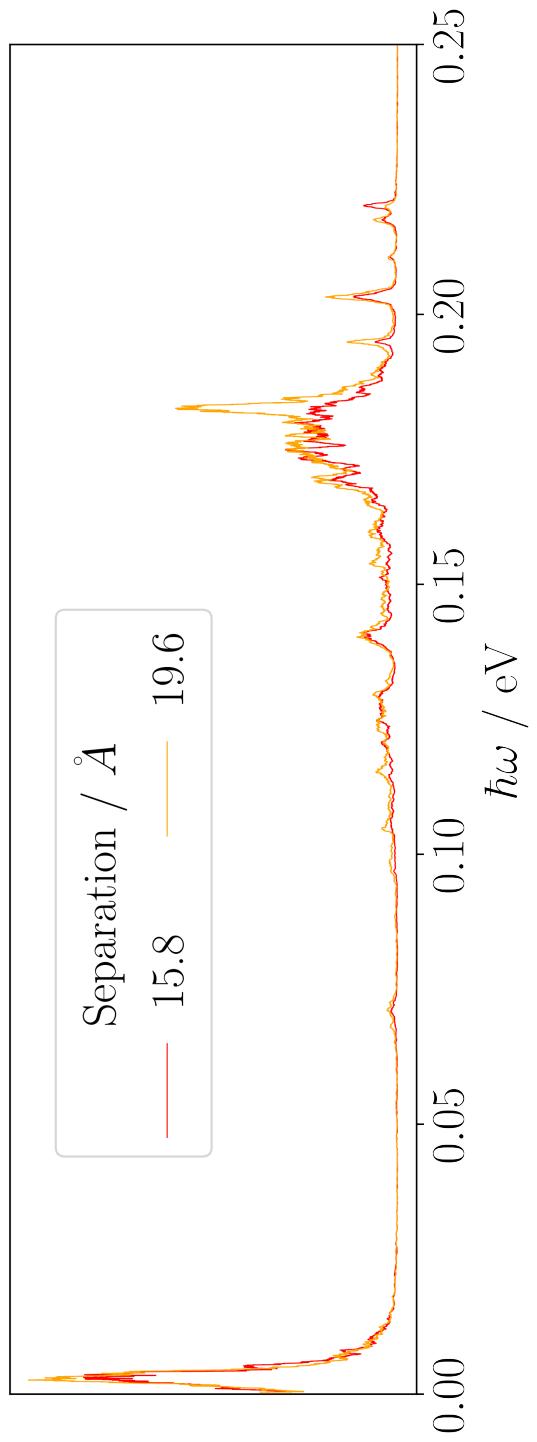


Figure 7.9: Spectral density of the separation between nearest neighbour sites in LH2.

7.4 Assigning Specific Motions

The observation that variations in exciton state energies are primarily due to intra-chromophore variations opens the question about which motions are causing these variations. Discussion of these spectra so far has been limited due to the lack of assignment of specific motions that are coupling to the environment. This section reports on some of the tests used to assign features in the spectral densities. This includes comparing the LH2 spectral densities to a spectral density of monomer chlorophyll embedded in diethyl ether, as well as a spectrum constructed from Huang-Rhys factors calculated with chl-xTB response properties and a GFN1-xTB hessian. Some speculative explanations of the lack of low frequency features are also given.

7.4.1 Ether system

So far only the environmental coupling of LH2 has been considered. Whilst some differences between the ring sites have been observed, it is not clear how other environments couple to the Q_y transition. A candidate environment of chlorophyll in diethyl ether was investigated, to match the previous chapters. The conclusion from comparing the spectral density of chlorophyll in diethyl ether would lie between two extremes - either there is no variation in the spectral density, or major variation. The first possibility would support the argument that the LH2 protein does not do anything discernable to chlorophyll spectral densities, whereas the second would imply that the protein environment does need to be considered more carefully.

The time series of chlorophyll geometries were taken from an MD simulation of a diethyl-ether embedded chlorophyll. This was constructed using a solvent box made with the packmol program with a single chlorophyll molecule in a 64 Å box with 1054 diethyl-ether molecules. The simulation was run with OpenMM using parameters for chlorophyll taken from the LH2 forcefield, and parameters for diethyl-ether taken from the General Amber ForceField (GAFF). The system was equilibrated for 60 ps, with a production workflow of 300 ps run. Structures were taken every 2 fs. A Langevin integrator set at 300 K was used with a timestep of 2 fs.

The Q_y transition was calculated for every frame of the MD trajectory. The spectral density of these transition energies was calculated with the same method as the previous spectra. The spectral density can be seen in figure 7.4.1.

The major peaks are found at similar frequencies to the site and state spectra, albeit with different magnitudes. The strongest feature in the diethyl-ether spectrum is at 0.21 eV, with a magnitude of 0.241 eV. The next strongest feature is at 0.165 eV, the frequency of the strongest peaks in the site and state spectra, but at a magnitude of 0.232 eV. The collection of features at 0.186 eV, 0.191 eV, 0.195 eV and 0.201 eV are also present with magnitudes of 0.15 eV, 0.069 eV, 0.08 eV and 0.103 eV.

It can be seen that the features in the spectral density for diethyl-ether embedded chlorophyll are similar to the features from the LH2 chlorophyll and exciton transitions. This implies that

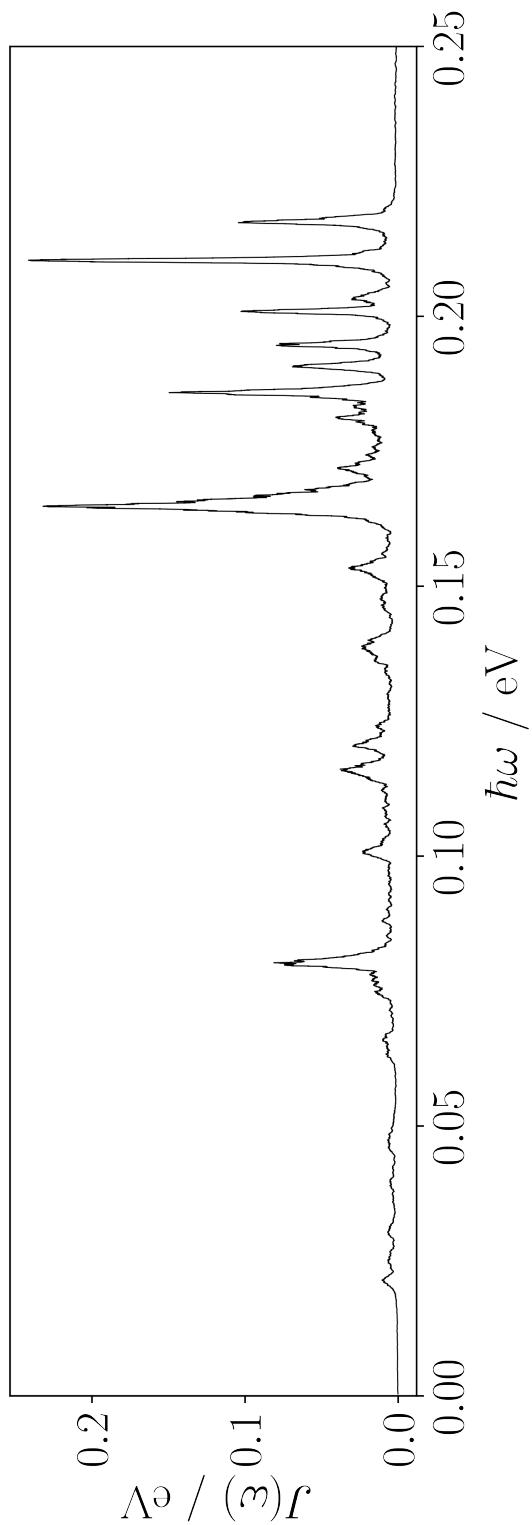


Figure 7.10: Spectral density of the Q_y transition for a chlorophyll in diethyl-ether.

the coupling of environment to energy fluctuations is due to mostly intrinsic properties of the chlorophyll geometry. There is still some variation in the magnitudes of the environmental coupling, however the lack of any features at different frequencies or major changes in coupling magnitudes imply little effect originating from the environment. Overall coupling between chlorophyll transition energies and the environment are controlled by intrinsic chlorophyll properties.

7.4.2 Huang Rhys Factors

These intrinsic properties that cause features in the spectral density would most likely be the coupling of internal vibrational motions of chlorophyll to the Q_y transition. A larger coupling of vibrational motion would imply greater variation in the Q_y transition energy, which would increase the magnitude of features in the spectral density. The coupling of normal modes to electron transition can be calculated with Huang-Rhys factors, defined by the difference between minima in the excited and ground state energy surfaces along normal mode coordinates. By comparing the Huang-Rhys factors of all normal modes of a chlorophyll molecule, it would be possible to identify which internal motions are responsible for spectral density features.

The normal modes used to calculate Huang-Rhys factors were calculated with a hessian calculation on an optimised single chlorophyll structure. It was found that rotations in the phytol tail caused issues in converging to an optimised geometry, attributed to the low energy barrier for C-C bond rotation. The phytol tail was removed to overcome this issue, with a hydrogen atom replacing the phytol group. The optimised geometry was then used to calculate normal modes with GFN1-xTB.

A scan of excitation energies was calculated for each normal mode, with the chlorophyll atoms being displaced along the vectors derived from the hessian of the optimised chlorophyll structure. The coordinate of the scan was defined as

$$(7.8) \quad q_i = \sqrt{\frac{\omega_i}{\hbar}} x_i^m$$

where ω_i is the angular frequency of the normal mode i and x_i^m is the displacement vector in mass weighted coordinates. The Q_y transition energy was calculated for a series of structures with successive values of q_i . Fits of the ground state and excited state energies were made with quadratic functions, from which it was possible to make estimates of the q value for a minimum ground state energy (q_{ground}) and excited state energy (q_{excited}).

From these it was possible to calculate the Huang-Rhys factors as

$$(7.9) \quad d = \frac{(q_{\text{excited}} - q_{\text{ground}})^2}{2}$$

. These Huang-Rhys factors were then used to construct a spectral density, using their absolute value for amplitude and the frequency of the corresponding normal mode as position in the frequency domain. A plot of this spectra is shown in figure 7.4.2.

The low frequency (<0.05 eV) modes in this spectra are clearly suppressed in LH2 and diethyl-ether environments. As these motions correspond to large scale deformations of the porphyrin ring, the high force constants might make these motions unobservable rather than any arguments based on the environment. It is fairly clear that they should not explain any features in the spectral densities.

The high frequency normal modes correspondence to spectral density features is less clear. Whilst some peaks in the Huang-Rhys spectrum are similar to the full spectral density, the overall change in frequency positions make comparison difficult. This could be due to the completely different vibrational modes present, however it is more likely that whole sections of normal modes have been shifted from the frequencies observed in the spectral densities, due to the differences in the forcefield and GFN1-xTB method. The motions themselves should be similar, even if the frequencies are not. This implies that the major features in the Huang-Rhys spectrum should correspond with major features in other spectral densities. These modes have been labeled in figure 7.4.2. All of these modes have significant movement in the N_A , N_B , N_C and N_D atoms, corresponding with a symmetry breaking along the Q_y dipole axis. This is shown in figure 7.4.2.

In total the Huang-Rhys factors do not offer a clean explanation of spectral density features. It could be that using normal modes calculated with the same forcefield method might improve the comparison, but this was not possible with the resources available. However looking at the atomic motions more explicitly could explain some of the features observed.

7.4.3 N Axes Deformation

The motion of the N_A - N_C and N_B - N_D axes inducing a D_{4h} - C_s symmetry break in the vibrational modes with the highest Huang-Rhys factors could suggest that many of the major spectral density features are due to this motion. As it is possible to construct a time series of a metric to describe this deformation, it would be possible to compare the spectral density of these motions to the transition energy spectral densities. The chosen metric was the ratio of N_A - N_C and N_B - N_D length, and the spectral density of this property is shown in figure 7.4.3, again without amplitude units as these would be physically meaningless. In order to account for all chlorophyll sites, the spectral density was calculated for each site and then averaged.

Whilst the major site/state feature at 0.165 eV is also present in this N axes spectrum, there is little correspondence in other peaks. The majority of peaks are in the 0.022-0.121 eV range, which is not populated in the transition energy spectra, and the only two corresponding peaks are at 0.165 eV and 0.205 eV, with the latter only present in B800 sites. Overall this metric offers little in explanation of which motions are responsible for transition energy spectral density features, with the exception of the major 0.165 eV peak. Whilst a little surprising that most

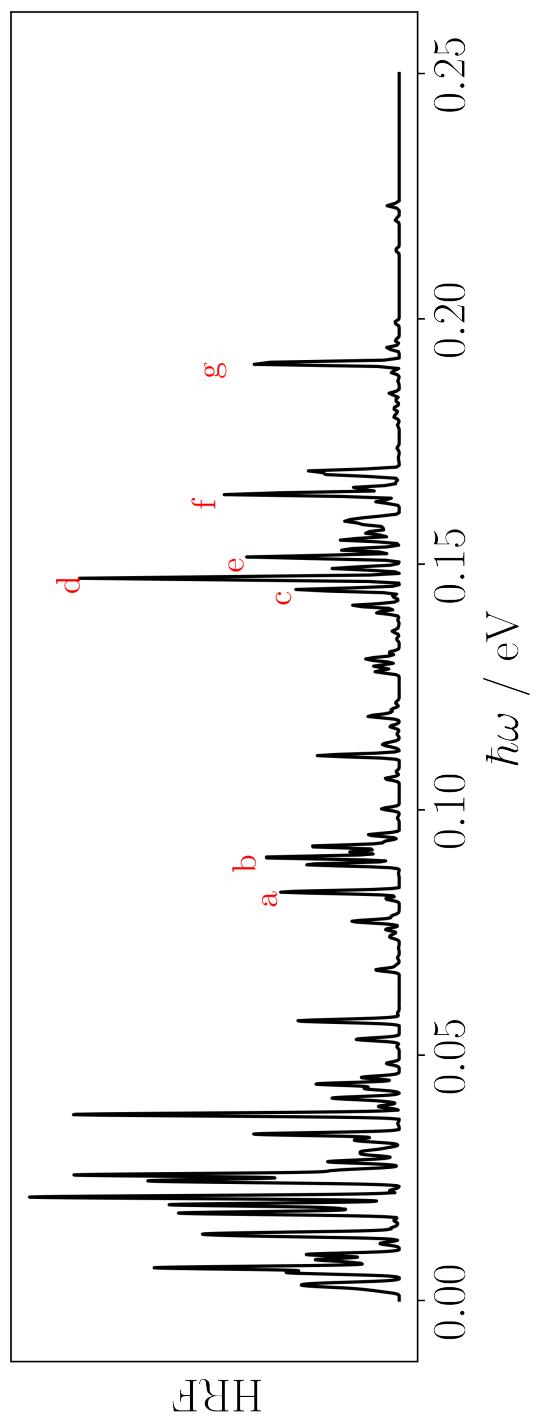


Figure 7.11: A simulated spectral density of the Q_y transition for chlorophyll constructed from Huang-Rhys factors and the frequency of normal modes. Labeled peaks are chosen as modes that may correspond to features in other chlorophyll spectral densities.

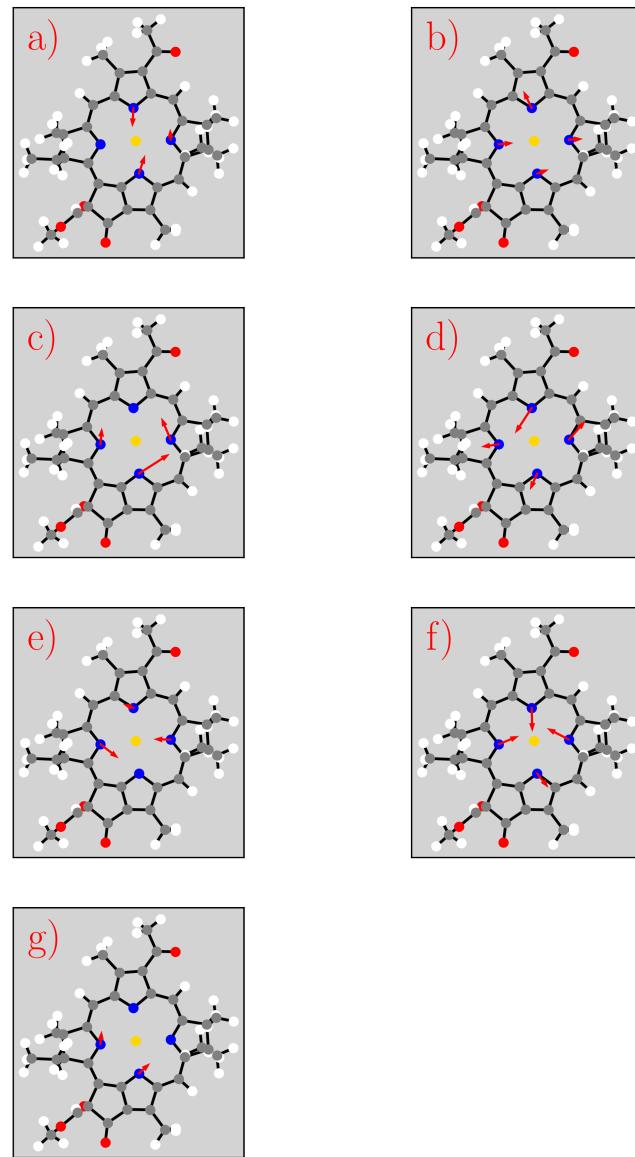


Figure 7.12: Motions of the four central nitrogen atoms in bacterial chlorophyll for the vibrational modes labelled in figure 7.4.2.

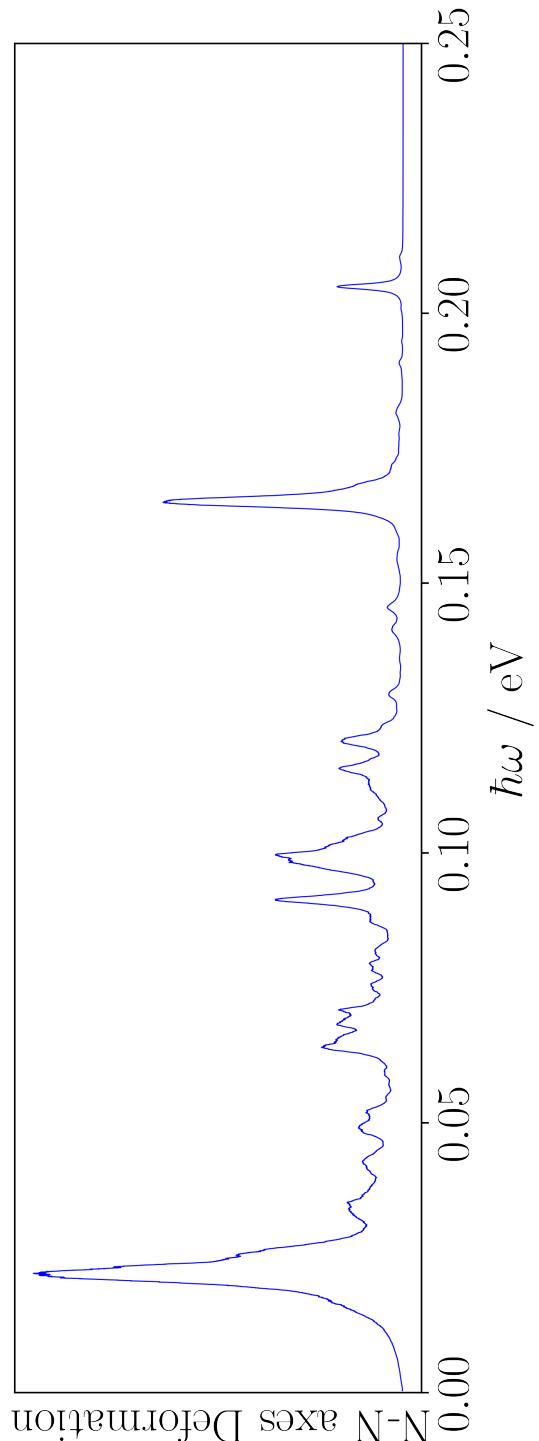


Figure 7.13: Spectral density of the ratio $\frac{|N_A - N_C|}{|N_B - N_D|}$, averaged over chlorophyll sites in LH2.

features must be due to over vibrational motions of chlorophyll, it is encouraging that the major feature corresponds to a D_{4h} - C_s symmetry breaking motion.

Full assignment of the spectral density features is outside the scope of this work, but it can be seen that there are two possible options for further investigation. One would be producing hessians from forcefield parameters, which would make a more compelling comparison of Huang-Rhys factors and spectral density features. The second would be a more thorough look at all atomic motions, creating a large series of spectral densities of geometry metrics. The issue with this investigation would still be the lack of correspondence between spectral density features and normal mode vibrations, however it may shed light on which chlorophyll atoms are particularly important to spectral density features.

7.5 conclusions

The work in this chapter shows that the chl-xTB exciton method can fulfill the criteria set out in the introduction when describing the scope of this project. Explicitly calculating the large number exciton states necessary for the spectral density was achieved due to the efficiency of using a semi-empirical method. Whilst other methods could have been used to achieve this, the accuracy against TD-DFT level data was a necessary factor in order to be able to make reliable conclusions about response property variance. The conclusions on exciton state spectral densities are based upon the good agreement between the LH2 site spectral densities and other reported works.

The exciton transition energy and coupling energy spectra clearly show that thermal environment coupling is mostly based around intra-chlorophyll variations. The coupling spectra show how low-frequency motions of the protein are a relatively unimportant factor in exciton state variation, and that even coupling variations originate at the monomer level than any aggregate factor.

Whilst these intra-chlorophyll variations are explored in some detail, it remains a challenge to fully assign the underlying motions for peaks at specific frequencies. The similarity of spectra from an LH2 environment and diethyl-ether environment strongly imply that these motions are intrinsic to chlorophyll, and that the environment only subtly changes the magnitudes of couplings and not the frequencies. The study into correlation between vibrational motions with significant Huang-Rhys factors and spectral density features gave some insight into these motions, but due to the differences in the hessian and forcefield method this is not conclusive. Additionally using geometry metrics such as the N axes deformation explain some features but not all. Studying more of these metrics, especially for atoms where the transition density is centred, may explain more features.

One obvious way to extend this study would be to calculate exciton states for a longer simulation. This would not affect any of the conclusions about the environmental coupling to exciton states in the high frequency region, but would test the absence of any low frequency

features. Similar as discussed above, the lack of protein motions may be due to the initial low energy crystal structure or a high energy barrier, but it could also be that these motions are slower than could be captured by the used resolution. Increasing what would count as a reasonable computation time would be necessary, for both the exciton states and calculating the autocorrelation and fourier transforms, however the workflow would not change.

Using chl-xTB to calculate spectral densities is not necessarily limited to LH2. For example spectra of the FMO light harvesting complex could be generated with little alteration to the workflow used here. However when looking at other systems some consideration may have to be given to the chl-xTB training data, making sure to include other conformations or chlorophyll molecule types. Considering the spectral densities from more systems would contribute to the conclusions about feature origins discussed here.

CONCLUSIONS

The discussion of results in the previous chapters have analysed whether the novel approach to designing a novel response method for LHCs has worked. In the first chapter it was found that while the more efficient Δ -SCF approach is viable for DFT based methods, the tight-binding Δ -xTB methods would not be accurate enough for LHC models. This was argued to be due to the parameterisation of the GFN-xTB methods used, as other tight-binding schemes do work well. From this the more specific chl-xTB method was designed and parameterised for Q_y transition properties. The specificity of this model is similar to the machine-learning models discussed in the introduction. chl-xTB proved to be highly accurate against the selected training data, although choices in the constructing the training data has shown some signs of limiting this method. This is discussed in below.

Applying this efficient response method to LHCs also proved successful. Dimer transition properties as well as the LH2 coupling values were well recovered by the chl-xTB framework, implying that conclusions about LHC (or other oligomer chlorophyll system) properties would be reliable for discussion.

Having established the usability of this efficient response method for LHCs, it was then possible to calculate novel properties. These were chosen to highlight how the chl-xTB exciton framework answers some of the issues discussion in the introduction with existing LHC models.

The first application was the rate of transition to the charge transfer state for chlorophyll dimers, for which it was necessary to have a model that could include embedding effects from different environments. This was straightforward to achieve with chl-xTB as the point-charge embedding terms can be added into the Fock matrix. This is in contrast to the machine-learning models where QM/MM effects must be present in the training data.

Calculating the spectral density of whole exciton state was the second application. This relied

on the accuracy of chl-xTB with respect to small changes in chlorophyll geometry, which may not be found in more general tight-binding methods (as shown in figure 5.2.4). It also required a highly efficient response method due to the large number of individual LH2 geometries needed for the spectra.

Both of these applications resulted in novel explanations of LHC phenomena. The rates of transition show how the LH2 protein purposefully inhibits the charge-transfer quenching mechanism of excited states when compared to a more stochastic environment such as diethyl-ether solvent. The spectral densities then investigated further how the environment couples to the exciton system, finding that intra-chlorophyll variations are far more important than inter-chlorophyll interactions, a surprising outcome given the delocalisation of exciton states across the site basis.

The investigation of novel efficient response methods and their applications reported here was not intended to be (and is not) an exhaustive search but more a proof-of-concept of how future models could be constructed. Some considerations for other models have been highlighted in the previous chapters, such as how the prediction of transition properties from chl-xTB is highly dependent on the training data. Other considerations are proposed below, sketching out possibilities for future work, based on some observations that have not been part of the discussion so far. This includes how more properties for LHCs could be calculated using the chl-xTB exciton framework. A workflow for designing similar methods for systems other than LH2 BChla chlorophyll is proposed, which may be of use in studies on systems similar to LHCs.

8.1 Efficient Response Methods

Testing the chl-xTB response method approximations on a different system would be the logical extension of this type of method. In chapter 5 it was posited that the success of this method was due to strong single-character transition, a feature that has been explored in other work as well. This is also corroborated by the accuracy of the DFT based Δ -SCF methods from section 4.2.4 as these are also single transition methods.

It is argued then that the response approximations and reparameterisation of GFN1-xTB may give good results for other single-character systems and transitions. These investigations would have a very similar method to that used in chapter 5, constructing a set of training data from a high level method and fitting the response method parameters. Due to the high level of accuracy achieved in chl-xTB from a relatively small set of training data, this reparamterisation could be easily done for other systems. If these systems are smaller then the training data could be made from higher level methods or from more data.

A logical candidate for this would be the Q_x transition in chlorophyll. This transition mostly characterised by the HOMO-1 - LUMO+1 transition so the approximations used in chl-xTB response would be applicable. The changes to the GFN1-xTB Hamiltonian parameters would

also be the same as the transition density of this transition is still centered on the porphyrin ring similar to the Q_y transition. Retraining for this transition then would be a straightforward investigation to do. This would be a useful method to construct for some models of LHCs, especially for those more complex than those looked at in this work. Some exciton systems expand the basis sites and transitions to include the Q_x transition, although this is only necessary to create a more detailed image of the higher energy exciton states.

Reparameterising for other chlorophyll systems would also be a good area for further investigation. As the LH2 protein only includes BChla molecules it was not necessary to construct models for other types of chlorophyll. However this would no longer be the case when looking at other chlorophyll systems such as solvated systems or other types of LHCs. Q band transition in other chlorophyll systems are similar to those found in BChla with changes found in the transition energies and the amount of HOMO-LUMO and HOMO-1 - LUMO+1 character. This is dependent on the functional groups that are attached to the porphyrin ring.

There would be two possible approaches to alter the chl-xTB as it is reported here for use on other chlorophyll systems and transitions. The first approach would be simple reparameterisation using the parameters already reported. Not altering which GFN1-xTB parameters are optimised would probably work best for a model that predicts a single system and transition as has been investigated here. A more encompassing approach would be to include Q_x transitions as well as other chlorophyll systems into one set of training data and optimise on model. Due to the diversity of this training data it would probably be necessary to reoptimise additional parameters, such as a better treatment of the atom types in the porphyrin ring functional groups (i.e. specific C and O parameters). The range of transition energies may also cause issues when using the objective function reported in section 5.2.1. Both the RMSE and R^2 values would increase with a more varied set of training data. It could be that the values of transition energies would form clusters around some set of means which would artificially raise the R^2 value for the whole training set and obscure the more detailed correlation for transition properties in one system. As stated in chapter 5 the correlation is an important metric for getting good response in predicted transition properties to geometry variations. Investigating the weights λ_n in the objective function (eq. 5.18) may solve this issue by getting the optimisation procedure to apply higher pressure to this metric. It is hard to argue whether a similar level of accuracy would be achieved by including differing systems and transitions however the benefits of having a method that could be applied to more than one specific case are obvious.

Going beyond chlorophyll, it would be possible to test this procedure for other systems. One candidate relevant to LHC studies would be carotenoid units (present in LH2) that participate in some electronic energy transfer to the chlorophyll system. Similar to the inclusion of Q_x transitions in the exciton framework some models also include carotenoids as basis sites again for a more detailed description of high energy states. It may be that these models would only be successful if the transitions are mostly single-character and have well defined transitions that

can be used in the objective function.

Expanding the chl-xTB method to other systems as transitions would be useful for calculating properties for more systems. For example, the spectral densities of other chlorophyll systems could be calculated and compared to those reported here. Using the chl-xTB exciton framework, as well as some extensions, for further LHC applications are discussed below.

8.2 Further Investigations into LHCs

As stated above the exciton framework could be extended with additional basis sites and transitions such as the carotenoid sites and Q_x transitions. These states are usually higher in energy and so many not couple much to the lower exciton states based on Q_y transitions. However they are important when considering all the mechanisms for electronic energy transfer between states. However to achieve these models with the similar methods used in this work would require reparameterisation as discussed above.

The chl-xTB exciton method could be used for many more LHCs studies than those reported here. One application would be calculating spectral densities of other LHC complexes, such as the FMO complex, to find whether these spectra are also be dominated by intrinsic chlorophyll properties or whether there is more effect from the protein manifold. The only objection to this investigation would be that, similar to the conclusion from section 5.3.2, the exclusive use of LH2 geometries in the training set data may have biased the model. This effect is almost negligible as the agreement of predicted spectra shown in figure 5.20 is very good and it is only the smaller Q_y peaks that are missing, in both the DFT reference methods as well as chl-xTB. Expanding the training set is achievable with some considerations already discussed above.

Another key benefit of the chl-xTB response approximations, and a design choice that has only lightly been discussed, is that the formalism is fairly simple compared to full linear-response or other tight-binding methods such as sTDA-xTB. Specifically the gradient of the response properties with respect to atomic positions would be substantially easier to calculate than these methods, as TD-DFT gradients would require solving couple-perturbed equations and sTDA-xTB gradients would have a gradient term in the basis functions due to their coordination number dependence. The gradient of the exciton system is constructed from these site gradients similar to the Hamiltonian. Whilst dynamics simulations of the LH2 exciton system have been done before with gradients, several approximations were required that would not be needed if using the chl-xTB framework. Additionally this approach required DFT level data which limits what could be calculated within reasonable expense. Using the chl-xTB framework would give a much more *ab initio* approach. Calculating the gradients of chl-xTB transition properties would be much easier than other methods. The gradient of the transition charges, for example, would require gradients of the MO coefficients from the altered GFN1-xTB method. However as this method is optimised for analytic geometry optimisations which also require this gradient term,

many programs would readily be able to provide this. Also as the other terms in calculating the transition energy are based on interatomic distances the gradient terms should be relatively easy to calculate.

Additionally as the chl-xTB exciton method is extremely efficient future investigations could be similar to previous studies where it would have been beneficial to just calculate a larger volume of properties. For example a larger model of the entire light-harvesting apparatus present in purple bacteria could be achieved, or investigating the effects on EET when multiple LH2 proteins are present. These investigations would not add anything more to the model but highlight how increased efficiency allows larger scale models to become feasible.

These suggestions for future work highlight the useful features of the chl-xTB monomer method as well as the chl-xTB exciton framework. As there are many potential areas for further research it is argued that these novel solutions to the problem of scale in LHC models introduced in chapter 2 will be useful for inspiring similar approach in the future. The efficiency of tight-binding methods has been well utilised and would allow for larger models to be possible. The more extendable formalism, achieved by being based on electronic structure methods and requiring few parameters, would allow for more many more properties and effects to be taken into account without the need for large change. Finally many of the transition properties are reliable due to the decent accuracy achieved by optimisation to a more specific training set than other methods, a feature which should also be true for cases beyond the Q_y chlorophyll transition. These approaches should therefore be useful tools for the future investigations in LHCs that are necessary to fully explore photosynthesis and light harvesting mechanisms.



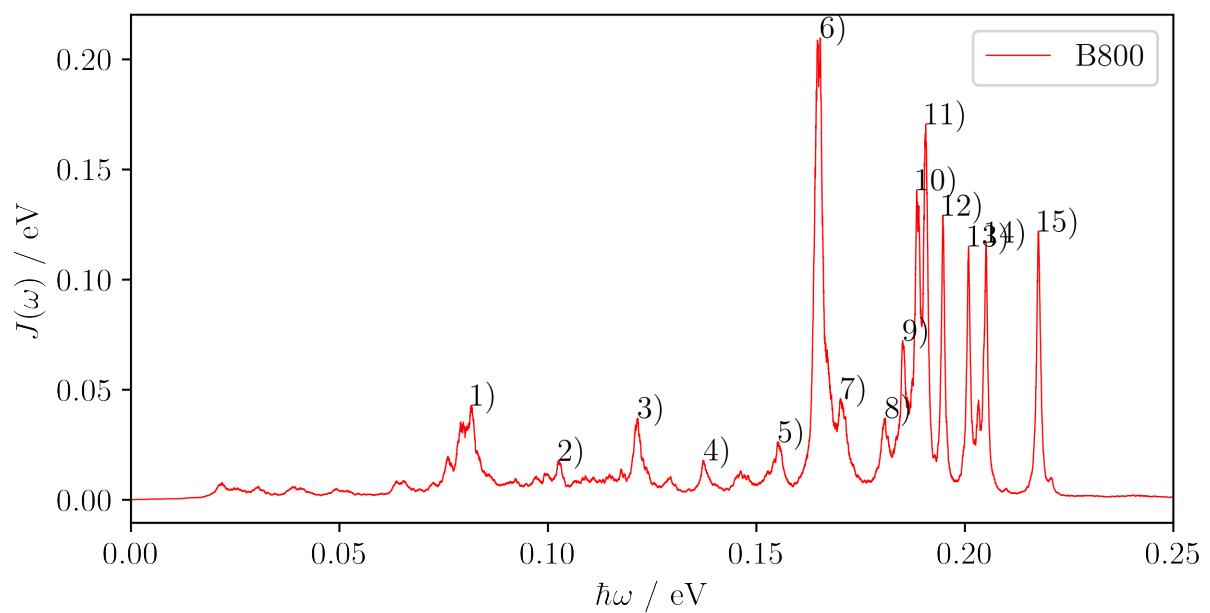
APPENDIX A

This appendix covers the common computational details of this work. Included are the software packages, hardware used. These are not exhaustive list, and additional details are provided in the main chapters. However, wherever implementations or methodology details are missing, the information will be found here.

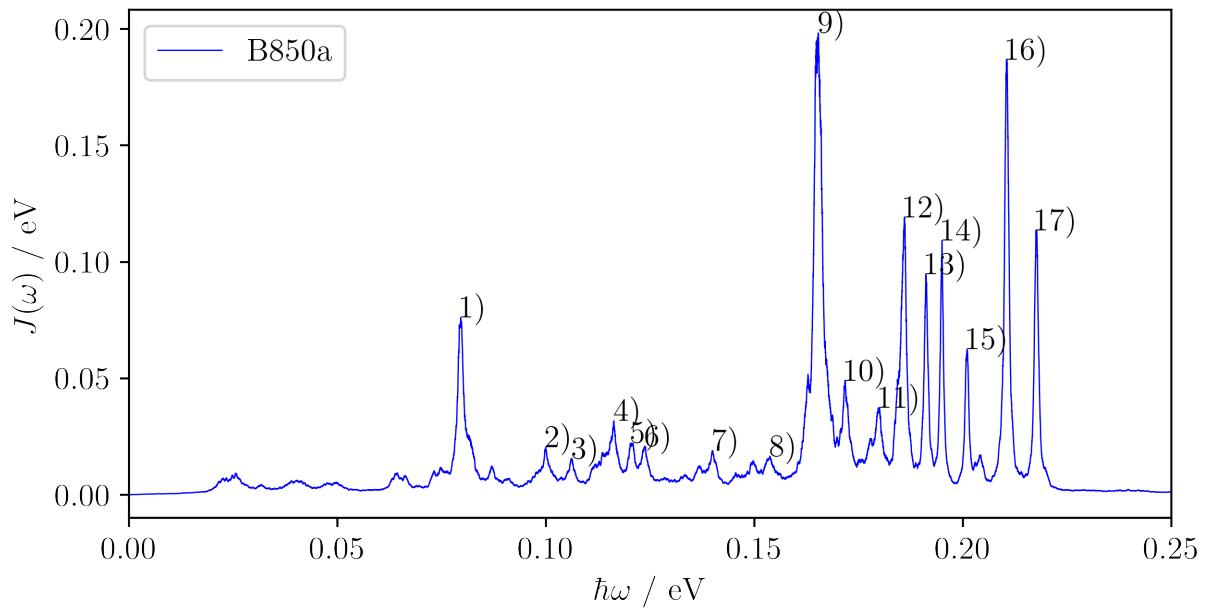
A.1 Electronic Structure Codes

This project has primarily used the QCORE software that is found as part of the ENTOS project. This is a software package for DFT and DFTB electronic structure calculations that has been written as a joint venture between the Miller group in California Institute of Technology and the Manby group in the University of Bristol. It is now being hosted by Entos Inc. It is a novel C++ implementation, with a focus on modularity, functional code and modern development practices to enable easier, cleaner and more reuseable code. All novel methods discussed in the chapters have been implemented in the QCORE package.

A.2 Computational Hardware

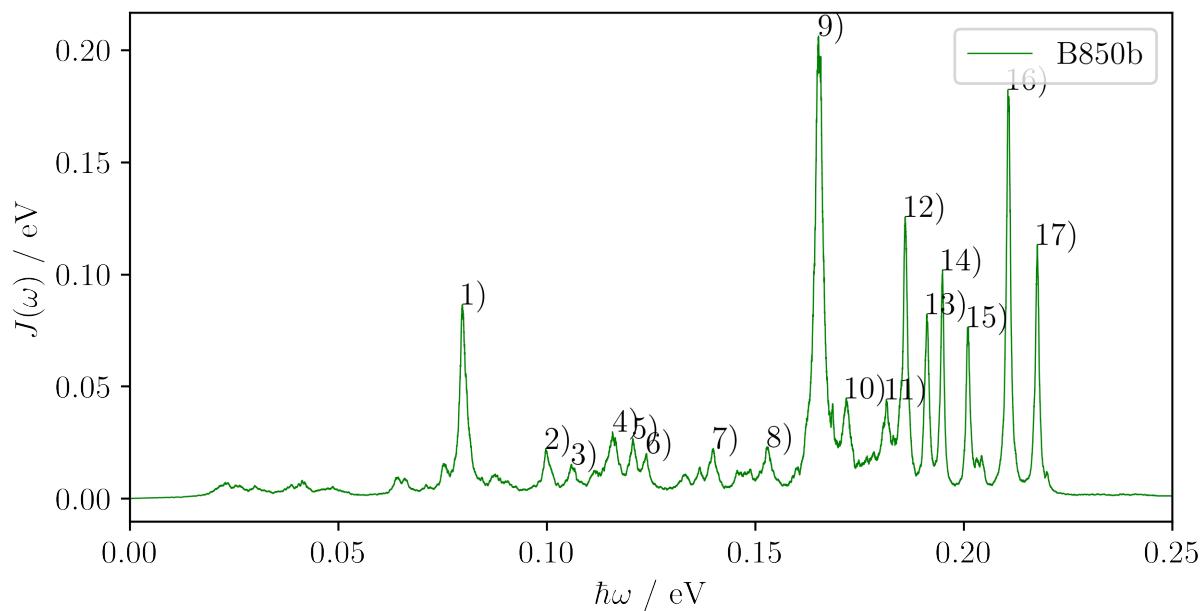
APPENDIX B**B.1 LH2 Data****B.1.1 Spectral Density Features****B.1.1.1 B800 ring sites**

Peak label	$\hbar\omega / \text{eV}$	$J(\omega) / \text{eV}$
0	0.082	0.043
1	0.102	0.018
2	0.122	0.037
3	0.137	0.018
4	0.155	0.026
5	0.165	0.210
6	0.170	0.046
7	0.181	0.037
8	0.185	0.072
9	0.188	0.141
10	0.191	0.171
11	0.195	0.129
12	0.201	0.115
13	0.205	0.117
14	0.218	0.122
15	0.401	0.012



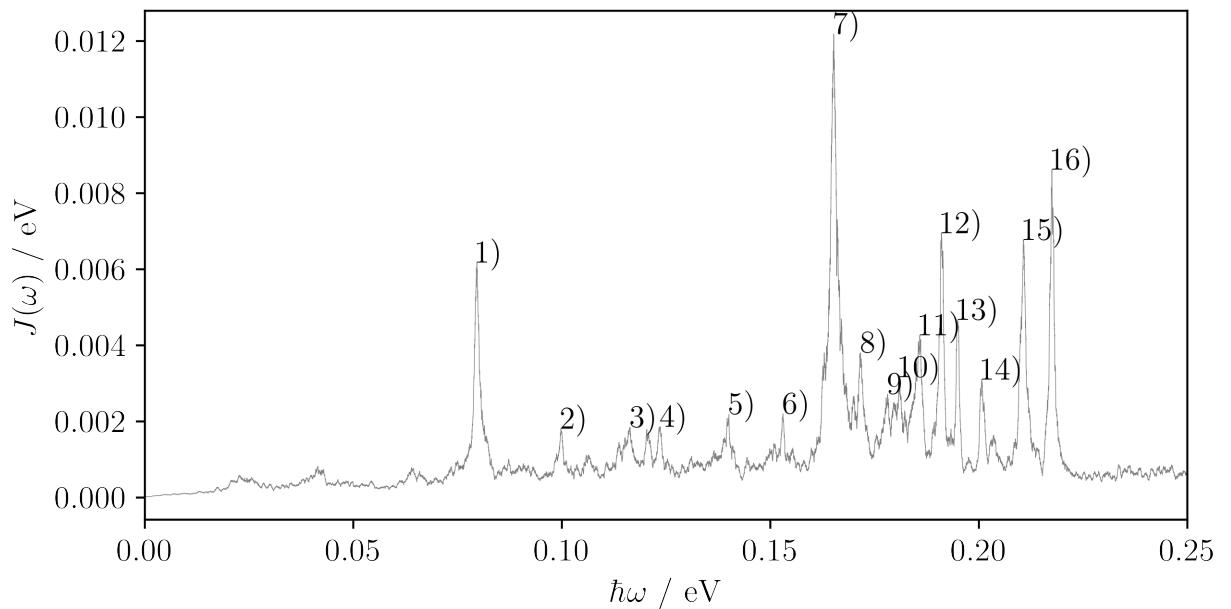
B.1.1.2 B850a ring sites

Peak label	$\hbar\omega$ / eV	$J(\omega)$ / eV
0	0.080	0.076
1	0.100	0.020
2	0.106	0.016
3	0.116	0.032
4	0.120	0.022
5	0.124	0.021
6	0.140	0.019
7	0.154	0.017
8	0.165	0.198
9	0.172	0.049
10	0.180	0.037
11	0.186	0.119
12	0.191	0.095
13	0.195	0.109
14	0.201	0.063
15	0.210	0.187
16	0.218	0.114
17	0.402	0.011



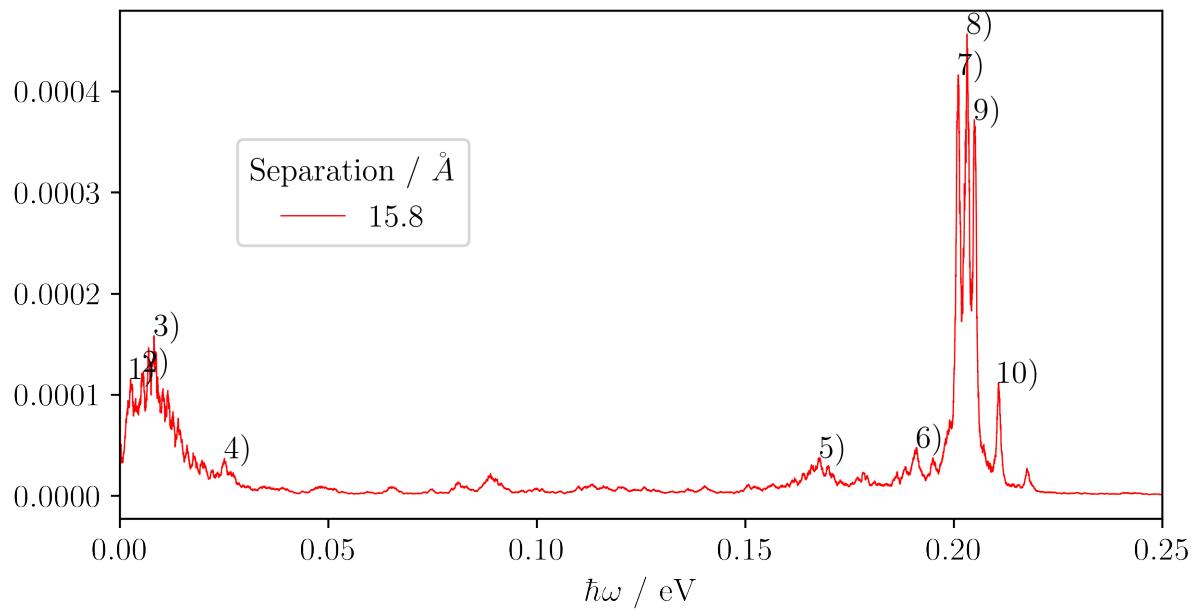
B.1.1.3 B850b ring sites

Peak label	$\hbar\omega / \text{eV}$	$J(\omega) / \text{eV}$
0	0.080	0.086
1	0.100	0.022
2	0.106	0.015
3	0.116	0.030
4	0.121	0.027
5	0.124	0.020
6	0.140	0.022
7	0.153	0.023
8	0.165	0.206
9	0.172	0.045
10	0.181	0.044
11	0.186	0.126
12	0.191	0.082
13	0.195	0.102
14	0.201	0.077
15	0.211	0.182
16	0.218	0.113
17	0.402	0.010



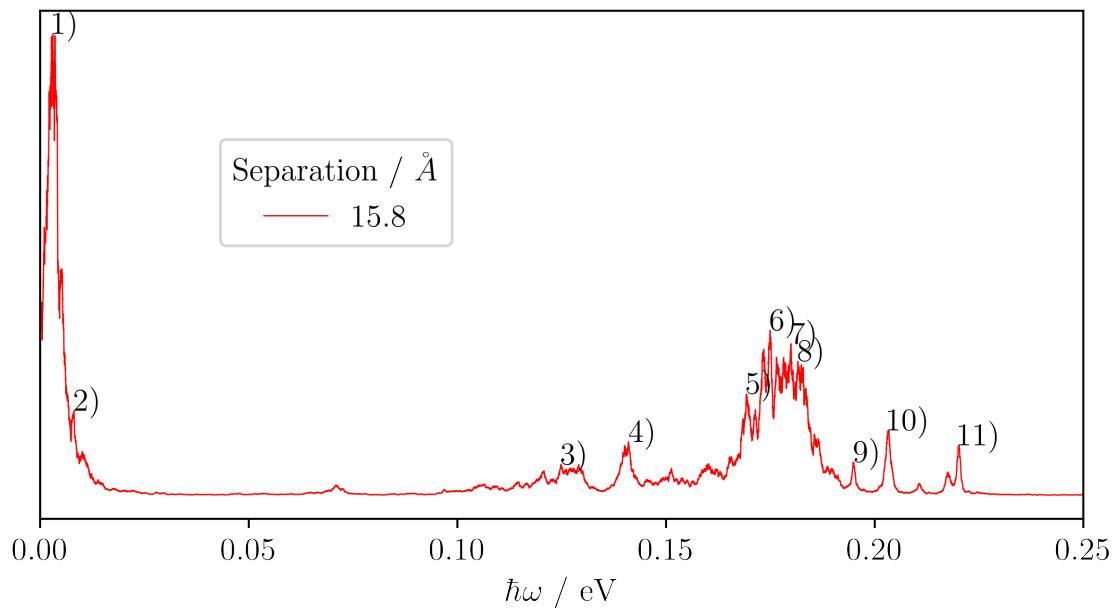
B.1.1.4 Exciton states

Peak label	$\hbar\omega / \text{eV}$	$J(\omega) / \text{meV}$
0	0.080	6.179
1	0.100	1.812
2	0.116	1.850
3	0.124	1.860
4	0.140	2.182
5	0.153	2.201
6	0.165	12.183
7	0.172	3.800
8	0.178	2.712
9	0.181	3.166
10	0.186	4.261
11	0.191	6.949
12	0.195	4.667
13	0.201	3.101
14	0.211	6.775
15	0.218	8.620
16	0.402	1.593



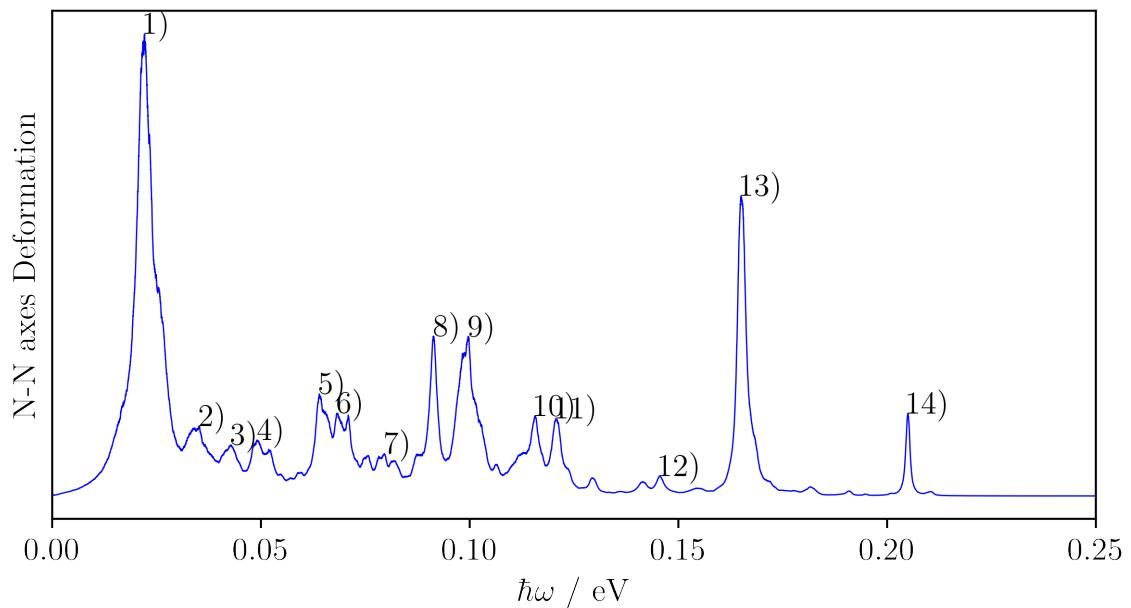
B.1.1.5 Hamiltonian Coupling Values

Peak label	$\hbar\omega / \text{eV}$	$J(\omega) / \text{meV}$
0	0.003	0.116
1	0.006	0.123
2	0.008	0.158
3	0.025	0.037
4	0.168	0.038
5	0.191	0.048
6	0.201	0.416
7	0.203	0.457
8	0.205	0.372
9	0.211	0.112
10	0.406	0.024



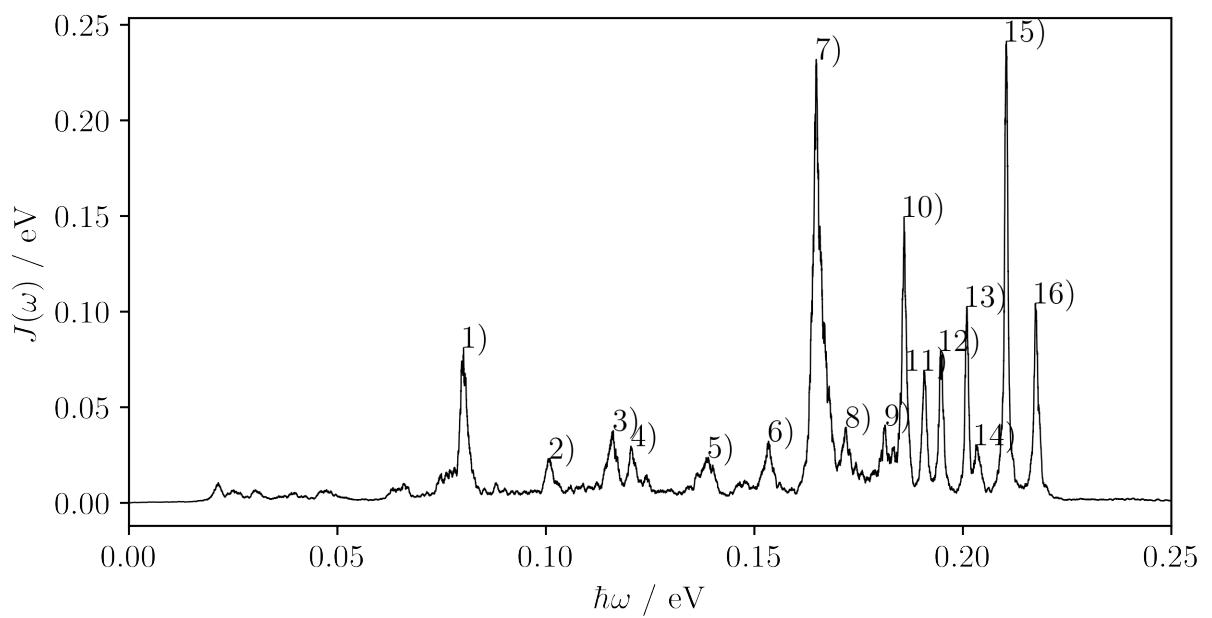
B.1.1.6 Interchromophore distances

Peak label	$\hbar\omega / \text{eV}$	rel. height
0	0.003	1.000
1	0.008	0.184
2	0.125	0.068
3	0.141	0.117
4	0.169	0.221
5	0.175	0.358
6	0.180	0.330
7	0.182	0.291
8	0.195	0.073
9	0.203	0.142
10	0.220	0.110



B.1.1.7 N Axes Deformation

Peak label	$\hbar\omega / \text{eV}$	rel. height
0	0.022	1.000
1	0.035	0.152
2	0.043	0.110
3	0.049	0.121
4	0.064	0.221
5	0.068	0.179
6	0.080	0.092
7	0.091	0.346
8	0.100	0.345
9	0.116	0.174
10	0.121	0.168
11	0.146	0.043
12	0.165	0.649
13	0.205	0.179

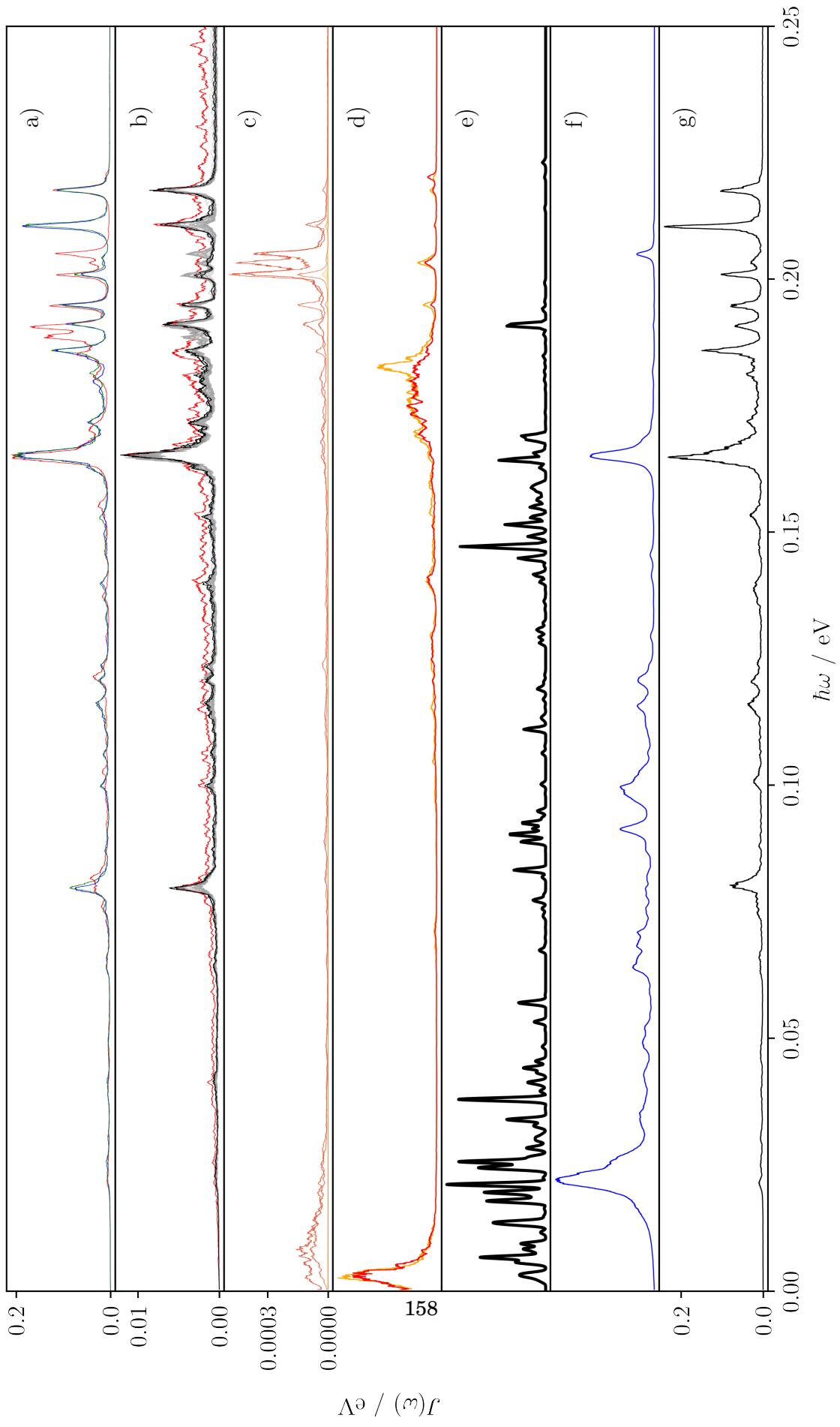


B.1.1.8 Diethyl-Ether system

Peak label	$\hbar\omega$ / eV	$J(\omega)$ / eV
0	0.080	0.081
1	0.101	0.023
2	0.116	0.038
3	0.120	0.030
4	0.139	0.024
5	0.153	0.032
6	0.165	0.232
7	0.172	0.040
8	0.181	0.041
9	0.186	0.150
10	0.191	0.069
11	0.195	0.080
12	0.201	0.103
13	0.203	0.030
14	0.210	0.241
15	0.217	0.104
16	0.402	0.028

B.1.1.9 All Spectra

APPENDIX B. APPENDIX B



BIBLIOGRAPHY

- [1] Maria Mallus, Yashoj Shakya, Jigneshkumar Dahyabhai, and Ulrich Kleinekathofer.
Environmental effects on the dynamics in the light-harvesting complexes LH2 and LH3
based on molecular simulations | Elsevier Enhanced Reader, 2018.
- [2] Robert E Blankenship.
Future Perspectives in Plant Biology Early Evolution of Photosynthesis 1.
Plant Physiol., 154:434–438, 2010.
- [3] Richard J. Cogdell and Tu C. Nguyen-Phan.
Photosynthesis | The Purple Photosynthetic Bacterial Light Harvesting System.
Encycl. Biol. Chem. Third Ed., 2:291–304, jan 2021.
- [4] Steffen Klamt, Hartmut Grammel, Ronny Straube, Robin Ghosh, and Ernst Dieter Gilles.
Modeling the electron transport chain of purple non-sulfur bacteria.
Mol. Syst. Biol., 4(1):156, jan 2008.
- [5] Sergei Tretiak, Chris Middleton, Vladimir Chernyak, and Shaul Mukamel.
Exciton Hamiltonian for the Bacteriochlorophyll System in the LH2 Antenna Complex of
Purple Bacteria.
2000.
- [6] Richard J Cogdell, Andrew Gall, and Jurgen Kohler.
The architecture and function of the light-harvesting apparatus of purple bacteria : from
single molecules to in vivo membranes.
- [7] Liam Cleary, Hang Chen, Chern Chuang, Robert J Silbey, and Jianshu Cao.
Optimal fold symmetry of LH2 rings on a photosynthetic membrane.
Biophys. Comput. Biol. Chem., 110(21):8537–8542, 2013.
- [8] Elad Harel and Gregory S Engel.
Quantum coherence spectroscopy reveals complex dynamics in bacterial light-harvesting
complex 2 (LH2).
109(3), 2012.

BIBLIOGRAPHY

- [9] G Mcdermott, M Prince, A A Freer, A M Hawthornthwaite-Lawless, M Z Papiz, R J Cogdell T Be, and N W Isaacs.
Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria.
Nature, 374:517, 1995.
- [10] Juergen Koepke, Xiche Hu, Cornelia Muenke, Klaus Schulten, and Hartmut Michel.
The crystal structure of the light-harvesting complex II (B800-850) from Rhodospirillum molischianum.
Structure, 4(5):581–597, 1996.
- [11] Vladislav Sla Ma, Lorenzo Cupellini, and Benedetta Mennucci.
Exciton properties and optical spectra of light harvesting complex II from a fully atomistic description †.
Phys. Chem. Chem. Phys, 22:16783, 2020.
- [12] Seogjoo Jang, Eva Rivera, and Daniel Montemayor.
Molecular level design principle behind optimal sizes of photosynthetic LH2 complex: Taming disorder through cooperation of hydrogen bonding and quantum delocalization.
J. Phys. Chem. Lett., 6(6):928–934, 2015.
- [13] Carles Curutchet and Benedetta Mennucci.
Quantum chemical studies of light harvesting.
Chem. Rev., 117(2):294–343, 2017.
- [14] Tihana Mirkovic, Evgeny E Ostroumov, Jessica M Anna, Rienk Van Grondelle, Govindjee, and Gregory D Scholes.
Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms.
2016.
- [15] Johannes Neugebauer.
Photophysical Properties of Natural Light-Harvesting Complexes Studied by Subsystem Density Functional Theory.
2008.
- [16] Vadim Cherezov, Jeffrey Clogston, Miroslav Z. Papiz, and Martin Caffrey.
Room to Move: Crystallizing Membrane Proteins in Swollen Lipidic Mesophases.
J. Mol. Biol., 357(5):1605–1618, apr 2006.
- [17] Benedetta Mennucci, Felipe Cardoso Ramos, Michele Nottoli, and Lorenzo Cupellini.
The molecular mechanisms of light adaption in light-harvesting complexes of purple bacteria revealed by a multiscale modeling.

10:9650, 2019.

- [18] J. Frenkel.
On The Transformation Of Light Into Heat In Solids.
Phys. Rev., 37:17–44, 1931.
- [19] A S Davydov.
Soviet Physics Uspekhi THE THEORY OF MOLECULAR EXCITONS.
Sov. Phys. Usp, 7:393–448, 1964.
- [20] Edoardo Cignoni, Vladislav Slama, Lorenzo Cupellini, and Benedetta Mennucci.
The atomistic modeling of light-harvesting complexes from the physical models to the computational protocol.
J. Chem. Phys., 156:120901, 2022.
- [21] Sandro Jurinovich, Lucas Viani, Carles Curutchet, and Benedetta Mennucci.
Limits and potentials of quantum chemical methods in modelling photosynthetic antennae.
Phys. Chem. Chem. Phys., 17:30783, 2015.
- [22] Carsten Olbrich and Ulrich Kleinekathofer.
Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II.
J. Phys. Chem. B, 114:12427, 2010.
- [23] Carles Curutchet, Jacob Kongsted, Aurora Mu~ Noz-Losa, Hoda Hossein-Nejad, Gregory D Scholes, and Benedetta Mennucci.
Photosynthetic Light-Harvesting Is Tuned by the Heterogeneous Polarizable Environment of the Protein.
J. Am. Chem. Soc, 133:58, 2011.
- [24] Carles Curutchet, Vladimir I Novoderezhkin, Jacob Kongsted, Aurora Mun, Rienk van Grondelle, Gregory D Scholes, Benedetta Mennucci, and Joan Xxiii.
Energy Flow in the Cryptophyte PE545 Antenna Is Directed by Bilin Pigment Conformation.
2012.
- [25] Oliviero Andreussi, Stefan Knecht, Christel M Marian, Jacob Kongsted, and Benedetta Mennucci.
Carotenoids and Light-Harvesting: From DFT/MRCI to the Tamm-Dancoff Approximation.
2015.
- [26] Thorsten Hansen and Luca De Vico.
Q y and Q x Absorption Bands for Bacteriochlorophyll a Molecules from LH2 and LH3.

BIBLIOGRAPHY

- 2019.
- [27] Vladimir V. Poddubnyy, Maxim I. Kozlov, and Ilya O. Glebov.
The origin of the red shift of Qy band of chlorophylls d and f.
Chem. Phys. Lett., 778, sep 2021.
- [28] Christoph Bannwarth, Caldeweyher Eike, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme.
Extended tight-binding quantum chemistry methods.
WIREs Comput. Mol. Sci., 11(2):1493, 2020.
- [29] Stefan Grimme and Christoph Bannwarth.
Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB).
J. Chem. Phys., 145:54103, 2016.
- [30] Aleksandr V Marenich, Steven V Jerome, Christopher J Cramer, and Donald G Truhlar.
Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases.
2012.
- [31] Johann Gasteiger and Mario Marsili.
A new model for calculating atomic charges in molecules.
Tetrahedron Lett., 19(34):3181–3184, 1978.
- [32] Stefan Grimme.
A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules.
J. Chem. Phys., 138:244104, 2013.
- [33] K. Nishimoto and N. Mataga.
Electronic Structure and Spectra of Some Nitrogen Heterocycles.
Zeitschrift fur Phys. Chemie Neue Folge, 12:335–338, 1957.
- [34] Kimio Ohno.
Some Remarks on the Pariser-Parr-Pople Method.
Theor. chim. Acta, 2:219–227, 1964.
- [35] G Klopman, Vol 86, J Linevsky, K S Seshadri, and D White.
A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules.
J. Am. Chem. Soc., 86(21):4550–4557, 1964.

- [36] Jakob Seibert, Jana Pisarek, Sarah Schmitz, Christoph Bannwarth, Stefan Grimme, and Dieter Cremer Memorial.
Molecular Physics An International Journal at the Interface Between Chemistry and Physics Extension of the element parameter set for ultra-fast excitation spectra calculation (sTDA-xTB) Extension of the element parameter set for ultra-fast excitation spec.
Mol. Phys., 117(9):1104–1116, 2019.
- [37] Liam Wilbraham, Enrico Berardo, Lukas Turcani, Kim E Jelfs, and Martijn A Zwijnenburg. High-Throughput Screening Approach for the Optoelectronic Properties of Conjugated Polymers.
J. Chem. Inf. Model., 58:16, 2018.
- [38] Shomik Verma, Miguel Rivera, and David O Scanlon. Machine learned calibrations to high- throughput molecular excited state calculations.
J. Chem. Phys., 156:134116, 2022.
- [39] Isabelle Heath-Apostolopoulos, Liam Wilbraham, and Martijn A Zwijnenburg. Computational high-throughput screening of polymeric photocatalysts: exploring the effect of composition, sequence isomerism and conformational degrees of freedom.
Faraday Discuss., 215:98, 2019.
- [40] Abhishek Sirohiwal, Romain Berraud-Pache, Frank Neese, Robert Izsak, and Dimitrios A Pantazis. Accurate Computation of the Absorption Spectrum of Chlorophyll a with Pair Natural Orbital Coupled Cluster Methods.
J. Phys. Chem., 124:8761–8771, 2020.
- [41] Marshall G Cory, Michael C Zerner, Xiche Hu, and Klaus Schulten. Electronic Excitations in Aggregates of Bacteriochlorophylls.
J. Phys. Chem. B, 102:7640–7650, 1998.
- [42] Xiche Hu, Thorsten Ritz, Ana Damjanovic', Damjanovic',, and Klaus Schulten. Pigment Organization and Transfer of Electronic Excitation in the Photosynthetic Unit of Purple Bacteria.
J. Phys. Chem. B, 101:3854–3871, 1997.
- [43] Clement Stross, Marc W. Van Der Kamp, Thomas A.A. Oliver, Jeremy N. Harvey, Noah Linden, and Frederick R. Manby. How static disorder mimics decoherence in anisotropy pump-probe experiments on purple-bacteria light harvesting complexes.
J. Phys. Chem. B, 120(44):11449–11463, 2016.

BIBLIOGRAPHY

- [44] Miroslav Z. Papiz, Steve M. Prince, Tina Howard, Richard J. Cogdell, and Neil W. Isaacs. The structure and thermal motion of the B800-850 LH2 complex from *Rps. acidophila* at 2.0 Å resolution and 100 K: New structural features and functionally relevant motions. *J. Mol. Biol.*, 326(5):1523–1538, 2003.
- [45] Pavlo O Dral. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.*, 11:2336–2347, 2020.
- [46] Jorg Behler. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.*, 13:17930–17955, 2011.
- [47] Julia Westermayr and Philipp Marquetand. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.*, 121:9873–9926, 2020.
- [48] K T Schütt, M Gastegger, A Tkatchenko, K.-R Müller, and R J Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.*, 10:5024, 2019.
- [49] Manas Sajjan, Junxu Li, Raja Selvarajan, Bc Shree, Hari Sureshbabu, Bd Sumit, Suresh Kale, Rishabh Gupta, Vinit Singh, and Sabre Kais. Quantum machine learning for chemistry and physics. *Chem. Soc. Rev.*, 51:6475, 2022.
- [50] Florian Häse, Stephanie Valleau, Edward Pyzer-Knapp, and Alan Aspuru-Guzik. Machine learning exciton dynamics. *Chem. Sci.*, 7:5139–5147, 2016.
- [51] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.*, 108:058301, 2012.
- [52] Gregoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvao Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Muller, and O Anatole Von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.*, 15:095003, 2013.
- [53] Shampa Raghunathan and U. Deva Priyakumar. Molecular representations for machine learning applications in chemistry. *Int. J. Quantum Chem.*, 122(7), apr 2022.

- [54] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams.
Learning internal representations by error propagation.
1986.
- [55] Ardavan Farahvash, Chee-Kong Lee, Qiming Sun, Liang Shi, and Adam P Willard.
Machine learning Frenkel Hamiltonian parameters to accelerate simulations of exciton dynamics.
J. Chem. Phys., 153:74111, 2020.
- [56] Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
Springer, New York, 2nd edition, 2009.
- [57] Carl Rasmussen and Christopher Williams.
Gaussian Processes for Machine Learning.
MIT Press, Cambridge, 2006.
- [58] Aaron Sisto, Clem Stross, Marc W. Van Der Kamp, Michael O'Connor, Simon McIntosh-Smith, Graham T. Johnson, Edward G. Hohenstein, Fred R. Manby, David R. Glowacki, and Todd J. Martinez.
Atomistic non-adiabatic dynamics of the LH₂ complex with a GPU-accelerated: Ab initio exciton model.
Phys. Chem. Chem. Phys., 19(23):14924–14936, 2017.
- [59] D Porezag, Th Frauenheim, Th Kohler, 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):12948–12957, 1994.
- [60] Pekka Koskinen and Ville Mäkinen.
Density-functional tight-binding for beginners.
Comput. Mater. Sci., 47(1):237–253, 2009.
- [61] Michael Gaus, Qiang Cui, and Marcus Elstner.
DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB).
J. Chem. Theory Comput., 7:931–948, 2011.
- [62] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker.
An overview of the Amber biomolecular simulation package.
Wiley Interdiscip. Rev. Comput. Mol. Sci., 3(2):198–210, mar 2013.

BIBLIOGRAPHY

- [63] Madushanka Manathunga, Yipu Miao, Dawei Mu, Andreas W Go, and Kenneth M Merz. Parallel Implementation of Density Functional Theory Methods in the Quantum Interaction Computational Kernel Program.
Cite This J. Chem. Theory Comput., 16:4315–4326, 2020.
- [64] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions.
J. Chem. Theory Comput., 15:1652–1671, 2019.
- [65] Stefan Grimme, Christoph Bannwarth, and Philip Shushkov. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1\text{--}86$).
J. Chem. Theory Comput., 13(5):1989–2009, 2017.
- [66] Philipp Pracht, Eike Caldeweyher, Sebastian Ehlert, and Stefan Grimme. A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules.
Chem. Eng. Sci., 17(preprint):1–19, 2019.
- [67] Sebastian Spicher and Stefan Grimme. Robust atomistic modeling of materials, organometallic and biochemical systems.
Angew. Chemie, 59:2–11, 2020.
- [68] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu.
J. Chem. Phys., 132:24103, 2010.
- [69] Eike Caldeweyher, Jan-Michael Mewes, Sebastian Ehlert, and Stefan Grimme. Extension and evaluation of the D4 London-dispersion model for periodic systems.
Phys. Chem. Chem. Phys., 22:8499, 2020.
- [70] Erich Runge and E K U Gross. Density-Functional Theory for Time-Dependent Systems.
Phys. Rev. Lett., 52:997–1000, 1984.
- [71] M A L Marques and E K U Gross. TIME-DEPENDENT DENSITY FUNCTIONAL THEORY.
Annu. Rev. Phys. Chem., 55:427–55, 2004.

- [72] P. Hohenberg and W. Kohn.
PHYSICAL REVIEW.
Phys. Rev., 136(3B):864, 1952.
- [73] Nanna Holmgaard List, Carles Curutchet, Stefan Knecht, Benedetta Mennucci, Jacob Kongsted, Joan Xxiii, and Barcelona Spain.
Toward Reliable Prediction of the Energy Ladder in Multichromophoric Systems: A Benchmark Study on the FMO Light-Harvesting Complex.
J. Chem. Theory Comput., 9:4928, 2013.
- [74] Adèle D. Laurent and Denis Jacquemin.
TD-DFT benchmarks: A review, sep 2013.
- [75] s. William J. Hunt and William A. Goddard.
Excited States of H₂O using improved virtual orbitals.
Chem. Phys. Lett., 3(6):414–418, jun 1969.
- [76] Andrew T B Gilbert, Nicholas A Besley, and Peter M W Gill.
Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM) †.
J. Phys. Chem., 112:13164, 2008.
- [77] C C J Roothaan.
New Developments in Molecular Orbital Theory.
Rev. Mod. Phys., 23(2):69, 1951.
- [78] Tiqing Liu, Wen-Ge Han, Fahmi Himo, I G Matthias Ullmann, Donald Bashford, Alexei Toutchkine, Klaus M Hahn, and Louis Noodleman.
Density Functional Vertical Self-Consistent Reaction Field Theory for Solvatochromism Studies of Solvent-Sensitive Dyes.
J. Phys. Chem. A, 108:3545–3555, 2004.
- [79] Jeppe Gavnholt, Thomas Olsen, Mads Engelund, and Jakob Schiøtz.
self-consistent field method to obtain potential energy surfaces of excited molecules on surfaces.
Phys. Rev. B, 78:075441, 2008.
- [80] Nicholas A Besley, Andrew T B Gilbert, and Peter M W Gill.
Self-consistent-field calculations of core excited states.
J. Chem. Phys., 130:124308, 2009.
- [81] Tim Kowalczyk, Shane R Yost, and Troy Van Voorhis.

BIBLIOGRAPHY

- Assessment of the D-SCF density functional theory approach for electronic excitations in organic dyes.
J. Chem. Phys., 134:164101, 2011.
- [82] Tino Gimon, Andrey Ipatov, Andreas Heßelmann, and Andreas Görling.
Qualitatively Correct Charge-Transfer Excitation Energies in HeH⁺ by Time-Dependent Density-Functional Theory Due to Exact Exchange Kohn-Sham Eigenvalue Differences.
J. Chem. Theory Comput., 5:27, 2009.
- [83] Per Olov Löwdin.
Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction.
Phys. Rev., 97(6):1474–1489, 1955.
- [84] V M Agranovich and D M Basko.
Molecular polaritons for controlling chemistry with quantum optics.
Mol. Orbital Heitler-London Methods J. Chem. Phys., 112:11, 2000.
- [85] Gregory D Scholes.
LONG-RANGE RESONANCE ENERGY TRANSFER IN MOLECULAR SYSTEMS.
Annu. Rev. Phys. Chem., 54:57–87, 2003.
- [86] Casper Steinmann and Jacob Kongsted.
Electronic Energy Transfer in Polarizable Heterogeneous Environments: A Systematic Investigation of Different Quantum Chemical Approaches.
J. Chem. Theory Comput., 11:4283–4293, 2015.
- [87] Xin Li, Robert M. Parrish, Fang Liu, Sara I.L. Kokkila Schumacher, and Todd J. Martínez.
An Ab Initio Exciton Model Including Charge-Transfer Excited States.
J. Chem. Theory Comput., 13(8):3493–3504, 2017.
- [88] Susannah Bourne Worster, Oliver Feighan, and Frederick R Manby.
Reliable transition properties from excited-state mean-field calculations.
J. Chem. Phys., 154:124106, 2021.
- [89] D. J Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch. Gaussian 16, Revision C.01, 2016.
- [90] Tom Ziegler, Arvi Rank, and Evert J Baerends.
On the Calculation of Multiplet Energies by the Hartree-Fock-Slater Method.

- Acta*, 43:261–271, 1977.
- [91] Carlo Adamo and Vincenzo Barone.
Towards reliable density functional methods without adjustable parameters: The PBE0 model.
J. Chem. Phys., 110:6158, 1999.
- [92] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs.
Fully optimized contracted Gaussian basis sets for atoms Li to Kr.
J. Chem. Phys., 97:2571, 1992.
- [93] Christof Hättig and Florian Weigend.
CC2 excitation energy calculations on large molecules using the resolution of the identity approximation.
J. Chem. Phys., 113:5154, 2000.
- [94] Arnim Hellweg, Sarah A Gruen Gruen, and Christof Hättig Hättig.
Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states.
Phys. Chem. Chem. Phys., 10:4119–4127, 2008.
- [95] Stefan Grimme and Christoph Bannwarth.
Supporting information for: Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB).
Technical report.
- [96] mcodev31/libmsym: molecular point group symmetry lib.
- [97] Yeongsu Cho, Sylvia J Bintrim, and Timothy C Berkelbach.
A simplified GW/BSE approach for charged and neutral excitation energies of large molecules and nanomaterials.
2021.
- [98] JunWoo Kim, Jonggu Jeon, Tai Hyun Yoon, and Minhaeng Cho.
Two-dimensional electronic spectroscopy of bacteriochlorophyll a with synchronized dual mode-locked lasers.
Nat. Commun., 11:6029, 2020.
- [99] Lu Zhang, Daniel-Adriano Silva, Yijing Yan, and Xuhui Huang.
Force Field Development for Cofactors in the Photosystem II.
J. Comput. Chem., 33:1969–1980, 2012.
- [100] H. Strain, M. Thomas, and J. Katz.

BIBLIOGRAPHY

Spectral Absorption Properties of Ordinary and Fully Deuteriated Chlorophylls a and b.
Biochem. Biophys. Acta., 75:306–311, 1963.