

---

---

# Efficient Protocols for Excited States of Large Biochromophores

---

---

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

The challenge with modelling the excited states of light harvesting complexes lies in scaling accurate approaches for large system sizes. Many common approaches to excited states are not viable on the scale of thousands of atoms, creating the need for more efficient methods. These methods come with compromises either in accuracy, extendability or simplicity. In spite of this, some models have found great success in being able to predict and explain the nuanced mechanisms of light harvesting complexes.

This work explores alternatives to existing methods that overcome some of the compromises. The design of these new methods requires an understanding of the key components of existing methods, which is achieved by comparing transition properties to high-level data from established methods. By introducing a new set of approximations to the standard response method framework and re-optimizing key parameters in the recently developed semi-empirical GFN-xTB method, a new approach is derived, referred to as Chl-xTB, that efficiently and accurately describes the  $Q_y$  transition in chlorophyll, the most important transition to model for light harvesting complexes.

The accuracy of Chl-xTB transition properties for monomer, dimer and oligomer chlorophyll systems is established with benchmarking and comparison to existing literature, using a Frenkel-Davydov exciton framework for systems beyond chlorophyll monomers. Chl-xTB is found to perform very well in these cases, giving confidence that it can be reliably used to predict light harvesting complex properties.

The successful application of the Chl-xTB method is showcased by calculating properties of the LH2 complex that are prohibitively expensive using existing techniques, such as electron transfer rates between charge-separated and vertical-excitation states, as well as spectral densities of properties that encompass the entire exciton network. These properties shed light on the role of the LH2 protein scaffold and its manipulation of the exciton system. For example it can be explicitly shown how the LH2 scaffold quells transition between vertical excitation to charge transfer states in chlorophyll dimers, a key mechanism in concentration quenching. Additionally, the spectral densities of exciton states show how environmental coupling to off-diagonal elements of the exciton Hamiltonian have little effect on the overall coupling to exciton states. The calculation of these novel properties and discussion of further applications of both the novel efficient response framework and the specific Chl-xTB method demonstrate a possible solution to the issues of efficiently modelling light harvesting complexes.



## DEDICATION AND ACKNOWLEDGEMENTS

In memory of Peter Feighan, who is missed every day.

This thesis would not be possible with the help and support of a great number of people, to whom I am very thankful.

Foremost are my supervisors, Prof. Fred Manby and Prof. Andrew Orr-Ewing, whose helpful advice on all elements of my project, from scientific challenges to practical skills, has been invaluable and without which this project would not exist or have been finished.

Equally I am very grateful to Dr. Susannah Bourne-Worster, who has provided so much time and effort in helping to realizing the full potential of these projects and others in the course of my PhD. I am deeply grateful for her support with all levels of this work.

I am indebted to the members of the Manby research group who taught me the technical skills necessary to carry out this work, especially Dr. Peter Bygrave, Dr. James Womack and Dr. Thomas Dresselhaus. I have greatly enjoyed the supportive and friendly atmosphere of this group, and the University of Bristol's Centre for Computational Chemistry in general - this environment has provided many stimulating conversations and I have found amiable encouragement through tough periods, especially with the support from Harry Stroud and Joe Crossley-Lewis.

Finally I would like to thank my family and friends, whose presence has been inspirational and set the foundations that led me to this work.



## 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
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Abbreviations</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Efficient Excited State Methods . . . . .	5
1.1.1 sTDA-xTB . . . . .	5
1.2 Statistical Methods . . . . .	7
1.2.1 Machine-Learning For Exciton Models . . . . .	8
1.3 GPU Acceleration . . . . .	12
1.4 Aims . . . . .	13
1.5 Outline . . . . .	13
<b>2 Background Theory</b>	<b>15</b>
2.1 Electronic structure . . . . .	15
2.1.1 Density Functional Theory . . . . .	15
2.1.2 Density Functional Tight Binding . . . . .	18
2.1.3 Extended Tight Binding . . . . .	22
2.2 Excited State Methods . . . . .	27
2.2.1 Linear Response TD-DFT . . . . .	27
2.2.2 Eigenvalue Difference . . . . .	30
2.2.3 $\Delta$ -SCF . . . . .	30
2.3 Excitation Energy Transfer and Frenkel-Davydov Exciton Hamiltonians . . . . .	33
<b>3 Mean-Field excited states</b>	<b>37</b>
3.1 Benchmarking $\Delta$ -SCF . . . . .	38
3.1.1 Reference Data and Test Sets . . . . .	38
3.1.2 Benchmarking Small Systems . . . . .	39
3.1.3 Addressing Non-orthogonality in $\Delta$ -SCF States . . . . .	40

---

**TABLE OF CONTENTS**

---

3.1.4	Δ-SCF Accuracy in LH2 Chlorophylls . . . . .	43
3.1.5	Benchmarking xTB Δ-SCF Methods . . . . .	46
3.2	Conclusions . . . . .	51
<b>4</b>	<b>Bespoke Chlorophyll Excited State Methods</b>	<b>53</b>
4.1	Response and Electronic Structure Theory Approximations . . . . .	54
4.1.1	Full TD-DFT Solutions . . . . .	54
4.1.2	Approximations to Full TD-DFT Solutions . . . . .	54
4.1.3	Integral Approximations with Monopole Expansions . . . . .	55
4.1.4	The $Q_y$ Transition . . . . .	57
4.1.5	Bespoke xTB Parameters for Chlorophyll . . . . .	59
4.1.6	Transition and Excited State Density from the Ground State . . . . .	61
4.2	Parameterization . . . . .	63
4.2.1	Objective Function . . . . .	63
4.2.2	Minimisation Algorithms . . . . .	64
4.2.3	Reference Data . . . . .	67
4.2.4	Results . . . . .	71
4.3	Cross-validation . . . . .	76
4.3.1	Vibrational Mode Coupling . . . . .	76
4.3.2	Predicting Absorption Spectra . . . . .	84
4.4	Conclusions . . . . .	85
<b>5</b>	<b>Exciton Models of Chlorophyll Dimers</b>	<b>89</b>
5.1	Frenkel Exciton Hamiltonian . . . . .	90
5.1.1	Exciton States . . . . .	90
5.2	Benchmarking the exciton model against full dimer excited states . . . . .	92
5.2.1	Exciton accuracy in LH2 Chlorophyll Dimers . . . . .	92
5.2.2	Systematic Rotation of Chlorophyll Dimer Orientations . . . . .	94
5.3	chl-xTB Excitons . . . . .	97
5.3.1	Benchmarking the Chl-xTB Exciton Framework . . . . .	97
5.4	Prediction of the Rate of Electron Transfer in Concentration Quenching . . . . .	106
5.5	Conclusions . . . . .	110
<b>6</b>	<b>Light Harvesting Complexes</b>	<b>113</b>
6.1	Calculating LH2 Excitons . . . . .	115
6.1.1	Molecular Dynamics . . . . .	115
6.1.2	Scaling computational performance . . . . .	115
6.1.3	Coupling Values . . . . .	116
6.1.4	Screening and Embedding . . . . .	116

---

TABLE OF CONTENTS

6.2	Long Timescale Spectral Densities of the LH2 Protein . . . . .	121
6.2.1	Spectral Densities of $Q_y$ Transition Energies at Individual Sites . . . . .	121
6.2.2	Spectral Densities of Exciton State Energy . . . . .	124
6.2.3	Spectral Densities of Exciton State Coupling . . . . .	127
6.3	Assigning Specific Motions . . . . .	130
6.3.1	Spectral Density of the $Q_y$ Transition from Chlorophyll in Diethyl Ether .	130
6.3.2	Assignment of Molecular Motion with Huang-Rhys Factors . . . . .	132
6.3.3	Assignment of Molecular Motion with Spectral Densities of $N$ Axes Deformation . . . . .	133
6.4	Conclusions . . . . .	137
<b>7</b>	<b>Conclusions</b>	<b>139</b>
7.1	Efficient Excited State Methods . . . . .	140
7.2	Further Investigations into LHCs . . . . .	142
<b>A</b>	<b>Appendix A</b>	<b>145</b>
A.1	Electronic Structure Codes . . . . .	145
A.2	Computational Hardware . . . . .	145
<b>B</b>	<b>Appendix B</b>	<b>147</b>
B.1	LH2 Data . . . . .	147
B.1.1	Spectral Density Features . . . . .	147
<b>Bibliography</b>		<b>161</b>



## LIST OF TABLES

<b>TABLE</b>	<b>Page</b>
3.1 Means and standard deviations of the errors between SCS-CC2 reference data and response methods for the whole test set in eV. The xtb4stda entry represents the eigenvalue difference method that uses the eigenvalues output from this program. . . . .	49
4.1 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 CAM-B3LYP/Def2-SVP TD-DFT data, for transition energies ( $\Delta E$ ) and transition dipole moments ( $ \mu $ ). . . . .	69
4.2 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. . . . .	74
B.1 Peak positions and heights for the spectral density of $Q_y$ transition at LH2 B800 sites.	148
B.2 Peak positions and heights for the spectral density of the $Q_y$ transition at LH2 B850a sites. . . . .	150
B.3 Peak positions and heights for the spectral density of $Q_y$ transitions at LH2 B850b sites.	151
B.4 Peak positions and heights for the spectral density of LH2 exciton transition energies.	153
B.5 Peak positions and heights for the spectral density of LH2 exciton couplings. . . . .	153
B.6 Peak positions and heights for the spectral density of LH2 Mg-Mg distances. . . . .	154
B.7 Peak positions and heights for the spectral density of LH2 N axes deformations. . . . .	155
B.8 Peak positions and heights for the spectral density of $Q_y$ transition of chlorophyll in explicit diethyl ether solvent. . . . .	157



## LIST OF FIGURES

<b>FIGURE</b>	<b>Page</b>
1.1 The trimer chlorophyll unit found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800) . . . . .	2
1.2 The complete chlorophyll system 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. . . . .	3
1.3 A schematic of a neural network showing how inputs, such as the (flattened) Coulomb matrix, are ordered into nodes (green and yellow) and can be used to generate outputs (red nodes). The blue nodes between these layers constitute the "hidden layers". Each arrow leading to a node represents a matrix multiplication and activation function application. The values in the output layer for models discussed in the text would be $Q_y$ transition energies or atom centered transition charges. . . . .	9
3.1 Transition energies $\Delta E$ from TD-DFT (black) and $\Delta$ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled. . . . .	40
3.2 Transition dipole magnitudes from TD-DFT (black) and $\Delta$ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference. . . . .	41
3.3 The absolute value of the error in transition dipole magnitude between $\Delta$ -SCF and EOM-CCSD, plotted against the $\Delta$ -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. . . . .	42
3.4 Transition energies from $\Delta$ -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. . . . .	43
3.5 Transition dipole magnitudes from $\Delta$ -SCF for the same 26 chlorophyll geometries as 3.4, 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. . . . .	44

## LIST OF FIGURES

---

3.6	A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the $T_d$ point group. . . . .	48
3.7	The distributions of errors compared to SCS-CC2 transition energies for the methods included in the $\Delta$ -xTB benchmarking. . . . .	50
4.1	Bacterial Chlorophyll a (BChla) with a model $Q_y$ transition dipole. . . . .	58
4.2	The HOMO orbital of Bchla from PBE0/Def2-SVP DFT. . . . .	59
4.3	The LUMO orbital of BChla from PBE0/Def2-SVP DFT. . . . .	60
4.4	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. . . . .	67
4.5	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. . . . .	68
4.6	Comparison of the $Q_y$ transition energies predicted from chl-xTB against PBE0/Def2-SVP values. . . . .	71
4.7	Comparison of the $Q_y$ transition dipole moments predicted from the reference methods as well as chl-xTB (red) against CAM-B3LYP/Def2-SVP values. . . . .	72
4.8	Comparison of the $Q_y$ transition dipole moments predicted from the reference methods as well as chl-xTB (red) against CAM-B3LYP/Def2-SVP values. . . . .	73
4.9	Distributions of the compute wall-times between TD-DFT with CAM-B3LYP/Def2-SVP and PBE0/Def2-SVP levels of theory, performed on 20 2.4 GHz Intel E5-2680 v4 CPUs (labelled <i>HPC</i> ). Wall-times for sTDA-xTB and Chl-xTB, performed on a desktop 2019 2.3 GHz Intel Core i5 Macbook Pro, are shown with the label <i>Desktop</i> . . . . .	74
4.10	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. . . . .	76
4.11	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. . . . .	77
4.12	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 smaller variance than in figure 4.11 led this metric to not be used in normal mode choices. . . . .	77
4.13	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 83rd normal mode (frequency $669.6\text{ cm}^{-1}$ ). . . . .	78
4.14	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 85th normal mode (frequency $701.7\text{ cm}^{-1}$ ). . . . .	79
4.15	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 88th normal mode (frequency $733.3\text{ cm}^{-1}$ ). . . . .	79
4.16	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 90th normal mode (frequency $745.5\text{ cm}^{-1}$ ). . . . .	80

---

4.17	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 91st normal mode (frequency $755.1\text{ cm}^{-1}$ ) . . . . .	80
4.18	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 129th normal mode (frequency $1105.0\text{ cm}^{-1}$ ) . . . . .	81
4.19	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 130th normal mode (frequency $1107.0\text{ cm}^{-1}$ ) . . . . .	81
4.20	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 132nd normal mode (frequency $1122.2\text{ cm}^{-1}$ ) . . . . .	82
4.21	Transition energies and dipole magnitudes for the $Q_y$ transition for geometries of truncated chlorophyll along the 135th normal mode (frequency $1142.9\text{ cm}^{-1}$ ) . . . . .	82
4.22	Predicted and experimental absorption spectrum of chlorophyll in diethyl ether, only in the $Q$ band region. Predicted spectra are shown without energy shifts on the left, and with an energy shift to match the experimental absorption maximum on the right. All spectra are normalized to have equal areas. . . . .	85
5.1	The correlation of lowest excitations in LH <sub>2</sub> 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). . . . .	93
5.2	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. . . . .	95
5.3	The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the $Q_z$ axis. . . . .	96
5.4	The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the $Q_x$ axis. . . . .	96
5.5	The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the $Q_y$ axis. . . . .	97
5.6	Distributions of the lowest 5 transition energies of chlorophyll dimers predicted by TD-DFT with the CAM-B3LYP functional (grey) and PBE0 functional (red). . . . .	98
5.7	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. . . . .	99
5.8	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. . . . .	101

---

## LIST OF FIGURES

5.9 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. . . . .	102
5.10 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. . . . .	102
5.11 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. . . . .	103
5.12 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. . . . .	104
5.13 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. . . . .	104
5.14 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. . . . .	105
5.15 Transition energies from exciton models constructed with CAM-B3LYP/Def2-SVP, PBE0/Def2-SVP TD-DFT, Chl-xTB and sTDA-xTB monomer properties, plotted against the corresponding transition energies from full dimer TD-DFT at a CAM-B3LYP/Def2-SVP level of theory. . . . .	105
5.16 Distributions of vertical energy gaps $\Delta E$ and fitted normal distribution functions for each solvated Chl dimer system and LH2 Chl dimers. . . . .	108
5.17 Free energy surfaces of the charge separated and photo-excited states of chlorophyll dimer systems, using the difference in state energies as the coordinate, for a series of separations. Free energy changes ( $\Delta A$ ) and reorganization energies ( $\lambda$ ) are shown for each separation. . . . .	109
5.18 Free energy surfaces of the charge separated and photo-excited states of LH2 chlorophyll dimer systems, using the same coordinate as figure 5.17. The red lines show FESs from LH2 chlorophyll dimers, with black lines showing FESs of solvated chlorophyll dimers. The large free energy change and movement of the charge separation state FES minimum to the left-hand side of the photo-excited FES imply the LH2 protein environment inhibits transfer to the charge separated state. . . . .	111

6.1 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 $\text{cm}^{-1}$ ) 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. . . . .	117
6.2 Simulated absorption spectra of LH2, with (dashed line) and without (solid line) a screening factor for point charge interaction. An experimental line reconstructed from Strain <i>et al.</i> [1] is plotted in black. The simulated spectra are shifted to match the position of the 850 nm peak. . . . .	119
6.3 Density contributions of site transitions to exciton states for the vacuum (left) and screened (middle) Hamiltonian, with the difference shown in on the right. . . . .	120
6.4 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> [2].	122
6.5 Average spectral densities for $Q_y$ transitions at sites in the B800, B850a and B850b rings. . . . .	123
6.6 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. . . . .	124
6.7 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. . . . .	125
6.8 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. . . . .	128
6.9 Spectral density of the separation between nearest neighbour sites in LH2. . . . .	129
6.10 Spectral density of the $Q_y$ transition for a chlorophyll in diethyl-ether. . . . .	131
6.11 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. . . . .	134
6.12 Motions of the four central nitrogen atoms in bacterial chlorophyll for the vibrational modes labelled in figure 6.11. . . . .	135
6.13 Spectral density of the ratio $\frac{ N_A-N_C }{ N_B-N_D }$ , averaged over chlorophyll sites in LH2. . . . .	136
B.1 Spectral density of the $Q_y$ transition at LH2 B800 sites with labelled peaks. . . . .	148
B.2 Spectral density of the $Q_y$ transition at LH2 B850a sites with labelled peaks. . . . .	149
B.3 Spectral density of the $Q_y$ transition at LH2 B850b sites with labelled peaks. . . . .	151
B.4 Spectral density of the LH2 exciton transition energies with labelled peaks. . . . .	152
B.5 Spectral density of the LH2 exciton coupling energies with labelled peaks. . . . .	154
B.6 Spectral density of the LH2 Mg-Mg distances with labelled peaks. . . . .	155

## LIST OF FIGURES

---

B.7	Spectral density of the LH2 N axes deformations with labelled peaks. . . . .	156
B.8	Spectral density of the $Q_y$ transition energies for chlorophyll in explicit diethyl ether, with labelled peaks. . . . .	157
B.9	A stacked version of all of the above spectra. . . . .	159

## LIST OF ABBREVIATIONS

**LHC** light harvesting complex

**RC** reaction centre

**LH1/LH2** light harvesting complex 1/2

**EET** excitation energy transfer

**BChla** bacterial chlorophyll  $\alpha$

**DFT** density functional theory

**TB** tight-binding

**DFTB** density functional tight binding

**SCF** self-consistent field

**LDA** local density approximation

**GGA** generalized gradient approximation

**LIST OF FIGURES**

---

**HF** Hartree-Fock

**KS** Kohn-Sham

**MO/AO** molecular orbital/ atomic orbital

**XC** exchange-correlation

**SAD** superposition of atomic densities

**DIIS** direct inversion in iterative subspace

**HOMO** highest occupied molecular orbital

**LUMO** lowest unoccupied molecular orbital

**QM/MM** quantum mechanics/molecular mechanics

**TD-DFT** time-dependent density functional theory

**TD-DFTB** time-dependent density functional tight binding

**MOM** maximum overlap method

**TDA** Tamm-Danoff approximation

---

LIST OF FIGURES

**SPA** single pole approximation

**xTB** extended tight binding

**GFN-xTB** geometries, frequencies and non-covalent xTB

**sTDA-xTB** simplified Tamm-Danoff extended tight binding

**chl-xTB** chlorophyll extended tight binding

**EHT** extended Hückel theory

**EN** electronegativity

**AES** anisotropic electrostatic

**AXC** anisotropic exchange-correlation

**EEQ** electronegativity equilibration

**(m)CN** (modified) coordination number

**RMSE** root mean squared error



## INTRODUCTION

**P**hotosynthesis 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. Photosynthetic organisms formed around 3.5 billion years ago [3]. Some of the most studied photosynthetic organisms include purple bacteria [4–10], which use light harvesting complexes (LHCs) to absorb and stabilize energy from light and transfer this energy to reaction centres (RCs) [11]. Chemical potential energy from the RCs is eventually used to produce adenosine triphosphate (ATP), the transportable energy source required for many biochemical processes. These complexes are extremely efficient, with the light harvesting complex 1 and 2 (LH1, LH2) found in purple bacteria being around 95% efficient at transferring the electronic energy to charge separation [12].

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 *Rhodoblastus acidophilus* is formed of a trimer unit with two bacterial chlorophyll *a* (BChla) chromophores in close proximity, with parallel porphyrin planes, and a third chromophore perpendicular at a greater separation (shown in figure 1.1) [13]. This unit is then repeated to form a circular structure with anywhere between an 8-10 fold symmetry depending on the species of bacterium (see figure 1.2) [2, 6]. The efficiency of LHCs is dependent on many factors including this structure, the conformations of the chromophores, and their effect on nuanced electronic energy transfer mechanisms [14].

Detailed computational study of these structures has been possible since the first reports on high level resolution of crystal structures [15, 16] and has produced a wealth of analysis. 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, including

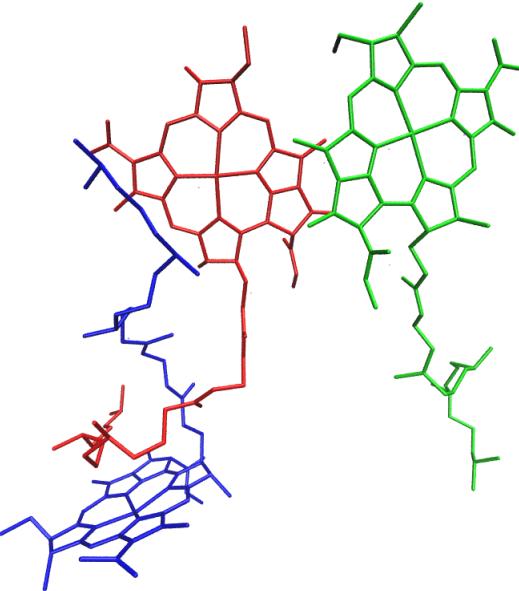


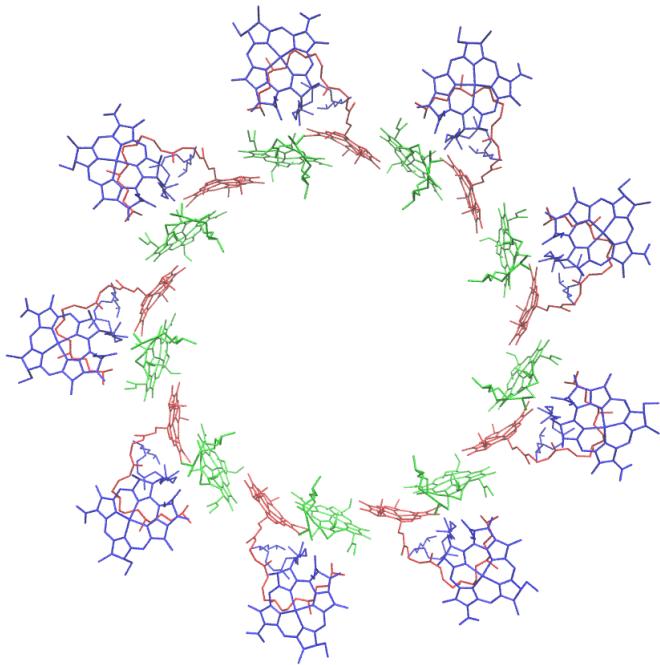
Figure 1.1: The trimer chlorophyll unit found in LH2, coloured by ring type (red and green for B850a and B850b, blue for B800).

insights into LH2 exciton states[17], energy transfer timescales[18, 19], excitation diffusion lengths[20], energetic fine-tuning by local effects such as H-bonding[21, 22] and more [5, 7, 23, 24]. These studies show that the effects of the protein which occur on the atomistic level are an important aspect to capture in LHC models.

Theoretical studies on LHCs can compliment experimental studies by altering factors that are not possible to control in lab-based studies. For example, two-dimensional electronic-vibrational (2DEV) spectra on the LH2 protein (in this case from spinach), produced experimentally [25], can probe the electronic excitation energy transfer between chlorophylls. Whilst many features in the protein scaffold can be proposed as origins for experimentally observed phenomena, such as amino acid side chain carbonyl groups for characteristic  $1670\text{ cm}^{-1}$  vibrational modes [26], complete assignment of all effects is not possible from experiment alone. Atomistic theoretical studies can pinpoint these properties, probing effects such as conformational distortions, hydrogen bonding, coordination state and polarization in site environments [25, 27].

Theoretical discussion of electronic energy transfer requires models for the excited states of LHCs. The ideal model would be a full electronic Hamiltonian of the entire LHC complex and it's environment. However due to the size of these systems it is usually necessary to reduce the number of degrees of freedom [2, 23]. For example in LH2 the chlorophyll molecules alone contain 3780 atoms with an additional 6000 atoms for the entire scaffold [28, 29]. Including a membrane and explicit solvent can quickly lead to system sizes in the region of 300,000 atoms [21]. Treating every electron in this system is not a sensible (or as yet possible) approach.

Evidently a naive approach of using electronic structure methods that describe molecular



**Figure 1.2:** The complete chlorophyll system 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.

excited states (e.g. DFT) is not tenable for LHC models. Instead exciton models are used. Excitons are defined as a bound state of an electron and electron hole, where energy can be transferred without the movement of charge [30, 31]. Characterizing excitons by the interaction between the electron and its hole gives two limits (i.e. a strong or weak interaction). A weak interaction (often in materials with a high dielectric constant) leads to Wannier-Mott excitons [32], common in semiconductors and bulk metallic materials. Conversely a strong interaction (low dielectric constant) give Frenkel excitons [33], common in organic systems (i.e. LHCs), which are more localized due to the strong force between electron and hole. As LHCs have low dielectric constants the Frenkel model is more appropriate. Often Frenkel exciton are modelled in a Davydov formalism, giving a Frenkel-Davydov exciton framework[34]. This framework constructs an exciton Hamiltonian from intra-site energies and inter-site interactions, reducing the degrees of freedom of the Hamiltonian from the entire complex to just single chromophores. A more formal description is given in the next chapter.

The Frenkel exciton Hamiltonian can either be constructed from time-independent parameters, fitted to experimental data or calculated theoretical values, or as functions of LHC geometry

and/or time. A good example of using static parameters to model LH2 is Tretiak's work on calculating LH2 ring energy transfer rates *et al.* [12]. In this work a collective electronic oscillator model was used to calculate chromophore energies and couplings, and subsequently build a Frenkel exciton framework which accurately reproduced B800-B850 and B800-B800 energy transfer rates. This model shows how a minimalist response theory and statical approach can produce physically meaningful results. However, to recover a truly atomistic treatment of LHCs it is necessary to take the functional (and dynamic) approach and treat every chromophore geometry variation, which can be far more expensive to calculate as Hamiltonians are required for every frame in a time series of LHC geometries. This approach also comes with the caveat 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 a time-averaged parameters. The resultant range of accuracy from geometry variations is  $\sim$ 0.2–0.4 eV [35, 36].

Often atomistic LHC models fall into a regular pattern, recently reviewed by Mennucci *et al.*[37]. 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 excited state 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 are highly dependent on the level of detail required for the exciton system (atomistic or coarse grain), and the number 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 [37]. Due to the size of chlorophyll molecules, using high level methods such as SCS-CC2 and EOM-CCSD would not be reasonable. Even with TD-DFT methods only a limited number of explicit Hamiltonians could be constructed before outstripping a reasonably accessible cost. 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 using low-level electronic structure and response calculations. Previous studies have suggested some solutions to this problem and these are discussed below, along with their limitations. These limitations sketch out the potential for a new kind of excited state LHC model, which is the main investigation of this work.

## 1.1 Efficient Excited State Methods

The most obvious solution to this scaling problem is to make the electronic structure and excited state calculations more efficient, often done at the expense of accuracy. Many recent studies use tight binding methods such as TD-DFTB or ZINDO to calculate transition properties [38–41]. 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 (for example Gaussian [42], DFTB+ [43], entos/qcore [44], ORCA [45, 46], and TeraChem [47, 48]). Comparisons 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 to investigate mechanisms in LHCs (for example the role of carotenoids [49], the interaction of excited states in the *Q* band [36] and the role of functional groups on excitation energies in chlorophylls *f*, *d* and *a* [50]) but it is necessary for benchmarking of these methods as not all give reliable results [51]. Particularly for studies on chlorophyll, TD-DFB and ZINDO are popular choices when requiring a large number of calculations [37].

Designing tight-binding models can be challenging due to the balance between accuracy and the number of parameters that need to be fit. This issue has been explored thoroughly by the xTB methods developed by Grimme *et al.* [52]. 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 LHC models, an outline and discussion of this method are given below.

### 1.1.1 sTDA-xTB

sTDA-xTB ("simplified Tann-Danoff Approximation - eXtended Tight Binding") is a method in the family of xTB methods developed by the Grimme and coworkers and is parameterised for transition properties [53]. 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 between 0.34 - 0.48 eV dependent on the benchmark.

Similar to other xTB methods, sTDA-xTB is based on tight-binding electronic structure that uses empirically fitted parameters and a minimal basis set. Discussed in more detail in chapter 2, tight-binding schemes assume small variations in electron density that remove the need to treat core electrons explicitly and often only consider valence electron effects. 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, basis set function coefficients in sTDA-xTB are dependent on coordination number of the atom centres. This makes basis functions far more flexible, which could only be achieved with fixed basis functions 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 approach reduces the cost of having larger basis sets, it makes calculating the gradients 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

$$(1.1) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left( k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left( h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

where  $\mu, \nu, l, l'$  are orbital and shell indices,  $k_\mu^l$  are shell-wise Hückel parameters,  $h$  are effective atomic-orbital energy levels,  $S_{\mu\nu}$  is the overlap of orbitals  $\mu$  and  $\nu$ ,  $k_T$  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) [54] charges for the XBS Fock matrix. These are calculated using Mulliken charges obtained from diagonalizing the Fock matrix with the VBS. The charges for the initial VBS Fock matrix are based on Gasteiger charges [55], modified by the parameterized 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. Diagonalize Fock matrix in the VBS to get the first set of Mulliken charges
3. Compute CM5 charges
4. Diagonalize 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 [56]. 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 (a more formal description is given in section 4.1.2.1). The second approximation is to use a Mataga-Nishimoto-Ohno-Klopman (MNOK) method to calculate two electron integrals instead of explicitly calculating them [57–59].

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

$$(1.2) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m},$$

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

$$(1.3) \quad \mathbf{C}' = \mathbf{S}^{\frac{1}{2}} \mathbf{C},$$

and  $\mu$  is an index that runs over the atomic orbitals (AO). The MO coefficients are the solution of diagonalizing the Fock matrix (from the Roothaan-Hall equations, see equation 2.60).

The approximations to full two electron integrals are given by damped charge-charge interactions - this approach will be discussed in more detail in chapter 4 as it is a crucial part of designing a new excited state method for chlorophyll systems.

Lastly the single particle excited space used to construct the full transitions 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 diagonalization whilst also capturing a broad enough range 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 [53, 60–63]. Whilst some early benchmarking of sTDA-xTB shows its ability to match experimental absorption spectra [53], and can be extended to predict experimental absorption spectra very well small metallo-organic complexes and [60], recent work has used sTDA-xTB to screen a wide range of compounds in high-throughput methods [61], as well as the basis for a  $\Delta$ -ML model (where the error between a high (TD-DFT) and low level (sTDA-xTB) method is calibrated) [62]. In this way 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 [64]. As sTDA-xTB prediction of chlorophyll systems has not been reported much before, especially with regard to conformer variations, 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 these to retain accuracy. This strategy is explored in the next section on statistical method based on high-level data to generate exciton Hamiltonian matrix elements.

## 1.2 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 [12, 65, 66]. These are referred to as 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, such as an estimate of the population

of excited states [65, 66] and B800-B850 energy transfers. As said, this approach breaks down when more atomistic considerations are necessary.

Over a long timescale where the full conformation space is well sampled, exciton Hamiltonians can be constructed from distributions of chromophore transition properties (i.e. excitation energies and transition densities) [67]. These properties are approximately 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 that can be limitlessly sampled to construct Hamiltonians without the need for explicit calculations on separate structures. The Hamiltonians could utilize 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, these methods are ill-suited for dynamic studies where structures at different times are correlated, which is an important consideration for LH2 [68]. Recently machine-learning methods have been reported that give time-dependent Hamiltonians but still without the need for expensive calculations.

### 1.2.1 Machine-Learning For Exciton Models

Machine-learning models have been used in many areas of computational chemistry, especially in areas where both large amounts of data and well-defined metrics make it easy to train methods [69–73]. For example, models that predict forcefield constants and atomization energies can be readily made as root mean square deviations of atomic positions and mean absolute errors in energies are well-defined metrics [74–76]. Similar to sTDA-xTB, ML models can be used for screening in high-throughput workflows - a particularly relevant example of this is on screening organic photovoltaics considering these systems have  $\pi$ -conjugation chemistry similar to chlorophyll [77]. 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 fitted to high level data. However the incorporation of atomic geometry information and machine-learning techniques can give time dependent models, as discussed below.

In 2016 Häse *et al.* reported on a multi-perceptron or neural network (NN) model that predicts the  $Q_y$  transition for chlorophyll molecules [78]. 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 "Coulomb matrix" [79, 80] is used as descriptor of the chlorophyll systems defined as

$$(1.4) \quad M_{AB} = \begin{cases} \frac{1}{2}Z^{2.4} & \text{for } A = B \\ \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} & \text{for } A \neq B \end{cases}$$

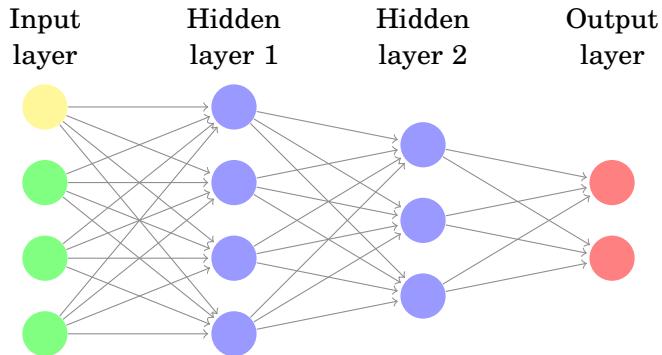


Figure 1.3: A schematic of a neural network showing how inputs, such as the (flattened) Coulomb matrix, are ordered into nodes (green and yellow) and can be used to generate outputs (red nodes). The blue nodes between these layers constitute the "hidden layers". Each arrow leading to a node represents a matrix multiplication and activation function application. The values in the output layer for models discussed in the text would be  $Q_y$ , transition energies or atom centered transition charges.

where  $Z_A$  is some measure of the atomic charge on atom  $A$ , and  $R_A$  is the position vector. It can be seen that the off-diagonal elements are simply the Coulombic interactions, and diagonal elements are a polynomial of the atomic charges. This descriptor is popular due to the similarity in information that an electronic structure calculation would start from, namely the positions and nuclear charges of atoms [81].

The Coulomb matrix is then used as an input for a neural network model. A neural network is a series of matrix operations applied to input data that overall acts as a non-linear function. In this way NN models are conceptually similar to biological neurons, which take in electrical signals and through some mechanism can be activated to send out a different signal. The outputs of each operation is described as a layer, with each element in a layer referred to as a node (i.e. nodes stack vertically into layers, layers stack horizontally into the network). The first and last layers are named input and output layers and any steps in-between referred to as "hidden" layers [82].

A layer can be any distinct vector of values. For example, flattening the Coulomb matrix into a 1D vector gives the input layer, and the  $Q_y$  transition energy and dipole moment the output vector (of 2 elements). The vector  $\mathbf{V}_{(n+1)}$  at layer index  $n + 1$  is given by multiplying the previous layer,  $\mathbf{V}_{(n)}$ , with a coefficient matrix  $c$  and applying some function  $f$

$$(1.5) \quad \mathbf{V}_{n+1} = f(\mathbf{c}_n \mathbf{V}_n)$$

where the coefficient matrices  $\mathbf{c}_n$  are fitted to give the smallest error between the output layer values and some target data (i.e.  $Q_y$  energies), and the function  $f$  can amplify beneficial values in nodes, mimicking the activation of biological neurons (giving the name activation functions).

The activation functions are also parameterised, with the most common functions being sinusoidal ( $f(\mathbf{V}_n) = \cos(m\mathbf{V}_n)$ ), sigmoid ( $f(\mathbf{V}_n) = \frac{m_1}{m_2 + e^{m_3 \mathbf{V}_n}}$ ), linear ( $f(\mathbf{V}_n) = m\mathbf{V}_n$ ) or rectified linear unit ( $f(\mathbf{V}_n) = \max(0, m\mathbf{V}_n)$ ). The coefficient matrices  $\mathbf{c}_n$ , and parameters used in the activation functions  $m$  or  $m_k$ , are fit by a back-propagation method, again parameterized - these fitting parameters are referred to as hyper-parameters and have to be optimized to give the best coefficient matrix using supervised learning techniques such as a systematic grid-search. Other considerations such as over-fitting also need to be taken into account. The brief explanation of NNs here glosses over more practical considerations, and creating these models takes in-depth knowledge and experience to achieve good results.

The H  se NN model predicted  $Q_y$  transition energies with around a 0.3 meV error for all of the 8 sites in the FMO complex [78]. This is exceptionally accurate, supporting the idea that models lacking in chemical nuance, but still given atomic positions and nuclear charges, have all the information necessary to predict transition energies. It is noted in this work that the root mean squared deviations (RSMD) of Nitrogen atom positions compared to the energy-minimized crystal structure correlate well with excited state properties. This implies that structural information is key to get transition properties correct and should be a primary consideration when discussing 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.*, utilized 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 [83]. KRR is another machine-learning method that can be understood as two processes [84]. 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

$$(1.6) \quad f'(\mathbf{X}) = \mathbf{X}\beta$$

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 Coulomb matrix, referred to as the feature matrix) and  $\beta$  is a set of fitted coefficients that minimize the value  $|f'(\mathbf{x}) - f(\mathbf{x})|$ . The matrix  $\beta$  can be found by minimizing the square of this value, 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 minimization of regression coefficients so that the inner products are not required, but requires a new function that compares the similarity of features. Ultimately, the function  $f'$  becomes

$$(1.7) \quad f'_{\text{KRR}} = \sum_j^{N_x} \beta_j k(\mathbf{x}, x_j)$$

where the  $N_x$  is the size of the feature vector  $\mathbf{x}$  (which in 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

$$(1.8) \quad k(\mathbf{x}, x_j) = \exp\left(\frac{-(\mathbf{x} - x_j)^2}{2\sigma^2}\right)$$

where  $\sigma$  is a fitted parameter [85]. Again these parameters are optimised by a systematic search through values, similar to the grid search referenced earlier.

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 that argues that transition charge based methods are accurate enough for almost all needs [86].

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 effects 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 optimized model to the QM/MM setup used. Additionally machine-learnt models would not be able to calculate any other properties than those they are taught - for example the models that return  $Q_y$  transition energies would not be able to return ground state energies, a trivial task for fully quantum methods like TD-DFT. In short, a machine learning method will only be as good as the method it is trained on and cannot predict anything not built into its architecture. There are also 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. For example the number of data points required to train the Häse and Farahvash models was on the order of  $10^4$  and  $10^5$  respectively. The compute walltime for such datasets would be in the hundreds or thousands of hours. In summary, whilst these models are accurate enough to reliably make time-dependent Hamiltonians they are limited by both their designs and their construction cost.

### 1.3 GPU Acceleration

An alternative approach to these problem would be to accelerate high-level calculations. This is possible with various established techniques such as multi-node parallelism (an early example is the parallelization of GAMESS [87]) but as this requires the exclusive use of many (dozens) of CPUs this is not a favourable option to take. Instead a better approach is to use graphical processing units (GPUs), which have been used to accelerate many different types of molecular chemistry simulations [88]. Programs written for GPUs can partition a limited number of basic operations over a massive number of parallel components, and this is exploited to parallelize the bottleneck of expensive calculations [89]. For example to accelerate TD-DFT calculations of chlorophyll excited states, the bottleneck would be evaluating the electron integrals necessary to construct the Casida equation (see section 2.2.1) Further afield, examples of GPU usages include the evaluating forces in MD simulations from classical equations as well as the matrix operations for NNs discussed above [90–92].

This approach does not require any parameter optimization 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 studies and is a popular way of drastically increasing system sizes while keeping computing time down [48]. One of the first and still popular electronic structure programs specifically designed for GPUs is TeraChem [47], with earlier programs focussing on molecular mechanics programs such as OpenMM [93] and Amber [94].

Previous work has used GPU-based programs to study LH<sub>2</sub>, using full TD-DFT calculations to construct Frenkel exciton Hamiltonians for every frame of an LH<sub>2</sub> MD simulation [95, 96]. The workflow was similar to exciton frameworks discussed above, using transition properties on sites to evaluate Hamiltonian elements - all of the transition properties were calculated using TD-DFT with the  $\omega$ PBEh functional and 6-31G basis set. These calculations were accelerated using GPU hardware with Terachem so the compute wall-time would be similar to tight-binding and machine-learning methods. With this acceleration it was also possible to calculate transition properties on larger aggregates of chlorophyll molecules, such as a combination of the two subunits shown in figure 1.1. There was an average error of 8 meV when benchmarking the exciton framework against TD-DFT calculations on these hexamer complexes [95]. Absorption spectra predicted by full TD-DFT and the exciton model were also very similar. The benefit of using a more efficient method to calculate TD-DFT data is showcased by the production of a non-adiabatic dynamic simulation with gradient terms of the exciton system being calculated on-the-fly or 300 fs. This required the construction of the Frenkel exciton Hamiltonian to be on the timescale of other gradient methods (i.e. force-fields or some semi-empirical tight-binding methods).

Whilst computing the electron integral terms with GPUs requires far less computing time, storing these values is a major issue [95]. This is due to the lightweight memory capacity on GPU cards, and even makes recalculation of some integrals more efficient than storage. This

means that a larger basis sets or more complicated density functionals may be too expensive to use. Therefore trying to calculate high-level data may be an issue for a more detailed study. It may be argued that accelerating the generation of high-level data would be better done by machine-learning methods. Similar to the other methods discussed, it is clear there are benefits and compromises with GPU acceleration.

## 1.4 Aims

The underlying issue of LHC models is clear - the systems are too large to explicitly calculate all transition properties required when using high level methods. Making well-chosen approximations has found success in solving some parts of this issue but ubiquitously at some other expense. The sTDA-xTB method (and other tight-binding methods) are efficient but generally not accurate enough to give reliable results like TD-DFT or higher-level methods. 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 GPUs could also be used but may still come up against memory issues which limits the extent of the basis set and density functional.

These shortcomings sketch out the need for a method that is efficient, accurate, extendable, memory light and easy to reproduce. The work presented in this thesis explores whether novel excited state methods in conjunction with efficient electronic structure methods could constitute a good method for LHCs models - this strategy is similar to the sTDA-xTB approach although with a more specialized scope. Most of this work is based on tight-binding approaches as these fulfill most of the criteria set out above except accuracy, making them the optimal starting place. In order to assay the usability of these novel approaches it is necessary to benchmark transition properties against established methods (explored in chapters 3 and 4) as well as provide case studies of calculating full LHC models (chapters 5 and 6). A description of all of the common theories used (DFT, DFTB/xTB, TD-DFT and Frenkel-Davydov models) is given in the next chapter.

## 1.5 Outline

The remainder of this thesis is structured into 6 more chapters. This chapter has introduced and discussed how efficient LHC models have been explored in previous work and the issues that have been discovered. The following chapter summarizes existing methods that are used throughout this work.

**Chapter 3** is an investigation of whether mean-field methods, such as  $\Delta$ -SCF and eigenvalue difference, can reliably predict transition properties and particularly addresses their suitability for treating for bulk chlorophyll transition properties. It also discusses how the underlying level of electronic structure theory affects the accuracy of transition properties.

**Chapter 4** uses these findings to inform the design of Chl-xTB, a novel approach to calculating biochromophore excited states. While it is not a general purpose method, its application to the  $Q_y$  transition of chlorophyll shows excellent performance. It fulfills the criteria set out above, obtaining transition properties with great efficiency whilst also achieving an accuracy comparable to TD-DFT methods.

**Chapter 5** demonstrates the application of chl-xTB to chlorophyll dimer 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 charge separation (which can potentially lead to unwanted fluorescence quenching) in chlorophyll dimers is suppressed by the LH2 protein scaffold.

**Chapter 6**, then reports on applying the novel method to the whole LH2 chlorophyll system. The high level of detail and efficiency is used to access low-frequency spectral densities of the entire LH2 complex, offering new insight into the effect of complex-wide vibrations on exciton transport dynamics.

The last chapter, **Chapter 7**, discusses the conclusions from the previous results chapters. Investigations such as applying the novel excited state 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.

## BACKGROUND THEORY

There are three main components to calculating atomic geometry dependent Frenkel-Davydov exciton models. First is the underlying electronic structure theory, such as density functional theory (DFT) or density functional tight binding (DFTB). The electronic structure theory is the basis for calculating excited states, providing MO coefficients for electron integrals, orbital energies and other properties. The second component is the excited state method, again such as time-dependent DFT (TD-DFT) or TD-DFTB. In this work other excited state methods are discussed so these theories are outlined in this chapter as well. The excited state methods provide the transition properties, such as transition densities, charges or dipoles and transition energies that are used to construct the exciton Hamiltonian. The third component is the exciton Hamiltonian itself, which depends on which energies to include and how to calculate the coupling parameters. In order to establish a basis for reporting and discussion in later chapters, outlines of each of these three components are given here.

### 2.1 Electronic structure

#### 2.1.1 Density Functional Theory

Density functional theory is ubiquitous in molecular chemistry simulations [97]. Often this theory is chosen for its good scaling with respect to system sizes as well as an appreciable accuracy against higher level methods. A brief overview of the DFT formalism is given here.

At its heart DFT is based on the two Hohenberg-Kohn theorems [98]. 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  $\rho_{GS}(r)$

$$(2.1) \quad E_{\text{GS}} : \rho_{\text{GS}}(r) \mapsto E[\rho_{\text{GS}}(r)],$$

where  $E[\rho_{\text{GS}}(r)]$  is the functional. The second theorem is closely related to the variation principle, stating that the *exact* ground state density also minimizes the total energy. This minimum corresponds to only one electron density. Whilst proven in principle, the exact functional of the electron density which gives this energy is unknown, and so various approximations have been made. One popular approximation is the Kohn-Sham approach, where electrons are approximated as non-interacting in order to tractably calculate the ground state density [99]. The total energy in the Kohn-Sham approach is the sum of functionals

$$(2.2) \quad E_{\text{tot}}[\rho(r)] = E_{T_S}[\rho(r)] + E_V[\rho(r)] + E_J[\rho(r)] + E_X[\rho(r)] + E_C[\rho(r)],$$

where the terms  $E_{T_S}[\rho(r)]$ ,  $E_V[\rho(r)]$ ,  $E_J[\rho(r)]$ ,  $E_X[\rho(r)]$ ,  $E_C[\rho(r)]$  correspond to the kinetic, (nuclear) potential, bare Coulombic, exchange and correlation interaction respectively. It is necessary to include the exchange and correlation terms as the Coulombic approximation does not include any treatment of spin effects. The solution for the minimum energy would satisfy the equation

$$(2.3) \quad \Delta \left[ E_{\text{tot}}[\rho(r)] - \mu \left( \int \rho(r) dr - N \right) \right] = 0$$

with the constraint that the total number of electrons  $N$  is conserved. The value of  $\mu$ , the Lagrange multiplier, is given by

$$(2.4) \quad \mu = \frac{\delta E_{\text{tot}}[\rho(r)]}{\delta \rho(r)},$$

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

$$(2.5) \quad \begin{aligned} \mu &= \frac{\delta E_{T_S}[\rho]}{\delta \rho(r)} + v_V[\rho(r)] + v_J[\rho(r)] + v_X[\rho(r)] + v_C[\rho(r)] \\ &= \frac{\delta E_{T_S}[\rho]}{\delta \rho(r)} + v_{KS}[\rho(r)], \end{aligned}$$

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

$$(2.6) \quad \left[ -\frac{1}{2} \nabla^2 + v_{KS}[\rho(r)] \right] \psi_i = \epsilon_i \psi_i$$

where  $\psi_i$  are one-electron wavefunctions,  $\epsilon_i$  are Lagrange multipliers to ensure orthonormality (also interpreted as orbital energies), and the kinetic energy term  $-\frac{1}{2}\nabla^2$  is given from the definition of the kinetic energy of non-interacting electrons as

$$(2.7) \quad E_{T_S} [\rho(r)] = -\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle.$$

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

$$(2.8) \quad \rho = \sum_i^N |\psi_i(r)|^2.$$

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 [100].

In practice a range of exchange-correlation functionals are employed dependent on the problem at hand. The simplest functionals 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 electron density and electron density gradient, using the generalized gradient approximation (GGA) [101]. Other examples include meta-GGAs, which go further and use the second derivative, as well as hybrid functionals, which use some fraction of the Hartree-Fock energy [102, 103]. Many exchange-correlation functionals require parameterization against high level data.

The first DFT calculations were done on periodic systems with infinite-domain plane wave functions used as a basis set. However molecular electronic structure packages often use basis functions centred on atomic positions [104], although this is not always the case. The one-electron wavefunctions can be written as linear combinations of basis functions  $\chi_j$

$$(2.9) \quad \psi_i = \sum_{\mu}^n c_{\mu i} \chi_{\mu}(r),$$

where  $\mu$  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  $\chi_{\mu}$  (also referred to as an atomic orbital or AO) for orbital (one-electron wavefunction)  $\psi_i$ . Rewriting the electron density gives the density matrix  $\mathbf{P}$

$$(2.10) \quad \rho = \sum_{\mu\nu} P_{\mu\nu} \chi_{\mu} \chi_{\nu},$$

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

$$(2.11) \quad P_{\mu\nu} = \sum_i^N c_{i\mu} c_{i\nu}.$$

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

$$(2.12) \quad F_{\mu\nu} = \frac{\delta E_{KS}}{\delta P_{\mu\nu}},$$

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

$$(2.13) \quad \mathbf{FC} = \mathbf{SC}\epsilon,$$

where  $\mathbf{S}$  is the overlap matrix and  $\epsilon$  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 [105]. As the Fock matrix is dependent on the MO coefficient  $\mathbf{C}$ , solving this matrix equation requires an iterative procedure to converge to a set of MO coefficients that can be calculated self-consistently. This procedure is referred to as a self-consistent field or SCF.

Having a convenient definition of the ground state electronic structure is useful in calculating other properties. As explained in section 2.2.1, 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.

### 2.1.2 Density Functional Tight Binding

In recent years there has been renewed interest in tight-binding (TB) methods, including tight binding methods derived from density functional theory (named density function tight binding or DFTB) [106]. Interest in DFTB and similar methods has grown in the last ~10 years as they are often more accurate than force-field methods, especially for complex metallic systems, and cheaper than DFT methods. Modern applications combine semi-empirical tight-binding methods with machine-learning to accelerate screening of potential drug candidates [107], materials [108] and organic electronics [109]. These methods approximate the density functional energy by expanding into a Taylor series, based on the density fluctuations  $\delta\rho$

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

with this series usually truncated between the first and third term[110, 111]. These terms are analogous to some of the terms in the Kohn-Sham functional. The first term, which does not contain density fluctuation terms, is called the band-structure energy and is given by the sum of energies of single particle wavefunction

$$(2.15) \quad E[\rho_0(r)] = \sum_i \langle \psi_i | H[\rho_0] | \psi_i \rangle,$$

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

$$(2.16) \quad \frac{1}{2} \int \int \frac{\delta^2 E[\rho(r)]}{\delta\rho(r)\delta\rho(r')} \Big|_{\rho_0} \delta\rho(r)\delta\rho(r') = \frac{1}{2} \int \int \frac{\delta^2 E_{XC}[\rho(r)]}{\delta\rho(r)\delta\rho(r')} + \frac{1}{|r-r'|} \Big|_{\rho_0} \delta\rho(r)\delta\rho(r'),$$

with the other Taylor expansion terms collected into what is known as the repulsive energy term  $E^{\text{rep}}$ . The common expression for DFTB (also known as DFTB2 or SCC-DFTB [112]) energy is then given as the sum

$$(2.17) \quad E^{\text{DFTB}} = \sum_i \langle \psi_i | H[\rho_0] | \psi_i \rangle + E^{\text{XC}} + E^{\text{rep}}.$$

However it is common to include other correctional terms in this framework, as discussed in section 2.1.3. It is also common to replace the continuous electron density with a point charge model. Calculating energy terms with a point charge model instead of electron integrals is key in accelerating TB methods. Charges are determined by first approximating the charge fluctuations as a sum of fluctuations centred on atomic positions

$$(2.18) \quad \delta\rho(r) = \sum_A \delta\rho_A(r),$$

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

$$(2.19) \quad \delta\rho_A(r) \approx \Delta q_A F_A^{00} Y^{00},$$

where  $F_A^{00}$  and  $Y^{00}$  are the multipole expansion coefficients -  $\Delta q_A$  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 can be seen in section 1.1.1. 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

$$(2.20) \quad E^{\text{XC}} = \frac{1}{2} \sum_{A,B} \Delta q_A \Delta q_B \gamma_{AB},$$

where the function  $\gamma_{AB}$  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  $\gamma$  function uses the chemical hardness of atoms to dampen this interaction. For self-interaction (i.e.  $\gamma_{AA}$ ), the energy term is equivalent to the second derivative of the energy with respect to the atomic partial charge, equal to the Hubbard parameter  $U_A$  or twice the chemical hardness  $\eta_A$

$$(2.21) \quad \begin{aligned} E_{AA}^{\text{XC}} &= \frac{1}{2} \Delta q_A^2 \gamma_{AA} \\ &= U_A \\ &= 2\eta_A. \end{aligned}$$

For inter-atomic interactions the  $\gamma$  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 [52]. 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, with an additional harmonic constraint to prevent the orbitals from becoming too diffuse. Often this confinement potential is scaled to be within twice the covalent radius of each atom.

Solving for the ground state wavefunction employs a similar Roothaan-Hall matrix equation, although with the Fock matrix defined by equation 2.17. 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 require 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. This again is key to obtain higher efficiency than DFT methods. One common example is the core energy functional, which is written in terms of the AO contributions as

$$(2.22) \quad E [\rho_0(r)] = \sum_i \sum_{\mu\nu} c_{i\mu} c_{i\nu} H_{\mu\nu}$$

where the elements of  $H_{\mu\nu}$ , the Hamiltonian in the atomic basis, can be precomputed as they are not 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 [111, 113–117]. Generally the accuracy of DFTB is found to be on par with or just below DFT [118–120], and higher order derivative terms can be used to achieve more accurate results. Often it is found that the repulsive energy term is the trickiest to get correct, requiring the most reference data for parameterisation [111]. 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 [121]. Originally DFTB was formulated in a restricted ansatz, using doubly occupied orbitals instead of spin-orbitals. This approach 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 the effects of using a spin-unrestricted ansatz, making predictions of excited states much more accurate.

Tight binding methods are usually best 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. Alternative solutions for dealing with the size of these systems include force-field methods, which do not use any quantum mechanics theory 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, metallic centers or the making and breaking of chemical bonds[94], which covers many interesting biochemical systems including LHCs. For these systems using tight-binding methods seems to be a good trade-off 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 this gap where DFTB methods are useful [122].

### 2.1.3 Extended Tight Binding

Recently the extended tight binding (xTB) family of methods, developed by Grimme and coworkers, have been presented as another semi-empirical tight binding solution to investigating large chemical systems [52, 53, 123–126]. 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 fewer pair-wise parameters, intending to emulate the ZDO type methods, whilst remaining efficient and accurate for large systems. The issue with pair-wise parameters is similar to over-fitting in semi-empirical models. For example if one component uses element wise parameters for atoms up to radon (the original cutoff of GFN1-xTB) then 86 parameters are required - if using pair-wise parameters then 7,396 parameters are needed. This will require more training data as the number of training features should be much greater than the number of parameters needed to be fit, making datasets more expensive to calculate. Fitting more parameters is also generally more difficult to optimize for many heuristics. Other issues 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. Minimizing pair-wise parameters is then beneficial in terms of training set size, performance and re-usability. The number of parameters required varies systematically across the family of GFN methods, decreasing with increasingly sophisticated descriptions of the electronic structure. For example GFN2 uses a quadrupole expansion for electrostatics, which, 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 computational 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 [127], D4 [128] or a modified D4 method) and a halogen bonding correction. First order terms are calculated with an extended Hückel theory, and second and higher order terms are calculated by isotropic electrostatic and exchange-correlation terms. In the following equations, 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 summarized as

$$(2.23) \quad \begin{aligned} E_{\text{GFN1-xTB}} &= E_{\text{disp}}^{(0)} + E_{\text{rep}}^{(0)} + E_{\text{XB}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\text{IES+IXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \\ &= E_{\text{disp}}^{\text{D3}} + E_{\text{rep}} + E_{\text{XB}}^{\text{GFN1}} + E_{\text{EHT}} + E_{\gamma} + E_{\Gamma}^{\text{GFN1}} \end{aligned}$$

$$(2.24) \quad \begin{aligned} E_{\text{GFN2-xTB}} &= E_{\text{disp}}^{(0,1,2)} + E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\text{IES+IXC}}^{(2)} + E_{\text{AES+AXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \\ &= E_{\text{disp}}^{\text{D4'}} + E_{\text{rep}} + E_{\text{EHT}} + E_{\gamma} + E_{\text{AEC}} + E_{\text{AXC}} + E_{\Gamma}^{\text{GFN2}} \end{aligned}$$

$$(2.25) \quad \begin{aligned} E_{\text{GFN0-xTB}} &= E_{\text{disp}}^{(0)} + E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + \Delta E^{(0)} \\ &= E_{\text{disp}}^{\text{D4}} + E_{\text{rep}} + E_{\text{EHT}} + E_{\text{EEQ}} + E_{\text{srh}} \end{aligned}$$

Common to all of these expressions is the extended Hückel theory energy term ( $E_{\text{EHT}}$ ), derived from first order density fluctuations. This energy is given by tracing the Hamiltonian

$$(2.26) \quad H_{\mu\nu}^{\text{EHT}} = \frac{1}{2} K_{AB}^{ll'} S_{\mu\nu} (H_{\mu\mu} + H_{\nu\nu}) X(EN_A, EN_B) \Pi(R_{AB}, l, l') Y(\eta_l^A \eta_{l'}^B)$$

with the valence one-electron density  $P_{\mu\nu}$

$$(2.27) \quad E_{\text{EHT}} = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{\text{EHT}},$$

where  $\mu, \nu$  are the indices of atomic orbitals. In the equation above  $A, B$  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 parameterized 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 electronegativity  $EN$  function,  $\Pi$  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), each interaction term is readily separable and so can either be considered or discarded for future parameterization work. For example, the  $\Pi$ ,  $Y$  and  $X$  terms deal with interactions that are more strongly influenced 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 [53] method, which is discussed below.

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

$$(2.28) \quad E_\gamma = \frac{1}{2} \sum_{A,B}^N \sum_{l \in A} \sum_{l' \in B} q_l q_{l'} \gamma_{AB, ll'},$$

where  $q_l$  are shell-resolved Mulliken partial charges. The  $\gamma$  operator describes short-range Coulombic interactions

$$(2.29) \quad \gamma_{AB, ll'} = \frac{1}{\sqrt{R_{AB}^2 + \eta_{AB, ll'}^{-2}}},$$

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

$$(2.30) \quad E_\Gamma = \frac{1}{3} \sum_A^N q_A^3 \Gamma_A,$$

where  $q_A$  is the atom partial charge (sum of the shell partial charges on that atom), and  $\Gamma_A$  is a different operator constructed from atom-wise parameters. These terms are analogous to the  $\gamma$  operators discussed in the general DFTB theory section (equation 2.20). 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

$$(2.31) \quad E_{\text{AES}} = E_{q\mu} + E_{q\Theta} + E_{\mu\mu}$$

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

$$(2.32) \quad E_{q\mu} = \frac{1}{2} \sum_{AB} f_3(R_{AB}) \left[ q_A \left( \mu_B^T \mathbf{R}_{BA} \right) + q_B \left( \mu_B^T \mathbf{R}_{AB} \right) \right]$$

$$(2.33) \quad E_{q\Theta} = \frac{1}{2} \sum_{AB} f_5(R_{AB}) \left[ q_A \mathbf{R}_{AB}^T \Theta_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \Theta_A \mathbf{R}_{AB} \right]$$

$$(2.34) \quad E_{\mu\mu} = \frac{1}{2} \sum_{AB} f_5(R_{AB}) \left( \mu_A^T \mu_B \right) R_{AB}^2 - 3 \left( \mu_A^T \mathbf{R}_{AB} \right) \left( \mu_B^T \mathbf{R}_{AB} \right)$$

respectively, where  $\mu_A$  is the dipole moment on atom A and where  $\Theta_A$  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

$$(2.35) \quad E_{\text{AXC}} = \sum_A \left( f_{XC}^{\mu_A} |\mu_A|^2 + f_{XC}^{\Theta_A} ||\Theta_A||^2 \right)$$

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  $E_{\text{rep}}$  term in DFTB. This is defined as

$$(2.36) \quad E_{\text{rep}} = \frac{1}{2} \sum_{A,B} \frac{Z_A^{\text{eff}} Z_B^{\text{eff}}}{R_{AB}} e^{-\sqrt{\alpha_A \alpha_B} (R_{AB})^{k_f}}$$

where  $Z_A^{\text{eff}}$  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  $\alpha_A$  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

$$(2.37) \quad E_{\text{disp}}^{\text{D3}} = -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(CN_A, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB})$$

where  $C_n$  are 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 [129, 130] is given by

$$(2.38) \quad f_{\text{BJ-damping}}^{(n)}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + \left( a_1 \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} + a_2 \right)^n}$$

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

$$(2.39) \quad E_{\text{disp}}^{\text{D4}} = -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(q_A, CN_A, q_B, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB})$$

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

$$(2.40) \quad \begin{aligned} E_{\text{disp}}^{\text{D4}'} &= -\frac{1}{2} \sum_{A,B} \sum_{n=6,8} s_n \frac{C_n(CN_A, CN_B)}{R_{AB}^n} f_{\text{BJ-damping}}^{(n)}(R_{AB}) \\ &+ s_9 \sum_{A,B,C} \frac{(3 \cos(\theta_{ABC}) \cos(\theta_{BCA}) \cos(\theta_{CAB}) + 1) C_9(CN_A, CN_B, CN_C)}{(R_{AB} R_{AC} R_{BC})^3} \\ &\times f_{\text{damping}}^{(n)}(R_{AB}, R_{AC}, R_{BC}) \end{aligned}$$

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.

Also present 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

(2.41)

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

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

$$(2.42) \quad f_{AXB\text{-damping}} = \frac{1}{2} \left( 1 - \frac{1}{2} \frac{R_{XA}^2 + R_{XB}^2 - R_{AB}^2}{|R_{XA}| |R_{XB}|} \right)^6$$

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 Roothan-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 diagonalization of the Fock matrix is required for a ground state solution. These charges are given by solving the matrix equation

$$(2.43) \quad \begin{pmatrix} \mathbb{A} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ q_{\text{tot}} \end{pmatrix}.$$

In this scheme,  $q_{\text{tot}}$  is the total charge of the system,  $\mathbf{X}$  is a vector of electronegativities given by

$$(2.44) \quad X_A = \kappa_A \sqrt{mCN_A} - EN_A$$

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. The diagonal (single atom) and off-diagonal (atom pair) matrix elements are given by

$$(2.45) \quad \mathbb{A}_{AA} = J_{AA} + \frac{2\gamma_{AA}}{\sqrt{\pi}}$$

$$(2.46) \quad \mathbb{A}_{AB} = \frac{\text{erf}(\sigma_{AB} R_{AB})}{R_{AB}},$$

where  $\sigma_{AB}$  is a geometric mean of the atomic radii  $\alpha_A$ . Solving for these charges then gives the energy as

$$(2.47) \quad E_{EEQ} = \mathbf{q}^T \left( \frac{1}{2} \mathbb{A} \mathbf{q} - \mathbf{X} \right)$$

These charges are used in place of Mulliken charges that would be found in other energy terms, and as only a single diagonalization of equation 2.43 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 optimize 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% 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 and accurate methods.

## 2.2 Excited State Methods

### 2.2.1 Linear Response TD-DFT

Linear response time-dependent DFT (TD-DFT) is a well established method for calculating excitation energies and transition properties from only ground state information [131]. It is formulated from the Runge-Gross theorem [132], 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 2.1.1. 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 over which the systems responds to the external perturbation short, 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 [133]. 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 [133].

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

$$(2.48) \quad i \frac{\delta}{\delta t} \Psi(r, t) = \hat{H}(r, t) \Psi(r, t)$$

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

$$(2.49) \quad \hat{H}(t) = \hat{H}^0 + v_{\text{ext}}(t)$$

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 (such as that created by a photon) and so is time  $t$  dependent. This Hamiltonian can then be used to describe a time-dependent set of Kohn-Sham equations [98]. The time-dependent Kohn-Sham Hamiltonian is given by

$$(2.50) \quad \hat{H}_{KS}[\rho(t)] = \hat{H}_{KS}^0[\rho] + v_H[\rho(t)] + v_{XC}[\rho(t)] + v_{\text{ext}}(t)$$

where  $v_H$  and  $v_{XC}$  are the Kohn-Sham Coulomb 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 functional of the potential function

$$(2.51) \quad \rho(t) = \rho[v_{\text{ext}}(t)].$$

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

$$(2.52) \quad \delta\rho(r, \omega) = \int d^3r' \chi(r, r', \omega) \delta v_{\text{ext}}(r', \omega),$$

which, when calculated from the Fourier transform of the time series (i.e. calculated in the frequency space) can be written explicitly for a Kohn-Sham system of electrons as

$$(2.53) \quad \chi_{KS}(r, r', \omega) = \sum_{\mu\nu} (f_\mu - f_\nu) \frac{\psi_\mu(r)\psi_\mu^*(r')\psi_\nu(r')\psi_\nu^*(r)}{\omega - (\epsilon_\mu - \epsilon_\nu) + i\eta}$$

where  $f_\mu$  is the occupation number of the orbital  $\psi_\mu$ , and  $\eta$  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 approach would

calculate the response of the electron density, and therefore the excited state, explicitly. However the issue with the theory given so far is that it does not identify which frequencies  $\omega$  give any appreciable density response i.e. where the excited states are. Gross *et al.* solved this issue by writing the (non-KS) response function as in a Lehmann representation, with a discretized set of states  $m$

$$(2.54) \quad \chi(r, r', \omega) = \lim_{\eta \rightarrow 0^+} \sum_m \left[ \frac{\langle 0 | \hat{n}(r) | m \rangle \langle m | \hat{n}(r') | 0 \rangle}{\omega - (E_m - E_0) + i\eta} - \frac{\langle 0 | \hat{n}(r') | m \rangle \langle m | \hat{n}(r) | 0 \rangle}{\omega + (E_m - E_0) + i\eta} \right]$$

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 [134]. It can be seen that the poles of the response function (the places where it equals zero) 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

$$(2.55) \quad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

where  $\omega$  are the excitation energies, and the vectors  $\mathbf{X}$  and  $\mathbf{Y}$  describe the electronic transitions in the basis of ground state molecular orbitals. This equation is often referred to as the Casida equation [135], with a full derivation also given in later reviews [136, 137]. The elements of matrices  $\mathbf{A}$  and  $\mathbf{B}$  are given by

$$(2.56) \quad A_{ia,jb}(\omega) = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \int dr \int dr' \psi_i^*(r) \psi_a(r) f_{XC}(r, r', \omega) \psi_i(r') \psi_a^*(r')$$

$$(2.57) \quad B_{ia,jb}(\omega) = \int dr \int dr' \psi_i^*(r) \psi_a(r) f_{XC}(r, r', \omega) \psi_i(r') \psi_a^*(r')$$

where  $i, j$  and  $a, b$  are occupied and virtual orbital indices respectively,  $\epsilon_i$  are the energies of the ground state orbitals, and  $\delta$  is the usual kronecker delta function. The  $\mathbf{A}$  matrix include occupied-to-virtual transition contributions, whilst the  $\mathbf{B}$  matrix includes virtual-to-occupied transitions - these are also loosely referred to as excitation and "de-excitation" contributions [138]. 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. To make these functionals computable this self-consistency is neglected, such that the  $\mathbf{A}$  and  $\mathbf{B}$  matrix elements can be calculated from ground state properties. These could be provided by both DFT and DFTB methods.

Using TD-DFT for LHC investigations come with some caveats. Fairly early on it was found that the charge-transfer artifacts can affect low energy ( $Q$  band) excitations [51]. There is also difficulty in obtaining good reference data - for example coupled-cluster methods are extremely

expensive for chlorophyll systems as they contain  $\sim 140$  atoms. A recent successful benchmarking of density functionals compares TD-DFT to DFT/MRCI - DFT/MRCI was chosen in this example as a reference method due to its good performance against *ab initio* methods for  $\pi$  conjugated systems including chlorophyll [139–142]. In the TD-DFT benchmarking, a range of different density functionals were used to calculate the  $Q_y$  transitions in chlorophylls from the FMO complex, and it was found that PBE0, CAM-B3LYP and B3LYP had RMSEs in transition energies of less than 0.05 eV against DFT/MRCI [143]. CAM-B3LYP also proved the best in reproducing the whole  $Q$  band ( $Q_x$  and  $Q_y$ ). With few alternatives, TD-DFT is a common choice for calculating chlorophyll excited states.

### 2.2.2 Eigenvalue Difference

One approximation to full response theory is the eigenvalue difference method. For eigenvalue difference method, it is assumed there is no response of the orbital energies and shapes when interacting with light. This approach can be recovered from the complete Cassida equation (equation 2.55) if the coupling elements (off-diagonal elements) in the **A** and **B** matrices are 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 ( $\epsilon_2$ ) and the orbital has been excited from ( $\epsilon_1$ )

$$(2.58) \quad \Delta E = \epsilon_2 - \epsilon_1.$$

Generally, eigenvalue difference methods are not seen as accurate response methods, but can offer a quick and easy initial value [144].

### 2.2.3 $\Delta$ -SCF

$\Delta$ -SCF [145, 146] predicts the transition energy  $\Delta E$  of a system as the difference of the single point energy  $E_n$  of two states:

$$(2.59) \quad \Delta E = E_2 - E_1.$$

Assuming that the excited state solution will be in a similar location to the ground state in the MO coefficient space, the ground-state MO coefficients can be used as an initial guess for the excited state. In its simplest form, the  $\Delta$ -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 the 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 only a local minimum, and so often is less reliable to find as a solution especially from the standard SAD initial guess [147] (the SAD or "superposition of atomic orbitals" is a standard initial guess of the ground state density, calculated as a combination of electronic densities calculated for each atom of the system in isolation.) For this reason it is often found that converging to the  $\Delta$ -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 (mixing of previous iterations' Fock matrices with the current Fock matrix to reduce or dampen "jumps" which may prevent convergence), DIIS methods (quadratic extrapolation of Fock matrices) and sometimes intermediate initial guess steps (i.e. using excited state populations with fractional occupations before integral occupations).

In early version of  $\Delta$ -SCF, 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 [148]. 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 [149]. Consequently, an alternative 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) [149]. The maximum overlap method starts with the normal SCF procedure, where each iteration produces new molecular orbital coefficients by solving the Roothaan-Hall equations, [105]

$$(2.60) \quad \mathbf{FC}^n = \mathbf{SC}^n \epsilon,$$

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

$$(2.61) \quad \mathbf{F} = f(\mathbf{C}^{n-1}).$$

At this point MOM differs to normal SCF by populating set of orbitals with the most similarity to the previous orbitals. The amount of similarity in orbitals can be estimated from their overlap

$$(2.62) \quad \mathbf{O} = (\mathbf{C}^{\text{old}})^{\dagger} \mathbf{SC}^{\text{new}},$$

which for a single orbital can be evaluated as a projection

$$(2.63) \quad p_j = \sum_i^n O_{ij} = \sum_v^N \left[ \sum_{\mu}^N \left( \sum_i^n C_{i\mu}^{\text{old}} \right) S_{\mu v} \right] C_{vj}^{\text{new}}$$

where  $\mu, \nu$  are orbital indices. This method can be used for any excited state, with the caveat that the orbital solution will most likely be in the same region in the solution space as the ground state solution. For a small number of low-lying states this is generally true, and so  $\Delta$ -SCF can be used to calculate a small spectrum of excited states [149]. A similar method referred to as iMOM or initial-MOM determines population by similarities to the initial orbital coefficients, which can be beneficial for some convergence issues [150].

$\Delta$ -SCF has been shown to be cheap alternative to TD-DFT and other higher level methods [151–153] without considerable losses of accuracy in certain cases especially for HOMO-LUMO transitions. Kowalczyk *et al.* showed that for a test set of small to medium  $\pi$  conjugated systems, including a unsubstituted porphyrin ring,  $\Delta$ -SCF can reproduce experimental HOMO-LUMO transitions energies about as well as PBE0 TD-DFT [154]. Additionally, as the excited state is given as solutions to SCF equations, the gradient of this solution with respect to atomic positions can be given similar to normal mean-field theory. These gradients would be much cheaper than TD-DFT or coupled cluster methods, which is advantageous for simulating dynamics [152] as well as a vibrationally resolved absorption spectra [154].

### 2.2.3.1 Transition Density and Dipole Moments

Transition density matrices are a well used tool in excited state analysis [155, 156]. Detailed interpretations can sometimes differ (e.g. for excited states from CIS/TDA and normal TD-DFT [157]) but in general transition density matrices describe the change in electronic structure in going from one state to another. Key to this discussion is the fact that observables, such as the transition dipole, can be calculated by tracing operators with the transition density matrix. For  $\Delta$ -SCF the transition density is calculated from the SCF solutions for the ground and excited states. The reduced one-particle transition density matrix  $\mathbf{D}^{21}$  can be written as

$$(2.64) \quad \mathbf{D}^{21} = |\Psi_1\rangle\langle\Psi_2|,$$

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 orbital coefficients  $\mathbf{C}^{(n)}$ , the transition density matrix is

$$(2.65) \quad \mathbf{D}^{21} = \mathbf{C}^{(2)} \text{adj}(\mathbf{S}^{21}) \mathbf{C}^{(1)\dagger},$$

where  $\mathbf{S}^{21}$  is an overlap matrix with elements

$$(2.66) \quad S_{jk}^{21} = \langle\phi_j^2|\phi_k^1\rangle.$$

Equation 2.66 is derived using Löwdin's rules for non-orthogonal determinants [158]. Observable properties of the excited state can be calculated by tracing appropriate operator with the transi-

### 2.3. EXCITATION ENERGY TRANSFER AND FRENKEL-DAVYDOV EXCITON HAMILTONIANS

---

tion density matrix. For example, the transition dipole moment is found by taking the trace of the transition density matrix with the dipole operator

$$(2.67) \quad \mu^{1 \rightarrow 2} = \text{Tr}(\hat{\mu}\mathbf{D}) = \text{Tr}(\hat{\mu}|\psi_1\rangle\langle\psi_2|) = \langle\Psi_2|\hat{\mu}|\Psi_1\rangle = \sum_{jk} \mu_{jk}^{21} \text{adj}(\mathbf{S}^{21})_{jk}$$

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

$$(2.68) \quad |\mathbf{S}^{21}| = \langle\Psi_2|\Psi_1\rangle.$$

## 2.3 Excitation Energy Transfer and Frenkel-Davydov Exciton Hamiltonians

Excitation energy transfer (EET) and subsequent rates of transfer describe the time-frame that energy, absorbed through a photon of light for LHCs, is shuttled between states. It is thought that light harvesting systems have evolved to fine these rates [6]. In regimes with strong coupling (ie large transfer), excitation energy can flow back and forth between different sites. The Frenkel-Davydov model [33, 34] is used in the regime where the excited electron and its "hole" have a strong interaction and so are localized on the same chromophore - this means mainly vertical excitations are included instead of charge transfer states (although charge transfer states are important mechanisms for the fluorescence quenching of chlorophyll excited states, as discussed in section 5.4).

For a dimer system with chromophores  $A$  and  $B$ , there are two possible states where one molecule is in the ground state and the other in the excited state - these are labelled  $|A^*B\rangle, |AB^*\rangle$  where  $A^*$  denotes molecule  $A$  in the excited state. The exciton state is defined by a linear combination of these two states

$$(2.69) \quad |\phi\rangle = c_1 |A^*B\rangle + c_2 |AB^*\rangle,$$

where  $c_n$  are the coefficients of each possible state. In systems beyond dimers other chromophore sites can be easily included in the basis states (i.e.  $|ABC^*\rangle, |ABCD^*\rangle$  etc.). Calculating the coefficients  $c_n$  requires diagonalising the exciton Hamiltonian which can be written as a sum of individual site Hamiltonians and inter-site interactions

$$(2.70) \quad \hat{H}_{\text{tot}} = \sum_m^N \hat{H}_m + \sum_m^N \sum_n^N (\hat{V}_{\text{el-el}} + \hat{V}_{\text{el-nuc}} + \hat{V}_{\text{nuc-nuc}}),$$

where  $N$  is the number of sites,  $\hat{H}_m$  is the electronic Hamiltonian for a single chromophore molecular  $m$  and  $\hat{V}_{\text{el-el}}$ ,  $\hat{V}_{\text{el-nuc}}$  and  $\hat{V}_{\text{nuc-nuc}}$  are the electron-electron, electron-nuclear and nuclear-nuclear interactions between two sites respectively. It is assumed that  $H_m$  satisfies the time-independent Schrödinger equation

$$(2.71) \quad \hat{H}_{a_m} |\phi_{a_m}\rangle = E_{a_m} |\phi_{a_m}\rangle$$

where  $a_m$  indexes the electronic states on chromophore  $m$ . The total Hamiltonian can be represented in a basis of eigenfunctions  $\phi_{a_m}$  of the single chromophore Hamiltonians as

$$(2.72) \quad \langle \Phi_a | \hat{H}_{\text{tot}} | \Phi_b \rangle = \sum_m^N \langle \Phi_a | \hat{H}_m | \Phi_b \rangle + \sum_m^N \sum_n^N \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle.$$

This Hamiltonian neglects nuclear-electron and nuclear-nuclear interactions in favour of treating the surrounding nuclear environment in the individual site Hamiltonians [159] - this includes the other chromophore sites as well as any other environments ie. the LH2 protein environment. Alternatively, this term can just be ignored.

It is also common to assume that only one excited state on each chromophore participates in the energy transfer, and this is the first excited state. This reduces the size of the basis set, and is known as the Heitler-London approximation [160]. Additionally it is assumed that the chromophore sites are well separated enough that the wavefunctions do not overlap [33]. Consequently equation 2.72 reduces to

$$(2.73) \quad \langle \Phi_a | \hat{H}_{\text{tot}} | \Phi_b \rangle = E_{a_m}^{(m)} \prod_l^N \delta_{al} \delta_{bl} + \sum_m^N \sum_n^N \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle.$$

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

$$(2.74) \quad \begin{aligned} \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle &= \sum_{i \in \mathbf{r}_m} \sum_{j \in \mathbf{r}_n} \langle \Phi_a | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \Phi_b \rangle \\ &= \left( \sum_{i \in m} \sum_{j \in n} \langle \phi_{a_m}^{(m)} \phi_{a_n}^{(n)} | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \phi_{b_m}^{(m)} \phi_{b_n}^{(n)} \rangle \right) \prod_{l \neq m, n}^N \delta_{a_l, b_l} \end{aligned}$$

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[159]

$$(2.75) \quad \langle \Phi_a | \hat{V}_{\text{el-el}} | \Phi_b \rangle = \int d\mathbf{r}_m \int d\mathbf{r}_n \frac{\rho_{a_m}(\mathbf{r}_m) \rho_{b_n}(\mathbf{r}_n)}{|\mathbf{r}_m - \mathbf{r}_n|}$$

### 2.3. EXCITATION ENERGY TRANSFER AND FRENKEL-DAVYDOV EXCITON HAMILTONIANS

---

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

$$(2.76) \quad \hat{H}_{\text{eff}} = \sum_{m=1}^N \epsilon_m + \sum_{m \neq n}^{N,N} V_{mn}$$

where the diagonal elements  $\epsilon_m$  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 reduce the transition density with a multipole expansion [161]. 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 [162]). Using the second term in this expansion gives a dipole-dipole interaction, referred to as the point-dipole method

$$(2.77) \quad V_{mn} = \frac{\mu_m \mu_n}{R_{mn}^3} - 3 \frac{(\mu_m \cdot \mathbf{R}_{mn})(\mu_n \cdot \mathbf{R}_{mn})}{R_{mn}^5}$$

where  $\mu_m$  is the transition dipole for site  $m$  and  $R_{mn}$  is the distance between sites  $m$  and  $n$ . Alternatively the transition density can also be reduced to atom centred charges. This gives the coupling as a Coulombic interaction

$$(2.78) \quad V_{mn} = \sum_{A>B, A \in m, B \in n} \frac{q_A q_B}{|\mathbf{R}_A - \mathbf{R}_B|}$$

where  $A, B$  are indices for atoms on sites  $m$  and  $n$  respectively,  $\mathbf{R}_A, \mathbf{R}_B$  are the position vectors for these atoms and  $q_A, q_B$  are the atom centered transition charges.

Other methods for calculating coupling interactions include as the extended dipole [163], transition charges from electrostatic potential [163, 164] and transition density cube methods (i.e. a grid approach of discretizing the continuous transition density) [165]. Additionally calculating the coupling value straight from equation 2.75 is possible and necessary for benchmarking, but is the most expensive option [86]. 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 ( $>20\text{\AA}$ ), all of these methods converge to the same value [86].

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.



## 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 and Prof. Fred Manby in March 2021 [166]. These sections include sections 3.1 from 3.1.1 up to but excluding 3.1.5.

This chapter investigates the accuracy of  $\Delta$ -SCF methods with both an ab initio DFT level of theory as well as with semi-empirical, tight-binding approximations.  $\Delta$ -SCF is a well-known method for calculating transition energies, as well as ionisation potentials, electron affinities and other properties [118, 148, 154, 167]. 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  $\Delta$ -SCF has shown to be appreciably accurate for porphyrin HOMO-LUMO transition energies [154], there has been little work on benchmarking the chlorophyll  $Q_y$  transition and its dependence on geometry variation with  $\Delta$ -SCF. As  $\Delta$ -SCF does not necessitate solving the computationally intensive Casida equation, it would be a good replacement for the more common TD-DFT calculations used to construct exciton models. However, as  $\Delta$ -SCF is not as cheap as semi-empirical models, it may not be able to calculate a full time series of exciton Hamiltonians in a reasonable time, but instead be better suited to generate training data for the statistical methods discussed in section 1.2.

Instead, combining  $\Delta$ -SCF with semi-empirical approaches may give the best of both methods, resulting in accurate transition properties with high wall-time efficiency. As exemplified with the sTDA-xTB methods, using tight-binding theories as a electronic structure basis for transition properties can work well when considering a large range of systems. As  $\Delta$ -SCF is conceptually simpler it would be easier to implement compared to other excited state methods, and more

tractable when expanding the formalism to include, for example, embedding effects or gradients of excited states [166], compared to existing semi-empirical methods. Ubiquitous to all uses of parameterised models outside their training space, the success of a tight-binding based  $\Delta$ -SCF method is dependent on the extensibility of the parameters it is based on. The element-wise parameterisation of the xTB methods addresses this problem, with parameters fitted with chemical meaning as well as end-result accuracy, and so a GFN-xTB based  $\Delta$ -SCF method (termed  $\Delta$ -xTB) was investigated for the purposes of efficiently calculating excited states.

There are several practical challenges to overcome before confidently using  $\Delta$ -SCF to generate large volumes of transition property data from LHC structures. As stated, accuracy is foremost, which can be examined through benchmarking to higher level methods. However the non-orthogonality of the ground and excited states, an artifact from using separate SCF solutions, breaks the origin-independence of transition dipole moments and leads to nonphysical results and unsuitability for exciton models which rely on these moments. Additionally, appropriately assigning the symmetry of excited states is less well developed than in other methods, further impacting the confidence of results. Solutions for these issues are proposed and assessed in this chapter to enhance the conclusions of the benchmarking.

## 3.1 Benchmarking $\Delta$ -SCF

### 3.1.1 Reference Data and Test Sets

Full scale chlorophyll molecules are too large to be able to calculate benchmark transition energies and dipole moments with high-level methods. For example, it would not be possible to test  $\Delta$ -SCF data against coupled cluster data (a "gold standard" in transition property benchmarking for many systems [168]) going straight to benchmarking chlorophyll transitions. Instead a stepwise approach was used, covering a large chemical space by using a range of molecules before focusing on chlorophyll. These molecules had fewer atoms and 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 to benchmark both the  $\Delta$ -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 [124], and as it was constructed to design new response methods it would contain a wide range of systems to appropriately benchmark the  $\Delta$ -SCF methods.

The test set consisted of 109 small molecules. Each system was closed-shell, contained 12 atoms or less, and contained 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 [42].

### 3.1.2 Benchmarking Small Systems

Transition properties for the small molecule test set were calculated using TD-DFT and  $\Delta$ -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-Abelian group. Symmetry labelling was also only available for TD-DFT calculations, as these were performed with Gaussian 16.  $\Delta$ -SCF calculations were done with the QCORE program.

As the  $\Delta$ -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:

$$(3.1) \quad \Delta E_S = 2\Delta E^{i,\alpha \rightarrow \alpha, \alpha} - \Delta E^{i,\alpha \rightarrow \alpha, \beta}$$

was used to correct for the true singlet excitation energy  $\Delta E_S$  [169].

The results of comparing transition energies and transition dipole magnitudes are shown in figures 3.1 and 3.2. Overall, the excitation energies calculated with  $\Delta$ -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 have a mean and standard deviation of 0.41 eV and 0.27 eV respectively. Transition dipoles were similarly accurate compared to the reference data, although  $\Delta$ -SCF performs slightly worse in this respect. The mean and standard deviation in the absolute value of transition dipole moment,  $|\mu|$ , were 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. (here using  $ea_0$  as the atomic unit). The outlier circled in figure 3.1 is an ethene dimer system, and shows the inability of  $\Delta$ -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  $\pi$ -antibonding orbitals, and are very close in energy. The HOMO orbitals are the same on both ethene molecules, being the  $\pi$ -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  $\Delta$ -SCF cannot include this mixed behaviour as it only considers single transitions.  $\Delta$ -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 3.2 is due to  $\Delta$ -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  $\Delta$ -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,  $\Delta$ -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.

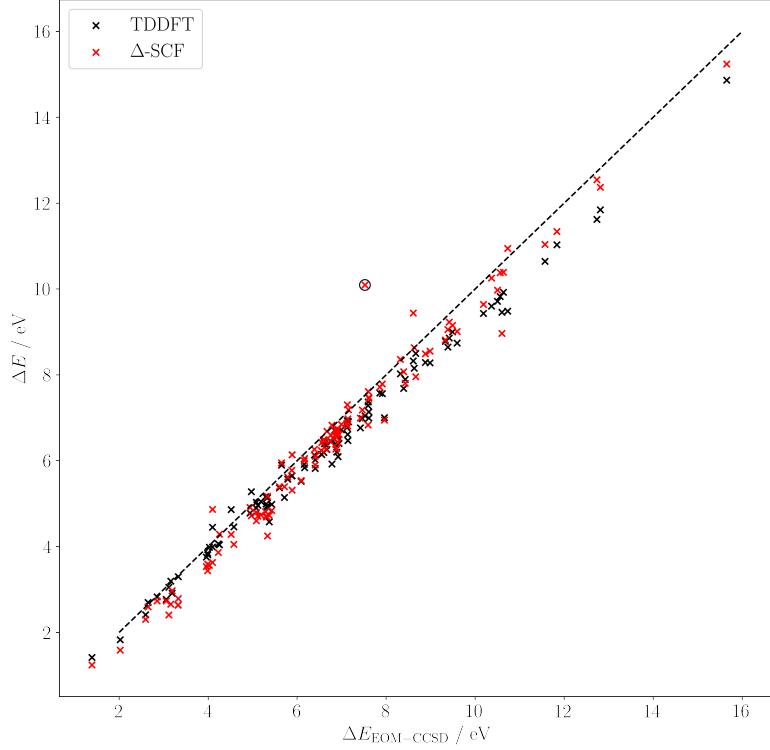


Figure 3.1: Transition energies  $\Delta E$  from TD-DFT (black) and  $\Delta$ -SCF (red) plotted against EOM-CCSD energies, with the line  $y = x$  (dashed) for reference. The ethene dimer outlier has been circled.

### 3.1.3 Addressing Non-orthogonality in $\Delta$ -SCF States

Generally the ground and excited states calculated for  $\Delta$ -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  $\Psi_1$  and  $\Psi_2$  to overlap such that the inner product,

$$(3.2) \quad S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle$$

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

$$(3.3) \quad q^{21} = \langle \Psi^2 | \Psi^1 \rangle$$

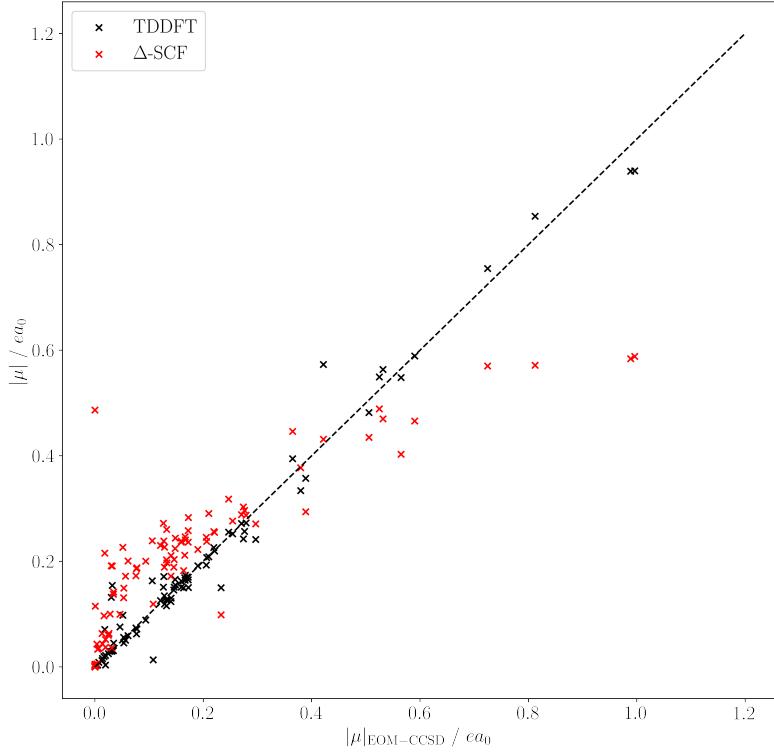


Figure 3.2: Transition dipole magnitudes from TD-DFT (black) and  $\Delta$ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line  $y = x$  (dashed) for reference.

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  $\Delta$ -SCF would always have this error.

In order to fix this issue, the standard transformation to symmetrically orthogonalize the two states was applied which also would preserve as much character of the original states as possible [170]. The transformation is given by

$$(3.4) \quad |\Psi_{\tilde{v}}\rangle = \sum_v |\Psi_v\rangle \left[ \mathbf{S}^{-\frac{1}{2}} \right]_{v\tilde{v}}$$

where  $\mathbf{S}$  here is a block matrix

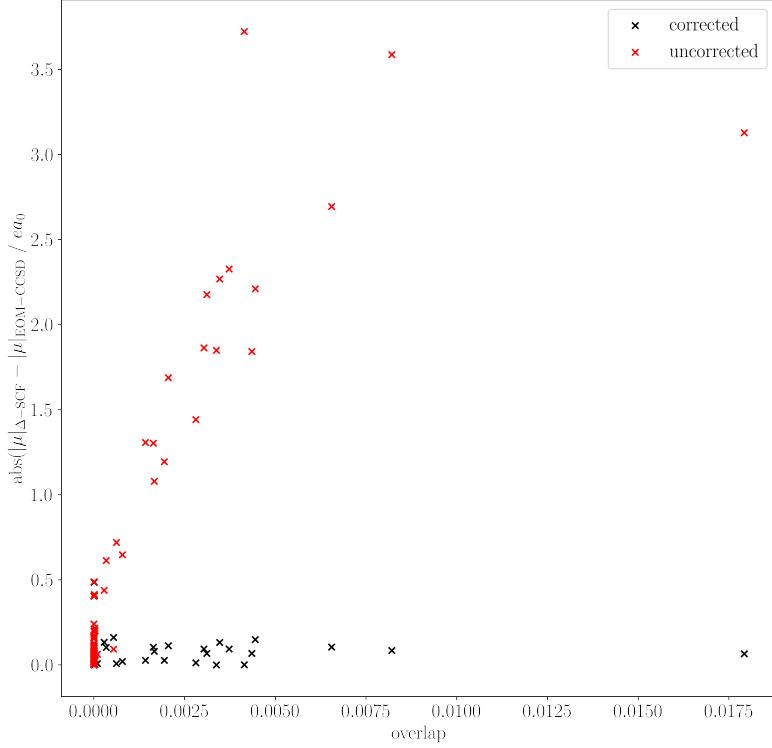


Figure 3.3: The absolute value of the error in transition dipole magnitude between  $\Delta$ -SCF and EOM-CCSD, plotted against the  $\Delta$ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown in red, whilst those with the symmetric orthogonalisation correction are shown in black.

$$(3.5) \quad \mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

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 3.3). The transition dipole for each molecule in the test set systems was calculated for molecules translated by 100 Å in the  $xyz$  direction. This would induce an error for the non-orthogonalized 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

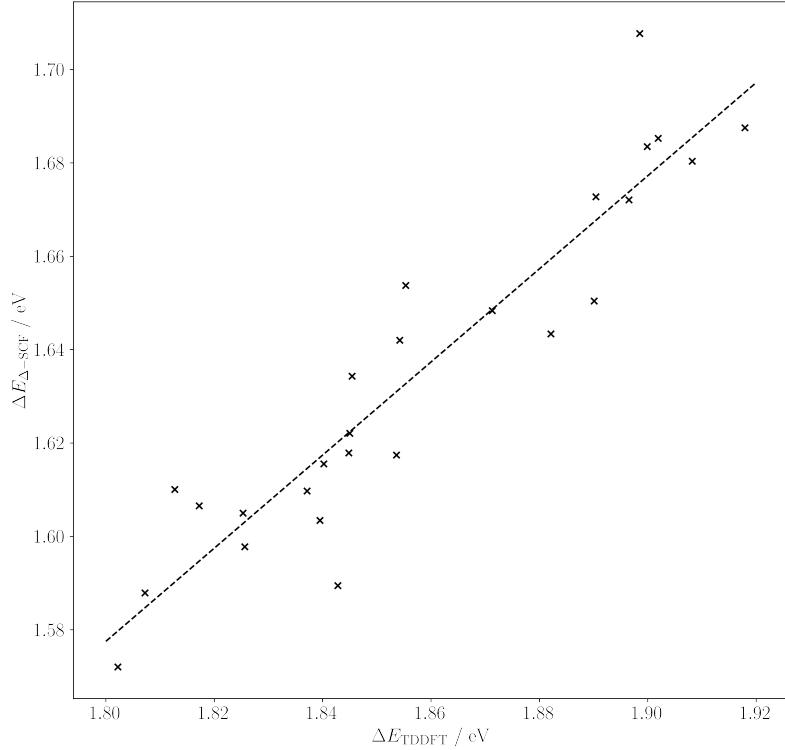


Figure 3.4: Transition energies from  $\Delta$ -SCF for 26 chlorophyll geometries from the LH2 protein of purple bacteria, plotted against energies from TD-DFT. The line of best fit ( $R^2 = 0.87$ ) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

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.

### 3.1.4 $\Delta$ -SCF Accuracy in LH2 Chlorophylls

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

The PBE0 functional with Def2-SVP basis set [171, 172] was used to calculate both TD-DFT and  $\Delta$ -SCF properties. This method has been used previously for BChla structures, and has been shown to be a good balance between accuracy and cost [67]. The high degree of correlation between  $\Delta$ -SCF and the TD-DFT results (shown in figures and 3.4 3.5) demonstrates that the

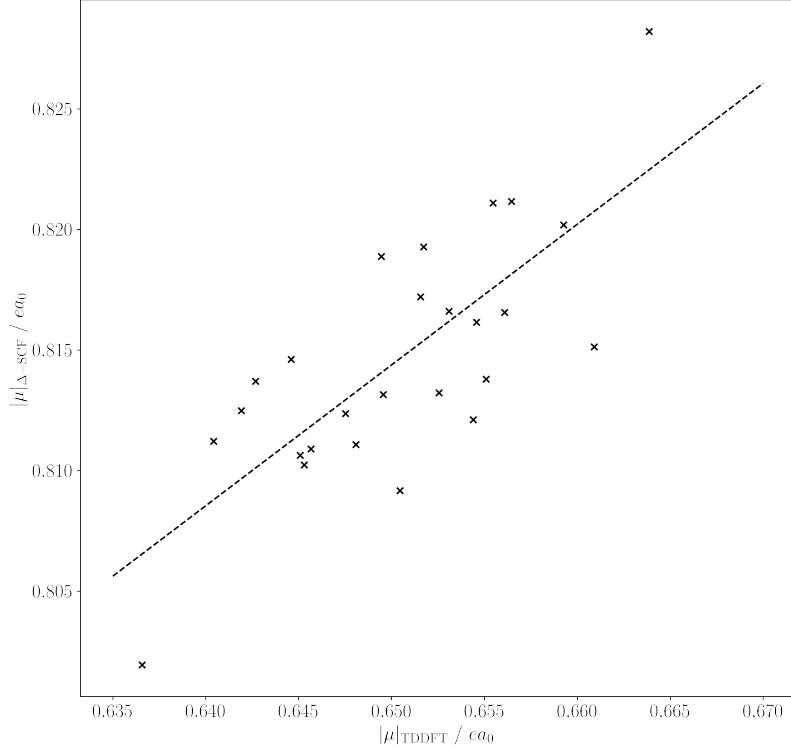


Figure 3.5: Transition dipole magnitudes from  $\Delta$ -SCF for the same 26 chlorophyll geometries as 3.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ( $R^2 = 0.57$ ) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

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 for the chlorophylls 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 is unclear whether this error might be from TD-DFT or  $\Delta$ -SCF. Additionally, there is a clear correlation between the transition dipole magnitudes from TD-DFT and  $\Delta$ -SCF, and so whilst quantitative results would not be possible, qualitative statements could be made from  $\Delta$ -SCF data. For example, whilst the exact value of a transition dipole moment at a given geometry may not be accurate, the change in transition dipole moment from several geometries would be similar in variation to TD-DFT calculated properties.

An important point 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.

### 3.1.5 Benchmarking xTB $\Delta$ -SCF Methods

$\Delta$ -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 the sTDA-xTB comparison it was necessary to use the spin-component-scaled second order coupled-cluster (SCS-CC2) [173, 174] reference data produced by Grimme et al. [53]. Two types of  $\Delta$ -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 states. A transition energy from  $\Delta$ -SCF based on GFN0-xTB energies would be functionally the same as an eigenvalue difference. The eigenvalue differences from sTDA-xTB, given by the `xtb4stda` program [53], were also included in this benchmarking. The `xtb4stda` program is the first of two programs that make up the full sTDA-xTB method, and runs a version of xTB which provides molecular orbital coefficients and energies for the sTDA program to use to construct and solve the truncated Casida equation.

The full list of methods included in the comparison is:

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

#### 3.1.5.1 Post-SCF Assignment of Symmetry

A source of error that has not been discussed in much detail so far is the assignment of transitions between different methods. A known problem of  $\Delta$ -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 problem could be seen in the symmetry of the excitation - if  $\Delta$ -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  $\Delta$ -SCF was not possible. Instead, transition dipole orientations and plots of MOs were used. Whilst a possible approach for  $\Delta$ -xTB, the difference between  $\Delta$ -xTB valence basis sets and full DFT basis sets caused issues. Instead the process of assigning symmetry to  $\Delta$ -SCF transitions was investigated in more detail.

Assigning symmetry to  $\Delta$ -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 is known at any point in the SCF procedure, and allows the Hamiltonian to be organized into a block matrix which is more efficient to diagonalize. This can be especially useful when solving for a large basis set or large system, as the matrix diagonalisation can be partitioned and parallelized over several cores or nodes on a cluster computer. However this approach works best if the system is highly symmetric, such as geometry 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

$$(3.6) \quad \tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{C}$$

and then summing the coefficients to obtain the character

$$(3.7) \quad P_A = \sum_v |\tilde{\mathbf{C}}_{A,\mu\nu}|$$

where  $\mu, \nu$  are indices for the atomic and molecular orbitals respectively. The symmetry label of a molecular orbital is the subspace in which this character is equal to 1, for cases with well defined assignment. However, in practice these character were mixed, and so the highest subspace character was taken as the assignment.

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

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

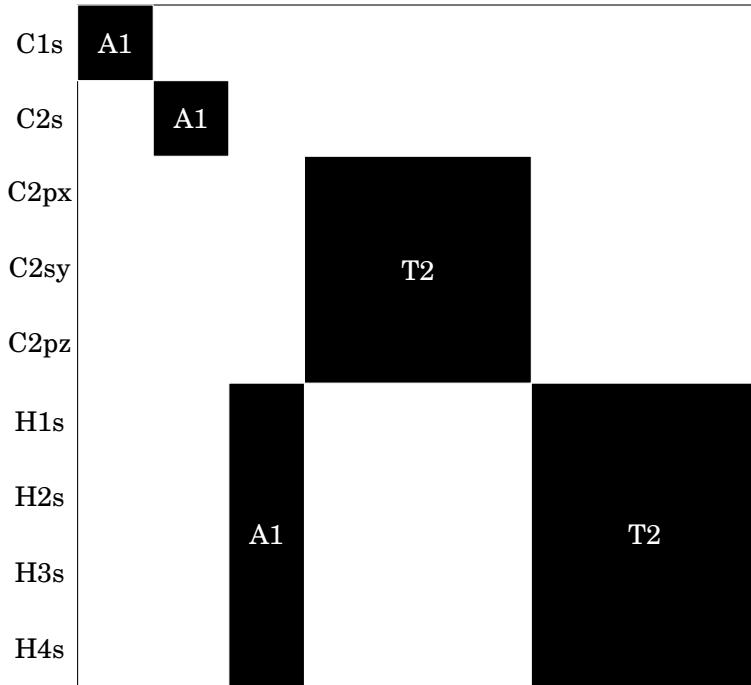


Figure 3.6: A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the  $T_d$  point group.

2. Get the symmetry adapted linear combinations (SALCs) of atomic orbitals for each subspace.
3. Use the SALCs to construct the transformation matrix  $T$ .
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 [175] 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  $A_1$  symmetry, and three occupied and unoccupied orbitals of  $T_2$  symmetry). A diagrammatic representation of the subspace characters of molecular orbitals is shown in figure 3.6. The overall wavefunction symmetry can then be expressed as the product of all MO symmetries, reduced with the reduction formula

$$(3.8) \quad n = \frac{1}{h} \sum_R \xi_r(R) \xi_i(R)$$

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
$\Delta$ -SCF CAM-B3LYP/aug-cc-pVTZ	-0.14	0.28
$\Delta$ -SCF PBE0/Def2-SVP	-0.62	0.50
TD-GFN1-xTB	0.27	1.47
TD-GFN0-xTB	-0.41	1.32
$\Delta$ -SCF GFN1-xTB	-0.12	2.11
$\Delta$ -SCF GFN0-xTB	-1.50	1.08
xtb4stda	4.39	1.26

Table 3.1: Means and standard deviations of the errors between SCS-CC2 reference data and response methods for the whole test set in eV. The `xtb4stda` entry represents the eigenvalue difference method that uses the eigenvalues output from this program.

where  $\xi_r, \xi_i$  are the reducible and irreducible representations respectively,  $h$  is the order of the group and  $R$  is the subspace. This method 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 problem was also observed for non-Abelian groups, where there would be degenerate subspaces, similar to the problem of assigning symmetry for the TD-DFT and EOM-CCSD transitions in the earlier benchmarking. Often the reduction of ground state wavefunctions gave non-physical answers.

Overall while improving some features of  $\Delta$ -SCF methods is in the scope of this project, this type of assignment is flawed for the LH<sub>2</sub> 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  $\Delta$ -SCF with this method. Hence while careful consideration was made for this problem, benchmarking of the  $\Delta$ -xTB methods could be expected to still have some errors associated with this issue. In the end the assignment of symmetry was based on the previously used inspection of transition dipole orientations and transition density plots.

### 3.1.5.2 $\Delta$ -xTB Benchmarking Results

The distributions of errors in SCS-CC2 data for each of the benchmarking methods for the entire test set are shown in figure 3.7. The means and standard deviations are reported in table 3.1.

Overall, both  $\Delta$ -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- $\Delta$ -xTB was -0.12 eV, and has a significant standard deviation of 2.11 eV. GFN0- $\Delta$ -xTB had a larger mean error of -1.50 eV, and whilst a slightly smaller standard deviation of 1.08 eV, this is still well

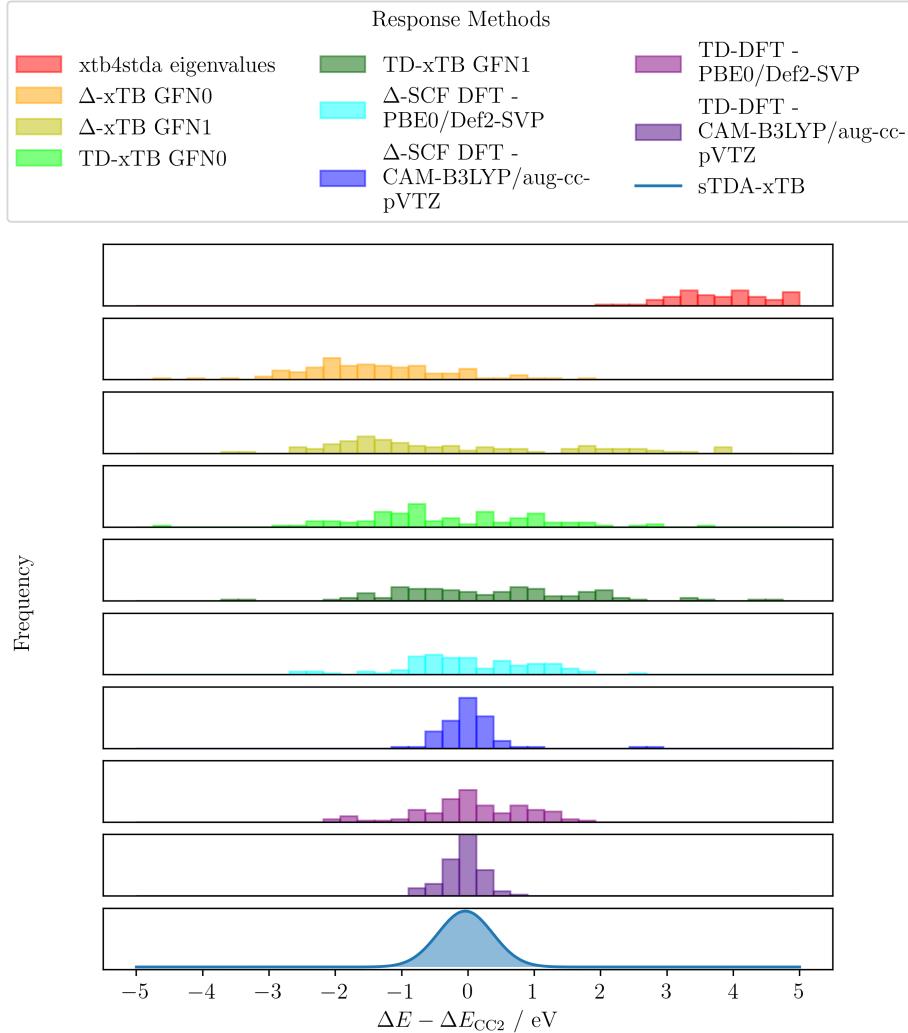


Figure 3.7: The distributions of errors compared to SCS-CC2 transition energies for the methods included in the  $\Delta$ -xTB benchmarking.

beyond a usable accuracy, given the fluctuations of  $Q_y$  energies are around 0.2 – 0.4 eV. The DFT methods, both the linear response and the  $\Delta$ -SCF methods, are still shown to be accurate at predicting excitation energies, with means and standard deviations within ranges reported in the previous sections.

From these results, it is argued that decreasing the level of theory for the electronic structure dramatically decreases the accuracy of computed transition properties. The highest level electronic structure method has the best accuracy. CAM-B3LYP/aug-cc-pVTZ TD-DFT and  $\Delta$ -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  $\Delta$ -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  $\Delta$ -SCF (0.28 eV to 0.50 eV). Again, the PBE0/Def2-SVP TD-DFT and  $\Delta$ -SCF are comparable, with standard deviations of 0.79 eV and 0.59 eV, although the mean for  $\Delta$ -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 poorer transition properties. Overall, the most inaccurate method is the eigenvalue difference methods from the `xtb4stda` method. The mean and standard deviation were 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)[53]. Arguably 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 may not be surprising. As opposed to 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 [52]. Whilst the lack of pair-wise parameters and a "top-down" parameterisation approach which give 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.

## 3.2 Conclusions

The transition properties of a test set of small molecules has been benchmarked with multiple  $\Delta$ -SCF , TD-DFT and high-level methods. It has been shown that DFT based  $\Delta$ -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 a chosen 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 was fixed by using a symmetric orthogonalisation of the two originally non-orthogonal states.

For a small set of BChla geometries, the same level of accuracy for transition energy could be found between  $\Delta$ -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  $\Delta$ -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 benchmarking, reliably obtaining and assigning transitions predicted from

$\Delta$ -SCF has proved to be an unsolved issue. Either  $\Delta$ -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. The best example of this is 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  $\Delta$ -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  $\Delta$ -SCF methods.

GFN-xTB based methods, termed  $\Delta$ -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  $\Delta$ -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  $\Delta$ -SCF methods, which have a simple gradient theory and are less expensive 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 electronic structure is sufficiently good. However, of the methods for which this criterion is met, namely DFT and not xTB, the accuracy in electronic structure is obtained at the cost of efficiency. DFT based  $\Delta$ -SCF is not the orders of magnitude faster than TD-DFT required to calculate properties for a large volume of chlorophyll structures. The efficiency of these calculations may be too low for a large exciton framework or for large monomer systems like Bchl<sub>a</sub>. Semi-empirical  $\Delta$ -SCF methods, which would be efficient enough, prove inaccurate in their current form. However, it is demonstrated that a correct electronic structure and a "top-down" parameterisation could make an accurate semi-empirical method, which is examined in the following chapter.

## BESPOKE CHLOROPHYLL EXCITED STATE METHODS

### Previous Published Work

All of the work presented in this chapter are also included in a paper published with Dr. Susannah Bourne-Worster and Prof. Fred Manby in January 2023 [176].

This chapter reports on designing and parameterising a novel method for transition properties, referred to as Chl-xTB. The framework and theory for the method is outlined in section 4.1. Parameterisation details are given in section 4.2, 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 4.3.

From the previous chapter it was found that the tight-binding based  $\Delta$ -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 with  $\Delta$ -SCF in general make these methods unreliable for many uses (e.g. for a correlated set of structure where success with every time-frame is required.)

However, as found in the previous chapter, the sTDA-xTB method does imply that a tight-binding based approach would work, the only issue is the parameterisation of the electronic structure 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 - this claim is substantiated in section 4.2.3. This chapter posits that a method that is specific to chlorophyll would solve this problem. The machine-learning models discussed in chapter 1 are similar in this respect, where models predicted only  $Q_y$  transition energies. Here it is argued that electronic structure methods could fit this purpose instead of statistical machinery, and would be much easier to parameterise and have more chemically significant parameters, dependent on

design choices in the excited state method.

By keeping the excited state method "light-weight" (i.e. with few parameters and minimal computation), the efficiency of these methods may outstrip current approaches and match machine-learning methods. Additionally the risk of over-fitting to training data is reduced when using fewer parameters - compare the training set sizes of ML methods, in the hundreds of thousands, to sTDA-xTB, which is in the hundreds. Fewer parameters also increase the reusability of methods, as these parameters can be easily re-optimised.

## 4.1 Response and Electronic Structure Theory Approximations

### 4.1.1 Full TD-DFT Solutions

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

$$(4.1) \quad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

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  $\omega$  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

$$(4.2) \quad A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2(ia|jb) - a_x(ij|ab) + (1 - a_x)(ia|f_{XC}|jb)$$

$$(4.3) \quad B_{ia,jb} = 2(ia|bj) - a_x(ib|aj) + (1 - a_x)(ia|f_{XC}|bj)$$

where indices  $a, b$  and  $i, j$  refer to virtual and occupied orbitals respectively,  $\epsilon_i$  is the orbital energy of orbital  $i$ ,  $a_x$  is the value of non-local Fock exchange in the XC functional  $f_{XC}$ , and  $\delta_{ij}$  is the Kronecker delta function. The integrals (here in Mulliken notation) are Coulomb 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, due to the Kronecker delta functions, only contributes to the diagonal elements.

### 4.1.2 Approximations to Full TD-DFT Solutions

#### 4.1.2.1 Tamm-Dancoff approximation and Diagonal Dominant A matrices

The first reduction to full linear response is the well-used Tamm-Dancoff approximation (TDA). TDA is one of the earliest approximations applied to full linear response theory, and outlines that

only virtual-occupied contributions are considered when finding roots of excitations [177]. The full casida equation in TDA theory is reduced to

$$(4.4) \quad \mathbf{AX} = \omega \mathbf{X}$$

where the definitions of elements are the same as above, but it should be noted that the solutions  $\mathbf{X}$  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 [177, 178].

As an eigenvalue problem, solving for  $\omega$  requires constructing and diagonalizing the full  $\mathbf{A}$  matrix. Seen in equation 2.56, the XC functional is dependent on  $\omega$  values and so iterating through diagonalisation is needed to find stable, self-consistent solutions.

The success of  $\Delta$ -SCF demonstrates that solving the roots of the entire Casida equation is sometimes unnecessary. In a similar spirit, a light-weight shortcut to solutions is the diagonal dominant approximation to eigenvalue solutions, where diagonal elements of a matrix are taken as an approximation to the real solutions. By neglecting the off-diagonal terms, this approach explicitly treats separate transitions as non-interacting, as the off-diagonal elements describe the coupling between transitions in an excitation. This approach is similar to the single-pole approximation (SPA) in this way [134, 179]. Excitation energies would be then be given by

$$(4.5) \quad \omega_{ia} \approx A_{ia,ia} = (\epsilon_a - \epsilon_i) + 2(ia|ia) - a_x(ii|aa) + (1 - a_x)(ia|f_{XC}|ia).$$

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 explored in the previous chapter.

This approximation reduces computational cost by removing the need to calculate all of the off-diagonal elements, as well as many 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 4.1.4.

### 4.1.3 Integral Approximations with Monopole Expansions

Whilst using a 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 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 by the orbital overlap  $\mathbf{S}$  for all orbitals  $p, q$  centred on atom  $A$

$$(4.6) \quad q_A^{mn} = \sum_{p \in A} D_{pq}^{mn} S_{pq}$$

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

$$(4.7) \quad q_A = Z_A - \sum_{p \in A} D_{pq} S_{pq}$$

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 [53]. For both Coulomb and exchange type integrals, this can be done with a short-range damped MNOK operators. The integral is approximated by

$$(4.8) \quad (pq|rs) \approx \frac{1}{2} \sum_A^N \sum_B^N q_A^{mm} q_B^{mn} \Gamma_{AB}$$

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 Coulomb type integrals, the operator  $\Gamma_{AB}$  is given by

$$(4.9) \quad \Gamma_{AB}^J = \left( \frac{1}{(R_{AB})^{y_J} + (a_x \eta)^{-y_J}} \right)^{\frac{1}{y_J}}$$

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

$$(4.10) \quad \eta(A) = \frac{\delta^2 E(A)}{\delta^2 N^2}$$

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

$$(4.11) \quad \Gamma_{AB}^K = \left( \frac{1}{(R_{AB})^{y_K} + \eta^{-y_K}} \right)^{\frac{1}{y_K}}$$

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

As the  $a_x$  parameter can account for many of the exchange effects, the density functional in equation 4.5 is also neglected to further reduce computational cost. The final form of the expression used to calculate excitation energies is then given by

$$(4.12) \quad \omega_{ia} = (\epsilon_a - \epsilon_i) + \sum_{A,B}^N \left( 2q_{ia}^A \Gamma_{AB}^K q_{ia}^B - q^A \Gamma_{AB}^J q^B \right)$$

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  $\Delta$ -xTB methods of the previous chapter, could not break into the accuracy needed to investigate chlorophyll systems well enough (discussed in more detail in 4.2.3). 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 section 3.1.5.1).

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.

#### 4.1.4 The $Q_y$ 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 analyzed in the literature [1, 64, 180–182]. It is also one of the most important transitions in light harvesting systems so an accurate treatment is necessary for high-performing LHC models.

The  $Q_y$  transition is one of the two transitions that make up the  $Q$  band in the absorption spectra of chlorophyll, the other being the  $Q_x$  transition [64]. It is well known that the  $Q_y$  transition is important for electronic energy transfer, and predicting both transition energies and dipole moments correctly is important in constructing frameworks that model this transfer [183]. The  $Q_y$  transition is mostly HOMO-LUMO in character (96%), with a small amount of HOMO-1 - LUMO+1 (remaining 4%) [184]. The analogous transition in the unsubstituted tetraphorphyrin ring has the transition dipole along the molecular axis defined by the N atoms [185], however due

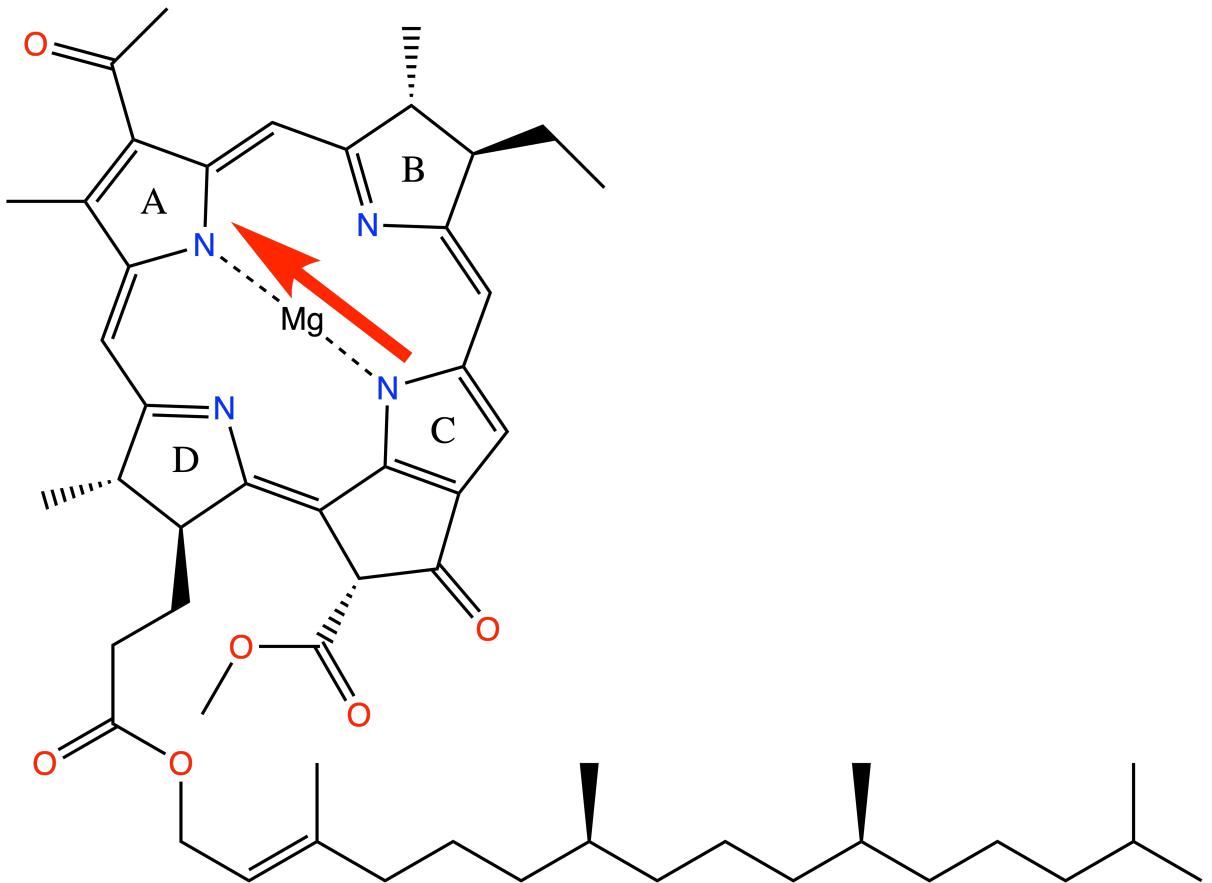


Figure 4.1: Bacterial Chlorophyll a (BChla) with a model  $Q_y$  transition dipole.

to the asymmetry introduced by substitutions and geometry deformations, BChla  $Q_y$  transitions have a deviation to this axis of around 12° [180]. In this was the  $Q_y$  transition in chlorophyll lies mostly along 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 delocalized 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 behavior seen in different versions of chlorophyll [180].

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 transition can be excluded from the transition character [184]. This is also supported by the results for the chlorophyll test set in the previous chapter as  $\Delta$ -SCF with its 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 so far. The lack of coupling elements that would

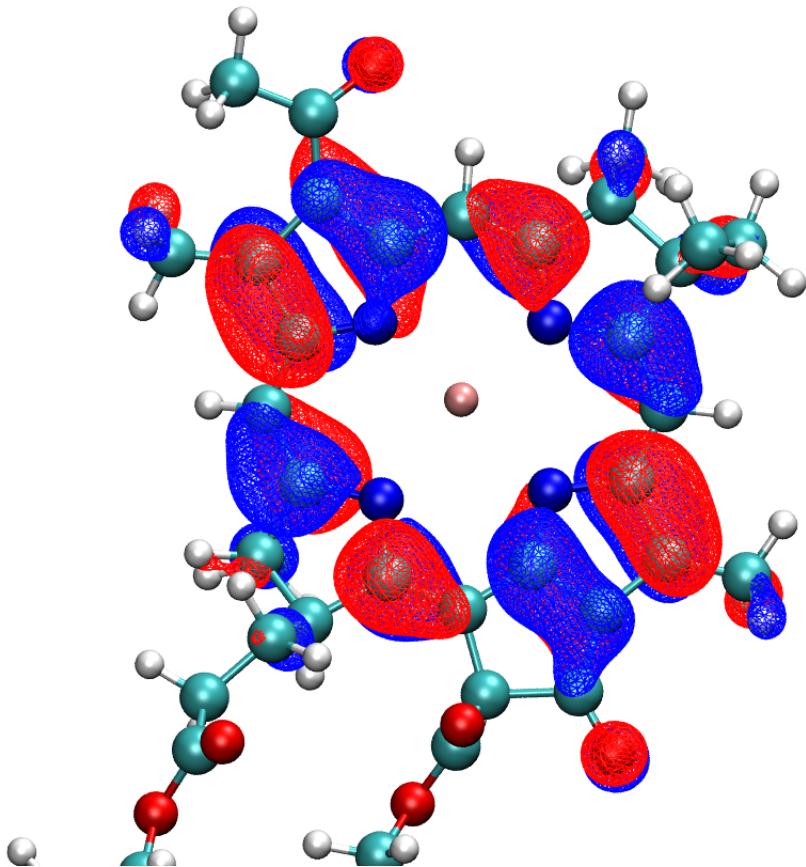


Figure 4.2: The HOMO orbital of Bchl-a from PBE0/Def2-SVP DFT.

appear in the  $\mathbf{A}$  matrix justifies the use of the diagonal dominant approximation.

Additionally, the well-defined transition makes 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 singular value this is ideal for autonomous optimisation, both for use in an objective function as well as for discarding outlier transitions.

#### 4.1.5 Bespoke xTB Parameters for Chlorophyll

As found in the last chapter, the underlying electronic structure can have a huge effect on the accuracy of transition properties. The GFN-xTB methods in particular, both linear response and  $\Delta$ -xTB, were 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  $\Delta$ -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

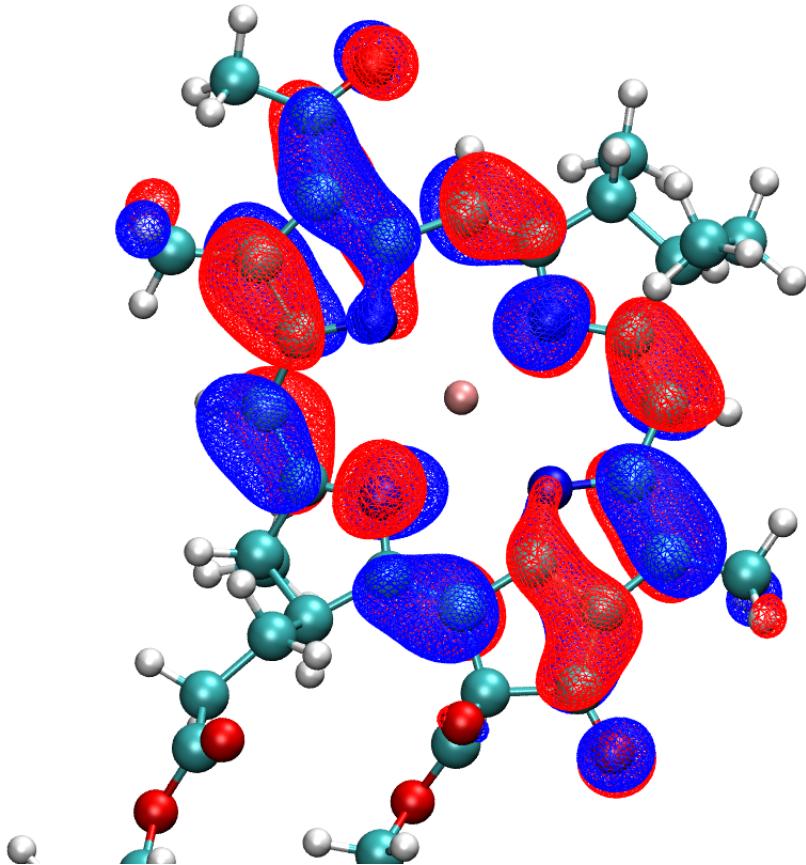


Figure 4.3: The LUMO orbital of BChla from PBE0/Def2-SVP DFT.

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, are also given in chapter 2. Only the charge dependent terms would have any effect on the partial and transition charges, thereby affecting 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. The planned uses for approach do not include calculating optimised geometries or hessians of chlorophyll in this work, as other methods would be expected to much perform better. Including more target properties into the parameter optimisation would decrease the accuracy to any one target, arguably 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.

The combination of specification to chlorophyll transition properties and using xTB electronic structure lends this specific method to be named Chl-xTB. However this is only a realization of an excited state method using the protocol outlined here. Discussion of application to other systems can be found in chapter 7.

#### 4.1.6 Transition and Excited State Density from the Ground State

Similar to  $\Delta$ -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

$$(4.13) \quad \mathbf{D}^{01} = |\Psi^0\rangle\langle\Psi^1|$$

with  $|\Psi^0\rangle$ ,  $|\Psi^1\rangle$  (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

$$(4.14) \quad \mathbf{D}^{(0)} = \sum_i n_i^{(0)} C_{ip}^{(0)} C_{iq}^{(0)}$$

$$(4.15) \quad \mathbf{D}^{(1)} = \sum_i n_i^{(1)} C_{ip}^{(0)} C_{iq}^{(0)}$$

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  $\Delta$ -SCF and eigenvalue difference methods, implying that the inclusion of the HOMO-1-LUMO+1 transition is key to accurately describing the transition density and its magnitude. To recover correct magnitudes, an additional parameter  $D_{scl}$  was included to scale the transition density (i.e.  $\tilde{D}^{Q_y} = D_{scl}D^{01}$  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.

## 4.2 Parameterization

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

### 4.2.1 Objective Function

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

$$(4.16) \quad f_{\text{RMSE}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_i^N (\Delta E_i - \Delta E_{i,\text{ref.}})^2}$$

where  $\mathbf{x}$  is the set of parameters,  $\delta E_i$ ,  $\delta E_{i,\text{ref.}}$  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

$$(4.17) \quad R^2 = 1 - \frac{\sum_i^N (\hat{y}_i - y_i)^2}{\sum_i^N (\hat{y}_i - \bar{y})^2}$$

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

$$(4.18) \quad f_{\text{full}}(\mathbf{x}) = \lambda_1 \text{RMSE}(\Delta E) + \lambda_2 \text{RMSE}(|\mu|) + \lambda_3 (1 - R^2(\Delta E)) + \lambda_4 (1 - R^2(|\mu|))$$

where  $\lambda_n$  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.

## 4.2.2 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 is necessary to use heuristics that can solve non-linear problems.

### 4.2.2.1 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. Minima in the complex space are found by propagating the simplex, either by expanding, contracting, shrinking or reflecting around a central value defined by the set of vertices. For example to find the parameters  $\mathbf{x}$  that correspond to a minimum of the function  $f(\mathbf{x})$

$$(4.19) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

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

$$(4.20) \quad f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$$

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

$$(4.21) \quad \mathbf{x}_r = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1})$$

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 is the best value found so far, and so the simplex should be expanded along the centroid-reflected vertex axis

$$(4.22) \quad \mathbf{x}_e = \mathbf{x}_0 + \gamma(\mathbf{x}_r - \mathbf{x}_0).$$

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

$$(4.23) \quad f(\mathbf{x}) = \begin{cases} \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_r - \mathbf{x}_0) & \text{if } f(\mathbf{x}_r) < f(\mathbf{x}_{n+1}) \\ \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_{n+1} - \mathbf{x}_0) & \text{otherwise } f(\mathbf{x}_r) \geq f(\mathbf{x}_{n+1}) \end{cases}$$

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

$$(4.24) \quad \mathbf{x}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$$

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 criterion 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.

#### 4.2.2.2 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

$$(4.25) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

however with an arbitrary amount of constraint functions  $c$

$$(4.26) \quad c_i(\mathbf{x}) = 0$$

$$(4.27) \quad c_j(\mathbf{x}) \leq 0$$

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  $\alpha_k$  are used to propagate the set of parameters by

$$(4.28) \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

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

$$(4.29) \quad \mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_n^m \lambda_n g_n(\mathbf{x})$$

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

$$(4.30) \quad \min_d f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 \mathcal{L}(\mathbf{x}_k, \lambda_k) d$$

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

$$(4.31) \quad \min_d \|(\mathbf{D}_k)^{\frac{1}{2}} (\mathbf{L}_k)^T d + (\mathbf{D}_k)^{-\frac{1}{2}} (\mathbf{L}_k)^{-1} \nabla f(\mathbf{x}_k)\|$$

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

$$(4.32) \quad \mathbf{L}_k \mathbf{D}_k (\mathbf{L}_k)^T = \mathbf{B}_k$$

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 visualization finding a parameter set  $\mathbf{x}$  corresponding to a local minimum of the Himmelblau function [186] (a standard benchmark for optimisation algorithms) is shown in figures 4.4 and 4.5. 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 4.2.4.

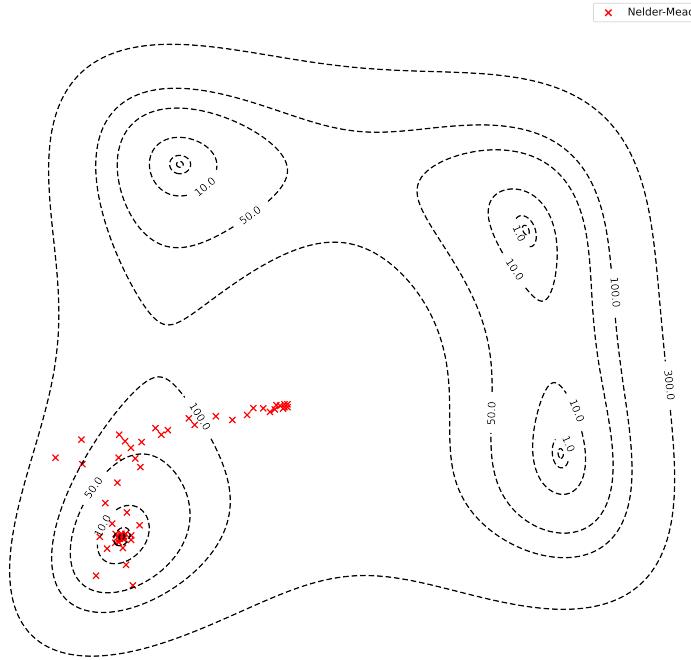


Figure 4.4: An example of the search through solution space using the Nelder-Mead method. The scatter points are positions of the Nelder-Mead simplex vertices, attempting to find the coordinates of the minimum in the Himmelblau function.

### 4.2.3 Reference Data

The geometries for the training set used to optimize the chl-xTB method were taken from molecular dynamics of the LH2 protein used by Manby *et al.* [67]. 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 give appropriate weighting to the differences in binding pockets. The stochastic collection of LH2 snapshots was 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. These data were chosen for the best accuracy-cost ratio, as well as having been previously used to investigate exciton properties for the LH2 system [67]. Additionally, from the outset it was unknown how much training data would be necessary, and so keeping potential future costs of expanding the training

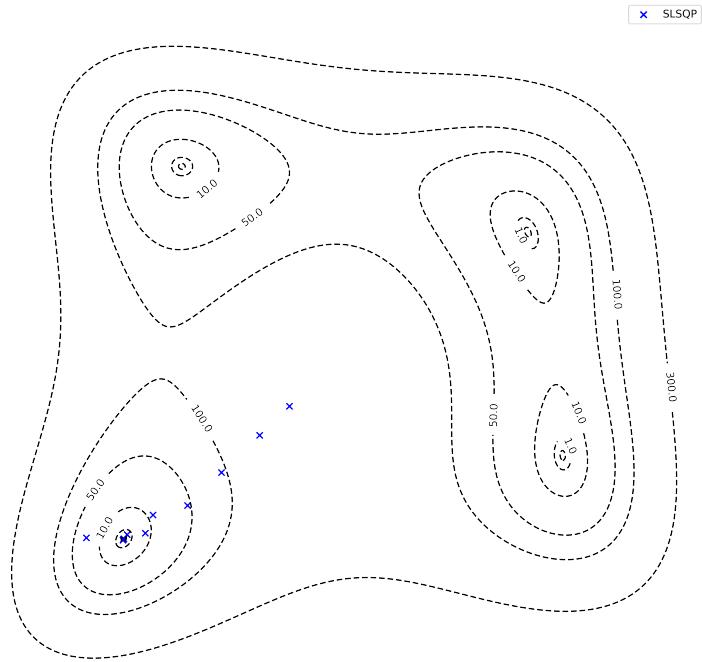


Figure 4.5: A similar example of the search through solution space, but using the SQSLP method. The scatter points are positions of the iterative solution vector  $\mathbf{x}$ , again attempting to find the coordinates of the minimum in the Himmelblau function.

data down was another factor in choosing this functional.

	RMSE( $\Delta E$ ) / eV	$R^2(\Delta E)$	RMSE( $ \mu $ ) / $ea_0$	$R^2( \mu )$
PBE0	0.147	0.795	0.172	0.919
$\omega$ -B97XD	0.053	0.972	0.043	0.993
BLYP	0.098	0.554	0.543	0.080
$\Delta$ -SCF	0.079	0.764	1.283	0.646
$\Delta\epsilon$	0.363	0.790	1.341	0.610
ZINDO	0.449	0.735	1.354	0.623
sTDA-xTB	0.517	0.258	0.843	0.151

Table 4.1: 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 CAM-B3LYP/Def2-SVP TD-DFT data, for transition energies ( $\Delta E$ ) and transition dipole moments ( $|\mu|$ ).

Transition properties were calculated with a range of methods, covering different levels of theory and response methods. These include an eigenvalue difference approach,  $\Delta$ -SCF , and TD-DFT with different levels of theory. This comparison was done so that the performance of any parameterisation results could be benchmarked with an idea about accuracy expectations.

The errors and correlations of the reference data are shown in table 4.1. The methods included in the benchmarking were  $\Delta$ -SCF , TD-DFT and eigenvalue difference all using the PBE0 functional and Def2-SVP basis set. Also included are  $\omega$ -B97XD and BLYP functionals with Def2-SVP basis sets, as a higher and lower level TD-DFT 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 [67]. Also included are ZINDO and sTDA-xTB to represent other efficient response methods.

There are large variations in correlation and RMSE values for the reference methods compared to CAM-B3LYP. This variation sets a reasonable expectation of how well a new method might perform. The  $\omega$ -B97XD functional performs best with an RMSE of 0.053 eV, and is the most correlated (of the TD-DFT methods) with an  $R^2$  value of 0.972. The other TD-DFT methods have similar accuracies ranging from 0.079 eV to 0.147 eV. Good accuracy is not followed by high correlation (see PBE0 with the highest RMSE but highest  $R^2$  past  $\omega$ -B97XD), which 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.  $\Delta$ -SCF performs as well as any of the TD-DFT methods with an RMSE and  $R^2$  values of 0.079 eV and 0.764 respectively. Whilst the eigenvalue difference method has a higher RMSE at 0.363 eV its correlation of 0.79 is

still similar to the TD-DFT methods. 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  $\Delta$ -SCF and eigenvalue difference methods is significantly higher (1.283 a.u. and 1.241 a.u. respectively) than the TD-DFT methods (0.172 a.u., 0.043 a.u. and 0.543 a.u. for PBE0,  $\omega$ -B97XD and BLYP respectively). The average magnitude of PBE0 transition dipoles is 2.751 a.u., with the average for  $\Delta$ -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 the direction of the 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 (0.919 and 0.993 for PBE0 and  $\omega$ -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 0.080, which is fully uncorrelated.  $\Delta$ -SCF and eigenvalue difference show a slight correlation at around 0.6-0.65.

Also included in the benchmarking was the semi-empirical method ZINDO. This had a poor accuracy but reasonable correlation, with RMSE and  $R^2$  values of 0.449 eV and 0.735 for transition energies, and 1.354 a.u. and 0.623 for transition dipoles. sTDA-xTB also performs poorly, with the highest RMSE at 0.517 eV and with a low correlation of 0.258. This supports the arguments made in 1.1.1 and demonstrates how this method is not suitable this study. These two methods act as a benchmark for how well a more efficient response method could be expected to perform.

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 at a usable level of accuracy. Other cross-validations are necessary, and will be discussed in section 4.3, 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.

#### 4.2.3.1 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 the training data. The sizes of each set were 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.

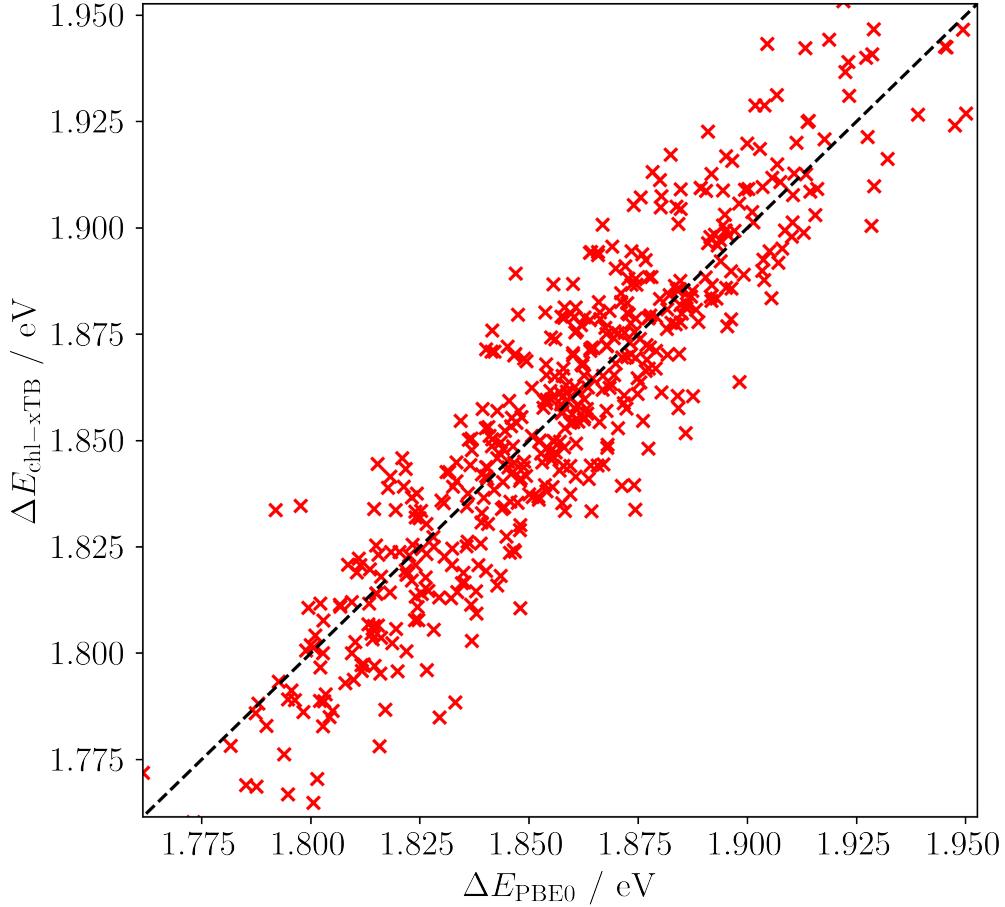


Figure 4.6: Comparison of the  $Q_y$  transition energies predicted from chl-xTB against PBE0/Def2-SVP values.

#### 4.2.4 Results

The chl-xTB method was parameterised to the PBE0/Def2-SVP training data, using the SLSQP method. Overall, chl-xTB performs 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 4.2. Chl-xTB with unoptimized parameters gave RMSEs and  $R^2$  values of 0.490 eV and 0.156 for excitation energies and 0.699 a.u. and 0.402 for transition dipole moments. The best performing set of parameters had an RMSE of excitation energy of 0.037 eV with an  $R^2$  value of 0.878, and an RMSE of transition dipole magnitude of 0.100 a.u. with an  $R^2$  value of 0.664. 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

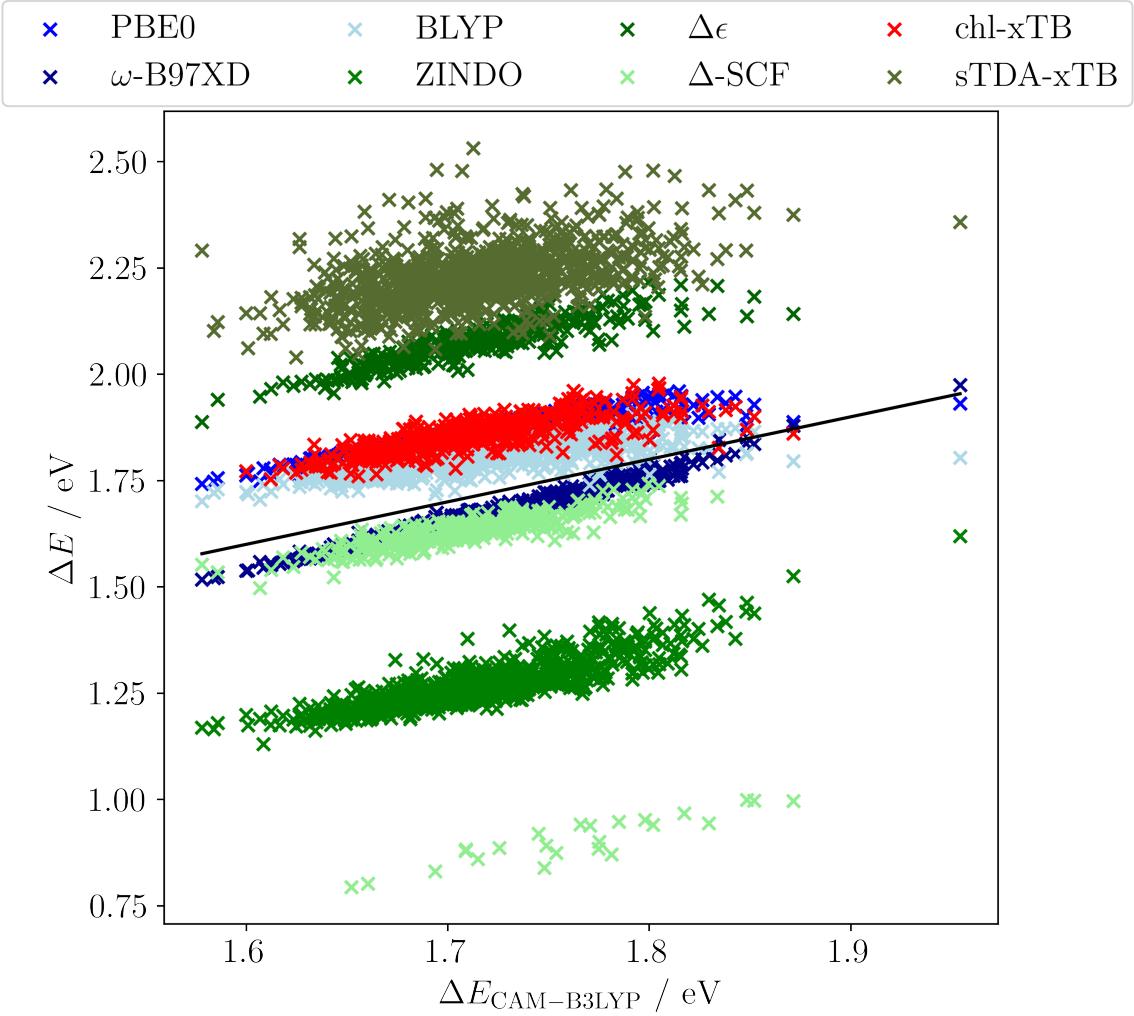


Figure 4.7: Comparison of the  $Q_y$  transition dipole moments predicted from the reference methods as well as chl-xTB (red) against CAM-B3LYP/Def2-SVP values.

the correlation in transition dipole magnitude is low, it is near to the expected correlation from  $\Delta$ -SCF and eigenvalue difference. It can also be seen in figure 4.8 that the variation in Chl-xTB transition dipole magnitude is much smaller than  $\Delta$ -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 [67], 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

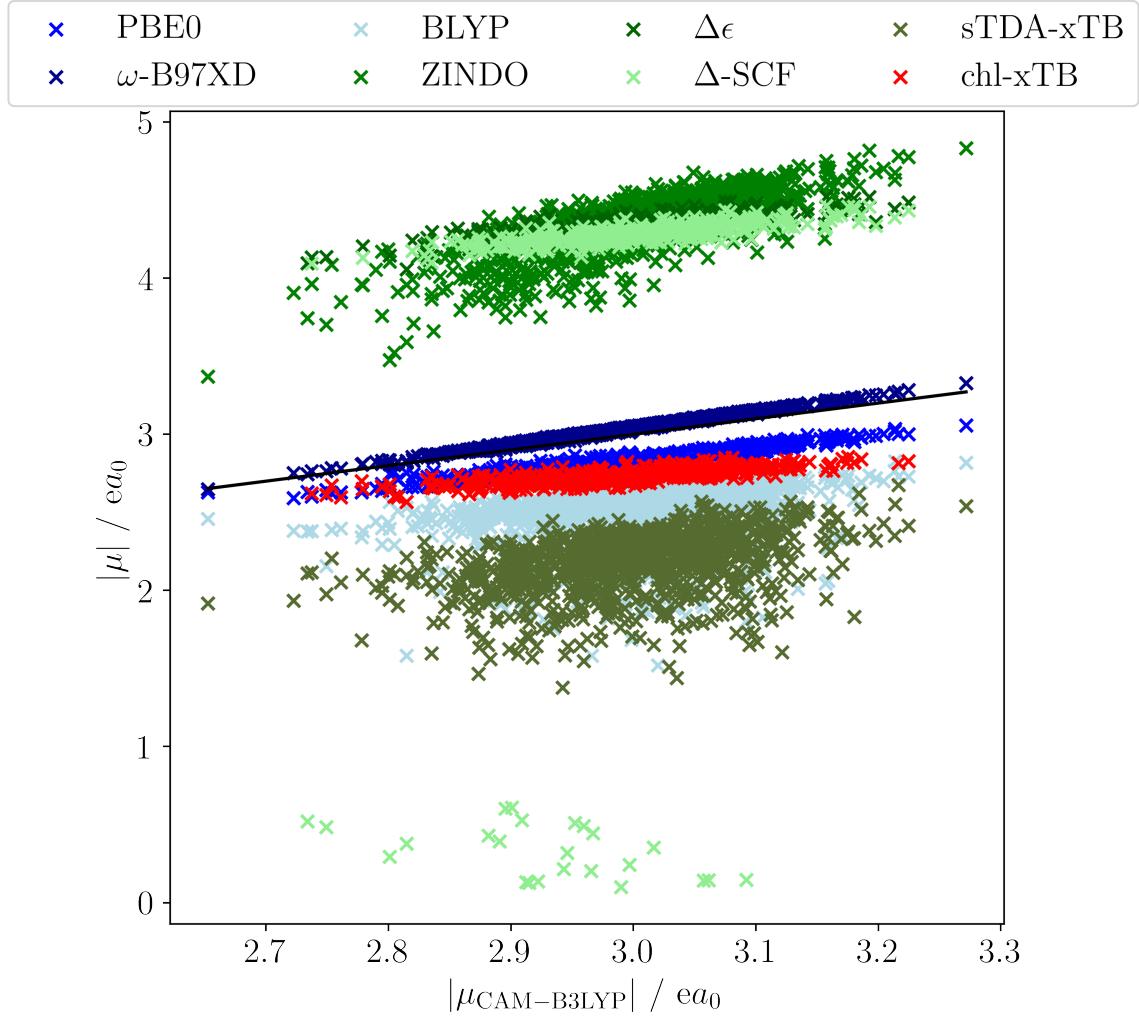


Figure 4.8: Comparison of the  $Q_y$  transition dipole moments predicted from the reference methods as well as chl-xTB (red) against CAM-B3LYP/Def2-SVP values.

accuracy as using both sets of parameters.

The initial guesses for parameters were the corresponding GFN1-xTB and sTDA-xTB parameters, 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 4.2), with the exception of the  $\alpha_x$  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 [187].

In figures 4.7 and 4.8, transition energies and dipole moments calculated using Chl-xTB, sTDA-xTB, and PBE0/Def2-SVP are shown against the CAM-B3LYP/Def2-SVP reference. Compared to this reference, the RMSEs in the transition energies are 0.116, and 0.147 eV for Chl-xTB,

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
$D_{\text{scale}}$	0.636	-

Table 4.2: 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.

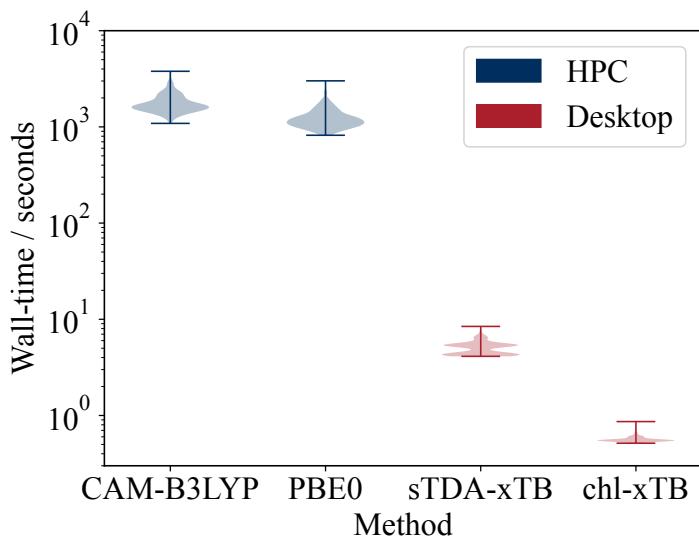


Figure 4.9: Distributions of the compute wall-times between TD-DFT with CAM-B3LYP/Def2-SVP and PBE0/Def2-SVP levels of theory, performed on 20 2.4 GHz Intel E5-2680 v4 CPUs (labelled *HPC*). Wall-times for sTDA-xTB and Chl-xTB, performed on a desktop 2019 2.3 GHz Intel Core i5 Macbook Pro, are shown with the label *Desktop*.

and PBE0/Def2-SVP, respectively. The corresponding spreads of error in each method, captured by  $R^2$ , are 0.637, and 0.795. For transition dipoles, the RMSEs for Chl-xTB, and PBE0/Def2-SVP are 0.267, and 0.172 a.u., respectively. The  $R^2$  values are 0.600, and 0.919. The similarity in these error statistics between Chl-xTB and PBE0/Def2-SVP reiterates Chl-xTB's ability to replicate the performance of the training method. By contrast, the precision and accuracy of sTDA-xTB are noticeably poorer. This is a result of optimizing the Chl-xTB parameters for a more specific problem. By taking such a targeted approach to parameterization, Chl-xTB overcomes the usual compromise between accuracy and computational cost. For the bacteriochlorophyll *a* geometries investigated in figures 4.7 and 4.8, the average times taken for a single CAM-B3LYP or PBE0 TD-DFT calculation were 1852 and 1293 s, respectively, using Gaussian1673 on 20 2.4 GHz Intel E5-2680 v4 CPUs in parallel. sTDA-xTB provides a significant computational advantage, with an average time for each monomer of 5.11 s on a 2019 2.3 GHz Intel Core i5 Macbook Pro. Chl-xTB achieves another order of magnitude speedup with an average runtime of 0.57 s. Distributions of the walltimes are shown in figure 4.9.

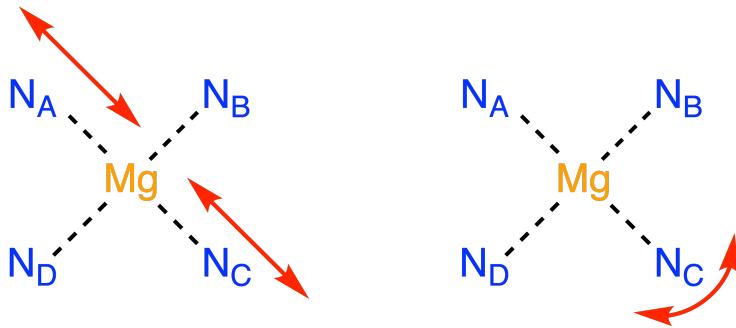


Figure 4.10: Normal modes of the nitrogen-magnesium centre of chlorophyll that have  $D_{4h}$ - $D_{2h}$  (left) and  $D_{4h}$ - $C_s$  (right) symmetry breaking components.

## 4.3 Cross-validation

### 4.3.1 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 to be 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 optimized geometry. The phytol tail in BChla (and chlorophyll in general) make geometry optimizations difficult due to the large number of degrees of freedom, and small energy barrier, in rotations along the carbon chain. Without an accurately optimized geometry for the normal mode hessian, the predicted displacement vectors for the normal modes would be meaningless. 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 has  $D_{4h}$  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 movements of  $N_A$ - $N_C$  atoms which induce  $D_{4h}$ - $D_{2h}$  and  $D_{4h}$ - $C_s$  symmetry breaking are shown in figure 4.10.

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

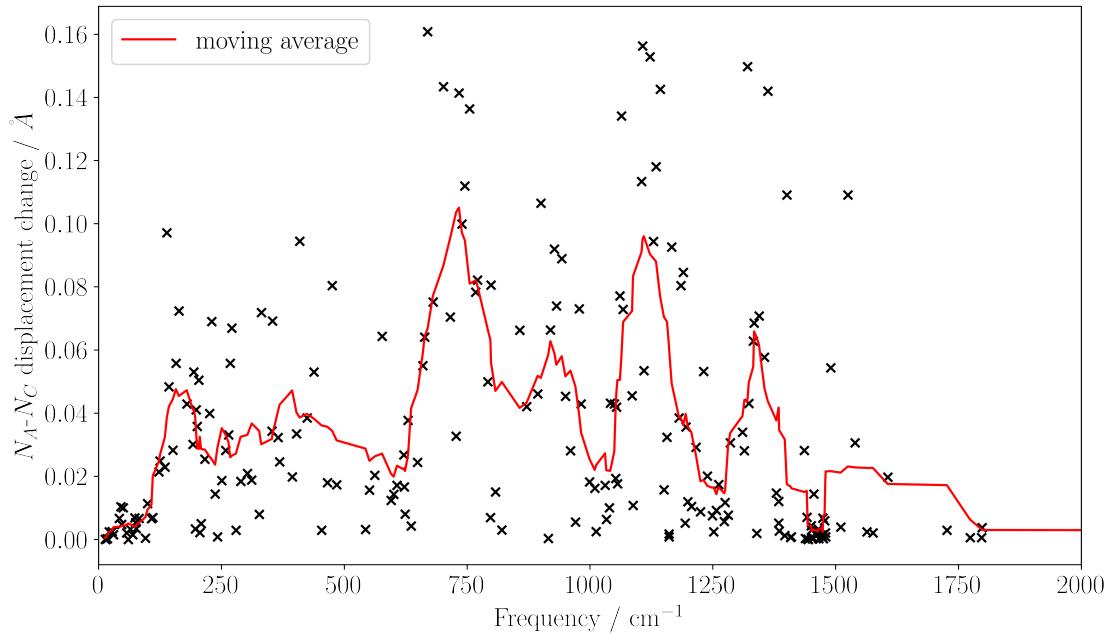


Figure 4.11: Change in the  $N_A$ - $N_C$  displacements along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytyl tail.

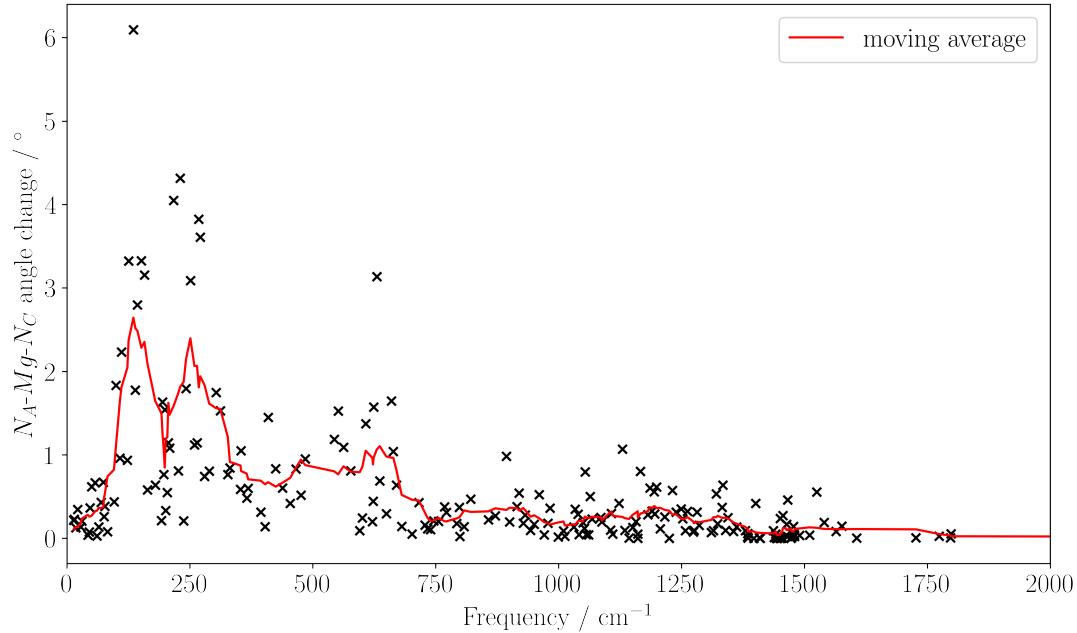


Figure 4.12: Change in the  $N_A$ -Mg- $N_C$  angle along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytyl tail. The smaller variance than in figure 4.11 led this metric to not be used in normal mode choices.

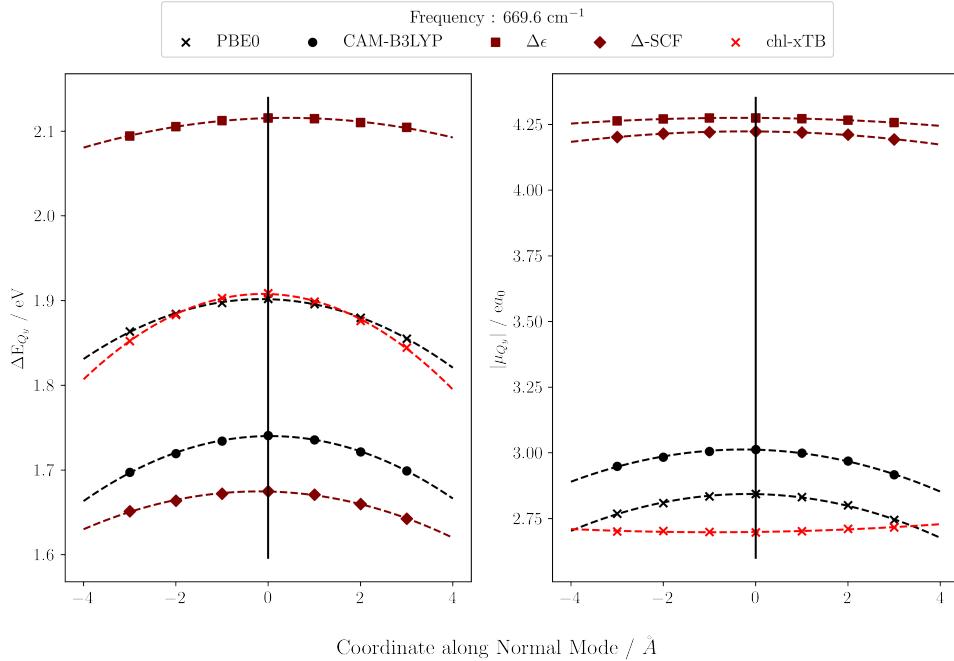


Figure 4.13: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 83rd normal mode (frequency  $669.6 \text{ cm}^{-1}$ ).

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 4.11. 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  $669.6, 701.7, 733.3, 745.5, 755.1, 1105.0, 1107.0, 1122.2, 1142.9, 1320.3, 1361.8 \text{ cm}^{-1}$ , which roughly correspond to previously identified normal modes with frequencies of around  $728$  and  $1156 \text{ cm}^{-1}$  [182].

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  $1 \text{ \AA}$ . This was done up to  $3 \text{ \AA}$  as it was found for most normal modes the energy difference between the optimised geometry and  $3 \text{ \AA}$  displaced geometry was greater than the thermal energy at  $300 \text{ K}$ . Alternatively, in an environment at  $300 \text{ K}$ , a chlorophyll system would not be expected to deform past these boundaries. At each increment, the  $Q_y$  transition energy and dipole were calculated using chl-xTB, as well as TD-DFT with PBE0/Def2-SVP and CAM-B3LYP/Def2-SVP levels of theory,  $\Delta$ -SCF and eigenvalue differences (both using PBE0/Def2-SVP). A quadratic fit was made for each of the response methods and is shown in figures 4.13 to 4.21. Some points for the  $\Delta$ -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.

### 4.3. CROSS-VALIDATION

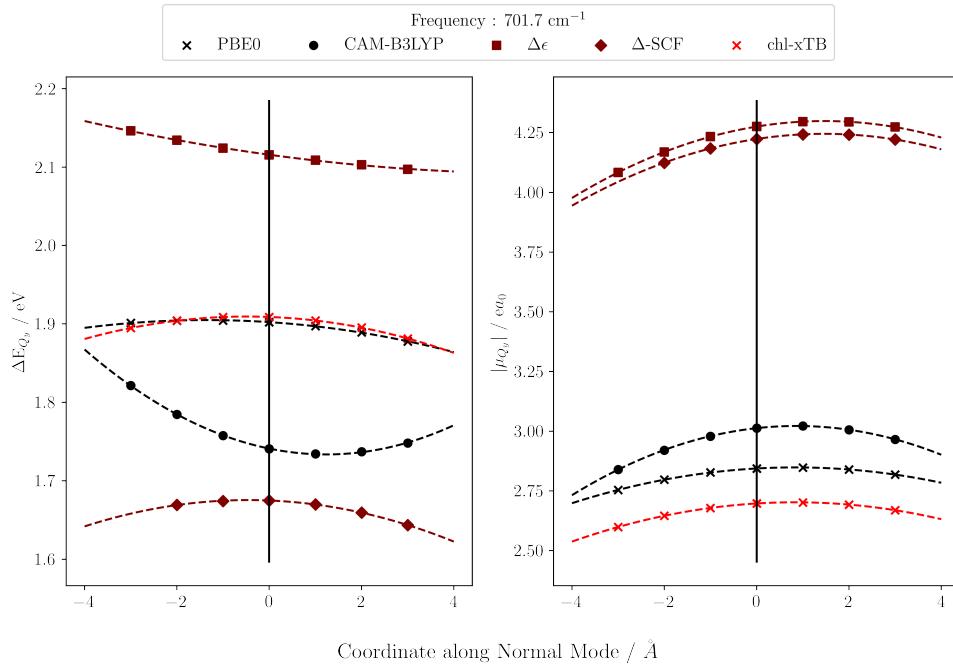


Figure 4.14: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 85th normal mode (frequency  $701.7 \text{ cm}^{-1}$ ).

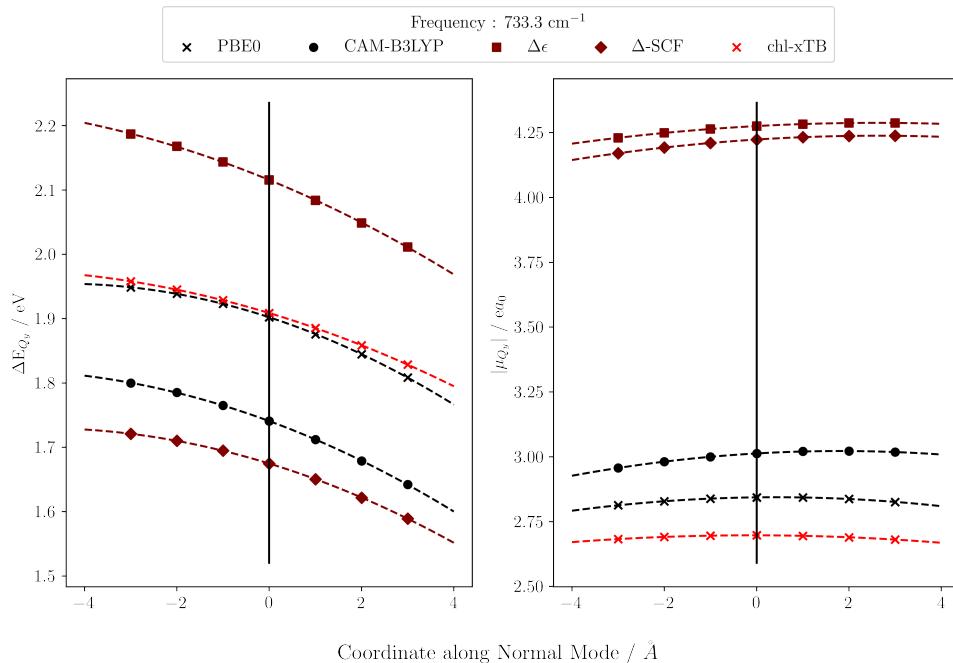


Figure 4.15: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 88th normal mode (frequency  $733.3 \text{ cm}^{-1}$ ).

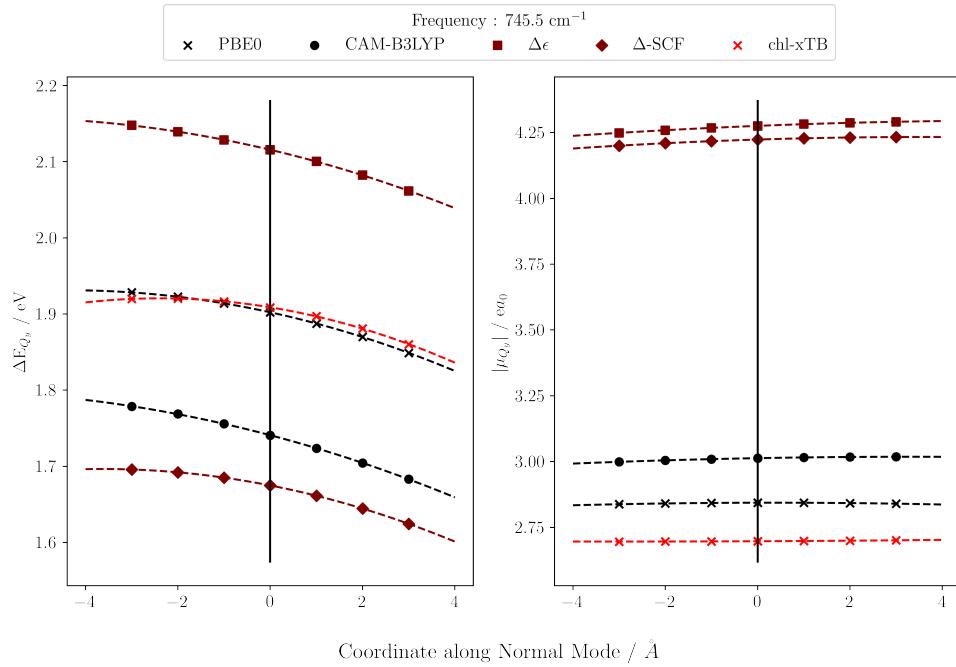


Figure 4.16: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 90th normal mode (frequency 745.5 cm<sup>-1</sup>).

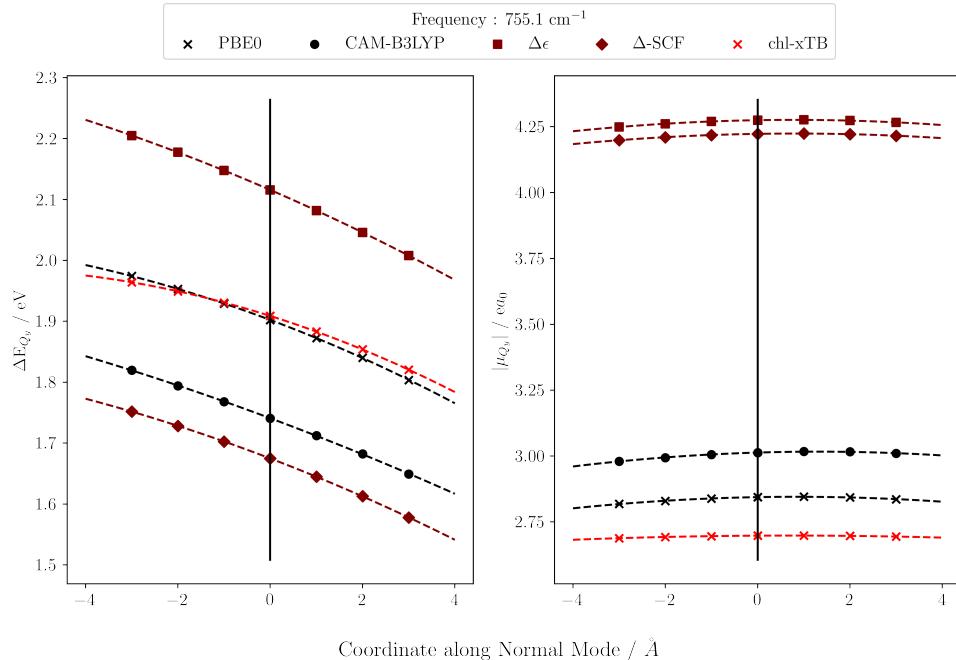


Figure 4.17: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 91st normal mode (frequency 755.1 cm<sup>-1</sup>).

### 4.3. CROSS-VALIDATION

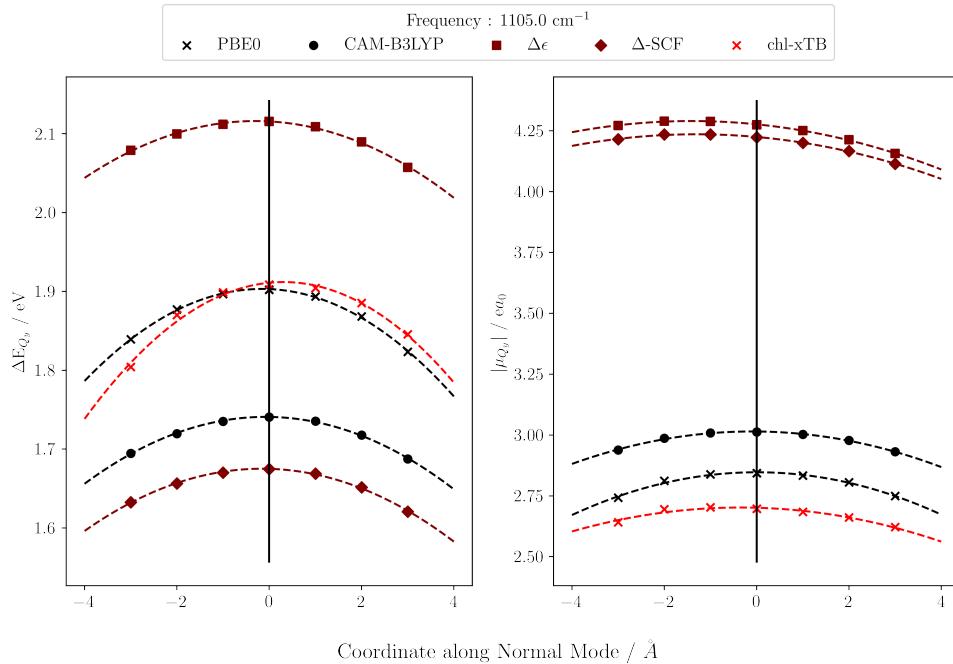


Figure 4.18: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 129th normal mode (frequency 1105.0 cm<sup>-1</sup>).

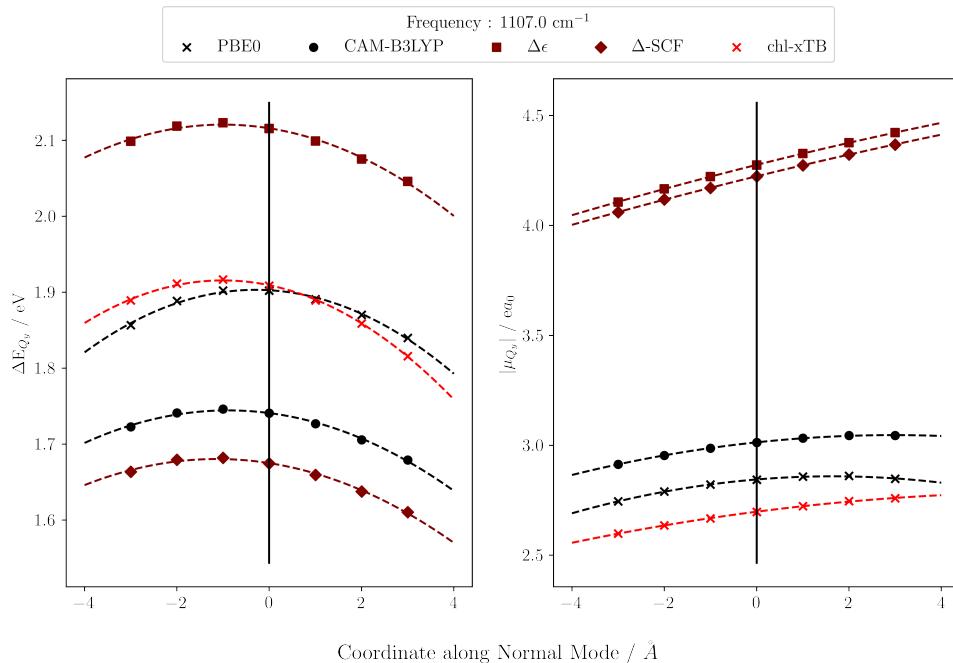


Figure 4.19: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 130th normal mode (frequency 1107.0 cm<sup>-1</sup>).

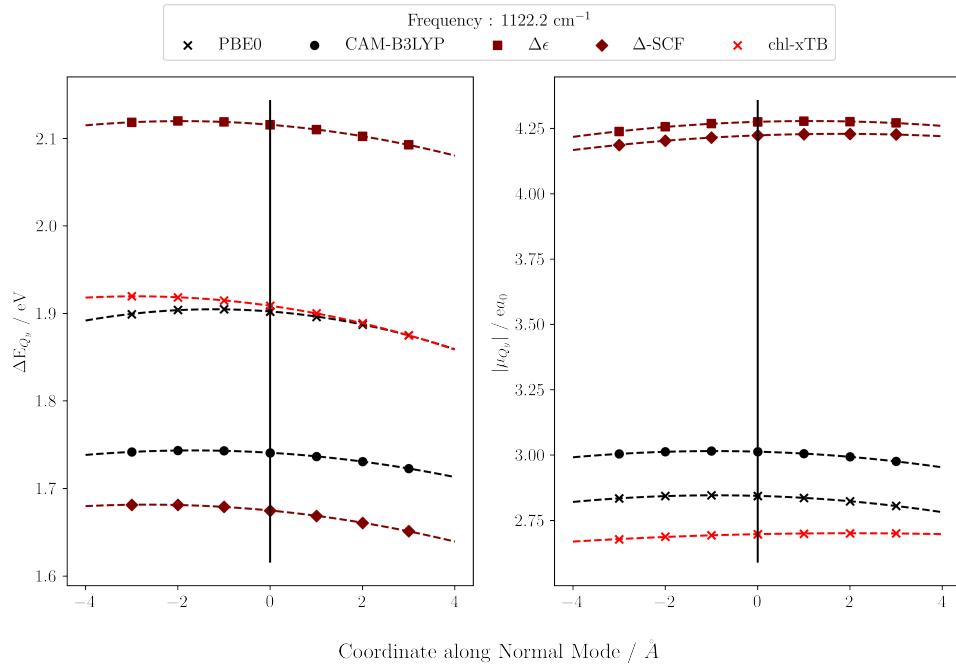


Figure 4.20: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 132nd normal mode (frequency 1122.2 cm<sup>-1</sup>).

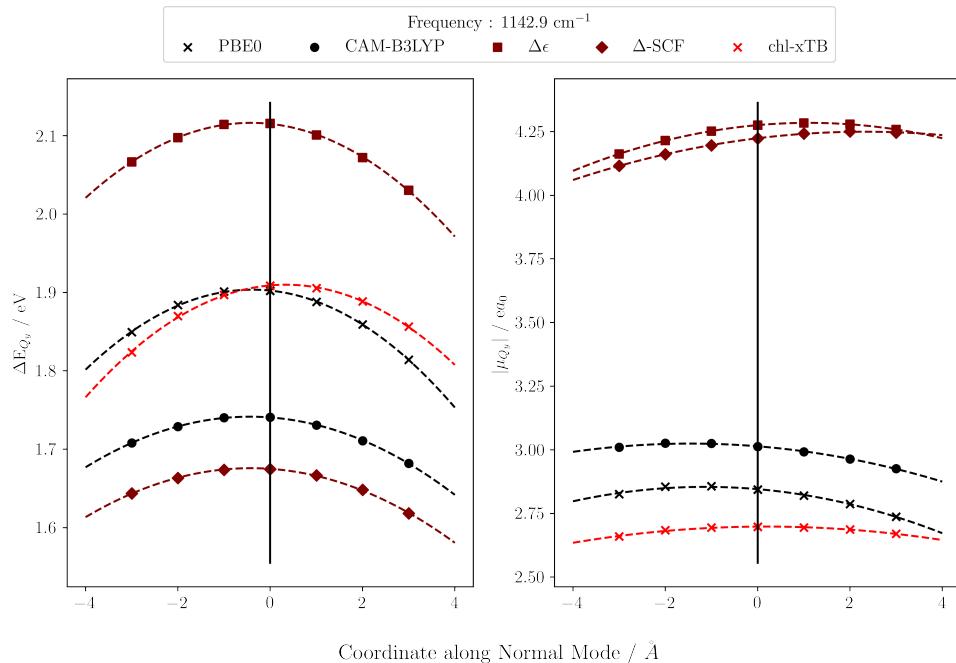


Figure 4.21: Transition energies and dipole magnitudes for the  $Q_y$  transition for geometries of truncated chlorophyll along the 135th normal mode (frequency 1142.9 cm<sup>-1</sup>).

Transition dipole magnitudes are predicted with worse accuracy, but still capture the general behavior 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 are well aligned between PBE0 and chl-xTB, especially compared to  $\Delta$ -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  $\Delta$ -SCF and eigenvalue difference being well above the region where PBE0 and CAM-B3LYP values sit.

### 4.3.2 Predicting Absorption Spectra

So far all of the benchmark 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 behavior 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 chlorophyll systems other than BChla. 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. Force-field parameters for the chlorophyll were taken from a bespoke parameterisation for photosystem II [188], with the rest of the system using the OpenForceField. The structure for chlorophyll was taken from the same source as the bespoke force-field, 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 time-step of 0.5 femtoseconds. The system energy was minimized before running a 10 ps equilibration. Frames were then taken from a 2 ns 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 were also taken from Katz *et. al* [1]. 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 reciprocal space spline term was calculated with the HelpME library. The absorption spectra for each method are shown as the absorption probability at excitation energies in nanometers, normalized such that the areas 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 4.22.

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 line-shape 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. Despite being trained purely on BChl *a* geometries (from *Rsp. Acidophilus*), it is able to predict the spectrum of solvated Chl *a* (from *Thermostichus vulcanus*) to within the same level of expected error (compared to PBE0). Although the chl-xTB line-shape is wider than the experimental spectrum, the CAM-B3LYP and PBE0 lines are also wider and so

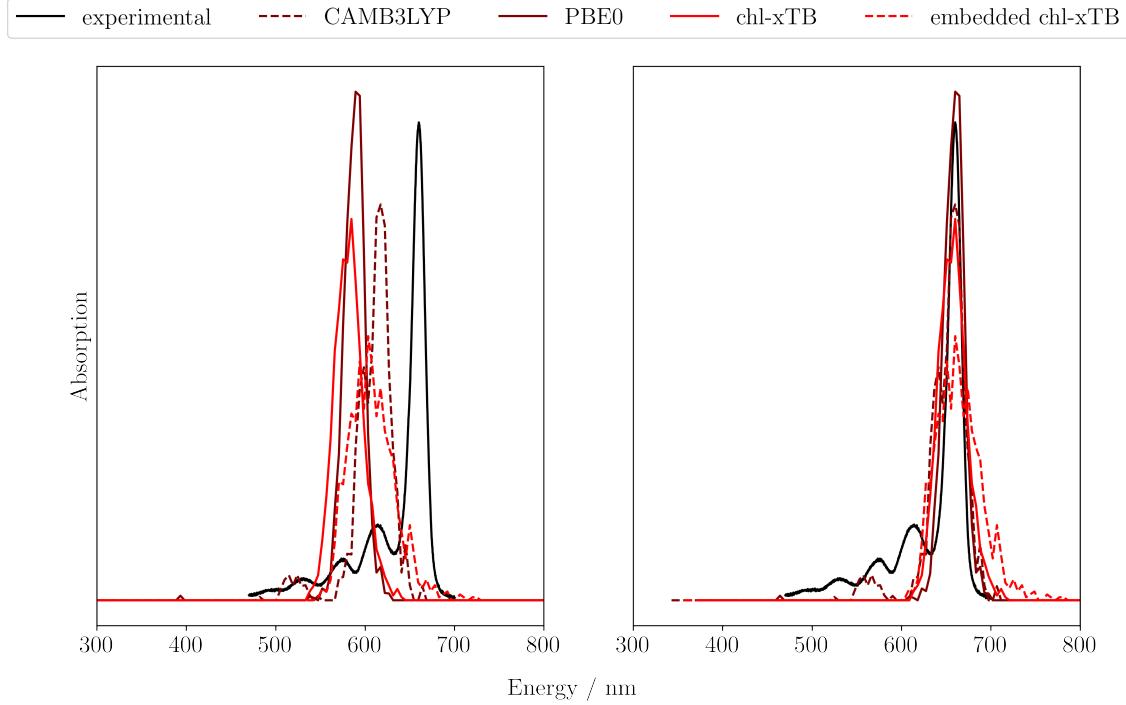


Figure 4.22: Predicted and experimental absorption spectrum of chlorophyll in diethyl ether, only in the  $Q$  band region. Predicted spectra are shown without energy shifts on the left, and with an energy shift to match the experimental absorption maximum on the right. All spectra are normalized to have equal areas.

chl-xTB is still within a reasonable expectation of accuracy.

All of the predicted spectra lack of the lower intensity peaks that are present in the experimental spectrum, resultant from 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 the current scope.

## 4.4 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 such as a lack of accuracy outside LH<sub>2</sub> conformations.

For alternative studies, it may not have been necessary to include both the altered xTB framework and the novel response approximations. If alternative studies require a lower efficiency (i.e. fewer calculations), a low level DFT calculation may have been good enough to achieve the required accuracy. This approach may have also been easier to parameterize, 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  $\Delta$ -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 other than of chlorophyll is outside the scope of this work, it would be beneficial to know how far the optimisation workflow can be pushed. With only a relatively small size of the training data required, reoptimization of the current parameters could be equally short as found with the approach here. 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. By understanding the limits of the protocol in greater detail, it would be known beforehand whether a candidate system and transition could be expected to work. Many studies that require large numbers of calculations could be bootstrapped in a similar fashion.

In terms of modelling light harvesting complexes, Chl-xTB fulfills all the required criteria set out in the introduction. It is 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 allows for good scaling on high performance computers, using chl-xTB in a highly parallelized program (discussed in more detail in chapter 6). 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 and chemically meaningful.

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

---

#### 4.4. CONCLUSIONS

PBE0 values, and well within the error between different high level DFT functionals. It could 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.



## EXCITON MODELS OF CHLOROPHYLL DIMERS

### Previous Published Work

Parts of the work presented in this chapter are also included in a paper published with Dr. Susannah Bourne-Worster and Prof. Fred Manby in January 2023. These are reported in section 5.4.

This chapter investigates whether an exciton framework based on Chl-xTB performs accurately enough for LHC models. Exciton coupling elements, describing the interaction of separate excited states in an oligomer chlorophyll network, are dependent on both transition energies as well as transition densities, as described in section 2.3.

Whilst the methods discussed in the introduction are good examples for calculating exciton Hamiltonian matrix elements, there are drawbacks that inhibit realizing a full, almost *ab initio*, LHC model. For the machine learning methods, this includes explicit embedding effects, as environment interactions can only be included implicitly in the training data. For TD-DFT, the cost of calculations necessitates cutbacks in detail. Additionally the complex interaction of excited states on separate sites makes gradient terms intractable, [67] if dynamic simulations are desired. Wide-reaching semi-empirical methods, such as sTDA-xTB and ZINDO, can be seen in the previous chapter to be too inaccurate to capture more minute details of the geometry dependence of  $Q_y$  transition properties. Given the energy proximity of excited states in the LH2 network, these inaccuracy would be exacerbated when carried forward into exciton Hamiltonians.

Using Chl-xTB as the engine for exciton models offers to solve many of these issues. For the monomer properties examined in the previous chapter, it can be seen to be nearly as accurate as TD-DFT is when comparing different functionals. Correct embedding effects, explicitly excluding from the training set, are accurately regained when using Chl-xTB to new species and environments. The  $\Delta$ -SCF /SPA and semi-empirical inspired approach to excited states has

kept computational efficiency high. Combined, these benefits anticipate LHC models that can accurately capture the interplay of the expansive environment (protein scaffold and solvent) and the complex quantum system inside, with an efficiency that can probe interesting and important phenomena.

A good example of such phenomena is the rate of conversion from vertical excited states (ES) to charge-separated (CS) states in embedded chlorophyll dimers. The ES→CS transfer pathway is an important proposed mechanism for fluorescence quenching in LHCs, which directly impacts the quantum efficiency of harvesting light. Understanding how the design of LHCs inhibits this pathway is key to designing similarly efficient artificial light harvesting materials.

Several other proposed mechanisms of fluorescence quenching have been already been excluded, being easier to model or experimentally determine. For example, impurity-formed trap states, inter-system crossing to triplet states and collisions with quenching quenching species have been ruled out by Porter *et al.* [189]. The concentration-independence of the spectral features of chlorophyll also rules out aggregatechlorophyll formation as a mechanism for quenching [189–191]. It is hypothesised that the CT state is the mostly likely candidate for a trap state, [190, 192, 193] the existence of which is indicated by fluorescence decay kinetics [194].

Currently, established methods are unsuitable for computing the energy surfaces of LHC excited states required to examine electron transfer rates of this complexity in great detail. It follows that, as the protein environment is highly influential to these rates, the environment must be included in the energy surface coordinate space. Including the environment, however introduces a huge number of degrees of freedom, which need to be accurately sampled to obtain energy surfaces. With an efficient, accurate and adaptable method in hand, realizing these states is possible.

The aim of this chapter is to go beyond monomer chlorophyll properties, and utilize the benefits of the protocol established in the previous chapter to realize more detailed models of LHCs. To establish accuracy in this expanded regime, the properties of exciton states, which have nuanced differences to monomer excited states, are examined against established methods. Specifically, accurate descriptions of the dependence of exciton energies with relative orientation and separation of chlorophylls require testing. Once established, this chapter showcases the benefits of the Chl-xTB exciton model for the problem of understanding LHC inhibition of charge-transfer fluorescence quenching.

## 5.1 Frenkel Exciton Hamiltonian

### 5.1.1 Exciton States

As stated in the introduction, a large majority of models of light harvesting systems utilize a Frenkel-Davydov Hamiltonian model of exciton states [7, 37], 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 2, the exciton states  $|\Psi\rangle$  can be constructed as a Hartree product of the states of individual sites (the monomers or chromophores in LHCs)

$$(5.1) \quad |\Psi\rangle = \Pi_m |\phi_m\rangle$$

where  $|\phi_m\rangle$  is the monomer state on site  $m$ . 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

$$(5.2) \quad |\Psi\rangle^i = |\phi_i\rangle^* \Pi_{m \neq i} |\phi_m\rangle$$

where  $|\phi_i\rangle^*$  is the excited state of monomer site  $i$ . States with more than one exciton are possible in this framework, but this is not usual for many light harvesting system models where the interaction between these states is weak. The Hamiltonian, including what is termed the ground state, where there are no excitons in the system, is given by

$$(5.3) \quad H = \begin{bmatrix} E_0 & V_{0,(1,1)} & \cdots & V_{0,(N,1)} \\ V_{0,(1,1)} & E_{(1,1)} & \cdots & V_{(1,1)(N,1)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{0,(N,1)} & V_{(1,1)(N,1)} & \cdots & E_{(N,1)} \end{bmatrix}$$

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

$$(5.4) \quad E_0 = \sum_m e_m + \sum_{m \neq n, A \in m, B \in n} \frac{q_m^A q_n^B}{r_{AB}}$$

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

$$(5.5) \quad E_{(m,1)} = e_m + \Delta e_m + \sum_{n \neq m} e_n + \sum_{n, A \in m, B \in n} \frac{q_m^{*A} q_n^B}{r_{AB}} + \sum_{n, p \neq m, A \in n, B \in p} \frac{q_n^A q_p^B}{r_{AB}}$$

where  $\Delta e_m$  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

$$(5.6) \quad V_{0,(m,1)} = \sum_{n,A \in m, B \in n} \frac{q_m^{\text{tr},A} q_n^B}{r_{AB}}$$

where charges marked  $q^{\text{tr}}$  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

$$(5.7) \quad V_{(m,1),(n,1)} = \sum_{A \in m, B \in n} \frac{q_m^{\text{tr},A} q_n^{\text{tr},B}}{r_{AB}}$$

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.

## 5.2 Benchmarking the exciton model against full dimer excited states

It has been well established in many other studies that the Frenkel exciton Hamiltonian can predict dimer excited state properties well [86, 95]. A similar type of study is 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, such as inter-chromophore distance and treatment of coupling of excited states, can be examined with two studies: first comparing the transition energies predicted by the exciton model against full TD-DFT for a set of chlorophylls from LH2; second examining a conformational scan of chlorophyll dimers along rotational axes, to pinpoint if errors may be due to any specific interactions.

### 5.2.1 Exciton accuracy in LH2 Chlorophyll 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 [67], with the phytyl tails removed for easier

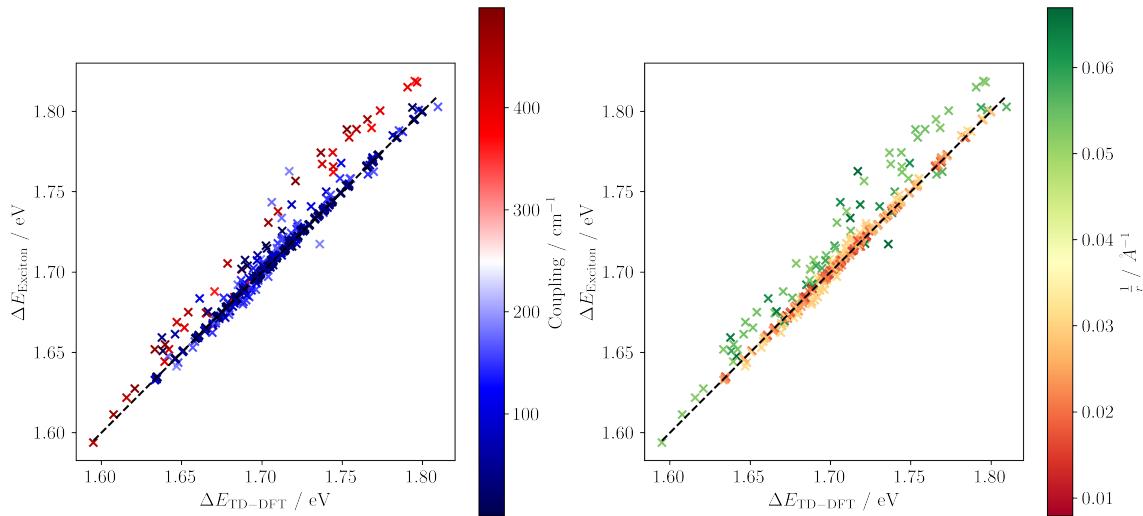


Figure 5.1: The correlation of lowest excitations in LH2 chlorophyll dimer systems predicted by the exciton model constructed from monomer TD-DFT and TD-DFT of the full dimer system. Scatter points are coloured by the coupling value between exciton states (left) and distance between magnesium centres (right).

calculation. Transition energies were calculated for the full dimer systems, as well as from the exciton model, constructed with 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. A scatter plot of these transition energies is shown in figure 5.1.

The RMSE between the exciton and full dimer CAM-B3LYP models is 7.3 meV, well below the threshold of LH2  $Q_y$  transition energy variation (0.2–0.4 eV), and the correlation between these two models is near perfect with an  $R^2$  value of 0.98. The systems in which there is any some can be seen to be cases with small distance between the monomers, with a smaller distance giving a higher error. There is a lower 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 at small separations there are effects that are not included in the exciton model. These effects could be from 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.

### 5.2.2 Systematic Rotation of Chlorophyll Dimer Orientations

Further investigations into the breakdown of the exciton model, especially resulting in inaccuracy in the coupling terms, require a more systematic approach than a stochastic collection of LH2 dimers. Scans of dimer conformations would systematically show whether qualitative descriptions of dimer coupling diverge between the full dimer and exciton models.

As explored in the introduction, many previous investigations treated inter-chromophore coupling with a point dipole interaction

$$(5.8) \quad v_{mn} = \frac{\vec{\mu}_m \cdot \vec{\mu}_n}{R_{mn}}$$

where  $\vec{\mu}_m$  is the dipole on site  $m$  and  $R_{mn}$  is the inter-chromophore distance, due to the large separation of chromophores in LHCs. From this expression it can be seen that there are two inter-chromophore coordinates. These coordinates are the angle between transition dipole moments, and distance between the monomers centres. As the results from the LH2 dimers above explores the relationship with distance, any unexplored effects would be seen through a scan of the relative dimer orientations.

Artificial dimer systems were constructed with controlled angles between the planes of the porphyrin rings and the  $N_A$ - $N_C$  or  $N_B$ - $N_D$  axis, dependent on which axis of rotation was used to subtend the dimers. 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 defined the separation, and as this axis is approximately the cross product of the  $N_A$ - $N_C$  axis and  $N_B$ - $N_D$  axis, which align with the  $Q_y$  and  $Q_x$  transition dipoles respectively, this axis is labelled  $Q_z$ . In a similar fashion, the axis of rotation along the  $N_A$ - $N_C$  axis is termed  $Q_y$  and along the  $N_B$ - $N_D$  axis is termed  $Q_x$ . 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 to 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 without the individual chromophores overlapping. The initial system and the axes of rotation can be seen in figure 5.2.

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 5.3, 5.4, 5.5. 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 trends in curvature are also similar. The only major discrepancy is seen in the 150-200 ° region in the  $Q_x$

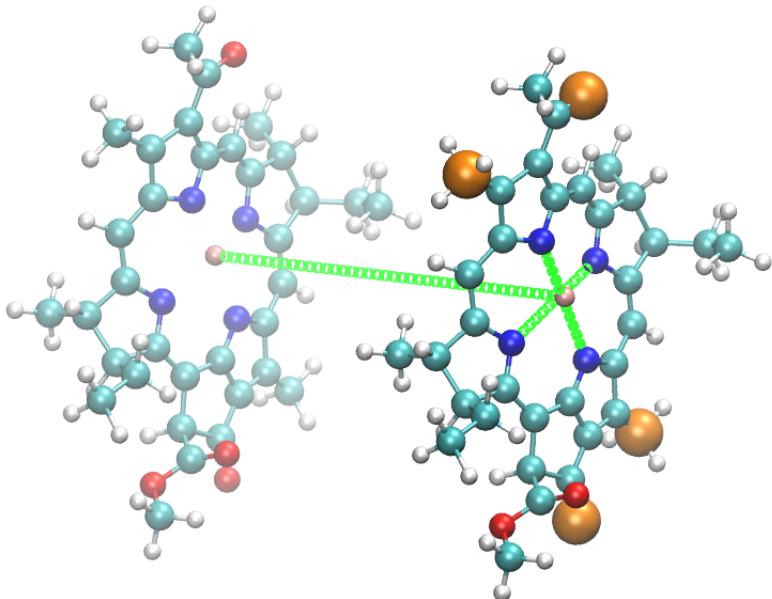


Figure 5.2: Diagram of the artificial dimer system, showing the axes of rotations in green and the chosen sites for calculating distance between functional groups in orange.

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 behavior 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.

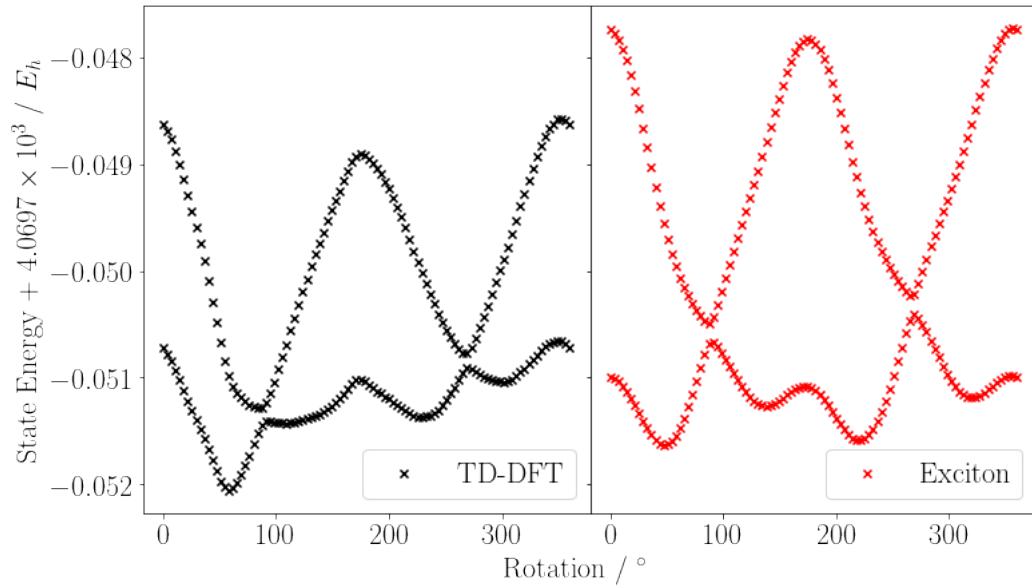


Figure 5.3: The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the  $Q_z$  axis.

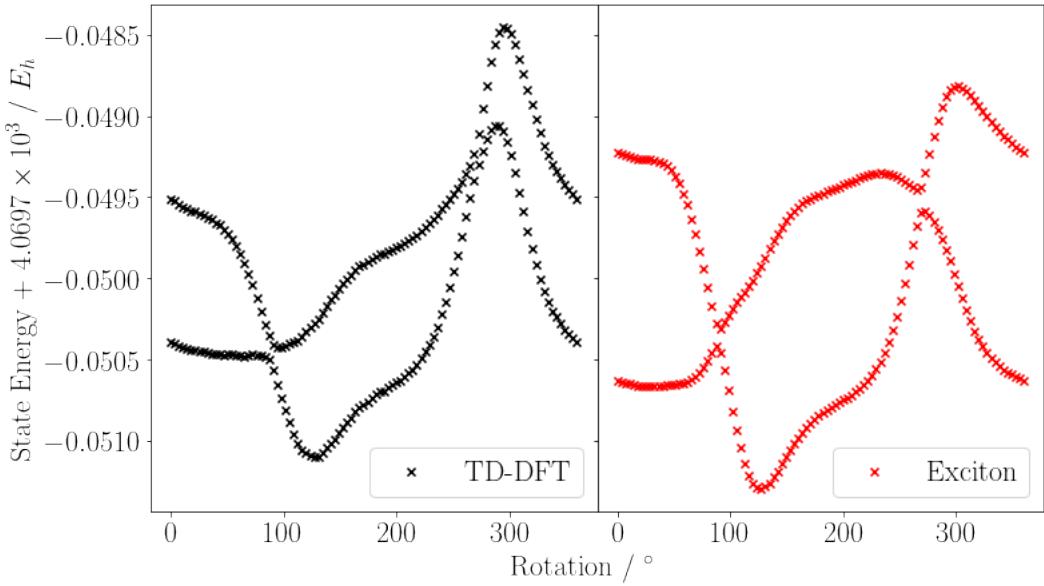


Figure 5.4: The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the  $Q_x$  axis.

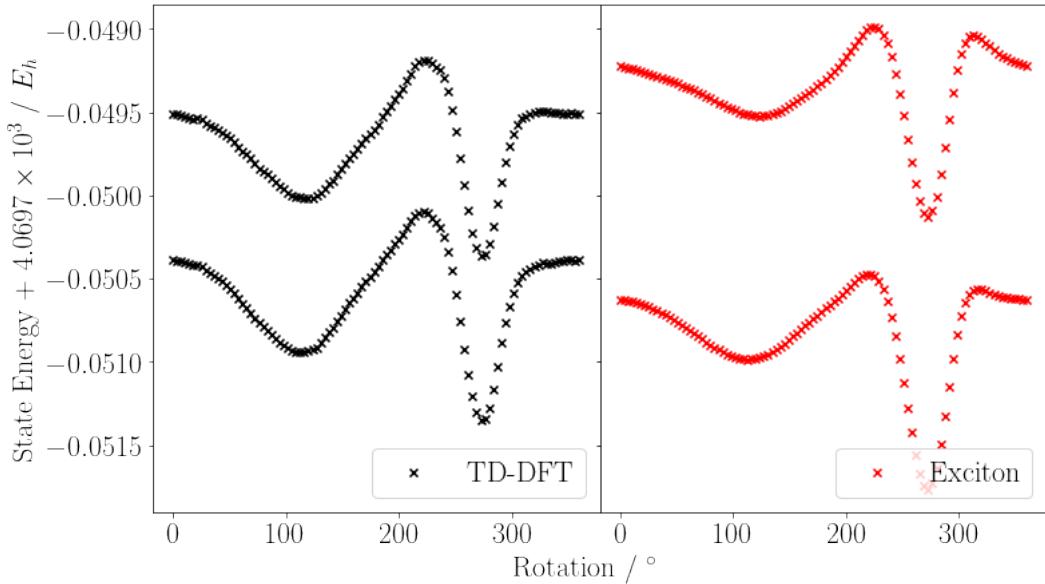


Figure 5.5: The dimer excited state energies predicted by full dimer TD-DFT CAM-B3LYP/Def2-SVP (left) and the exciton model (right) for chlorophyll dimer geometries with one monomer rotated around the  $Q_y$  axis.

## 5.3 chl-xTB Excitons

### 5.3.1 Benchmarking the Chl-xTB Exciton Framework

#### 5.3.1.1 Discarding PBE0

While in the previous section it was possible to compare excited state energies with the same CAM-B3LYP functional for both 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 transition energies had systematic errors, which are 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, attributed to the long-range corrections. This effect can be seen in the distribution of transition energies, shown in figure 5.6. Here the individual  $Q_y$  and higher energy transitions that are well resolved by CAM-B3LYP are not well resolved when using PBE0. Transition energies from PBE0 are uncorrelated to CAM-B3LYP transitions, as seen in figure 5.7, due to this charge-transfer artifact. Hence the PBE0 dimer data does 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.

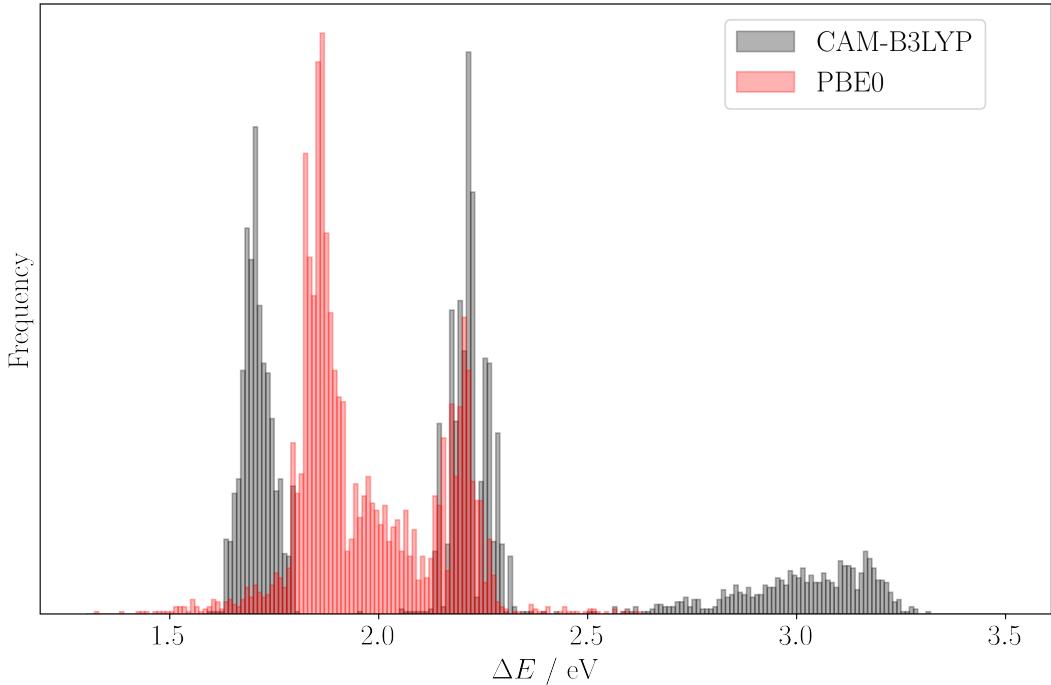


Figure 5.6: Distributions of the lowest 5 transition energies of chlorophyll dimers predicted by TD-DFT with the CAM-B3LYP functional (grey) and PBE0 functional (red).

### 5.3.1.2 Assignment of States

In the comparison between CAM-B3LYP full dimer TD-DFT and exciton model, both method gave excited states with the same energy ordering. Assigning and comparing the same excited states was therefore 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 a state with the exciton localized on monomer A as the lower energy states, 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 that this effect apparent.

It is 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 was localised. This centre  $\mathbf{c}$  was calculated by taking the average of the charge positions  $\mathbf{r}$  weighted by the absolute transition charge value  $|q^{\text{tr}}|$

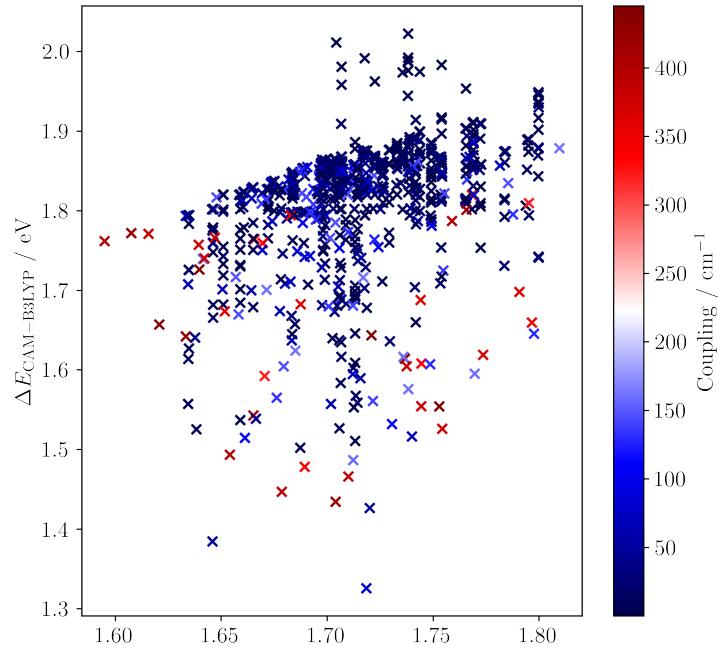


Figure 5.7: Scatter of the two lowest transitions from PBE0 TD-DFT against CAM-B3LYP TD-DFT, coloured by the coupling value from the exciton model using PBE0 monomer data.

$$(5.9) \quad \mathbf{c} = \frac{\sum_i |q_i^{\text{tr}}| \mathbf{r}_i}{\sum_i |q_i^{\text{tr}}|}$$

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 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 localized on monomer A.

### 5.3.1.3 Comparison

Figure 5.8 shows transition energies from 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

5.1). For the PBE0 model, the RMSE and  $R^2$  is 0.296 eV and 0.75 respectively, with the Chl-xTB model performing similarly with RMSE and  $R^2$  of 0.210 eV and 0.66 respectively. Whilst the correlation is not as good as the CAM-B3LYP exciton model, 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 RMSEs 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 5.1, it was concluded that the correlation of close proximity of monomers, related to high coupling values, to the overestimation of transition energy implied that the exciton model was the leading cause of error. Hence it is implied 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. The origin of error being from the monomer method then explains the behavior 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 3. 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 would still be expected to perform well in investigations on light harvesting complexes.

#### 5.3.1.4 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 benchmarking could not be done 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

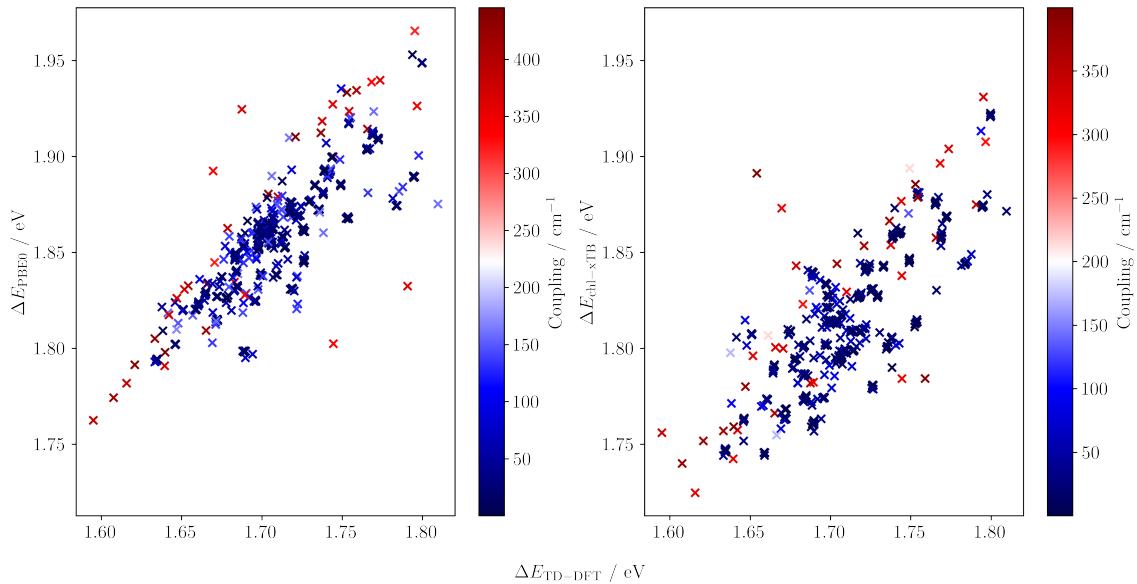


Figure 5.8: Transition energies of the exciton states calculated using PBE0 TD-DFT monomer data (left) and chl-xTB monomer data (right) against CAM-B3LYP full dimer TD-DFT transition energies, coloured by the coupling values.

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 conformations is fairly narrow, especially for nearest neighbours. In contrast, in this study the intra-chromophore geometry is constant, and the inter-chromophore geometry is changed. Using static monomer geometries 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 is in the coupling values, which were predicted much larger than CAM-B3LYP analogues in earlier versions of chl-xTB. Observing that the transition dipoles were also much larger in these earlier versions of Chl-xTB, an artifact also observed in  $\Delta$ -SCF and the eigenvalue difference methods of the previous chapters, the transition density matrix scaling factor was then included. Without removing the variation due to intra-chromophore geometry this effect may not have been clear.

As the exciton framework has been shown to break down at small separations, it was postu-

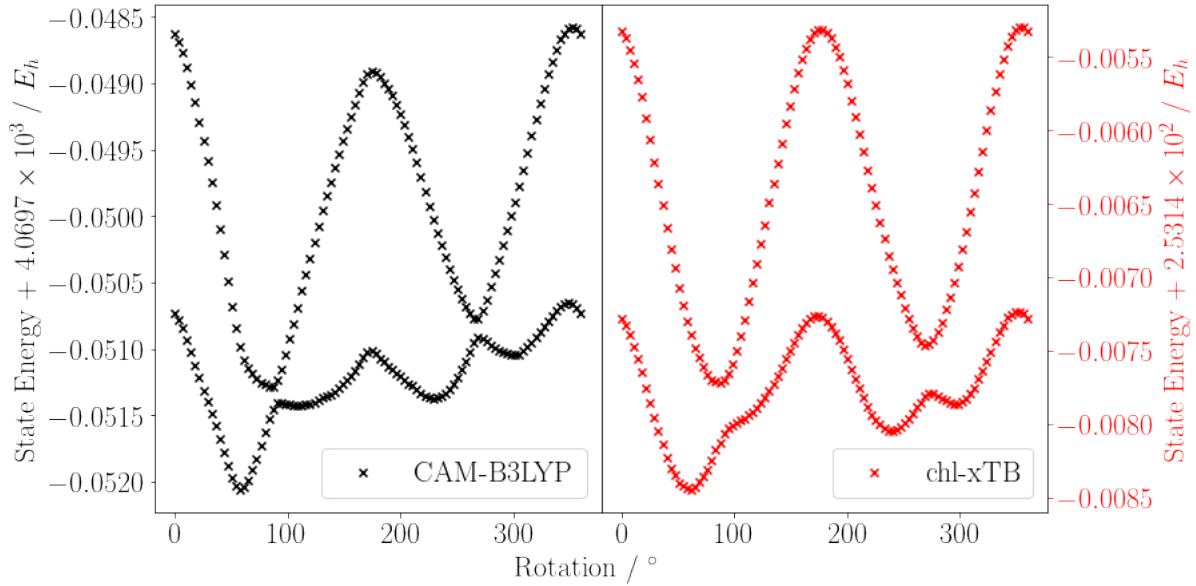


Figure 5.9: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the  $Q_z$  axis.

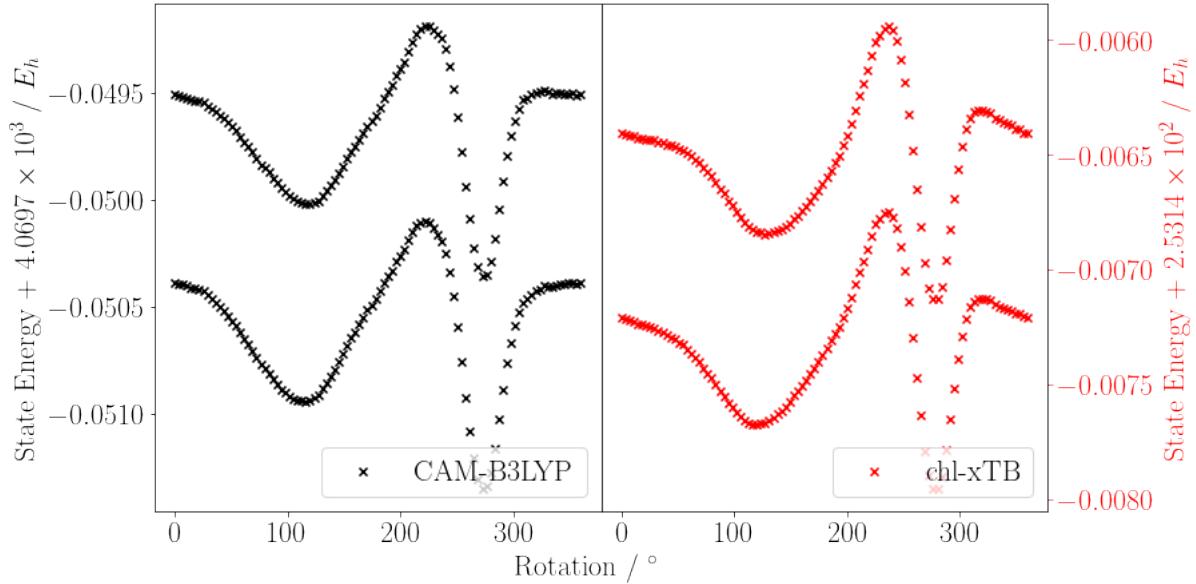


Figure 5.10: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the  $Q_y$  axis.

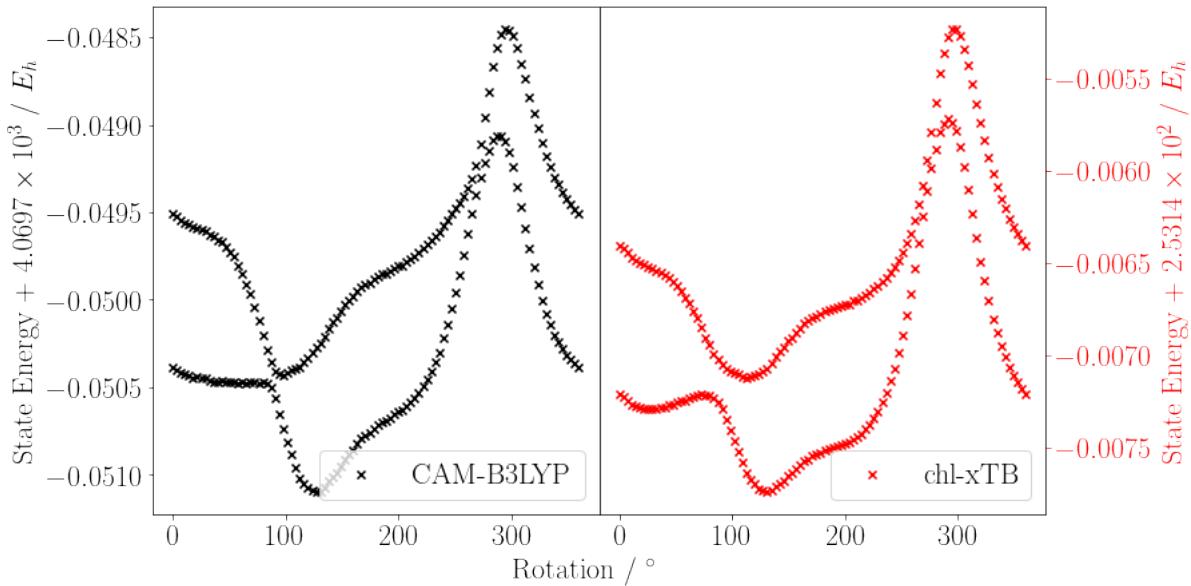


Figure 5.11: Dimer excited state energies predicted by full dimer TD-DFT (left) and the exciton model (right) constructed from chl-xTB monomer data for chlorophyll dimer geometries with one monomer rotated around the  $Q_x$  axis.

lated 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 centres of functional groups which have the most amount of transition density are shown in figure 5.2 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 distance of the closest functional groups on the porphyrin ring is shown in figures 5.12, 5.14 and 5.13. 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.

Both the accuracy in monomer properties, as well a chemically meaningful formalism has enabled Chl-xTB exciton frameworks to be perform well when compared against TD-DFT methods.

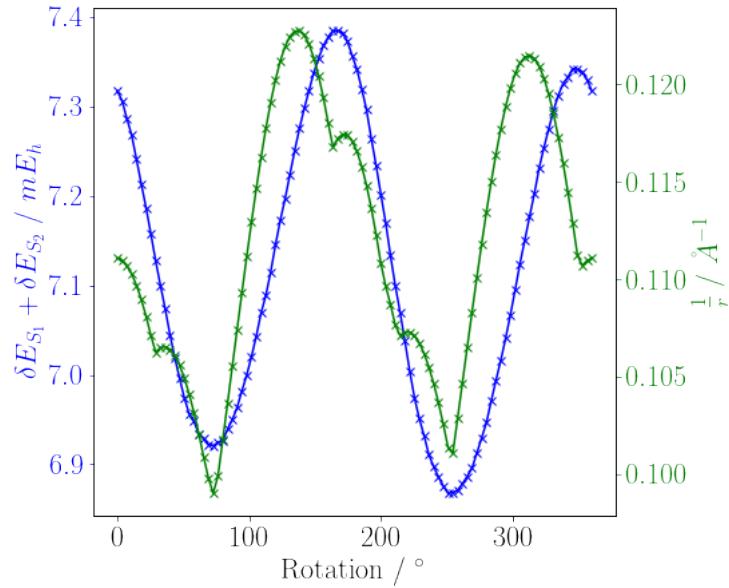


Figure 5.12: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the  $Q_z$  axis.

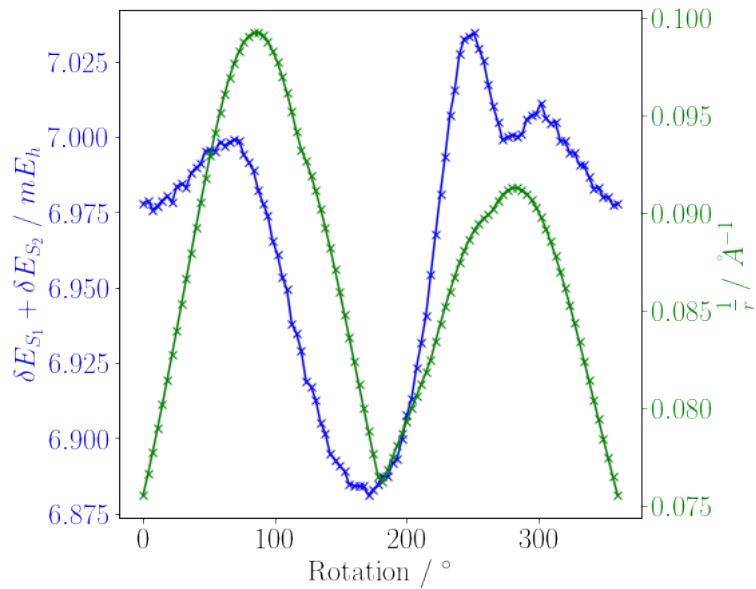


Figure 5.13: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the  $Q_y$  axis.

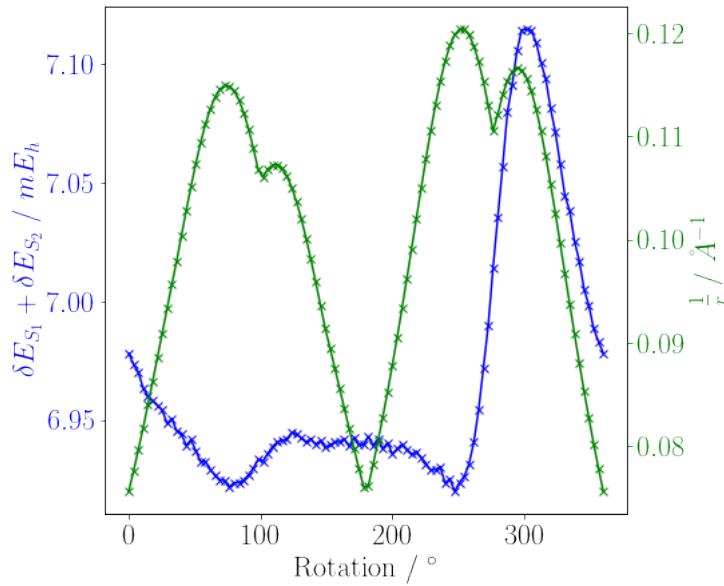


Figure 5.14: Profiles of the sum of errors between chl-xTB exciton transition energies and full CAM-B3LYP TD-DFT (blue), alongside the greatest reciprocal distance of the functional group sites at each angle (green) for rotations along the  $Q_x$  axis.

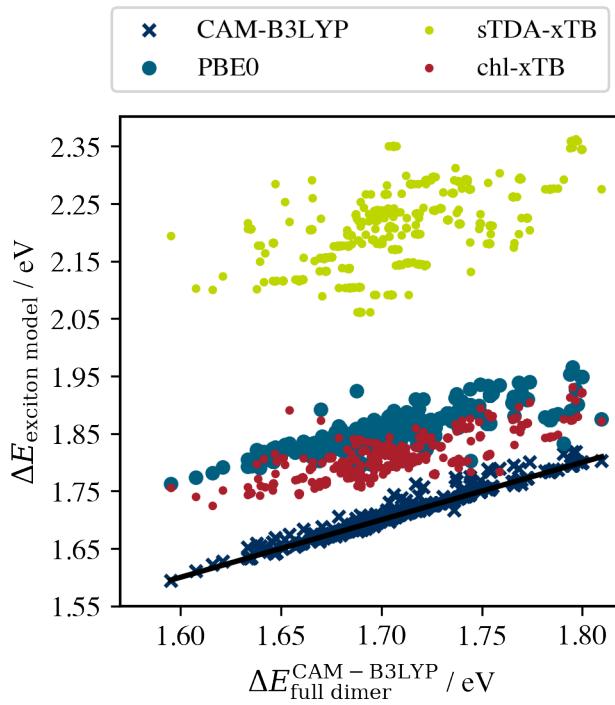


Figure 5.15: Transition energies from exciton models constructed with CAM-B3LYP/Def2-SVP, PBE0/Def2-SVP TD-DFT, Chl-xTB and sTDA-xTB monomer properties, plotted against the corresponding transition energies from full dimer TD-DFT at a CAM-B3LYP/Def2-SVP level of theory.

Alternative approaches based on a similar semi-empirical approach, such as using sTDA-xTB to predict exciton transition energies, are outperformed by this framework. Transition energies for the LH2 dimer test-set from section 5.2.1 were calculated using sTDA-xTB for chlorophyll monomer transition properties. Only a point-dipole coupling method was available with this framework, as transition densities or atom-resolved transition charges are not readily accessible. These energies are shown in figure 5.15.

## 5.4 Prediction of the Rate of Electron Transfer in Concentration Quenching

Efficiently calculating exciton network energies opens new avenues of modelling LHC properties. With more efficient and accurate calculations, coordinate spaces with large numbers of degrees of freedom can be confidently sampled, extracting the underlying chemical and physical structures that define features in these spaces. More specifically, utilizing the more efficient sampling that Chl-xTB provides, it is possible to map the energy surfaces of embedded chlorophyll dimer states, where the dimers are actively affected by the surrounding solvent (evidenced by the discrepancy in fluorescence quenching seen in solvated chlorophyll and chlorophyll networks in LH2 [195]). From these surfaces some of the rate constants of the proposed mechanism for fluorescence quenching can be predicted and systematically varied with solvent and protein environment. It should be noted that the high number of degrees of freedom in these surfaces directly originate from the effect of environment on chlorophyll geometry conformation. Therefore, a systematic scan of a reaction coordinate would not be possible, due to the sheer number of atoms involved, and accurate construction of these surfaces requires the three main benefits of the Chl-xTB framework (efficiency, accuracy, and inclusion of embedding effects).

According to Marcus theory [196], the rate constant of electron transfer from the ES to CS state depends upon the free energy change  $\Delta A$  and reorganization energy  $\lambda$  of said transition, and in the following way:

$$(5.10) \quad k_{\text{CS}} = \frac{2\pi}{\hbar} |H_{AB}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} \exp\left(\frac{-(\lambda + \Delta A)^2}{4\lambda k_B T}\right),$$

where  $H_{AB}$  is the coupling value between the ES and CS states,  $k_B$  is the usual Boltzmann constant and  $T$  is the temperature.  $\Delta A$  and  $\lambda$  are calculated from the minima of the ES and CS free energy surfaces (FESs) (see figure 5.17). Due to the dependence of these surfaces on both the internal geometry of the dimers and the surrounding solvent system, it is not possible to discern a simple reaction coordinate to perform systematic scans. However, because each electronic state varies differently as a function of chromophore and solvent configuration, the changing energy gap between states ( $E_{\text{ES}} - E_{\text{CS}}$ , where  $E_{\text{ES}}$ ,  $E_{\text{CS}}$  are the energies of the photoexcited and charge-separated state respectively) can be used as a proxy for the reaction coordinate. FESs

## 5.4. PREDICTION OF THE RATE OF ELECTRON TRANSFER IN CONCENTRATION QUENCHING

---

can be reconstructed from a distribution of these proxy values, calculated from a large volume of stochastic chlorophyll dimer geometries.

The distribution of the differences  $E_{ES} - E_{CS}$  were well fit by a normal distribution function

$$(5.11) \quad P(\Delta E) = \exp\left(-\frac{1}{2}\left(\frac{\Delta E - \mu}{\sigma}\right)^2\right)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of the distribution respectively, and  $\Delta E$  is the excitation energy for either the ES or CS state. From this, the FESs are quadratic, having the form

$$(5.12) \quad V(\Delta E) = \frac{kT}{2} \left(\frac{\Delta E - \mu}{\sigma}\right)^2.$$

Assuming the FESs are quadratic is consistent with studies on bacteriochlorophyll pairs in the photosynthetic reaction centre of *Rb. Sphaeroides* [197].

The rate constant  $k_{ES-CS}$  was calculated for a series of chlorophyll dimers at a constant separation, with MD simulations used to generate thermally distributed geometries of the dimer systems. A constraint was used to keep a dimer pair 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, following the work by Beddard and Porter which suggested that a critical distance of ~ 10 Å is required for the chlorophyll dimer to start exhibiting the charge-separation trap states proposed for quenching at a competitive rate [198]. For each separation, the dimer systems were explicitly solvated in 100 Å diethyl ether boxes. These boxes were prepared using **Packmol**, and force-field parameters were taken from OpenForceField version 1.3.0. Each system was equilibrated for 10 ps at 300 K with a Langevin Integrator, with a production simulation of 500 ps performed afterwards.

A  $k_{ES-CS}$  rate constant was also calculated for a pair of chlorophylls embedded in the LH2 protein. The average separation of nearest neighbours in LH2 is ~ 9 Å. The LH2 geometries were generated using the same MD method as the diethyl ether systems, but using the LH2 structure and force-field developed by Ramos *et al.* [21]. For both sets of environments, snapshots were taken every 1 ps for 500 total geometries.

Dimer excited states were calculated for all of the geometries, using the Frenkel exciton Hamiltonian and Chl-xTB. To account for environmental effects, the solvent and other chlorophyll(s) were included in point charges interactions (with charge values taken from the force-fields used) to polarize the Chl-xTB calculations. Means  $\mu$  and standard deviations  $\sigma$  of the symmetric normal distribution in equation 5.11 were fit to the series of lowest energy exciton states, to reconstruct the FES. Fitting  $\mu$  and  $\sigma$  was done with the `scipy.stats` module. Good fits were achieved with this method, justifying the assumption that the FESs follow a quadratic form (see figure 5.16).

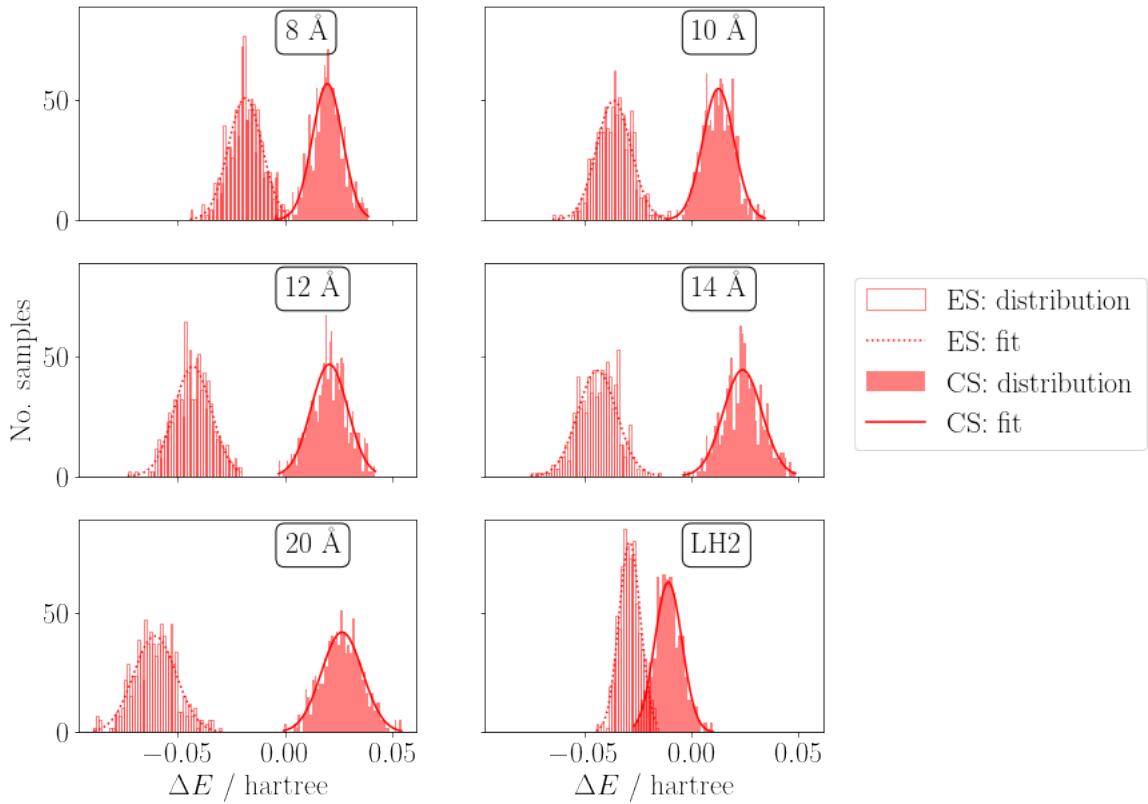


Figure 5.16: Distributions of vertical energy gaps  $\Delta E$  and fitted normal distribution functions for each solvated Chl dimer system and LH2 Chl dimers.

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 [199, 200], 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 phytyl tail removed were used to make the FODFT calculations feasible.

The rate constant calculated for the 10 Å dimer, in diethyl ether, was  $0.03 \text{ ns}^{-1} \pm 0.01 \text{ ns}^{-1}$ . As the fluorescence rate  $k_f$  is given as  $0.2 \text{ ns}^{-1}$  [201], charge separation is a competitive pathway for quenching vertically excited states. The  $0.03 \text{ ns}^{-1}$  value also assumes a full relaxation of the geometries to a new equilibrium. Taking values for  $\Delta A$  and  $\lambda$  from the reaction coordinate at initial photo-excitation (i.e. where the ground state dimer distribution would sit) gives a rate constant of  $0.04 \text{ ns}^{-1}$ . Other considerations, such as the destabilization of the CS state from overestimated polarisation in the FF methods or the constant value for  $H_{AB}$ , 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 increase in  $\lambda$  and  $\Delta A$  values for 8 Å, 10 Å, 12 Å, 14 Å and 20 Å shown in figure

## 5.4. PREDICTION OF THE RATE OF ELECTRON TRANSFER IN CONCENTRATION QUENCHING

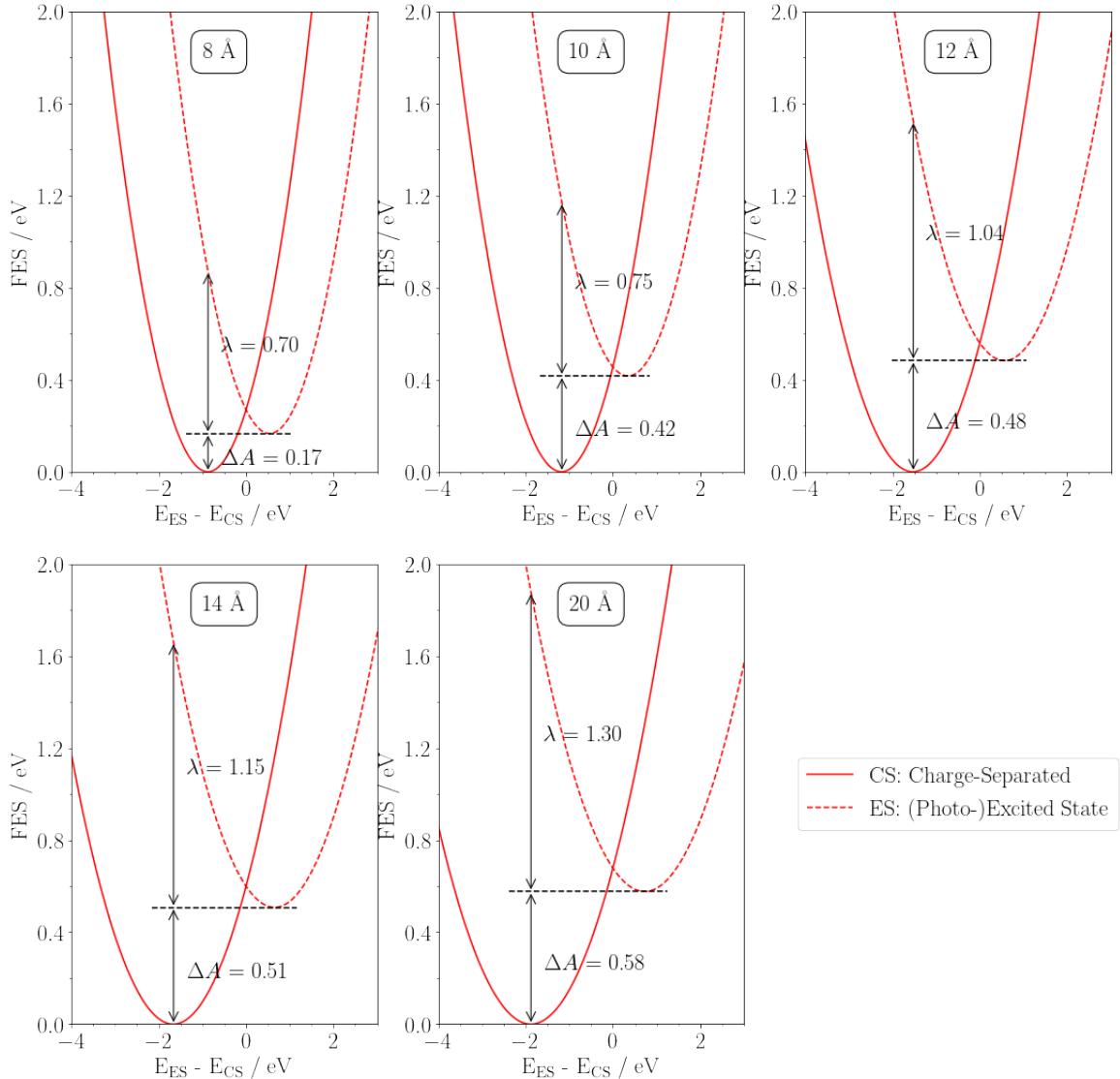


Figure 5.17: Free energy surfaces of the charge separated and photo-excited states of chlorophyll dimer systems, using the difference in state energies as the coordinate, for a series of separations. Free energy changes ( $\Delta A$ ) and reorganization energies ( $\lambda$ ) are shown for each separation.

5.17. As the systems become more separated and both  $\Delta A$  and  $\lambda$  increase, the rate constant is reduced exponentially. This is contrary to the previously held assumption that there is a critical distance where this rate drops to zero [202], although the exponential drop off in rate will be sharp. Additionally, the coupling value  $H_{AB}$  will also decrease with distance. This decrease is due to the dependence on overlap between the donor and acceptor orbitals in the electron transfer, which decreases with distance again exponentially [203–205].

Comparing the solvated 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 below the calculated critical distance for charge-separation/quenching, where transition to the CS state is most competitive[198]. However, as can be seen in figure 5.18, the free energy change is much higher for LH2 than for even the 20 Å case, attributed to a destabilization 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 biasing of FESs illustrates how the LH2 protein scaffold actively reduces the rate constant for quenching that would otherwise be present at such close separations.

## 5.5 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 showcase the accuracy of the Chl-xTB exciton method. The quantitative accuracy of transition energies for LH2 dimers is on the order expected from the training data. As Chl-xTB and PBE0 predict excitonic energies with similar levels of accuracy to CAM-B3LYP data, it is not expected that improvements to Chl-xTB optimization methods would have an appreciable effect in this specific test. The qualitative trends in excited state energies for artificial dimer geometries imply that Chl-xTB is good at predicting variations in excitonic energy due to geometry changes advantageous to the main focus of this work.

It can be seen how Chl-xTB design choices have created a method that can be readily applied to light harvesting 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 benchmarking other systems could be that the training data range would need to be extended, similar to the conclusion of the previous chapter. Some of the errors in the Chl-xTB exciton model may be improved by using other monomer interaction methods beyond the point charge interaction, such as using MNOK operators similar to the work done in the previous chapters, although with different parameters.

The benefits of the efficient Chl-xTB exciton method is illustrated in predicting the rate constants of transitions between excited states in chlorophyll dimers. This study has provided predictions of quenching mechanisms rates in chlorophyll dimers that take into account the atomistic properties of the LH2 protein scaffold. By including these effects it has been demonstrated that the protein scaffold biases against charge-separation quenching pathways when compared to dimers in solution phase, a clear indication of how the LH2 complex maximizes beneficial EET. As a case study, the application of the Chl-xTB exciton method demonstrates its potential uses for light harvesting studies that may not be feasible with other excited state methods. Similar to the arguments in section 4.4, it is the ability to reproduce high level data that is a main benefit

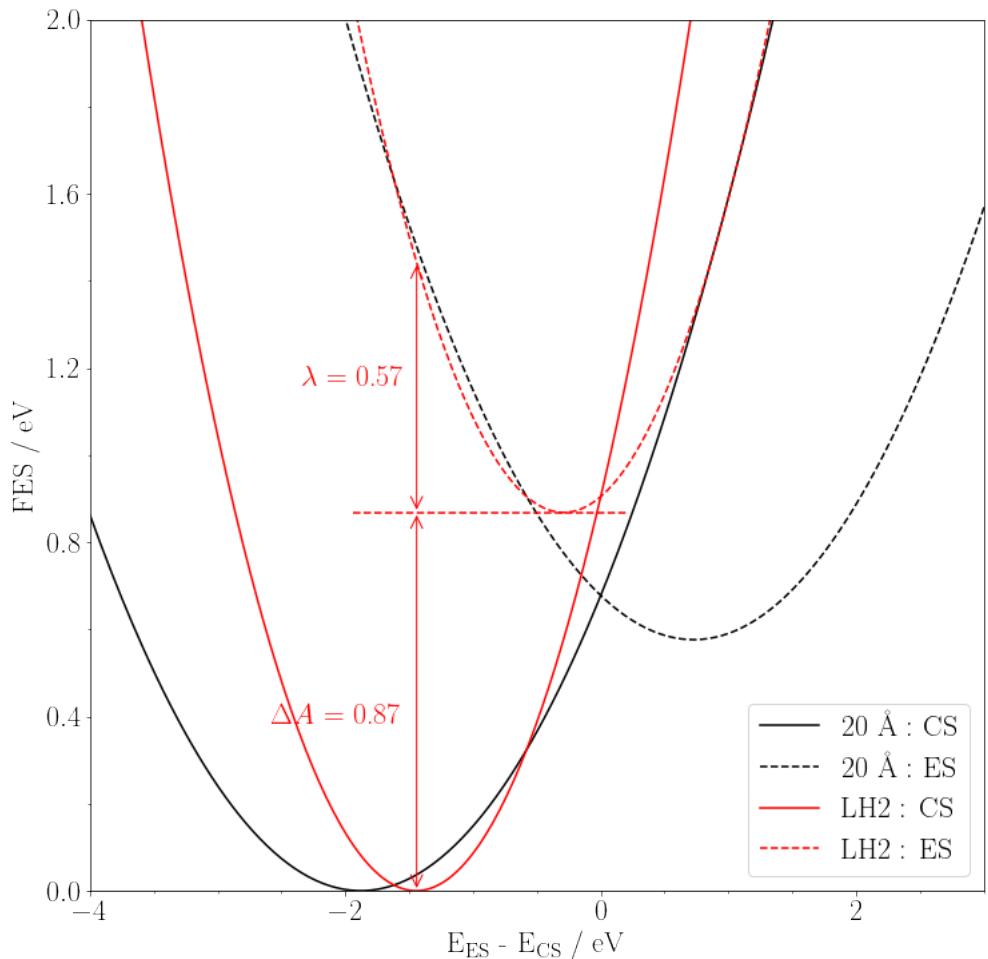


Figure 5.18: Free energy surfaces of the charge separated and photo-excited states of LH2 chlorophyll dimer systems, using the same coordinate as figure 5.17. The red lines show FESs from LH2 chlorophyll dimers, with black lines showing FESs of solvated chlorophyll dimers. The large free energy change and movement of the charge separation state FES minimum to the left-hand side of the photo-excited FES imply the LH2 protein environment inhibits transfer to the charge separated state.

of this kind of method.

## LIGHT HARVESTING COMPLEXES

### Previous Published Work

All of the work presented in this chapter are also included in a paper published with Dr. Susannah Bourne-Worster and Prof. Fred Manby in January 2023 [176].

In this chapter further large scale properties of LHCs are explored with the Chl-xTB exciton framework. As opposed to the previous chapter where only the chlorophyll dimer case was investigated, this chapter examines the interplay between the entire chlorophyll exciton network and the protein environment.

As seen in the previous chapter, the protein scaffold and chromophore geometry conformation have direct effect on the rates of EET. Similarly dynamic properties of these systems will effect EET, as the properties of chlorophyll chromophores in the LH2 complex are not static. Motion (e.g., vibrations) of the chromophores themselves and of the protein complex constantly changes the conformational energy of the chromophores, their interactions with each other, and their interactions with their protein environment, creating continual fluctuations in the excitonic energies of the chromophore network. These fluctuations drive relaxation processes, which lead to spectroscopic line-broadening and heavily influence the ability of the chromophore antenna network to efficiently transport energy.

This important environmental influence can be encapsulated in a single parameter called the spectral density, which describes the coupling between vibrations (of the chromophores and environment) and excitonic energies.

The spectral density describes how the oscillations in the surrounding environment can effect a property. Calculating spectral densities for transitions energies is usually used when looking at light harvesting complexes to assign functions to the protein structure, as well as necessary for population dynamics [2].

This important environmental influence can be encapsulated in a single parameter called the spectral density, which describes the coupling between vibrations (of the chromophores and environment) and excitonic energies. Spectral densities are commonly defined as the half-sided Fourier transform of the autocorrelation function of a time series for the property of interest (usually a specific energy). For example, the spectral density for a transition energy  $\Delta E$  is

$$(6.1) \quad J(\omega) = \frac{\beta\omega}{\pi} \int_0^\infty dt C(t) \cos(\omega t),$$

where  $\omega$  is frequency and

$$(6.2) \quad C(t_j) = \frac{1}{N-j} \sum_{k=1}^{N-j} \Delta E(t_j + t_k) \Delta E(t_k)$$

is the autocorrelation function in the time domain. The indices  $j, k$  label frames in the time series of  $\Delta E$  (usually a series of energies calculated along consecutive frames of a molecular dynamics trajectory) and  $N$  is the total number of frames. The prefactor  $\frac{\beta\omega}{\pi}$  in equation 6.1 is required to maintain detailed balance, with  $\beta = \frac{1}{k_B T}$  and  $T = 300K$ .

The ability to capture coupling to vibrational modes of a particular frequency depends on the length ( $N$ ) and frequency ( $1/(t_k - t_j)$ ) of sampling across the time series. Specifically, the upper frequency limit is  $\frac{1}{2(t_k - t_j)}$ , whilst the lower limit is  $\frac{1}{2N}$ . In order for the spectral density to cover high-frequency vibrations (e.g. vibrations of hydrogen bonds), as well as low frequency vibrations (e.g. large-scale motion of the protein complex), transition energies have to be sampled over the long time scale of protein motion (10s – 100s of picoseconds) at a femtosecond or even sub-femtosecond resolution.

With current electronic structure methods and computing power, this is an almost prohibitively large volume of data to acquire. Consequently, most spectral densities calculated from MD trajectories only cover a small frequency range. Recent examples for bilin [206] and Nile red chromophores [207] (both smaller than bacteriochlorophyll), calculated at a TD-DFT level of theory (B3LYP/aug-cc-pVDZ and B3LYP/cc-pVTZ for bilin chromophores and CAM-B3LYP/6-31G\* for Nile red chromophores), sample high frequency molecular motion with a sampling rate of  $0.5 \text{ fs}^{-1}$  but do not simulate beyond a few picoseconds (6 ps and 40 ps for bilin and Nile red simulations respectively) so cannot capture lower frequency modes.

They nevertheless required months of GPU time. Switching to lower level electronic structure methods, such as ZINDO, give access to longer timescales (Mallus *et al.* achieve 300 ps at 2 fs intervals[2]) but still remain too expensive to treat the entire exciton framework collectively. In the absence of affordable *ab initio* spectral densities[208–210], but even these can be difficult to obtain and only offer an approximate, coarse-grained picture.

The speed and accuracy of Chl-xTB, particularly in describing the relative excitation energies of closely related geometries (see figures 4.13 through to 4.21), addresses this issue by making it feasible to sample transition energies over long timescales at high resolution. Using this method

it becomes possible to calculate spectral densities that not only cover a broader frequency range than was previously possible but also for the exciton state energies of the entire LH2 antenna complex (rather than just the transition energies on individual chromophore sites).

This chapter tries to answer two questions about environmental frequency coupling to the exciton states. The first is whether low frequency modes are present 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 reveal 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 individual chlorophyll molecules (inside the LH2 protein). 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. 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 biomimetic systems.

## 6.1 Calculating LH2 Excitons

### 6.1.1 Molecular Dynamics

Exciton states were calculated from a series of structures of LH2 and chlorophyll generated by molecular dynamics simulations run with the OpenMM package. The force-field and geometry were the same as used by Ramos *et al.* [21]. The system was equilibrated for 60 ps, before starting a production workflow of 300 ps, with frames taken every 2 fs similar to the work done by Mallus *et al.* [2]. A Langevin integrator set at 300 K was used with a timestep of 2 fs. Non-bonded interactions were treated with a particle mesh Ewald method.

### 6.1.2 Scaling computational performance

Exciton states were calculated with the same method as used in the previous chapter (encapsulated in equation 5.3), but extending beyond the dimer system to include all chlorophyll sites. The elements in the exciton Hamiltonian were calculated using Chl-xTB properties. It was necessary to implement a highly parallelized version of the code to calculate Chl-xTB properties efficiently. 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 into a separate program to construct the exciton Hamiltonian. As the high performance computing machines available had around 20 cores, the compute wall-time of each exciton run was on the order of a single chl-xTB calculation.

The main bottleneck in analysing these states was the volume of Chl-xTB properties needed, as it was found construction and diagonalization 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 individual Chl-xTB calculations on distinct geometries to construct the exciton Hamiltonian. The time for each Chl-xTB calculation is  $\approx 0.5$  s (see figure 4.9), so this presents about 46 days of serial CPU time. This was significantly reduced by using the highly parallelized program, to  $\sim 30$  hours. Whilst having 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 be a more efficient use of resources than the approach used here.

### 6.1.3 Coupling Values

A summary of the coupling values, broken down by ring-interaction type, are shown in figure 6.1. 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  $350\text{ cm}^{-1}$ , which corresponds well with previously used values between  $238$  and  $771\text{ cm}^{-1}$  [18]. 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 estimates of inter-dimer couplings being  $100\text{ cm}^{-1}$  less than intra-dimer couplings [211–214]. Inter-ring couplings are much lower, with distributions around  $35\text{ cm}^{-1}$  and  $10\text{ cm}^{-1}$ , consistent with previous estimates of between  $24$  and  $31\text{ cm}^{-1}$  [165].

### 6.1.4 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 `he1PME` 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  $\sim 1$  second to  $\sim 40$  seconds. Additionally the PME functions required multiple cores in order to get the best performance, which competed with the parallelization 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.* [2], 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

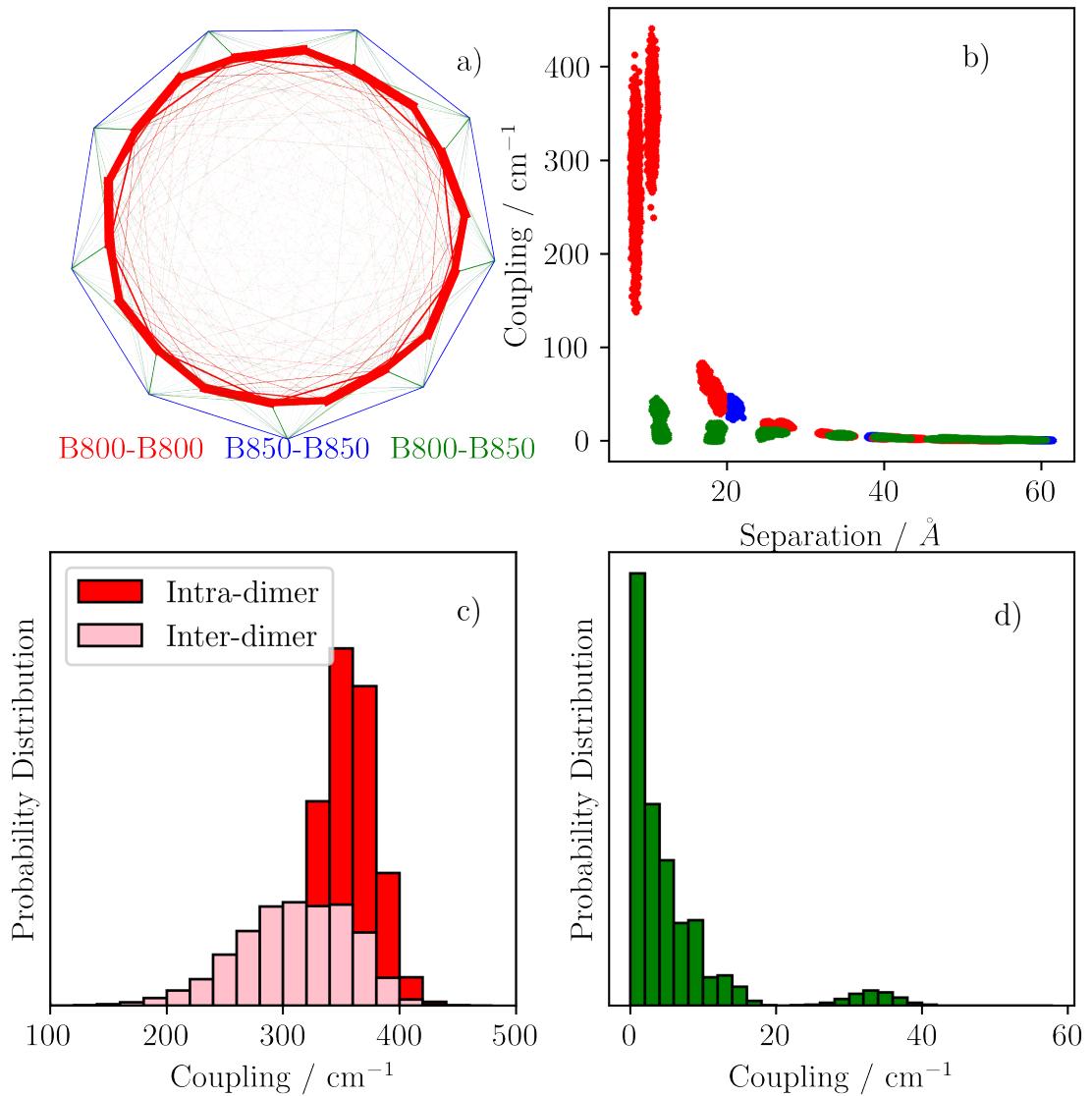


Figure 6.1: a) a diagrammatic representation of coupling strengths and interactions between sites specified by ring labels, with thicker lines indicating stronger exciton coupling interactions. b) Exciton coupling values (in  $\text{cm}^{-1}$ ) plotted as a function of distance between chlorophyll sites, 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.

$$(6.3) \quad V_{(m,1),(n,1)} = \sum_{A \in m, B \in n} \frac{f}{4\pi\epsilon_0} \frac{q_m^{\text{tr},A} q_n^{\text{tr},B}}{r_{AB}}$$

where the definitions of variables are the same for equation 5.7. The scaling factor  $f$  is given by

$$(6.4) \quad f = A \exp(-BR_{ij}) + f_0$$

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.

#### 6.1.4.1 Absorption Spectra

Absorption spectra of LH2 are usually simulated by calculating the intensities 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  $\Psi_0$  to a (one) exciton state  $\Psi_k$  is given by

$$(6.5) \quad I_k \propto E_k |\langle \Psi_k | \hat{e} \cdot \mu | \Psi_0 \rangle|^2$$

where  $E_k$  is the energy of the state  $\langle \Psi_k | H | \Psi_k \rangle$ ,  $\hat{e}$  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

$$(6.6) \quad I_k \propto \left| \sum_{j=1}^N c_{kj} \hat{e} \cdot \hat{\mu}_j \right|^2$$

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  $\hat{\mu}_j$  is the transition dipole moment of chlorophyll  $j$

$$(6.7) \quad \hat{\mu}_j = \langle \phi_j^{(1)} | \mu | \phi_j^{(0)} \rangle$$

As this intensity is dependent on the orientation of the unit vector  $\hat{e}$ , the average intensity can be found by either calculating the intensities for a large distribution of randomized 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.

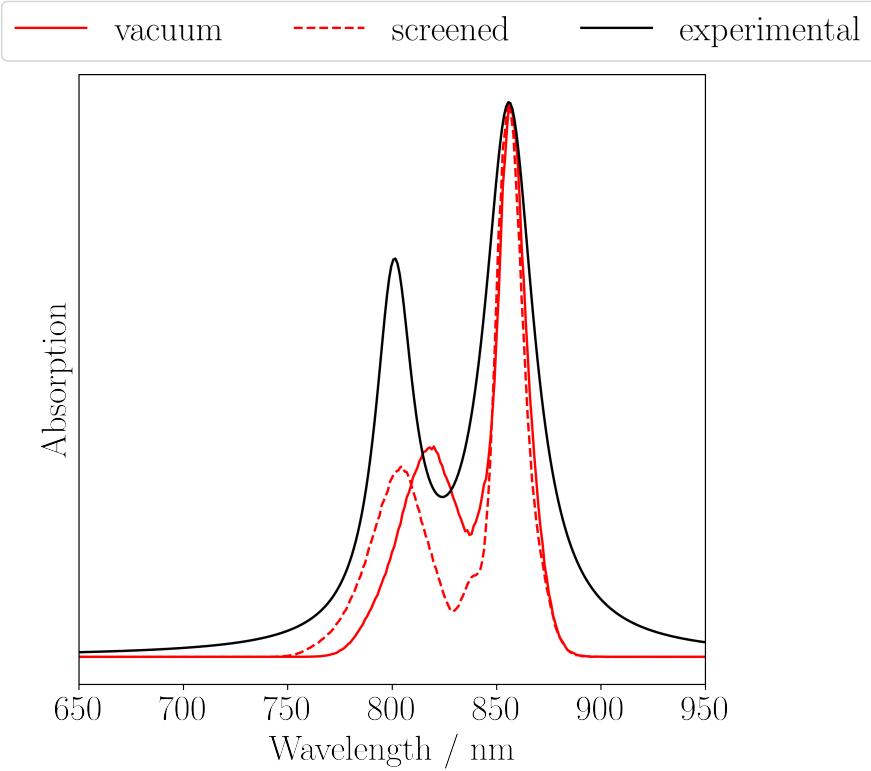


Figure 6.2: Simulated absorption spectra of LH<sub>2</sub>, with (dashed line) and without (solid line) a screening factor for point charge interaction. An experimental line reconstructed from Strain *et al.* [1] is plotted in black. The simulated spectra are shifted to match the position of the 850 nm peak.

The simulated LH<sub>2</sub> spectra with and without screening factors, alongside the experimental spectrum, are shown in figure 6.2. 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 [21, 67]. 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 implies that the only effect of using the screening terms is to lower the B800 energy states to better fit the experimental spectrum. This would be due to the inclusion of B850 site character into exciton states localized 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.

#### 6.1.4.2 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

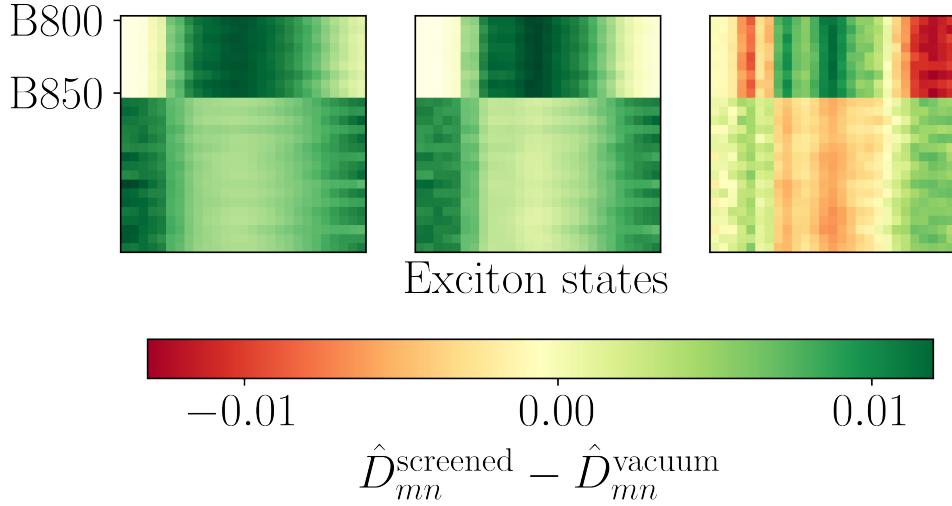


Figure 6.3: Density contributions of site transitions to exciton states for the vacuum (left) and screened (middle) Hamiltonian, with the difference shown on the right.

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 6.3 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 localized on B800 and B850 sites. The plot of the density difference shows how density is localized more on B800 sites for exciton states calculated with the screening factor. This is also true for the B850 sites. However it does not reduce the delocalization 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, they may not change the overall qualitative behavior 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.

## 6.2 Long Timescale Spectral Densities of the LH2 Protein

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.

### 6.2.1 Spectral Densities of $Q_y$ Transition Energies at Individual 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 benchmarking was done for a B800 site of LH2 chlorophyll, and the spectrum can be seen in figure 6.4.

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.* [2], and uses the same axis scaling (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 force-fields 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 [2], and so difference in the force-field 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 spectrum shown in figure 6.5 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 in 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 magnitude of 0.185 eV, is not present in the B800 sites. Similarly the 0.205 eV (height 0.117 eV)

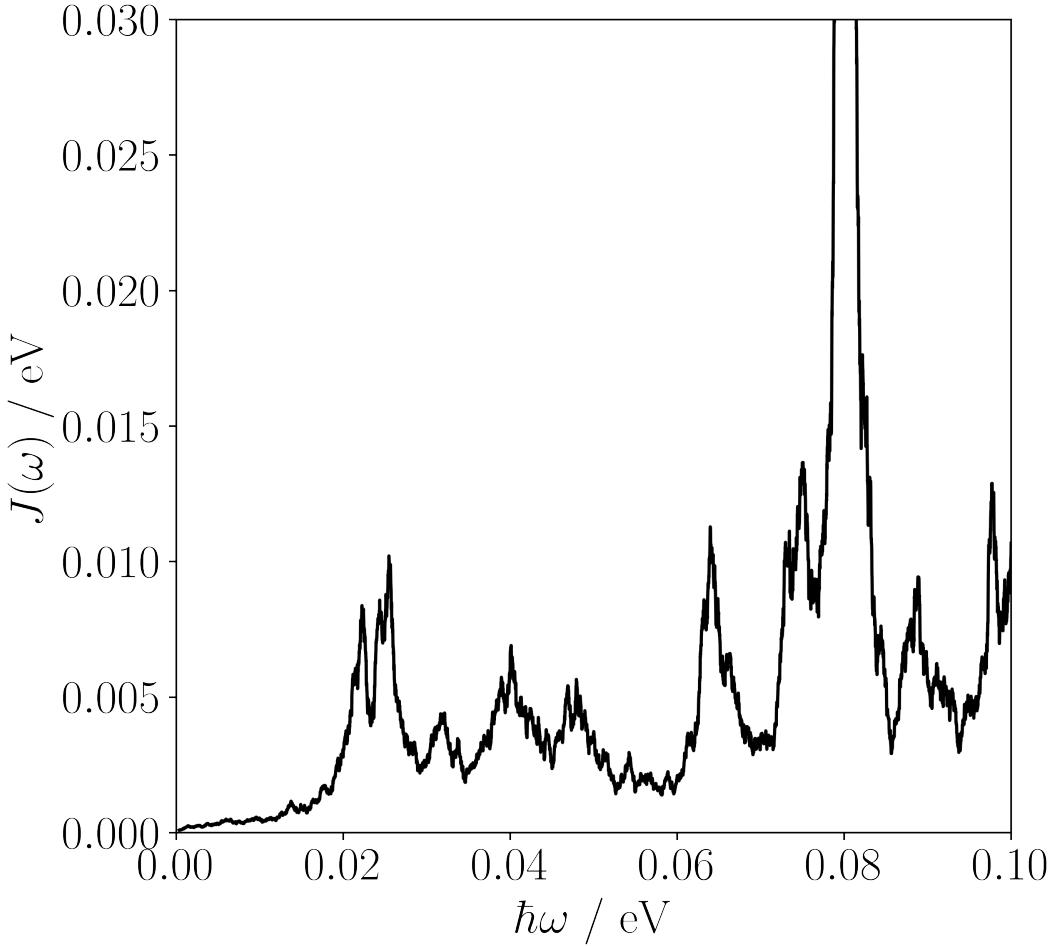


Figure 6.4: Spectral density of the  $Q_y$  transition for a chlorophyll in the B800 ring in LH2. Axis scaling and units are chosen to best reproduce the spectrum reported by Mallus *et al.*[2].

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 leads to greater variations in transition energy [39].

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 better workflows in calculating spectral densities.

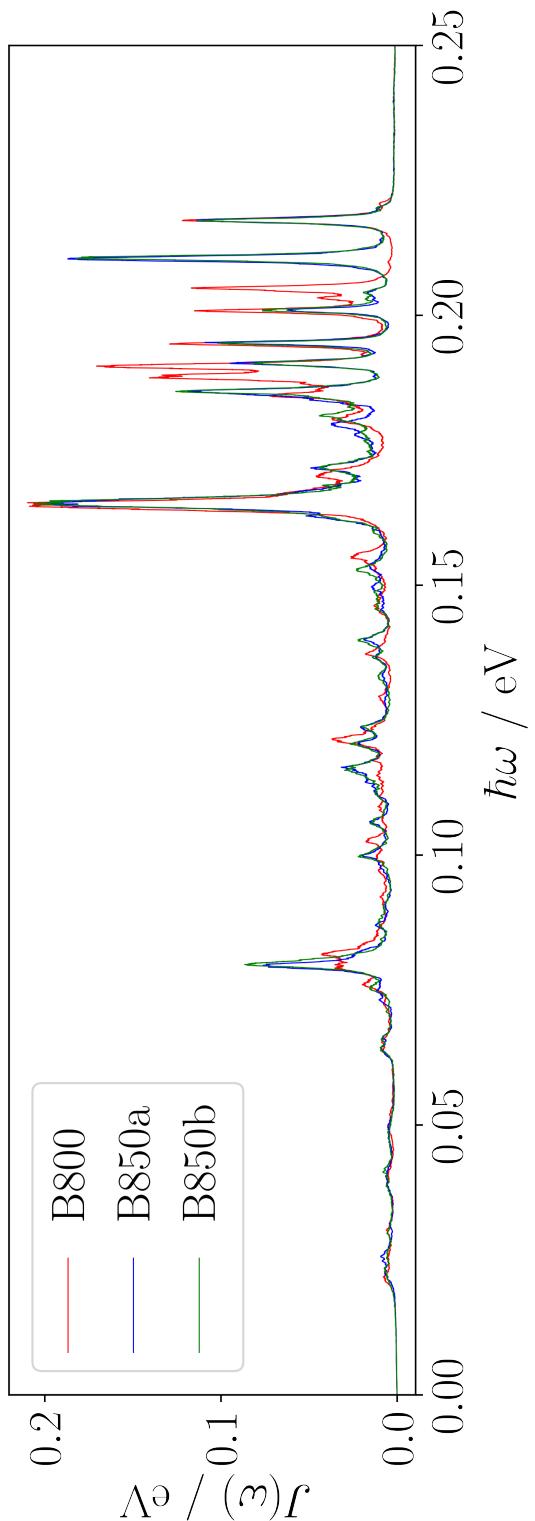


Figure 6.5: Average spectral densities for  $Q_y$  transitions at sites in the B800, B850a and B850b rings.

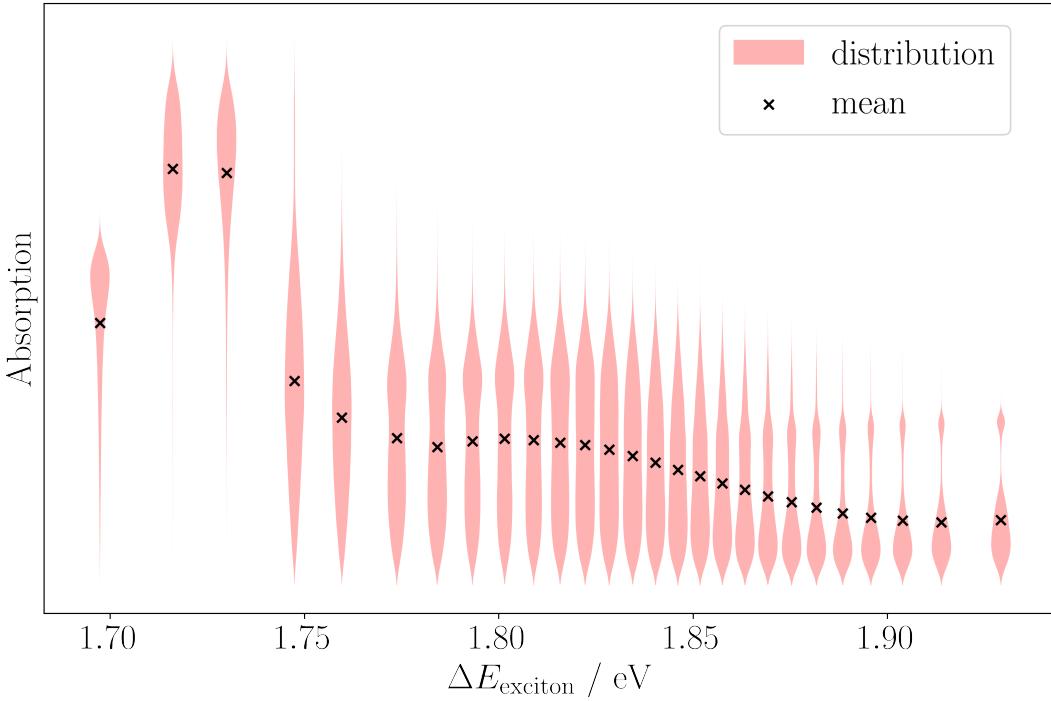


Figure 6.6: Violin plots of absorption probabilities for exciton states, positioned by the average transition energy from the ground state. The width of each distribution indicates the density of values at absorption probabilities, with crosses marking the mean value.

### 6.2.2 Spectral Densities of Exciton State Energy

This section reports on the calculation of 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 6.7 have been weighted in colour by their time-average absorption probabilities (shown in figure 6.6) to clearly present the features in the spectra. From figure 6.6 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

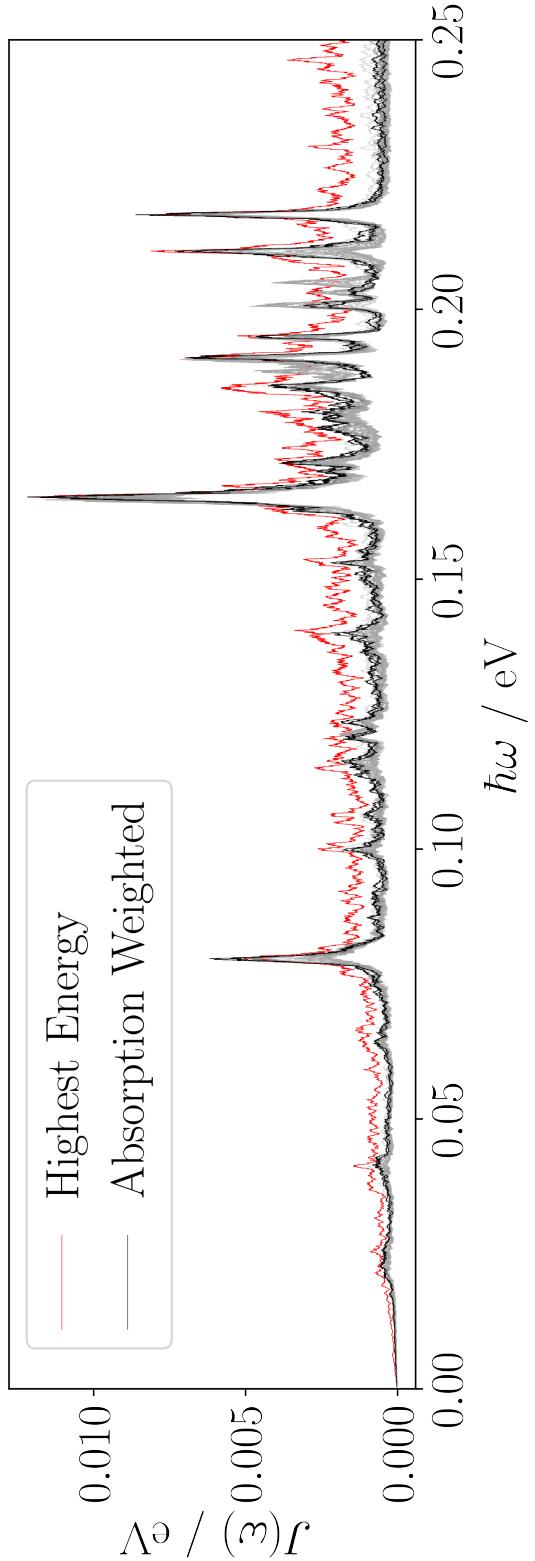


Figure 6.7: Spectral densities of exciton transition energies of LH2, weighted by the averaged absorption probability. The highest energy state is indicated in red. This state has a higher fluctuation in energy, causing excess noise in the spectral density.

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 increases 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 localized at each site. This localization of effects 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.

### 6.2.3 Spectral Densities of Exciton State Coupling

The spectral densities of the coupling terms were calculated with the same method as the state and site transition energies, and are shown in figure 6.8. 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 of 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 inter-chromophore distance (reported without units as these are not physically meaningful). These spectra can be seen in 6.9.

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.

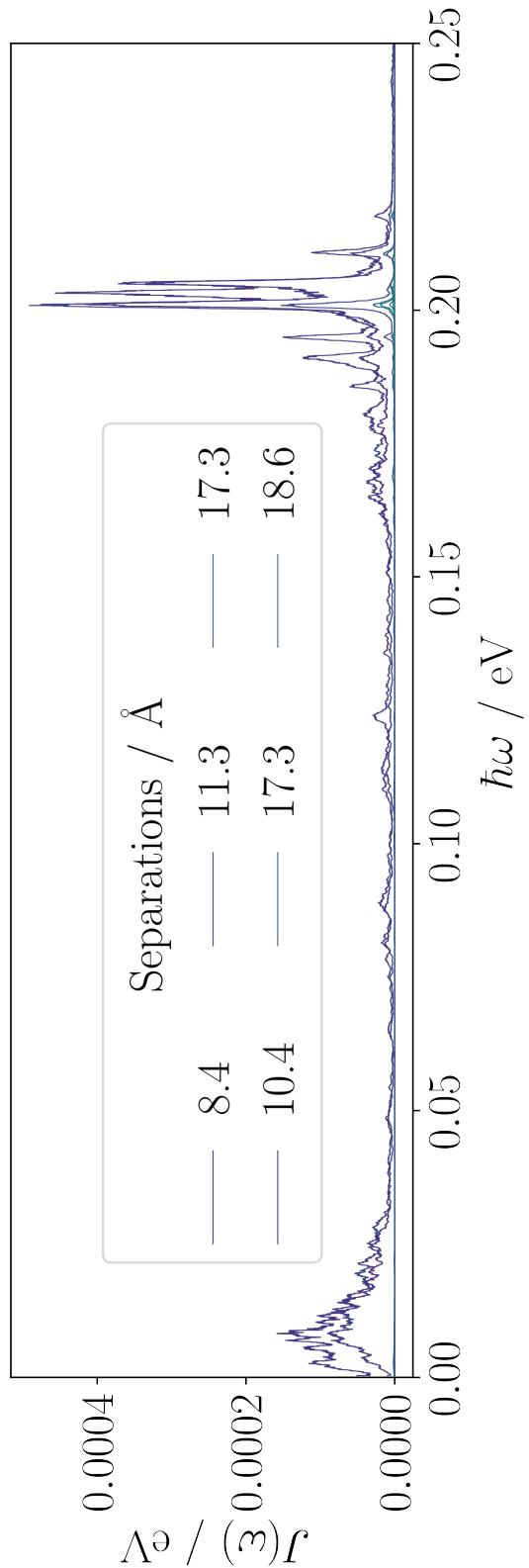


Figure 6.8: Spectral density of exciton state coupling values (i.e. off diagonal elements of the exciton Hamiltonian) of LH2 coloured by average distance between chlorophyll sites.

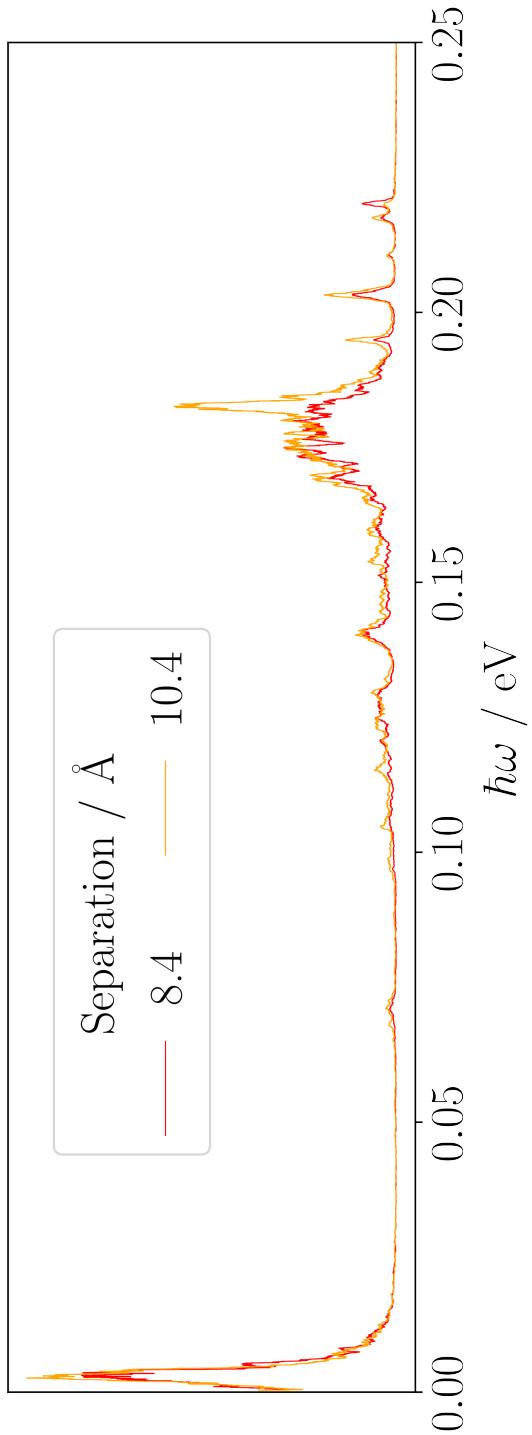


Figure 6.9: Spectral density of the separation between nearest neighbour sites in LH2.

## 6.3 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. These tests include 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.

### 6.3.1 Spectral Density of the $Q_y$ Transition from Chlorophyll in Diethyl Ether

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 was taken from an MD simulation of a diethyl ether embedded chlorophyll. This simulation 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 force-field, 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 time step 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 6.10.

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

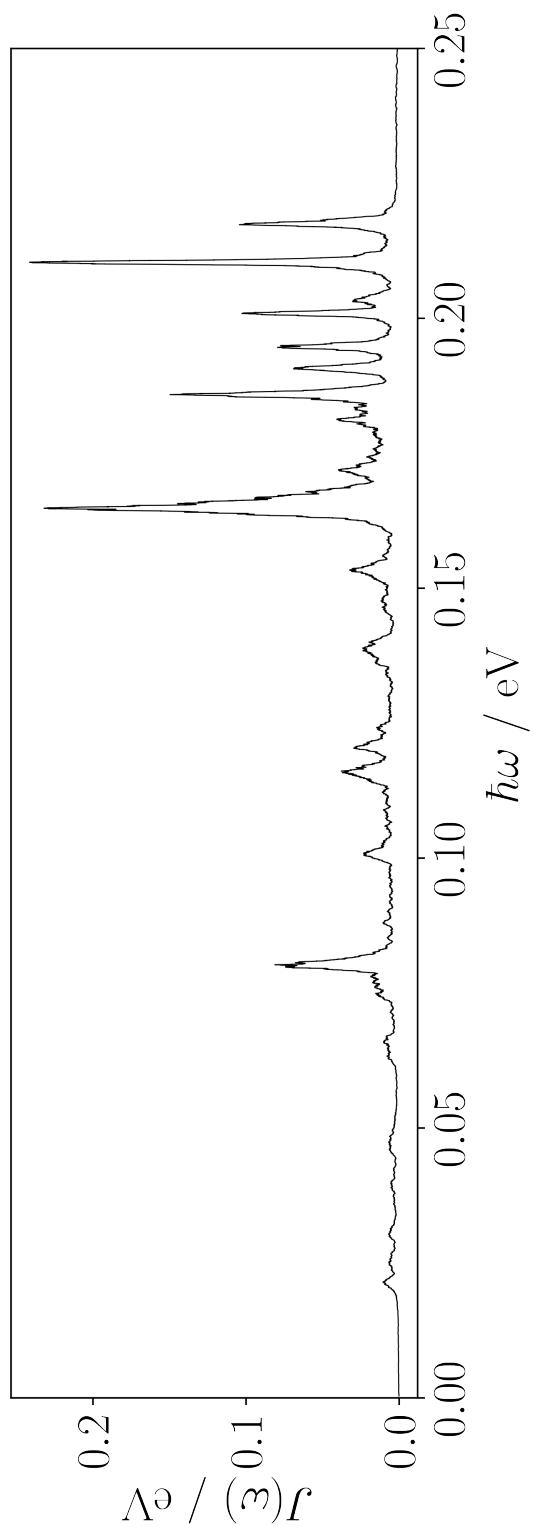


Figure 6.10: Spectral density of the  $Q_y$  transition for a chlorophyll in diethyl-ether.

are similar to the features from the LH2 chlorophyll and exciton transitions. This similarity implies that 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 couplings between chlorophyll transition energies and the environment are controlled by intrinsic chlorophyll properties.

### 6.3.2 Assignment of Molecular Motion with 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 transitions 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

$$(6.8) \quad q_i = \sqrt{\frac{\omega_i}{\hbar}} x_i^m$$

where  $\omega_i$  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 ( $q_{\text{ground}}$ ) and excited state energy ( $q_{\text{excited}}$ ).

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

$$(6.9) \quad d = \frac{(q_{\text{excited}} - q_{\text{ground}})^2}{2}$$

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 spectrum is shown in figure 6.11.

The low frequency ( $<0.05$  eV) modes in this spectrum 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 effects from the environment. It is fairly clear that these motions should not correspond to 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 change 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 force-field and GFN1-xTB method. The motions themselves should be similar, even if the frequencies are not. Assuming a similarity in motion 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 6.11. 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. These symmetry breaking modes are shown in figure 6.12.

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.

### 6.3.3 Assignment of Molecular Motion with Spectral Densities of 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 6.13, 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

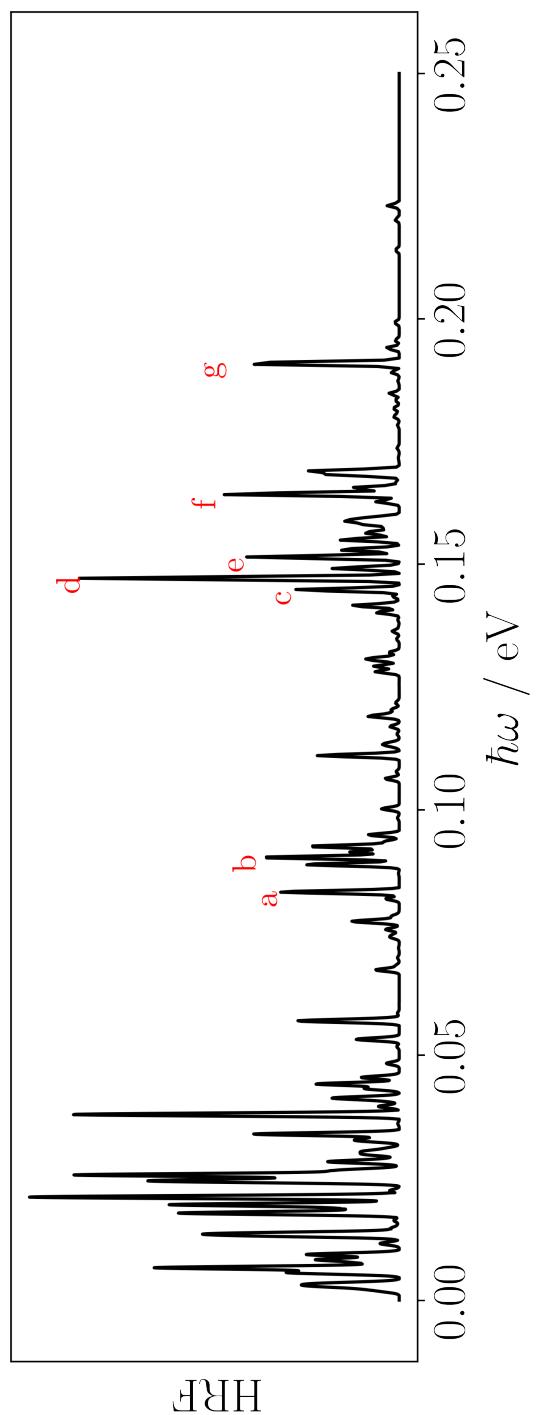


Figure 6.11: A simulated spectral density of the  $Q_y$  transition for chlorophyll constructed from Huang-Rhys factors and the frequency of normal modes. Labeled peaks are chosen as modes that may correspond to features in other chlorophyll spectral densities.

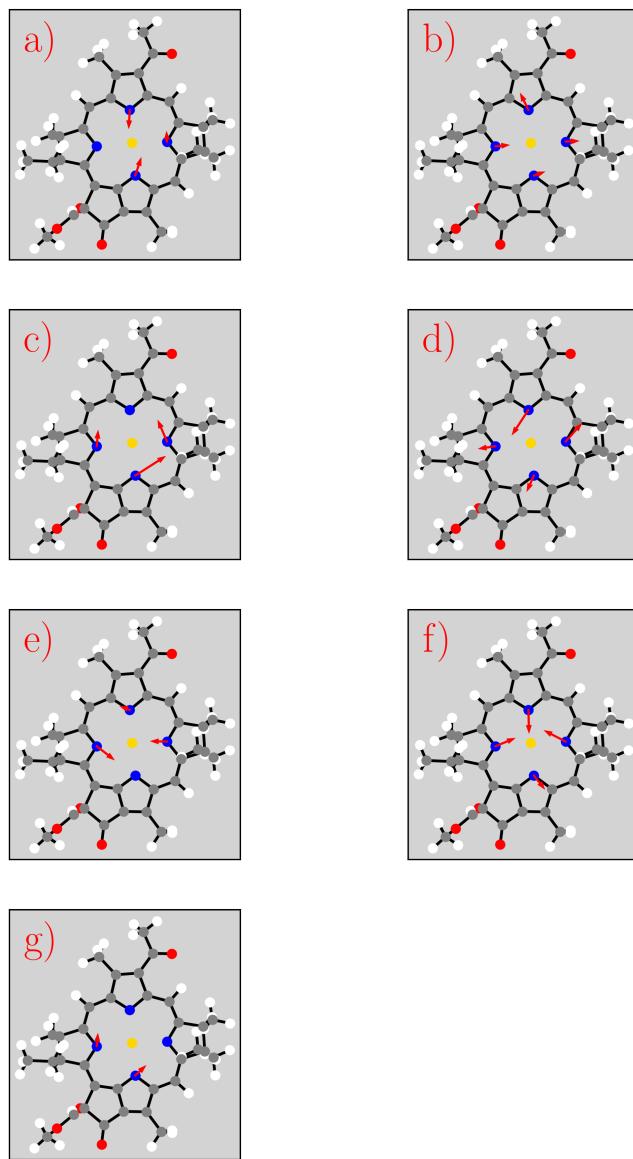


Figure 6.12: Motions of the four central nitrogen atoms in bacterial chlorophyll for the vibrational modes labelled in figure 6.11.

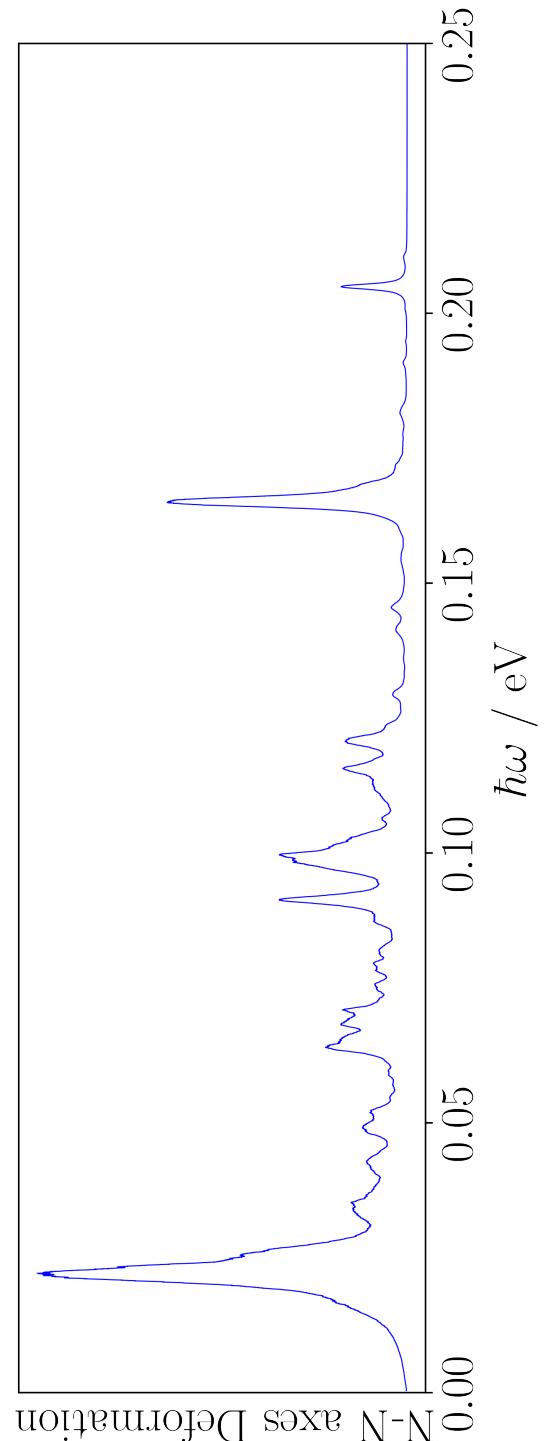


Figure 6.13: Spectral density of the ratio  $\frac{|N_A - N_C|}{|N_B - N_D|}$ , averaged over chlorophyll sites in LH2.

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 it is a little surprising that most features must be due to other 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 force-field 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.

## 6.4 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 of 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 coupling variations originate at the monomer level than aggregate effects. This extends the observation made in previous work that the Qy excitation energy of individual chlorophyll sites remains essentially constant when intramolecular BChl vibrations are “frozen out” (i.e., the BChl geometry is artificially fixed during a molecular dynamics simulation) to leave only the effect of large-scale protein motion [215].

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. A known issue with full assignment is also the mismatch between the quantum mechanical methods used for excited states and the classical force-field methods used for MD [207]. 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 force-field 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 to the discussion 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 discussions of results in the previous chapters analyzed whether novel response methods for LHCs are viable alternatives to other methods proposed in the literature. In the first chapter it was found that while the more efficient  $\Delta$ -SCF approach is accurate when using DFT based methods,  $\Delta$ -xTB methods are not accurate enough for LHC models. This shortcoming was argued to be due to the ground-state property parameterization of the GFN-xTB methods as other tight-binding schemes do work well when parameterized for transition properties. Running with this conclusion, it followed that a bespoke parameterisation of the tight-binding electronic structure theory and excited state formalism may be beneficial, giving the chl-xTB method designed and parameterized specifically for  $Q_y$  transition properties. The specificity of this model is similar to the machine-learning models discussed in the introduction, but relies on electronic structure theory rather than any machine-learning method. It is also much easier to parameterize. Chl-xTB proved to be highly accurate against the selected training data, although choices in constructing the training data set have shown some signs of limiting this method.

While targeted parameterization gives Chl-xTB a very specific application, the underlying physical model means it can be more easily extended or interfaced into larger multiscale simulations than machine learning models, which are strictly limited to describing the properties and situations they were trained for. This protocol also makes far more efficient use of parameters than a typical ML model, requiring only a handful of constants that all have specific chemical meaning (compared to tens of thousands used in a supervised learning model such as the one by Häse *et al.*)[78].

Using this efficient response method for work with LHCs also proved successful. Dimer transition properties as well as the LH<sub>2</sub> coupling values were well recovered by the Chl-xTB framework,

implying that conclusions about LHC (or other oligomer chlorophyll system) properties would be reliable.

Having established the usability of this efficient response method for LHCs, it was then possible to calculate novel properties. The free energy surfaces and spectral densities posed a challenge to established methods for excited states due to the massive volume of calculations required, coupled with the accuracy and extensive detail necessary.

Both of the applications of the Chl-xTB exciton framework have overcome these challenges and resulted in novel observations of LHC phenomena. The rates of ES-CS transition show how the LH2 protein purposefully inhibits the charge-separation quenching mechanism of vertical excited states when compared to a more disorganised 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. This outcome is surprising given the emphasis on accuracy in inter-chromophore and environment-chromophore coupling.

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.

## 7.1 Efficient Excited State Methods

Testing the Chl-xTB excited state approximations on different systems would be the logical extension of this type of method. In chapter 4 it was posited that the success of this method was due to a 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  $\Delta$ -SCF methods from section 3.1.4 as these are also single transition methods.

It is argued then that the response approximations and reparameterisation of Chl-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 4, 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 reparameterisation 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 strategy this would be the  $Q_x$  transition in chlorophyll. This transition is dominated 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 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 would only be necessary to create a more detailed image of the higher energy exciton states.

Reparameterizing 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 transitions 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 reoptimize the Chl-xTB parameters for uses on other chlorophyll systems and transitions. The first approach would be simple reparameterization using the parameters already reported. Not altering which GFN1-xTB parameters are optimized 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 optimize one model. Due to the diversity of this training data it would probably be necessary to reoptimize 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 4.2.1. Both the RMSE and  $R^2$  values would increase when a more varied set of training data is used but this increase would not be reflective of greater accuracy. 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 4 the correlation is an important metric for getting a good description of the relationship between transition properties and geometry variations. Investigating the weights  $\lambda_n$  in the objective function (eq. 4.18) may solve this issue by getting the optimization 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 [49, 216].

Future applications to molecules other than chlorophyll is facilitated by the deliberate use of widely available computational tools in the parameterization workflow (the open-source GFN1-xTB implementation and the SLSQP optimization procedure, which is available in most standard statistical software packages, such as Python’s SciPy). The approximation of a diagonally dominant  $\mathbf{A}$  matrix used to derive transition energies limits the approach to transitions that are dominated by a single orbital excitation. While scaling the transition density can recover a limited amount of mixed-transition character, we would not expect transitions consisting of a mix of degenerate excitations to work well. However, transitions with single-excitation character are relevant to a number of interesting biological molecules. For example, both the S2 excited state in carotenoids and the Q band in bilins (two of the other important pigments in LH2 complexes) are almost entirely HOMO  $\rightarrow$  LUMO transitions in character[217–220]. In general, larger molecules that are too expensive to treat with TD-DFT (and therefore require a cheaper method) are often not symmetric enough to have much degeneracy, so are more likely to have transitions that would be suitable for describing with the type of protocol outlined here.

Another obvious extension of the Chl-xTB protocol is to retrain it to a different reference method, such as the long-range corrected hybrid functional CAM-B3LYP, although this may require changes to equation 4.9 and 4.11 to account for bifurcating the Coulomb and exchange interactions into short- and long-range terms. Altering the formalism to reflect the structure of more complicated functionals would not preclude the approximations that make Chl-xTB fast and efficient. Thus we anticipate that, with only minor adjustments, the Chl-xTB protocol can be adapted to different levels of theory to suit the system of interest.

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.

## 7.2 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 may 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 [216]. To achieve these models with the similar methods used in this work would require reparameterization as discussed above.

The Chl-xTB exciton method without modifications could still be used for many more LHC

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 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 4.3.2, by only training to LH2 bacteriochlorophyll structures, which are constrained by protein binding pockets, the current form of Chl-xTB has the potential to misrepresent other geometries that would occur in different environments. This effect is almost negligible as the agreement of predicted spectra shown in figure 4.22 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 excited state approximations, and a design choice that has only been lightly 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 other methods. For example TD-DFT gradients would require solving couple-perturbed equations and sTDA-xTB gradients would have basis function gradient terms 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 [96]. 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 optimized for analytic geometry optimizations which also require this gradient term, many programs would be able to readily 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 for multiple LH2 proteins. 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 1 will be useful for inspiring similar approach in the future. The efficiency of tight-

## CHAPTER 7. CONCLUSIONS

---

binding methods has been well utilized and would allow larger models to be possible. The more extendable formalism, based on electronic structure methods and requiring few parameters, would allow 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 optimization 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 such as the software packages and hardware used. These are not exhaustive list, and additional details are provided in the main chapters.

### A.1 Electronic Structure Codes

This project has primarily used the QCORE software for electronic structure calculations. This includes all of the chl-xTB and  $\Delta$ -xTB calculations, as well as some of the  $\Delta$ -SCF calculations. At time of writing the only implementation of chl-xTB and  $\Delta$ -xTB are only found in QCORE, although other packages can do more general  $\Delta$ -SCF calculations. QCORE is a program that is 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.

Other calculations, including TD-DFT and ZINDO calculations, were done with the GAUSSIAN program [42].

### A.2 Computational Hardware

Much of the method development and analysis was done on a 2017 MacBook Pro with an 2.3 GHz Intel Core i5 processor. Running small scale calculations was also done on this machine.

Larger calculations (e.g. MD simulations, TD-DFT on full chlorophyll, chl-xTB parameter optimization) utilized the high performance computing systems at the University of Bristol. This

## APPENDIX A. APPENDIX A

---

work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bristol.ac.uk/acrc/>.



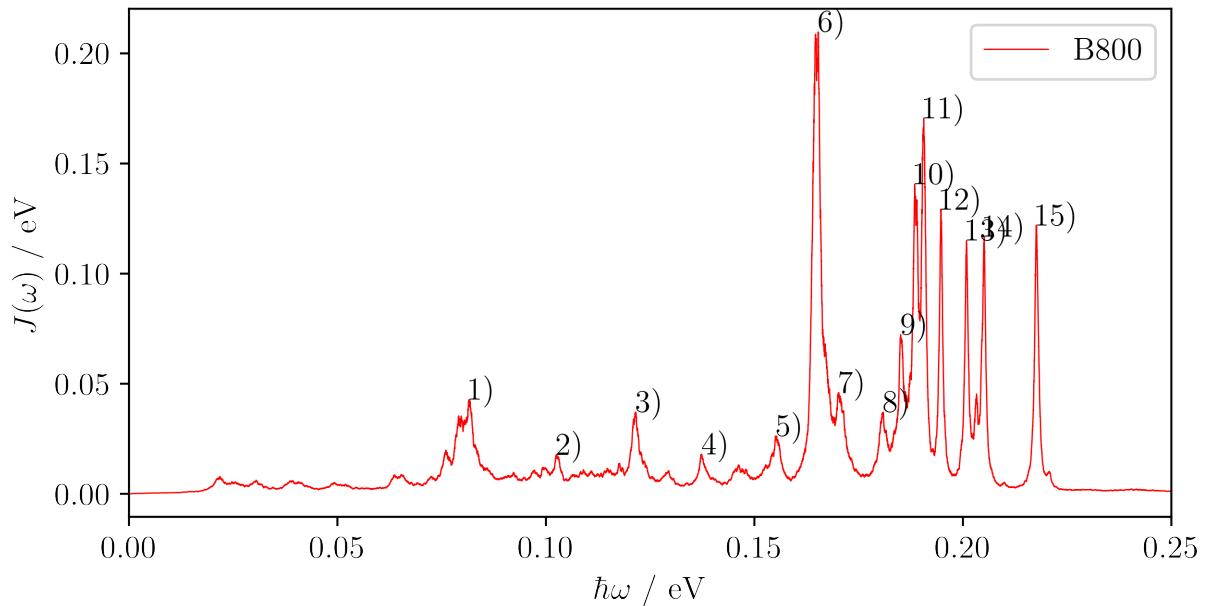
## APPENDIX B

### B.1 LH2 Data

This appendix collates ancillary figures of the spectral densities reported in chapter 6, alongside the frequencies and magnitudes of the important features.

#### B.1.1 Spectral Density Features

##### B.1.1.1 B800 ring sites

Figure B.1: Spectral density of the  $Q_y$  transition at LH2 B800 sites with labelled peaks.

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

Table B.1: Peak positions and heights for the spectral density  $Q_y$  transition at LH2 B800 sites.

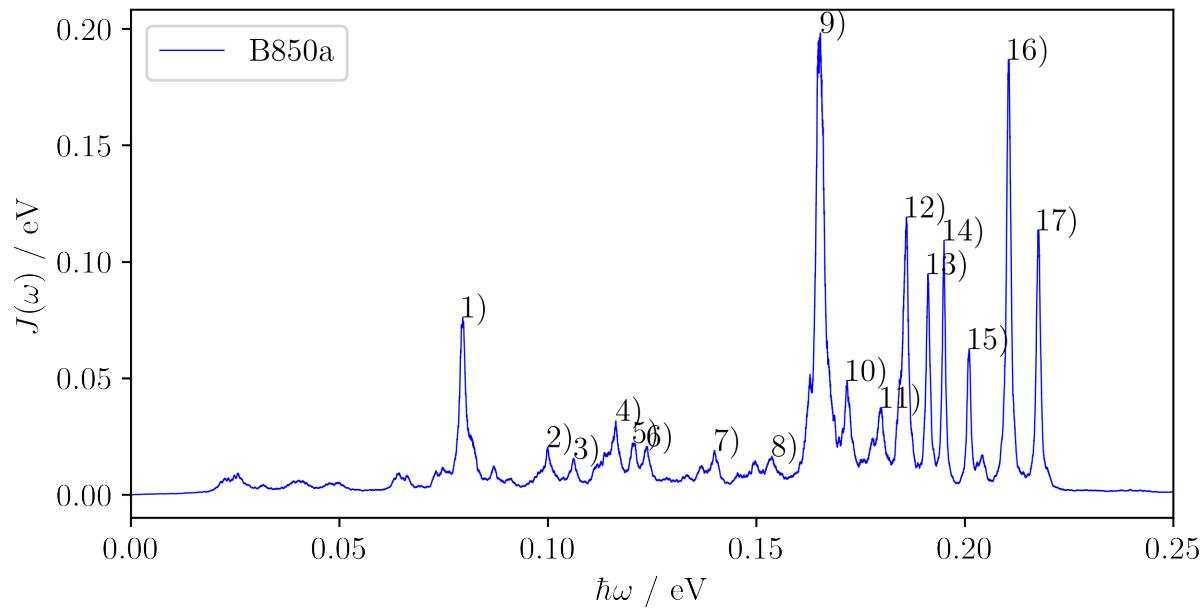


Figure B.2: Spectral density of the  $Q_y$  transition at LH2 B850a sites with labelled peaks.

### 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

Table B.2: Peak positions and heights for the spectral density of the  $Q_y$  transition at LH2 B850a sites.

### B.1.1.3 B850b ring sites

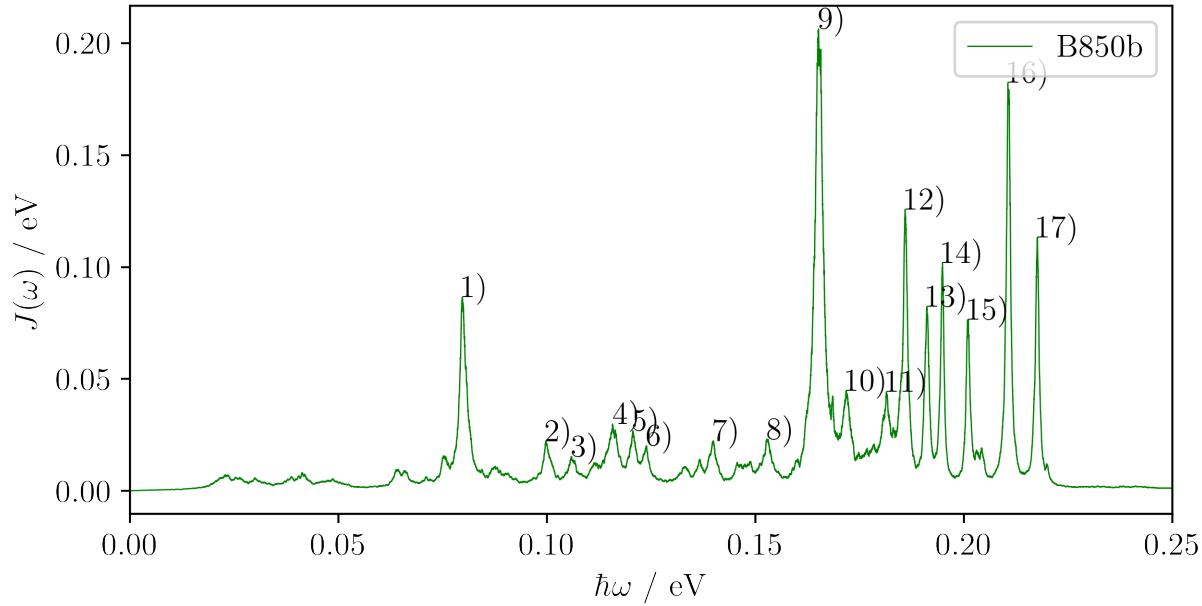


Figure B.3: Spectral density of the  $Q_y$  transition at LH2 B850b sites with labelled peaks.

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

Table B.3: Peak positions and heights for the spectral density  $Q_y$  transitions at LH2 B850b sites.

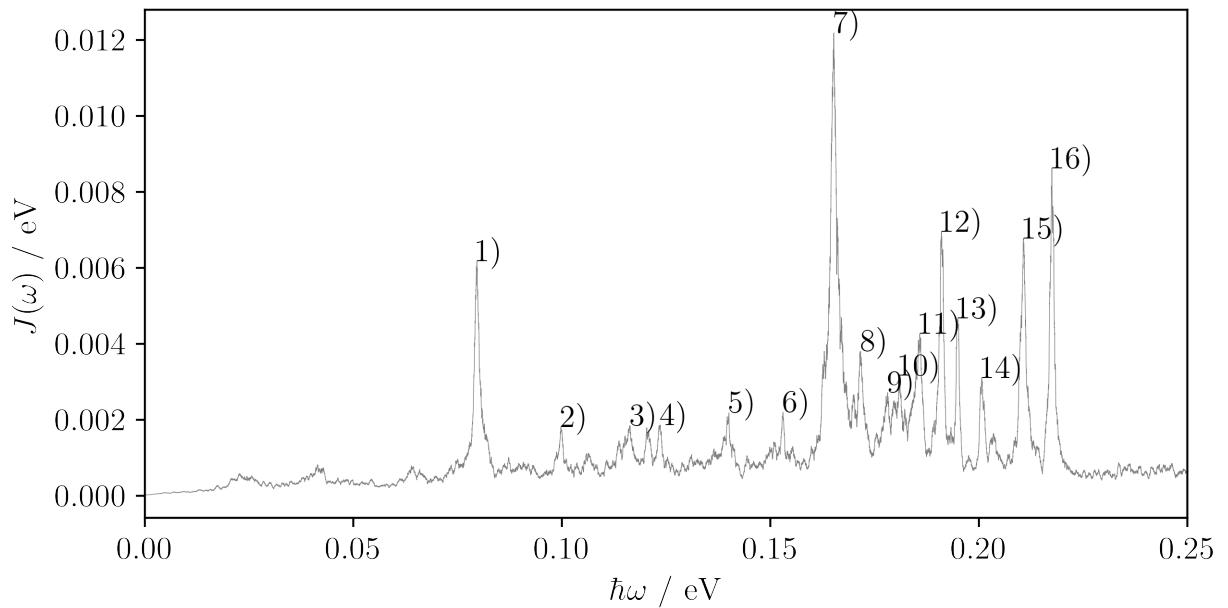


Figure B.4: Spectral density of the LH2 exciton transition energies with labelled peaks.

#### B.1.1.4 Exciton states

Peak label	$\hbar\omega$ / eV	$J(\omega)$ / 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

Table B.4: Peak positions and heights for the spectral density of LH2 exciton transition energies.

Peak label	$\hbar\omega$ / eV	$J(\omega)$ / 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

Table B.5: Peak positions and heights for the spectral density of LH2 exciton couplings.

**B.1.1.5 Hamiltonian Coupling Values**

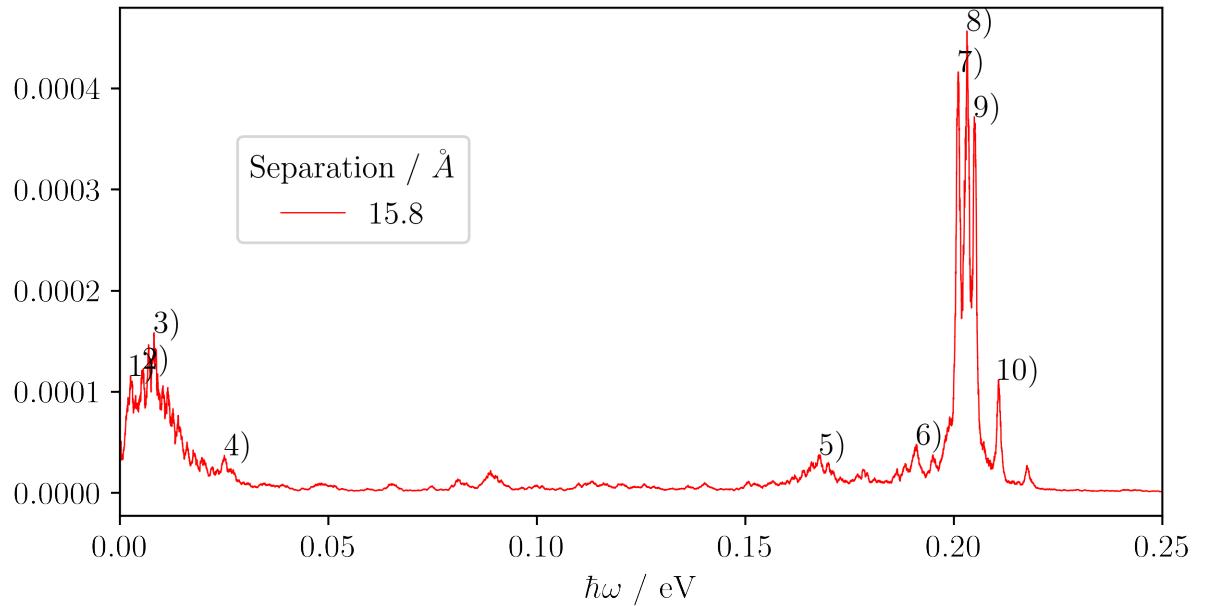


Figure B.5: Spectral density of the LH2 exciton coupling energies with labelled peaks.

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

Table B.6: Peak positions and heights for the spectral density of LH2 Mg-Mg distances.

### B.1.1.6 Interchromophore distances

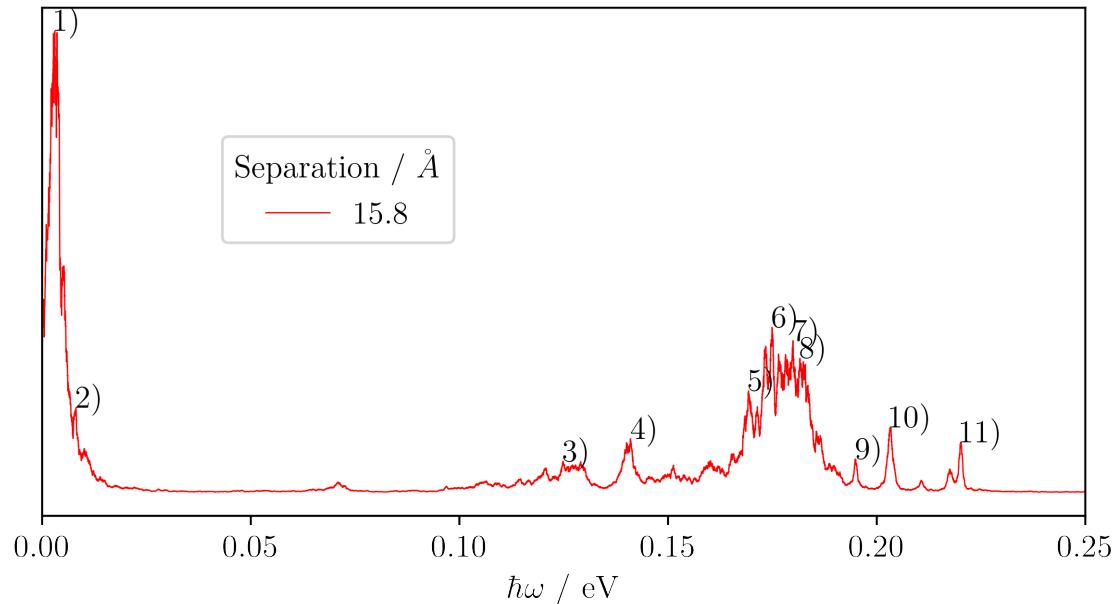


Figure B.6: Spectral density of the LH2 Mg-Mg distances with labelled peaks.

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

Table B.7: Peak positions and heights for the spectral density of LH2 N axes deformations.

#### B.1.1.7 N Axes Deformation

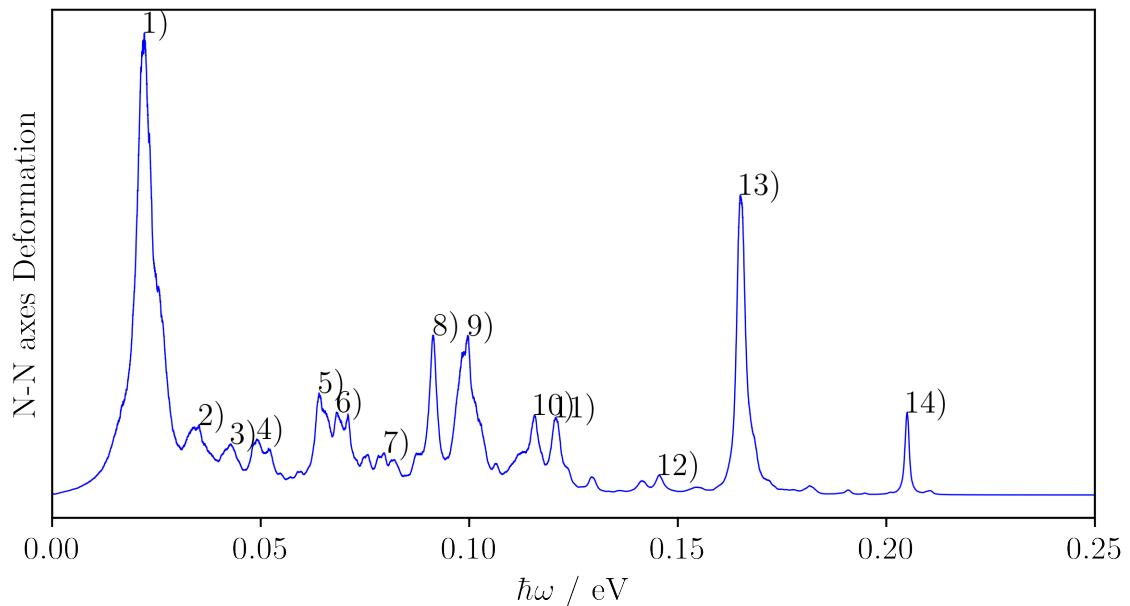


Figure B.7: Spectral density of the LH2 N axes deformations with labelled peaks.

#### B.1.1.8 Diethyl-Ether system

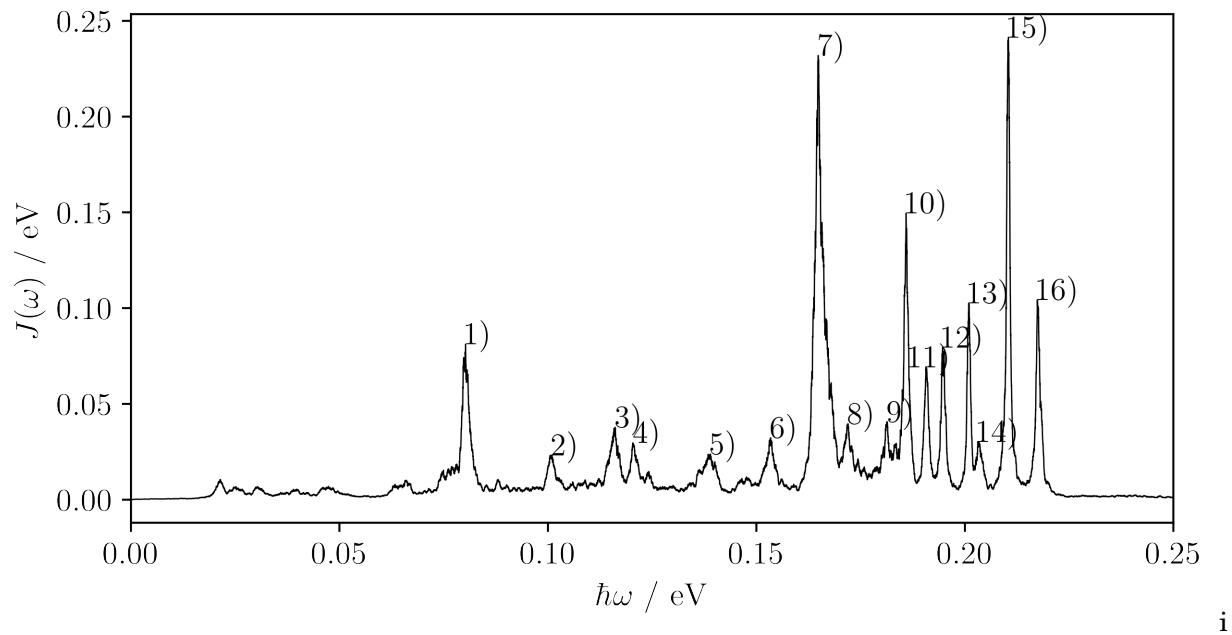


Figure B.8: Spectral density of the  $Q_y$  transition energies for chlorophyll in explicit diethyl ether, with labelled peaks.

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

Table B.8: Peak positions and heights for the spectral density of  $Q_y$  transition of chlorophyll in explicit diethyl ether solvent.

**B.1.1.9 All Spectra**

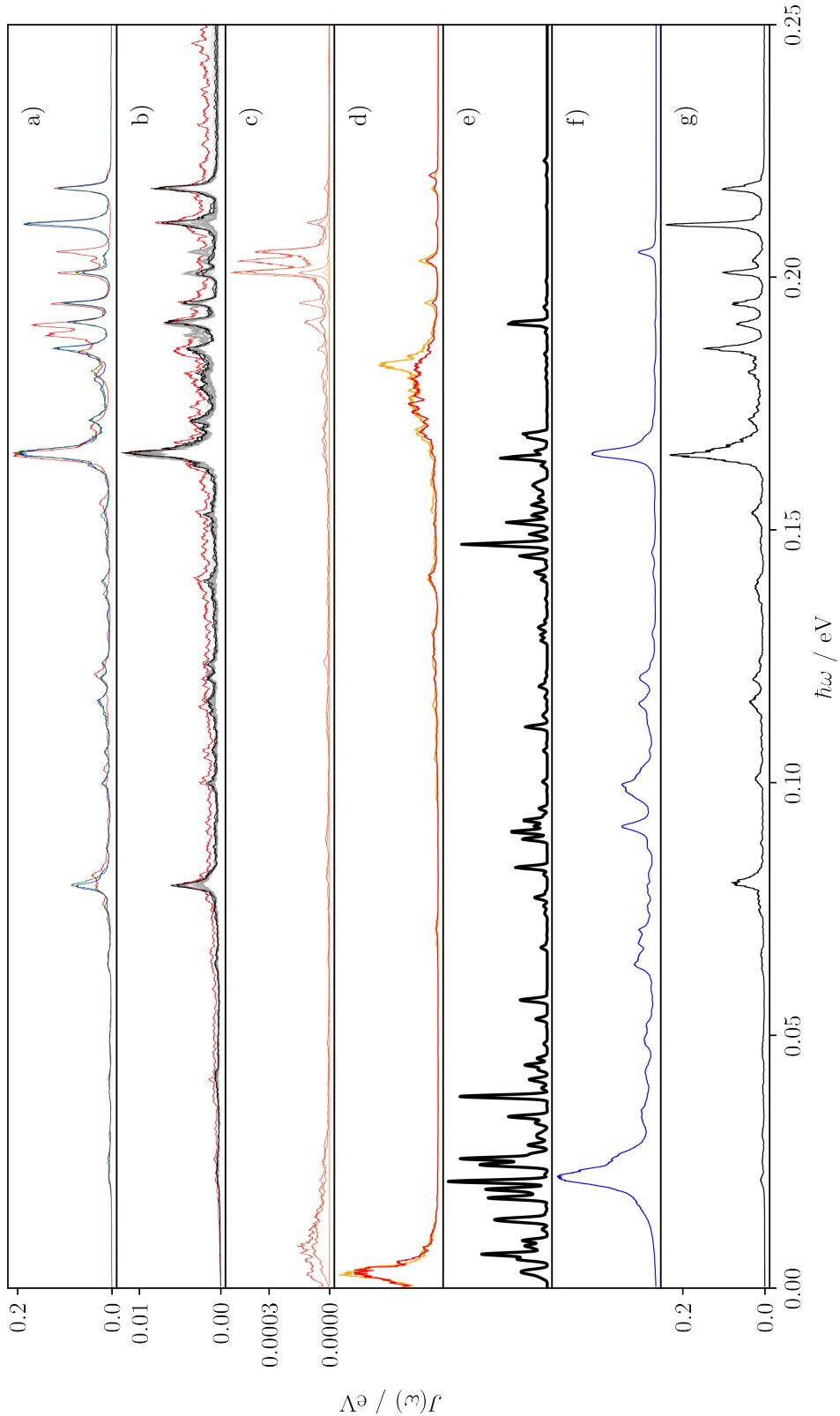


Figure B.9: A stacked version of all of the above spectra.



## BIBLIOGRAPHY

- [1] H. Strain, M. Thomas, and J. Katz.  
Spectral Absorption Properties of Ordinary and Fully Deuteriated Chlorophylls a and b.  
*Biochem. Biophys. Acta.*, 75:306–311, 1963.
- [2] Maria Mallus, Yashoj Shakya, Jigneshkumar Dahyabhai, and Ulrich Kleinekathöfer.  
Environmental effects on the dynamics in the light-harvesting complexes LH2 and LH3  
based on molecular simulations.  
*Chem. Phys.*, 515:141–151, 2018.
- [3] Robert E Blankenship.  
Future Perspectives in Plant Biology Early Evolution of Photosynthesis 1.  
*Plant Physiol.*, 154:434–438, 2010.
- [4] Richard J. Cogdell and Tu C. Nguyen-Phan.  
Photosynthesis I The Purple Photosynthetic Bacterial Light Harvesting System.  
*Encycl. Biol. Chem. Third Ed.*, 2:291–304, jan 2021.
- [5] 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 Or-  
ganisms.  
*Chem. Rev.*, 117(2):249–293, 2017.
- [6] Liam Cleary, Hang Chen, Chern Chuang, Robert J Silbey, and Jianshu Cao.  
Optimal fold symmetry of LH2 rings on a photosynthetic membrane.  
*PNAS*, 110(21):8537–8542, 2013.
- [7] Carles Curutchet and Benedetta Mennucci.  
Quantum chemical studies of light harvesting.  
*Chem. Rev.*, 117(2):294–343, 2017.
- [8] Villy Sundström, Tõnu Pullerits, and Rienk Van Grondelle.  
Photosynthetic Light-Harvesting: Reconciling Dynamics and Structure of Purple Bacterial  
LH2 Reveals Function of Photosynthetic Unit.

## BIBLIOGRAPHY

---

- J. Phys. Chem. B*, 103:2327–2346, 1999.
- [9] Ana Damjanović, Ioan Kosztin, Ulrich Kleinekathöfer, and Klaus Schulten.  
Excitons in a photosynthetic light-harvesting system: A combined molecular dynamics,  
quantum chemistry, and polaron model study.  
*Phys. Rev. E*, 65:031919, 2002.
- [10] Carolin König and Johannes Neugebauer.  
Quantum chemical description of absorption properties and excited-state processes in  
photosynthetic systems.  
*Chem. Phys. Chem.*, 13(2):386–425, 2012.
- [11] 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.
- [12] Sergei Tretiak, Chris Middleton, Vladimir Chernyak, and Shaul Mukamel.  
Exciton Hamiltonian for the Bacteriochlorophyll System in the LH2 Antenna Complex of  
Purple Bacteria.  
*J. Phys. Chem. B*, 104:4519–4528, 2000.
- [13] 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.  
*Q. Rev. Biophys.*, 39(3):277–324, 2006.
- [14] Elad Harel and Gregory S Engel.  
Quantum coherence spectroscopy reveals complex dynamics in bacterial light-harvesting  
complex 2 (LH2).  
*PNAS*, 109(3):706–711, 2012.
- [15] G. McDermott, Steve M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz,  
R. J. Cogdell, and N. W. Isaacs.  
Crystal structure of an integral membrane light-harvesting complex from photosynthetic  
bacteria.  
*Nature*, 374:517, 1995.
- [16] 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.

- [17] Tessa R Calhoun, Naomi S Ginsberg, Gabriela S Schlau-Cohen, Yuan-Chung Cheng, Matteo Ballottari, Roberto Bassi, and Graham R Fleming.  
Quantum Coherence Enabled Determination of the Energy Landscape in Light-Harvesting Complex II.  
*J. Phys. Chem. B*, 113(51):16291–16295, 2009.
- [18] Parveen Akhtar, Paweł J. Nowakowski, Wenda Wang, Thanh Nhut Do, Songhao Zhao, Giuliano Siligardi, Győző Garab, Jian Ren Shen, Howe Siang Tan, and Petar H. Lambrev.  
Spectral tuning of light-harvesting complex II in the siphonous alga Bryopsis corticulans and its effect on energy transfer dynamics.  
*Biochim. Biophys. Acta - Bioenerg.*, 1861(7):148191, jul 2020.
- [19] Vladimíra Moulisová, Larry Luer, Sajjad Hoseinkhani, Tatas H.P. Brotosudarmo, Aaron M. Collins, Guglielmo Lanzani, Robert E. Blankenship, and Richard J. Cogdell.  
Low light adaptation: Energy transfer processes in different types of light harvesting complexes from Rhodopseudomonas palustris.  
*Biophys. J.*, 97(11):3019–3028, dec 2009.
- [20] Kapil Amarnath, Doran I G Bennett, Anna R Schneider, Graham R Fleming, Jim Barber, Victor S Batista, and Thomas Renger.  
Multiscale model of light harvesting by photosystem II in plants.  
*PNAS*, 113(5):1156–1161, 2016.
- [21] 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.  
*Chem. Sci.*, 10:9650, 2019.
- [22] Daniel Montemayor, Eva Rivera, and Seogjoo J Jang.  
Computational Modeling of Exciton-Bath Hamiltonians for Light Harvesting 2 and Light Harvesting 3 Complexes of Purple Photosynthetic Bacteria at Room Temperature.  
*J. Phys. Chem. B*, 122:3815–3825, 2018.
- [23] 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.
- [24] 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.

## BIBLIOGRAPHY

---

- [25] Nicholas H C Lewis, Natalie L Gruenke, Thomas A A Oliver, Matteo Ballottari, Roberto Bassi, and Graham R Fleming.  
Observation of Electronic Excitation Transfer Through Light Harvesting Complex II Using Two-Dimensional Electronic-Vibrational Spectroscopy.  
*J. Phys. Chem. Lett.*, 7:4197–4206, 2016.
- [26] Andreas Barth.  
The infrared absorption of amino acid side chains.  
*Prog. Biophys. Mol. Biol.*, 74:141–173, 2000.
- [27] Nicholas H C Lewis, Hui Dong, Thomas A A Oliver, and Graham R Fleming.  
A method for the direct measurement of electronic site populations in a molecular aggregate using two-dimensional electronic-vibrational spectroscopy.  
*J. Chem. Phys.*, 143:124203, 2015.
- [28] Johannes Neugebauer.  
Photophysical Properties of Natural Light-Harvesting Complexes Studied by Subsystem Density Functional Theory.  
*J. Phys. Chem. B*, 112:2207–2217, 2008.
- [29] 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.
- [30] D. P. Craig and S. H. Walmsley.  
*Excitons in molecular crystals: theory and applications*.  
Benjamin, New York, 1968.
- [31] Gregory D Scholes and Garry Rumbles.  
Excitons in nanoscale systems.  
*Nat. Mater.*, 5:683–696, 2006.
- [32] Gregory H Wannier.  
The Structure of Electronic Excitation Levels in Insulating Crystals.  
*Phys. Rev.*, 52:191–197, 1937.
- [33] J. Frenkel.  
On The Transformation Of Light Into Heat In Solids.  
*Phys. Rev.*, 37:17–44, 1931.
- [34] A S Davydov.  
Soviet Physics Uspekhi THE THEORY OF MOLECULAR EXCITONS.  
*Sov. Phys. Usp.*, 7:393–448, 1964.

- [35] Thomas Renger, Mohamed El-Amine Madjet, Marcel Schmidt, Julian Adolphs, and Frank Mü.  
Structure-based modeling of energy transfer in photosynthesis.  
*Photosyn. Res.*, 116:367–388, 2013.
- [36] Thorsten Hansen and Luca De Vico.  
Qy and Qx Absorption Bands for Bacteriochlorophyll a Molecules from LH2 and LH3.  
*J. Phys. Chem. A*, 123(25):5283–5292, 2019.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.  
*J. Phys. Chem. B*, 117:4263–4273, 2013.
- [42] 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.
- [43] B Hourahine, B Aradi, and V Blum.  
DFTB+, a software package for efficient approximate density functional theory based atomistic simulations.

## BIBLIOGRAPHY

---

- J. Chem. Phys.*, 152:124101, 2020.
- [44] Frederick R Manby, Thomas F Miller Iii, Peter J Bygrave, Feizhi Ding, Thomas Dresselhaus, Fidel A Batista-Romero, Alexander Buccheri, Callum Bungey, Sebastian J R Lee, Rocco Meli, Kaito Miyamoto, Casper Steinmann, Takashi Tsuchiya, Matthew Welborn, Timothy Wiles, and Zack Williams.  
entos: A quantum molecular simulation package.  
2019.
- [45] Frank Neese.  
The ORCA program system.  
*Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(1):73–78, jan 2012.
- [46] Frank Neese.  
Software update: the ORCA program system, version 4.0.  
*Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(1), jan 2018.
- [47] Stefan Seritan, Christoph Bannwarth, and B Scott Fales.  
TeraChem: Accelerating electronic structure and ab initio molecular dynamics with graphical processing units.  
*J. Chem. Phys.*, 152:224110, 2020.
- [48] Stefan Seritan, Christoph Bannwarth, Bryan S. Fales, Edward G. Hohenstein, Christine M. Isborn, Sara I.L. Kokkila-Schumacher, Xin Li, Fang Liu, Nathan Luehr, James W. Snyder, Chenchen Song, Alexey V. Titov, Ivan S. Ufimtsev, Lee Ping Wang, and Todd J. Martínez.  
TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics.  
*Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 11(2), mar 2021.
- [49] Oliviero Andreussi, Stefan Knecht, Christel M Marian, Jacob Kongsted, and Benedetta Mennucci.  
Carotenoids and Light-Harvesting: From DFT/MRCI to the Tamm-Dancoff Approximation.  
*J. Chem. Theory Comput.*, 11:655–666, 2015.
- [50] 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.
- [51] Mats G. Dahlbom and Jeffrey R. Reimers.  
Successes and failures of time-dependent density functional theory for the low-lying excited states of chlorophylls.  
*Mol. Phys.*, 103(6-8):1057–1065, 2005.

- [52] 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.
- [53] 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.*, 145:054103, 2016.
- [54] 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.  
*J. Chem. Theory Comput.*, 8:527–541, 2012.
- [55] Johann Gasteiger and Mario Marsili.  
A new model for calculating atomic charges in molecules.  
*Tetrahedron Lett.*, 19(34):3181–3184, 1978.
- [56] Stefan Grimme.  
A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules.  
*J. Chem. Phys.*, 138:244104, 2013.
- [57] K. Nishimoto and N. Mataga.  
Electronic Structure and Spectra of Some Nitrogen Heterocycles.  
*Zeitshrift für Phys. Chemie Neue Folge*, 12:335–338, 1957.
- [58] Kimio Ohno.  
Some Remarks on the Pariser-Parr-Pople Method.  
*Theor. chim. Acta*, 2:219–227, 1964.
- [59] G. Klopman.  
A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules.  
*J. Am. Chem. Soc.*, 86(21):4550–4557, 1964.
- [60] 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.

## BIBLIOGRAPHY

---

- Mol. Phys.*, 117(9):1104–1116, 2019.
- [61] 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*, 58:2450–2459, 2018.
- [62] Shomik Verma, Miguel Rivera, and David O Scanlon. Machine learned calibrations to high- throughput molecular excited state calculations.  
*J. Chem. Phys.*, 156:134116, 2022.
- [63] 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.
- [64] Abhishek Sirohiwal, Romain Beraud-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.
- [65] 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.
- [66] 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.
- [67] 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.
- [68] 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.

- [69] Pavlo O Dral.  
Quantum Chemistry in the Age of Machine Learning.  
*J. Phys. Chem. Lett.*, 11:2336–2347, 2020.
- [70] Jorg Behler.  
Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations.  
*Phys. Chem. Chem. Phys.*, 13:17930–17955, 2011.
- [71] Julia Westermayr and Philipp Marquetand.  
Machine Learning for Electronically Excited States of Molecules.  
*Chem. Rev.*, 121:9873–9926, 2020.
- [72] 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.
- [73] 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.
- [74] 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.*, 108:58301, 2012.
- [75] Sebastian Dick and Marivi Fernandez-Serra.  
Learning from the density to correct total energy and forces in first principle simulations.  
*J. Chem. Phys.*, 151:144102, 2019.
- [76] Christoph Scherer, Rene Scheid, Denis Andrienko, and Tristan Bereau.  
Kernel-Based Machine Learning for Efficient Simulations of Molecular Liquids.  
*Cite This J. Chem. Theory Comput.*, 16:3194–3204, 2020.
- [77] Zhi-Wen Zhao, Yun Geng, Alessandro Troisi, and Haibo Ma.  
Performance Prediction and Experimental Optimization Assisted by Machine Learning for Organic Photovoltaics.  
*Adv. Intell. Syst.*, 4(6):2100261, jun 2022.
- [78] Florian Häse, Stephanie Valleau, Edward Pyzer-Knapp, and Alan Aspuru-Guzik.  
Machine learning exciton dynamics.  
*Chem. Sci.*, 7:5139–5147, 2016.

## BIBLIOGRAPHY

---

- [79] 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.
- [80] 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.
- [81] Shampa Raghunathan and U. Deva Priyakumar. Molecular representations for machine learning applications in chemistry. *Int. J. Quantum Chem.*, 122(7), apr 2022.
- [82] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. In *Parallel Distrib. Process. Explor. Microstruct. Cogn. Found.*, pages 318–362. MIT Press, 1986.
- [83] 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.*, 153:074111, 2020.
- [84] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, 2009.
- [85] Carl Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006.
- [86] Elise P Kenny and Ivan Kassal. Benchmarking Calculations of Excitonic Couplings between Bacteriochlorophylls. *J. Phys. Chem. B*, 120:25–32, 2016.
- [87] Graham D. Fletcher, Michael W. Schmidt, Brett M. Bode, and Mark S. Gordon. Distributed data interface in GAMESS. *Comput. Phys. Commun.*, 128(1):190–200, jun 2000.
- [88] Mohit Pandey, Michael Fernandez, Francesco Gentile, Olexandr Isayev, Alexander Tropsha, Abraham C Stern, and Artem Cherkasov. The transformational role of GPU computing and deep learning in drug discovery. *Nat. Mach. Intell.*, 4:211–221, 2022.

- [89] Simon McIntosh-Smith, Charles Gillan, Nico Sanna, Stan Scott, and Thomas Steinke. Special issue of the Journal of Parallel and Distributed Computing (JDPC) on novel architectures for high-performance computing. *J. Parallel Distrib. Comput.*, 73(11):1415–1416, 2013.
- [90] Ivan S Ufimtsev and Todd J Martínez. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.*, 4:222–231, 2008.
- [91] Mark S. Friedrichs, Peter Eastman, Vishal Vaidyanathan, Mike Houston, Scott Legrand, Adam L. Beberg, Daniel L. Ensign, Christopher M. Bruns, and Vijay S. Pande. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.*, 30(6):864–872, apr 2009.
- [92] Xin Wu, Axel Koslowski, and Walter Thiel. Semiempirical Quantum Chemical Calculations Accelerated on a Hybrid Multicore CPU-GPU Computing Platform. *J. Chem. Theory Comput.*, 8:2272–2281, 2012.
- [93] Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, Rafal P Wiewiora, Bernard R Brooks, and Vijay S Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Comput. Biol.*, 2017.
- [94] 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.
- [95] Aaron Sisto, David R Glowacki, and Todd J Martinez. Ab Initio Nonadiabatic Dynamics of Multichromophore Complexes: A Scalable Graphical-Processing-Unit-Accelerated Exciton Framework. *Acc. Chem. Res.*, 47:2857–2866, 2014.
- [96] 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 LH<sub>2</sub> complex with a GPU-accelerated: Ab initio exciton model. *Phys. Chem. Chem. Phys.*, 19(23):14924–14936, 2017.

## BIBLIOGRAPHY

---

- [97] Neepa T Maitra.  
Perspective: Fifty years of density-functional theory in chemical physics.  
*J. Chem. Phys.*, 144:220901, 2016.
- [98] P. Hohenberg and W. Kohn.  
Inhomogeneous Electron Gas.  
*Phys. Rev.*, 136(3B):864, 1964.
- [99] W Kohn and L J Sham.  
Self-Consistent Equations Including Exchange and Correlation Effects.  
*Phys. Rev.*, 140(4A):1133, 1965.
- [100] Susi Lehtola.  
Assessment of Initial Guesses for Self-Consistent Field Calculations. Superposition of Atomic Potentials: Simple yet Efficient.  
*J. Chem. Theory Comput.*, 15:1593–1604, 2019.
- [101] David C. Langreth and M. J. Mehl.  
Beyond the local-density approximation in calculations of ground-state electronic properties.  
*Phys. Rev. B*, 28(4):1809, 1983.
- [102] Jianwei Sun, Martijn Marsman, Adrienn Ruzsinszky, Georg Kresse, and John P Perdew.  
Improved lattice constants, surface energies, and CO desorption energies from a semilocal density functional.  
*Phys. Rev. B*, 83(4):121410–121411, 2011.
- [103] Axel D Becke.  
A new mixing of Hartree-Fock and local density-functional theories.  
*J. Chem. Phys.*, 98(2):1372, 1993.
- [104] S. F. Boys.  
Electronic wavefunctions I. A general method of calculation for the stationary states of any molecular system.  
*Proc. R. Soc. London A*, A200:542, 1950.
- [105] C C J Roothaan.  
New Developments in Molecular Orbital Theory.  
*Rev. Mod. Phys.*, 23(2):69, 1951.
- [106] 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):12947–12957, 1994.
- [107] Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R. Manby, and Thomas F Miller III.  
OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features.  
*J. Chem. Phys.*, 153:124111, 2020.
- [108] Ganesh Hegde and R Chris Bowen.  
Machine-learned approximations to Density Functional Theory Hamiltonians.  
*Sci. Rep.*, 7:42669, 2017.
- [109] Omer H. Omar, Marcos Del Cueto, Tahereh Nemataram, and Alessandro Troisi.  
High-throughput virtual screening for organic electronics: a comparative study of alternative strategies.  
*J. Mater. Chem. C*, 9:13557, 2021.
- [110] 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.
- [111] Pekka Koskinen and Ville Mäkinen.  
Density-functional tight-binding for beginners.  
*Comput. Mater. Sci.*, 47(1):237–253, 2009.
- [112] M Elstner, D Porezag, G Jungnickel, J Elsner, M Haugk, Th Frauenheim, S Suhai, and G Seifert.  
Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties.  
*Phys. Rev. B*, 58(11):7260, 1998.
- [113] Nir Goldman and Laurence E Fried.  
Extending the Density Functional Tight Binding Method to Carbon Under Extreme Conditions.  
*J. Phys. Chem. C*, 116:2198–2204, 2012.
- [114] Carlo Andrea Rozzi, Filippo Troiani, Ivano Tavernelli, Kang Lan, Shijie Xie, Xiangji Cai, Chenfeng Cao, and Shi-Yao Hou.  
Atomistic theory of transport in organic and inorganic nanostructures.  
*Rep. Prog. Phys.*, 67:1497–1561, 2004.

## BIBLIOGRAPHY

---

- [115] Christof Köhler, Gotthard Seifert, and Thomas Frauenheim.  
Density functional based calculations for Fen (n<32).  
*Chem. Phys.*, 309(1):23–31, feb 2005.
- [116] T. A. Niehaus, M. Rohlfing, F. Della Sala, A. Di Carlo, and T. A. Frauenheim.  
Quasiparticle energies for large molecules: A tight-binding-based Green's-function approach.  
*Phys. Rev. A*, 71:022508, 2005.
- [117] Wen Ge Han, Marcus Elstner, K. J. Jalkanen, Thomas Frauenheim, and Sándor Suhai.  
Hybrid SCC-DFTB/molecular mechanical studies of H-bonded systems and of N-acetyl-(L-Ala)<sub>n</sub> N'-methylamide helices in water solution.  
*Int. J. Quantum Chem.*, 78(6):459–479, 2000.
- [118] V. Lutsker, B. Aradi, and T. A Niehaus.  
Implementation and benchmark of a long-range corrected functional in the density functional based tight-binding method.  
*J. Chem. Phys.*, 143:184107, 2015.
- [119] Maja Gruden, Ljubica Andjelkovic, Akkarapattiakal Kuriappan Jissy, Stepan Stepanović, Matija Zlatar, Qiang Cui, and Marcus Elstner.  
Benchmarking density functional tight binding models for barrier heights and reaction energetics of organic molecules.  
*J. Comput. Chem.*, 38(25):2171–2185, sep 2017.
- [120] Van Quan Vuong, # Jissy, Akkarapattiakal Kuriappan, Maximilian Kubillus, Julian J Kranz, Thilo Mast, Thomas A Niehaus, Stephan Irle, and Marcus Elstner.  
Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules.  
*J. Chem. Theory Comput.*, 14:115–125, 2018.
- [121] Patrick Melix, Augusto Faria Oliveira, Robert Rüger, and Thomas Heine.  
Spin polarization in SCC-DFTB.  
*Theor. Chem. Acc.*, 135(9):1–6, 2016.
- [122] 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.
- [123] 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.
- [124] 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.
- [125] 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.
- [126] Sebastian Spicher and Stefan Grimme.  
Robust atomistic modeling of materials, organometallic and biochemical systems.  
*Angew. Chemie*, 59:2–11, 2020.
- [127] 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.
- [128] 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.
- [129] Erin R Johnson and Axel D Becke.  
A post-Hartree-Fock model of intermolecular interactions.  
*J. Chem. Phys.*, 123:154104, 2005.
- [130] Axel D Becke and Erin R Johnson.  
A density-functional model of the dispersion interaction.  
*J. Chem. Phys.*, 123:6158, 2005.
- [131] Adèle D. Laurent and Denis Jacquemin.  
TD-DFT benchmarks: A review.  
*Int. J. Quantum Chem.*, 113(17):2019–2039, sep 2013.
- [132] Erich Runge and E K U Gross.  
Density-Functional Theory for Time-Dependent Systems.  
*Phys. Rev. Lett.*, 52:997–1000, 1984.
- [133] M A L Marques and E K U Gross.

## BIBLIOGRAPHY

---

- TIME-DEPENDENT DENSITY FUNCTIONAL THEORY.  
*Annu. Rev. Phys. Chem.*, 55:427–55, 2004.
- [134] M Petersilka, U J Gossmann, and E K U Gross.  
Excitation Energies from Time-Dependent Density-Functional Theory.  
*Phys. Rev. Lett.*, 76(8):1212, 1996.
- [135] M. E. Casida.  
*Recent Advances in Density Functional Methods*.  
World Scientific Singapore, 1 edition, 1995.
- [136] Mark E. Casida and Tomasz A. Wesolowski.  
Generalization of the Kohn-Sham equations with constrained electron density formalism  
and its time-dependent response theory formulation.  
*Int. J. Quantum Chem.*, 96(6):577–588, feb 2004.
- [137] Miguel A. L. Marques, Neepa T. Maitra, Fernando M. S. Nogueira, Eberhard K. U. Gross,  
and Angel Rubio.  
*Fundamentals of Time-Dependent Density Functional Theory*.  
Springer, New York, 2012.
- [138] M E Casida and M Huix-Rotllant.  
Progress in Time-Dependent Density-Functional Theory.  
*Annu. Rev. Phys. Chem.* 2012, 63:287–323, 2012.
- [139] Christel M Marian and Natalie Gilka.  
Performance of the Density Functional Theory/ Multireference Configuration Interaction  
Method on Electronic Excitation of Extended  $\pi$ -Systems.  
*J. Chem. Theory Comput.*, 4:1501–1515, 2008.
- [140] Serhiy Perun, Jçrg Tatchen, and Christel M Marian.  
Singlet and Triplet Excited States and Intersystem Crossing in Free-Base Porphyrin:  
TDDFT and DFT/MRCI Study.  
*ChemPhysChem*, 9:282–292, 2008.
- [141] Mihajlo Etinski, Jörg Tatchen, and Christel M Marian.  
Time-dependent approaches for the calculation of intersystem crossing rates.  
*J. Chem. Phys.*, 134:154105, 2011.
- [142] Andreas B J Parusel and Stefan Grimme.  
A Theoretical Study of the Excited States of Chlorophyll a and Pheophytin a.  
*J. Phys. Chem. B*, 104:5395–5398, 2000.

- [143] Nanna Holmgård List, Carles Curutchet, Stefan Knecht, Benedetta Mennucci, and Jacob Kongsted.  
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.
- [144] Tino Gimon, Andrey Ipatov, Andreas Heßelmann, and Andreas Görling.  
Qualitatively Correct Charge-Transfer Excitation Energies in HeH<sup>+</sup> by Time-Dependent Density-Functional Theory Due to Exact Exchange Kohn-Sham Eigenvalue Differences.  
*J. Chem. Theory Comput.*, 5:27, 2009.
- [145] R O Jones and O Gunnarsson.  
The density functional formalism, its applications and prospects.  
*Rev. Mod. Phys.*, 61(3):689, 1989.
- [146] A. Hellman, B. Razaznejad, and B. I. Lundqvist.  
Potential-energy surfaces for excited states in extended systems.  
*J. Chem. Phys.*, 120:4593, 2004.
- [147] J. Almlöf, K. Faegri, and K. Korsell.  
Principles for a direct SCF approach to LCAO-MO ab-initio calculations.  
*J. Comput. Chem.*, 3(3):385–399, 1982.
- [148] s. William J. Hunt and William A. Goddard.  
Excited States of H<sub>2</sub>O using improved virtual orbitals.  
*Chem. Phys. Lett.*, 3(6):414–418, jun 1969.
- [149] 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.
- [150] Giuseppe M J Barca, Andrew T B Gilbert, and Peter M W Gill.  
Simple Models for Difficult Electronic Excitations.  
*J. Chem. Theory Comput.*, 14:1501–1509, 2018.
- [151] 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.

## BIBLIOGRAPHY

---

- [152] Jeppe Gavnholz, Thomas Olsen, Mads Engelund, and Jakob Schiøtz.  
Δ self-consistent field method to obtain potential energy surfaces of excited molecules on  
surfaces.  
*Phys. Rev. B*, 78:075441, 2008.
- [153] 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.
- [154] Tim Kowalczyk, Shane R Yost, and Troy Van Voorhis.  
Assessment of the ΔSCF density functional theory approach for electronic excitations in  
organic dyes.  
*J. Chem. Phys.*, 134:054128, 2011.
- [155] R. McWeeny.  
Some Recent Advances in Density Matrix Theory.  
*Rev. Mod. Phys.*, 32(2):335–369, 1960.
- [156] Sergei Tretiak and Shaul Mukamel.  
Density Matrix Analysis and Simulation of Electronic Excitations in Conjugated and  
Aggregated Molecules.  
*Chem. Rev.*, 102:3171–3212, 2002.
- [157] Thibaud Etienne.  
Transition matrices and orbitals from reduced density matrix theory.  
*J. Chem. Phys.*, 142:4775, 2015.
- [158] 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 configura-  
tional interaction.  
*Phys. Rev.*, 97(6):1474–1489, 1955.
- [159] Gregory D Scholes.  
LONG-RANGE RESONANCE ENERGY TRANSFER IN MOLECULAR SYSTEMS.  
*Annu. Rev. Phys. Chem.*, 54:57–87, 2003.
- [160] 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.
- [161] 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.
- [162] 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.
- [163] M. E. Madjet, A. Abdurahman, and T. Renger.  
Intermolecular Coulomb Couplings from Ab Initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers.  
*J. Phys. Chem. B*, 110:52, 2006.
- [164] Thomas Renger.  
Theory of excitation energy transfer: from structure to function.  
*Photosyn. Res.*, 102:471–485, 2009.
- [165] Brent P Krueger, Gregory D Scholes, and Graham R Fleming.  
Calculation of Couplings and Energy-Transfer Pathways between the Pigments of LH2 by the ab Initio Transition Density Cube Method.  
*J. Ph*, 102:5378–5386, 1998.
- [166] Susannah Bourne Worster, Oliver Feighan, and Frederick R Manby.  
Reliable transition properties from excited-state mean-field calculations.  
*J. Chem. Phys.*, 154:124106, 2021.
- [167] U Terranova and D R Bowler.  
Δ Self-Consistent Field Method for Natural Anthocyanidin Dyes.  
*J. Chem. Theory Comput.*, 9:3181–3188, 2013.
- [168] Alexander G Donchev, G Taube, Elizabeth Decolvenaere, Cory Hargus, Robert T McGibbon, Ka-Hei Law, Brent A Gregersen, Je-Luen Li, Kim Palmo, Karthik Siva, Michael Bergdorf, John L Klepeis, and David E Shaw.  
Quantum chemical benchmark databases of gold-standard dimer interaction energies.  
*Sci. Data*, 8(55):1, 2021.
- [169] 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.
- [170] Per-Olov Löwdin.

## BIBLIOGRAPHY

---

- On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals.  
*J. Chem. Phys.*, 18:365, 1950.
- [171] Carlo Adamo and Vincenzo Barone.  
Towards reliable density functional methods without adjustable parameters: The PBE0 model.  
*J. Chem. Phys.*, 110:6158, 1999.
- [172] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs.  
Fully optimized contracted Gaussian basis sets for atoms Li to Kr.  
*J. Chem. Phys.*, 97:2571, 1992.
- [173] 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.
- [174] Arnim Hellweg, Sarah A. Grün, and Christof Hättig.  
Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states.  
*Phys. Chem. Chem. Phys.*, 10:4119–4127, 2008.
- [175] Marcus Johansson.  
libmsym: molecular point group symmetry library, 2018.
- [176] Oliver Feighan, Frederick R Manby, Susannah Bourne-Worster, and ) Affiliations.  
An efficient protocol for excited states of large biochromophores.  
*J. Chem. Phys.*, 158(2):024107, 2023.
- [177] So Hirata and Martin Head-Gordon.  
Time-dependent density functional theory within the Tamm-Dancoff approximation.  
*Chem. Phys. Lett.*, 314(3-4):291–299, dec 1999.
- [178] M Yoshimine, A D Mclean, B Liu, and G S Hammond.  
Toward a Systematic Molecular Orbital Theory for Excited States.  
*J. Am. Chem. Soc.*, 96(20):334, 1992.
- [179] H. Appel, E. K.U. Gross, and K. Burke.  
Double-pole approximation in time-dependent density functional theory.  
*Int. J. Quantum Chem.*, 106(14):2840–2847, nov 2006.
- [180] M. Belén Oviedo and Cristian G. Sánchez.

- Transition Dipole Moments of the Q<sub>y</sub> Band in Photosynthetic Pigments.  
*J. Phys. Chem. A*, 115:12280–12285, 2011.
- [181] Giuseppe Zucchelli, Robert C. Jennings, Flavio M. Garlaschi, Gianfelice Cinque, Roberto Bassi, and Oliviero Cremonesi.  
The calculated in vitro and in vivo chlorophyll a absorption bandshape.  
*Biophys. J.*, 82(1):378–390, 2002.
- [182] 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.
- [183] V Zazubovich, I Tibe, and G J Small.  
Bacteriochlorophyll a Franck-Condon Factors for the S<sub>0</sub> f S<sub>1</sub>(Q<sub>y</sub>) Transition.  
*J. Phys. Chem. B*, 105:12410–12417, 2001.
- [184] Keisuke Saito, Koji Mitsuhashi, and Hiroshi Ishikita.  
Dependence of the chlorophyll wavelength on the orientation of a charged group: Why does the accessory chlorophyll have a low site energy in photosystem II?  
*J. Photochem. Photobiol. A Chem.*, 402:112799, nov 2020.
- [185] M. Fragata, B. Norden, and T. Kurucsev.  
Linear Dichroism (250-700 nm) of chlorophyll a and pheophytin a oriented in a lamellar phase of glycerylmonoctanoate/H<sub>2</sub>O. Characterization of electronic transitions.  
*Photochem. Photobiol.*, 47(1):133–143, 1988.
- [186] David M. Himmelblau.  
*Applied nonlinear programming*.  
MgGraw-Hill, New York, 1972.
- [187] 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.  
*J. Chem. Theory Comput.*, 18(6):3438–3446, 2022.
- [188] 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.
- [189] A. R. Kelly and G. Porter.  
Model systems for photosynthesis I Energy transfer and light harvesting mechanisms.  
*Proc. R. Soc. London A*, 315:149–161, 1970.

## BIBLIOGRAPHY

---

- [190] A. R. Kelly and L. K. Patterson.  
Model systems for photosynthesis II Concentration quenching of chlorophyll b fluorescence in solid solutions.  
*Proc. R. Soc. London A*, 324:117–126, 1971.
- [191] G. S. Beddard, S. E. Carlin, and G. Porter.  
Concentration quenching of chlorophyll fluorescnece in bilayer lipid vesicles and liposomes.  
*Chem. Phys. Lett.*, 43(1):27–32, 1976.
- [192] V. P. Gutschick.  
Concentration Quenching in Chlorophyll-a and Relation to Functional Charge Transfer in vivo.  
*J. Bioenerg. Biomembr.*, 10:153–170, 1978.
- [193] G. R. Seely.  
The energetics of electron-transfer reactions of chlorophyll and other compounds.  
*Photochem. Photobiol.*, 27:639–654, 1978.
- [194] M. J. Yuen, L. L. Shipman, J. Katz, and C. Hindman.  
Concentration Quenching of fluorescence from chlorophyll-a, pheophytin-a, pyropheophytin-a and their covalently linked pairs.  
*Photochem. Photobiol.*, 32(3):281–296, 1980.
- [195] W F Watson and Robert Livingston.  
Self-quenching of nitrobenzoxadiazole labeled phospholipids in lipid membranes.  
*J. Chem. Phys.*, 18(6):802, 1950.
- [196] R. A. Marcus.  
CHEMICAL AND ELECTROCHEMICAL ELECTRON-TRANSFER THEORY.  
*Annu. Rev. Phys. Chem.*, 15:155–196, 1964.
- [197] A. Warshel and W. W. Parson.  
Dynamics of biochemical and biophysical reactions: Insight from computer simulations.  
*Q. Rev. Biophys.*, 34(4):563–679, 2001.
- [198] G.S. Beddard and G. Porter.  
Concentration quenching in chlorophyll.  
*Nature*, 260:366–367, 1976.
- [199] Harald Oberhofer and Jochen Blumberger.  
Revisiting electronic couplings and incoherent hopping models for electron transport in crystalline C 60 at ambient temperatures.  
*Phys. Chem. Chem. Phys.*, 14:13846–13852, 2012.

- [200] Harald Obenhofer, Karsten Reuter, and Jochen Blumberger.  
Charge Transport in Molecular Materials: An Assessment of Computational Methods.  
*Chem. Rev.*, 117:10319–10357, 2017.
- [201] Seymour Steven Brody.  
Fluorescence lifetime, yield, energy transfer and spectrum in photosynthesis, 1950-1960 \*.  
*Photosynth. Res.*, 73:127–132, 2002.
- [202] G. S. Beddard and G. Porter.  
Concentration quenching in chlorophyll.  
*Nature*, 260:367, 1976.
- [203] Adam Kubas, Fruzsina Gajdos, Alexander Heck, Harald Oberhofer, Marcus Elstner, and Jochen Blumberger.  
Electronic couplings for molecular charge transfer: benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations. II.  
*Phys. Chem. Chem. Phys.*, 17:14342–14354, 2015.
- [204] David M Silver and Klaus Ruedenberg.  
Overlap Integrals over Slater-Type Atomic Orbitals.  
*Coulomb Integr. over Slater-Type At. Orbitals J. Chem. Phys.*, 49(10):4301–4306, 1968.
- [205] N. C. Handy, M. T. Marron, and H. J. Silverstone.  
Long-Range Behaviour of Hartree-Fock Orbitals.  
*Phys. Rev.*, 180(1):45–48, 1969.
- [206] Samuel M Blau, Doran I G Bennett, Christoph Kreisbeck, Gregory D Scholes, and Aíán Aspuru-Guzik.  
Local protein solvation drives direct down-conversion in phycobiliprotein PC645 via incoherent vibronic transport.  
*PNAS*, 115(15):E3342–E3350, 2018.
- [207] Tim J. Zuehlsdorff, Hanbo Hong, Liang Shi, and Christine M. Isborn.  
Influence of Electronic Polarization on the Spectral Density.  
*J. Phys. Chem. B.*, 124:531–543, 2020.
- [208] Akihito Ishizaki and Graham R Fleming.  
Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature.  
*PNAS*, 106(41):17255–17260, 2009.
- [209] Thomas Renger and R A Marcus.

## BIBLIOGRAPHY

---

- On the relation of protein dynamics and exciton relaxation in pigment-protein complexes:  
An estimation of the spectral density and a theory for the calculation of optical spectra.  
*J. Chem. Phys.*, 116:9997, 2002.
- [210] Daniele Padula, Myeong H Lee, Kirsten Claridge, and Alessandro Troisi.  
Chromophore-Dependent Intramolecular Exciton-Vibrational Coupling in the FMO Complex: Quantification and Importance for Exciton Dynamics.  
*J. Phys. Chem. B*, 121:10026–10035, 2017.
- [211] M. H. C. Koolhaas, G. Van Der Zwan, R. N. Frese, and R. Van Grondelle.  
Red Shift of the Zero Crossing in the CD Spectra of the LH2 Antenna Complex of Rhodopseudomonas acidiphila: A Structure-Based Study.  
*J. Phys. Chem. B*, 101:7262–7270, 1997.
- [212] Ralph Jimenez, Srivatsan N Dikshit, Stephen E Bradforth, and Graham R Fleming.  
Electronic Excitation Transfer in the LH2 Complex of Rhodobacter sphaeroides.  
*J. Phys. Chem.*, 100:6825–6834, 1996.
- [213] Mirianas Chachisvilis, Oliver Ku, Tõnu Pullerits, and Villy Sundstro.  
Excitons in Photosynthetic Purple Bacteria: Wavelike Motion or Incoherent Hopping?  
*J. Phys. Chem. B*, 101:7275–7283, 1997.
- [214] Gregory D Scholes and Graham R Fleming.  
On the Mechanism of Light Harvesting in Photosynthetic Purple Bacteria: B800 to B850 Energy Transfer.  
*J. Phys. Chem. B*, 104:1854–1868, 2000.
- [215] Kirsten Claridge, Daniele Padula, and Alessandro Troisi.  
How fine-tuned for energy transfer is the environmental noise produced by proteins around biological chromophores?  
*Phys. Chem. Chem. Phys.*, 20:17279, 2018.
- [216] Dario Polli, Giulio Cerullo, Guglielmo Lanzani, Sandro De Silvestri, Hideki Hashimoto, and Richard J. Cogdell.  
Carotenoid-bacteriochlorophyll energy transfer in LH2 complexes studied with 10-fs time resolution.  
*Biophys. J.*, 90(7):2486–2497, 2006.
- [217] Dustin A Tracy, Sebastian Fernandez-Alberti, Sergei Tretiak, and Adrian E Roitberg.  
Adiabatic Excited-State Molecular Dynamics with an Explicit Solvent: NEXMD-SANDER Implementation.  
*J. Chem. Theory Comput.*, 18:5213–5220, 2022.

- [218] Ricardo A Matute, Renato Contreras, and Leticia Gonz Alez.  
Time-Dependent DFT on Phytochrome Chromophores: A Way to the Right Conformer.  
*J. Phys. Chem. Lett.*, 1:796–801, 2010.
- [219] Gregory D Scholes, Francesca Fassioli, Raymond Dinshaw, and Paul C Arpin.  
Photosynthetic light harvesting: excitons and coherence.  
*J. R. Soc. Interface*, 11:20130901, 2013.
- [220] Emanuele Coccia, Daniele Varsano, and Leonardo Guidoni.  
Ab Initio Geometry and Bright Excitation of Carotenoids: Quantum Monte Carlo and Many  
Body Green's Function Theory Calculations on Peridinin Chart 1. Sketch Representa-  
tion of Peridinin (PID) and of the Reduced Peridinin Model (PID1).  
*J. Chem. Theory Comput.*, 10:501–506, 2014.

