
Atomistic Modelling of Aggregate Chlorophyll Systems

By

OLIVER J. H. 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.

JANUARY, 2022

Word count: 0

ABSTRACT

Here goes the abstract

DEDICATION AND ACKNOWLEDGEMENTS

Here goes the dedication.

AUTHOR'S DECLARATION

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

SIGNED: DATE:

TABLE OF CONTENTS

	Page
List of Tables	xi
List of Figures	xiii
1 Outline	1
2 Introduction	3
2.1 Quantum Exploits in Light Harvesting Systems	3
2.1.1 Electronic Energy Transfer	3
2.1.2 Coherence	3
2.2 Light-Matter Response	3
2.3 Electronic Structure for Large Systems	3
2.4 The Aim	3
3 Literature Review	5
4 Mean-Field excited states	7
4.1 Theory	8
4.1.1 Δ -SCF	8
4.1.2 Eigenvalue Difference	9
4.1.3 Transition Density and Dipole Moments	10
4.1.4 Semi-empirical extensions	11
4.2 Benchmarking	13
4.2.1 Reference Data and test set	13
4.2.2 Small Systems	13
4.2.3 Non-orthogonality	17
4.2.4 LHII Chlorophyll	19
4.2.5 xTB methods	22
4.3 Conclusions	28
5 Chlorophyll specific methods	31

TABLE OF CONTENTS

5.1	Theory	31
5.1.1	Full solutions TD-DFT equation	31
5.1.2	Approximations to Solutions	32
5.1.3	Integrals approximations	33
5.1.4	Q_y Transition	35
5.1.5	Changes to xTB parameters	39
5.1.6	Transition and Excited State Density from the Ground State	39
5.2	Parameterization	41
5.2.1	Objective Function	41
5.2.2	Minimisation Algorithms	42
5.2.3	Reference Data	45
5.2.4	Results	50
5.3	Cross-validation	54
5.3.1	Vibrational Mode Coupling	54
5.3.2	Absorption Spectra	62
5.4	Conclusions	63
6	Beyond Monomer Chlorophyll	67
6.1	Theory	67
6.1.1	Exciton States	67
6.2	Same-functional Benchmarking	69
6.2.1	LH2 Dimers	69
6.2.2	Rotation	71
6.3	chl-xTB Excitons	77
6.3.1	Full Dimer comparison	77
6.4	Concentration Quenching	90
6.5	Conclusions	92
7	Light Harvesting Complexes	95
7.1	LH2	95
7.1.1	Spectral Density Method	96
7.2	Calculating LH2 Excitons	96
7.2.1	Screening and Embedding	97
7.3	Spectra	99
7.3.1	Sites	99
7.3.2	Exciton states	100
7.4	Assigning Specific Motions	103
7.4.1	Ether system	103
7.4.2	Huang Rhys Factors	104

TABLE OF CONTENTS

7.5 conclusions	107
8 Discussion	109
8.1 Transition Property Approximations	109
8.2 Further Investigations into LHII	109
8.3 Coherence	109
A Appendix A	111
A.1 Electronic Structure Codes	111
A.2 Computational Hardware	111
Bibliography	113

LIST OF TABLES

TABLE	Page
4.1 Mean and standard deviations of the errors, in eV, against SCS-CC2 reference data. The <code>xtb4stda</code> entry represents the eigenvalue difference method that uses the eigenvalues output from this program.	27
5.1 ref data vals	48
5.2 optimized parameters from SLSQP procedure.	50

LIST OF FIGURES

FIGURE	Page
4.1 Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.	15
4.2 Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.	16
4.3 The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown in red, whilst those with the symmetric orthogonalisation correction are shown in black.	18
4.4 Transition energies from Δ -SCF for 26 chlorophyll geometries from the LHII 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.	20
4.5 Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 4.2.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.	21
4.6 A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.	24
4.7 The distributions of errors compared to SCS-CC2 transition energies for the methods included in the Δ -xTB benchmarking.	26
5.1 Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.	36
5.2 The HOMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.	37
5.3 The LUMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.	38
5.4 Positions of the Nelder-Mead simplex vertices for the HimmelBlau function.	46
5.5 Positions of iterative solution vector \mathbf{x} from SLSQP for the HimmelBlau function.	47
5.6 caption	51
5.7 caption	52

LIST OF FIGURES

5.8	Normal modes of the nitrogen-magnesium centre of chlorophyll that have D_{4h} - D_{2h} (left) and D_{4h} - C_s (right) symmetry breaking components.	54
5.9	Change in the N_A - N_C displacements along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail.	56
5.10	Change in the N_A - Mg - N_C angle along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail. The small smaller variance than in figure 5.3.1 led to this metric to not be used in normal mode choices.	57
5.11	Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with different response theories.	62
5.12	Chlorophyll in diethyl ether.	64
6.1	caption	74
6.2	caption	75
6.3	caption	76

OUTLINE

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

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

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

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

CHAPTER 1. OUTLINE

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

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

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

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

INTRODUCTION

Naturally occurring light harvesting systems present an interesting scientific challenge. With near perfect efficiency, the energy from a photon will be taken and transferred to a reaction centre, leading to charge transfer processes that culminate in powering biological systems. Making models that can predict and explain these effects are key to making similarly efficient photovoltaic systems.

2.1 Quantum Exploits in Light Harvesting Systems

Begins a section.

2.1.1 Electronic Energy Transfer

Begins a subsection.

2.1.2 Coherence

2.2 Light-Matter Response

2.3 Electronic Structure for Large Systems

2.4 The Aim

C H A P T E R

3

LITERATURE REVIEW

MEAN-FIELD EXCITED STATES

Previous Published Work

Some of the work presented in this chapter forms part of a paper published with Dr Susannah Bourne-Worster, in March 2021[31]. The account given in this chapter includes research that was also reported in this publication, namely the parts of section 4.2 from 4.2.1 up to but excluding 4.2.5.

This chapter investigates the accuracy of Δ -SCF methods with both an ab initio DFT level of theory as well as with semi-empirical, tight-binding approximations. The reduced cost and moderate accuracy of these methods make them an ideal substitute for full TD-DFT or high-level methods when investigating large systems like chlorophyll. Transition properties were calculated for a range of molecules, as well as for a small set of chlorophyll geometries, using variety of different basis sets, density functionals, response methods and electronic structure methods. Most of the work was compared to either high-standard EOM-CCSD or SCS-CC2 reference, from which conclusions were made on the accuracy of each method. The non-orthogonality issue of the ground and excited states was also investigated for the mean-field Δ -SCF method. Assigning the symmetries of transitions was also investigated to improve the workflow of comparing Δ -SCF and TD-DFT results. It was found that while the DFT methods give reasonable results, the semi-empirical Δ -SCF methods are not as accurate.

Δ -SCF is a well-known method for calculating transition energies, as well as ionisation potentials, electron affinities and other properties. Although slightly less accurate than other high-level response theories, it is much more efficient, requiring only two SCF solutions for the two states of a transition. In the context of calculating large volumes of transition properties for the chlorophyll sites of LHII, this efficiency makes it a good candidate for investigation.

Previous work in the literature has discussed the benefits of the Δ -SCF method for large scale

systems. Notably

4.1 Theory

4.1.1 Δ -SCF

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

$$(4.1) \quad \Delta E = E_2 - E_1$$

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

The issue with finding the excited state solution is that the variation principle and SCF iterative procedure will try to find the global minimum, which is the ground state. The excited state is a local minimum, and so often is less reliable to find as a solution, especially from the standard SAD initial guess. For this reason it is often found that converging to the Δ -SCF excited state will fail. Even when using the ground state as an initial guess with excited state occupations, normal SCF procedure may still collapse back to the ground state. Usually it is necessary to include additional changes to the SCF procedure, such as Fock damping, alternative DIIS methods and sometimes intermediate initial guess steps.

Initially, the excited state was calculated by relaxing the orbitals which contain the excited electron and hole in the ground state space, so that the excited state and ground state are orthogonal [17]. 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 [9]. Alternatively, an SCF like method was proposed, where instead of populating orbitals according to the Aufbau principle, orbitals which most resemble the previous iteration's orbitals should be occupied. This is known as the maximum overlap method (MOM). In the maximum overlap method, each iteration in an SCF procedure produces new molecular orbital coefficients by solving the Roothaan-Hall equations [27], generally given as an eigenvalue problem:

$$(4.2) \quad \mathbf{FC}^n = \mathbf{SC}^n \epsilon$$

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

$$(4.3) \quad \mathbf{F} = f(\mathbf{C}^{n-1})$$

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

$$(4.4) \quad \mathbf{O} = (\mathbf{C}^{\text{old}})^\dagger \mathbf{S} \mathbf{C}^{\text{new}}$$

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

$$(4.5) \quad p_j = \sum_i^n O_{ij} = \sum_v^N \left[\sum_\mu^N \left(\sum_i^n C_{i\mu}^{\text{old}} \right) S_{\mu v} \right] C_{vj}^{\text{new}}$$

where μ, v are orbital indices. the set of orbitals with the highest projection p_j are then populated with electrons. This method can be used for any excited state, with the caveat that the orbital solution will most likely be in the same region as the ground state solution. For a small number of low lying states, this is generally true, and so Δ -SCF can be used to calculate a small spectrum of excited states [9].

Δ -SCF has been shown to be cheap alternative to TD-DFT and other higher level methods [4, 8, 21], without considerable losses of accuracy in certain cases, especially for HOMO-LUMO transitions [20]. Additionally, as the excited state is given as solutions to SCF equations, the gradient of this solution can be given by normal mean-field theory. These gradients would be much cheaper than TD-DFT or coupled cluster methods, which is advantageous for simulating dynamics [8].

4.1.2 Eigenvalue Difference

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

$$(4.6) \quad \Delta E = \epsilon_e - \epsilon_g$$

. Additionally, transition properties can be calculated by constructing transition density matrices from the ground state orbitals such that needing only a single SCF optimization is required. Generally, eigenvalue difference methods are not seen as accurate response methods, but can offer a quick and easy initial value [10].

4.1.3 Transition Density and Dipole Moments

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

$$(4.7) \quad \mathbf{D}^{21} = |\Psi_1\rangle\langle\Psi_2|$$

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

$$(4.8) \quad \mathbf{D}^{21} = \mathbf{C}^{(2)} \text{adj}(\mathbf{S}^{21}) \mathbf{C}^{(1)\dagger}$$

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

$$(4.9) \quad S_{jk}^{21} = \langle\phi_j^2|\phi_k^1\rangle$$

. The dependence on the adjunct of the overlap can be understood using Löwdin's normal rules for non-orthogonal determinants [22]. In the same way, the transition dipole moment is given by

$$(4.10) \quad \langle\Psi_2|\hat{\mu}|\Psi_1\rangle = \sum_{jk} \mu_{jk}^{21} \text{adj}(\mathbf{S}^{21})_{jk}$$

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

$$(4.11) \quad |\mathbf{S}^{21}| = \langle\Psi_2|\Psi_1\rangle$$

The general definition of the transition dipole

$$(4.12) \quad \mu^{1 \rightarrow 2} = \langle\Psi_2|\hat{\mu}|\Psi_1\rangle$$

can be expressed with this transition density matrix as:

$$(4.13) \quad \begin{aligned} \langle\Psi_2|\hat{\mu}|\Psi_1\rangle &= \text{tr}(\hat{\mu}|\Psi_1\rangle\langle\Psi_2|) \\ &= \text{tr}(\hat{\mu}\mathbf{D}^{21}) \end{aligned}$$

4.1.4 Semi-empirical extensions

A main investigation of this chapter is whether the range of DFT methods that could be used for Δ -SCF and eigenvalue difference methods could be extended by using tight-binding methods to predict transition properties. This mainly focused on the recently published GFN-xTB method, parameterized by the Grimme group [14]. This method has been parameterized for geometries, frequencies and non-covalent interactions, and uses an extended version of Hückel theory. The name GFN-xTB is an acronym for "Geometries, Frequencies, Non-Covalent - eXtended Tight Binding". This method was chosen as a similar method has already been published that calculates transition properties, the precursor to the GFN-xTB methods. This is the sTDA-xTB method. Additionally it was convenient the GFN-xTB method was already implemented in the QCORE package. This significantly reduced the amount of effort required for this project, with other users and developers who could help with implementation of this new method.

4.1.4.1 sTDA-xTB

sTDA-xTB ("simplified Tann-Danoff Approximation - eXtended Tight Binding") is another method in the family of xTB methods developed by the Grimme group, and is parameterised for transition properties [13]. The accuracy in calculating transition energies with this method is very good, with the error compared to high-level method, such as SCS-CC2, being around 0.3 - 0.5 eV.

Similar to other xTB methods, the sTDA-xTB method is a tight-binding method that uses empirically fitted parameters and a minimal basis set. It was trained on a test set of highly accurate coupled cluster and density functional theory excitation energies, as well as atomic partial charges for inter-electronic interactions.

Unlike other xTB methods, coefficients in the basis set for sTDA-xTB are dependent on the D3 coordination number. This makes basis functions far more flexible, which would usually be achieved with fixed basis functions by using diffuse or other additional orbitals in the basis set. Additionally, it uses two sets of parameterized basis sets - a smaller valence basis set (VBS) and an extended basis set (XBS). Whilst this reduces the cost of having larger basis sets, it makes calculating the gradient of transition properties much more difficult. This motivates the work on designing an alternative method with more tractable gradients, instead of using this already established method.

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, as opposed to an SCF method, to obtain molecular orbital coefficients. It is given by,

$$(4.14) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left(k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left(h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

where μ, ν, l, l' are orbital and shell indices, k_μ^l are shell-wise Hückel parameters, h are effective atomic-orbital energy levels, $S_{\mu\nu}$ is the overlap of orbitals μ and ν , k_T is a global constant and \hat{T} is the kinetic energy operator. The charges used in the inter-electronic repulsion function are given by charge model 5 (CM5) [23] charges for the XBS Fock matrix. These are calculated using Mulliken charges obtained from diagonalising the Fock matrix with the VBS. The charges for the initial VBS Fock matrix are based on Gasteiger charges [7], modified by the parameterised electronegativities of atoms in the system.

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

1. Calculate modified Gasteiger charges for the first initial guess
2. Diagonalise Fock matrix in the VBS to get the first set of Mulliken charges
3. Compute CM5 charges
4. Diagonalise Fock matrix in the VBS again for final set of Mulliken charges.
5. Recalculate CM5 charges with this final set, and diagonalize the Fock matrix in the XBS.

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-Danoff Approximation [11]. There are several approximations made between full linear response theory and the sTDA method. First is the Tamm-Danncoff approximation, where the **B** matrix is ignored. The second approximation is to use monopole approximations with Mataga-Nishimoto-Ohno-Klopman (MNO) operators instead of explicit 2 electron integral as well as neglecting the density functional term.

Transition charges are used to calculate these MNO integrals. The charge q_{nm}^A centred on atom A associated with the transition from $n \rightarrow m$, are computed using a Löwdin population analysis:

$$(4.15) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m}$$

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

$$(4.16) \quad \mathbf{C}' = \mathbf{S}^{\frac{1}{2}} \mathbf{C}$$

and μ is an index that runs over the atomic orbitals (AO). The MO coefficients are the solution of diagonalising the Fock matrix, similar to equation 4.2.

Approximations to full 2 electron integrals are given by charge-charge interaction damped by the MNO[24][25][19] functions. For exchange and coloumb type integrals, difference exponents

are used, along with an additional free parameter to recover the amount of Fock exchange mixing in the original matrix element equation. These will be discussed in more detail in the next chapter, as they are a crucial part of designing a new response method for chlorophyll systems.

Third is the truncation of single particle excited space that is used to construct the \mathbf{A} matrix. This reduces the number of elements that need to be calculated, and so reduces the time taken for diagonalisation, whilst also capturing a broad enough spectrum of excitation energies. The sTDA-xTB has many of the same goals as this project, except in one respect, which is the gradient theory. As the sTDA-xTB method still requires constructing and diagonalizing the \mathbf{A} matrix, albeit with a tight-binding method for molecular orbital coefficients, the gradient of the transition properties would still be difficult to calculate.

4.2 Benchmarking

Full scale chlorophyll molecules are too large to be able to calculate a high-level benchmark. A test set of small molecules, which would cover the same range of elements as found in organic chromophores and biological molecules, was then chosen as it would be able to benchmark both the Δ -SCF methods as well as TD-DFT. With an accurate idea of how reliable TD-DFT is, it would be possible to then compare Δ -SCF to TD-DFT for chlorophyll systems, as both methods have usable scaling to calculate transition properties for larger systems,

The test set chosen was previously used by the Grimme group to parameterise and test the sTDA-xTB method [14], and was constructed with the same goals as this work.

4.2.1 Reference Data and test set

The test set consisted of 109 small molecules. Each system was closed-shell, contained 12 atoms or less, and contained on H, C, N, O and F atoms. The size and specificity of this test set was chosen to minimise any other factors that could cause errors in analysing the results, whilst still covering all relevant elements and chemical environments for organic and biological chromophores.

Reference data was calculated as the three lowest energy singlet excited states, using EOM-CCSD with an aug-cc-pVTZ basis set. These results were generated using the Gaussian 16 program [6].

4.2.2 Small Systems

Transition properties for this test set were calculated using TD-DFT and Δ -SCF , both using a CAM-B3LYP functional and aug-cc-pVTZ basis set. The transitions were assigned to the EOM-CCSD results by comparing transition dipoles, energies and the character of the MOs involved in the transitions. Where the symmetries could be assigned, these were also used, but this was not the case for all systems as many were unsuccessful in labelling symmetry or defaulted to a

non-Abelain group. Symmetry labelling was also only available for TD-DFT calculations, as these were performed with Gaussian 16. Δ -SCF calculations were done with the QCORE program.

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

$$(4.17) \quad \Delta E_S = 2\Delta E^{i,\alpha \rightarrow a, \alpha} - \Delta E^{i,\alpha \rightarrow a, \beta}$$

was used to correct for the true singlet excitation energy ΔE_S [33].

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

Overall, the excitation energies calculated with Δ -SCF are as accurate at predicting EOM-CCSD energies as TD-DFT. The mean error is 0.35 eV, with a standard deviation of 0.25 eV. This is a marginal improvement on the TD-DFT results, which has a mean and standard deviation of 0.41 eV and 0.27 eV respectively. Transition dipoles were similarly accurate to the reference data, although Δ -SCF performs slightly worse in this respect. The mean and standard deviation in the absolute value of transition dipole moment, $|\mu|$, was 0.07 a.u. and 0.08 a.u. respectively. For TD-DFT, the mean and standard deviation were 0.03 a.u. and 0.06 a.u. (the atomic unit here being equal to ea_0). The outlier circled in figure 4.1 is an ethene dimer system, and shows the inability of Δ -SCF to capture a mixed excited state. The two LUMO orbitals in this dimer system are in-phase and out-of-phase combinations of the π -antibonding orbitals, which are very close in energy. The HOMO orbitals are the same on both ethene molecules, being the π -bonding orbitals, which are degenerate in energy. The first excited state is predicted by TD-DFT and EOM-CCSD to be a mix of these two close HOMO-LUMO transitions. However Δ -SCF cannot include this mixed behaviour. Δ -SCF predicts two transitions of 7 eV and 10 eV, whilst TD-DFT and EOM-CCSD predict 7.5 eV. The outlier in figure 4.2 is due to Δ -SCF finding a different but still valid description of the transition dipole. This point corresponds to a stretched benzene system, where the HOMO-1, HOMO, LUMO and LUMO+1 orbitals are all degenerate. The Δ -SCF transition dipole magnitude agrees with the value of an equally mixed HOMO - LUMO+1 and HOMO-1 - LUMO transition, which given the degeneracy is an equally valid description of the transition.

In summary, Δ -SCF can be seen to accurately predict transition properties to a EOM-CCSD level of accuracy with as much success as TD-DFT, except in cases of mixed transitions. It might then be expected a tight-binding method, with good electronic structure treatment, could also be accurate whilst drastically reducing the cost of calculation.

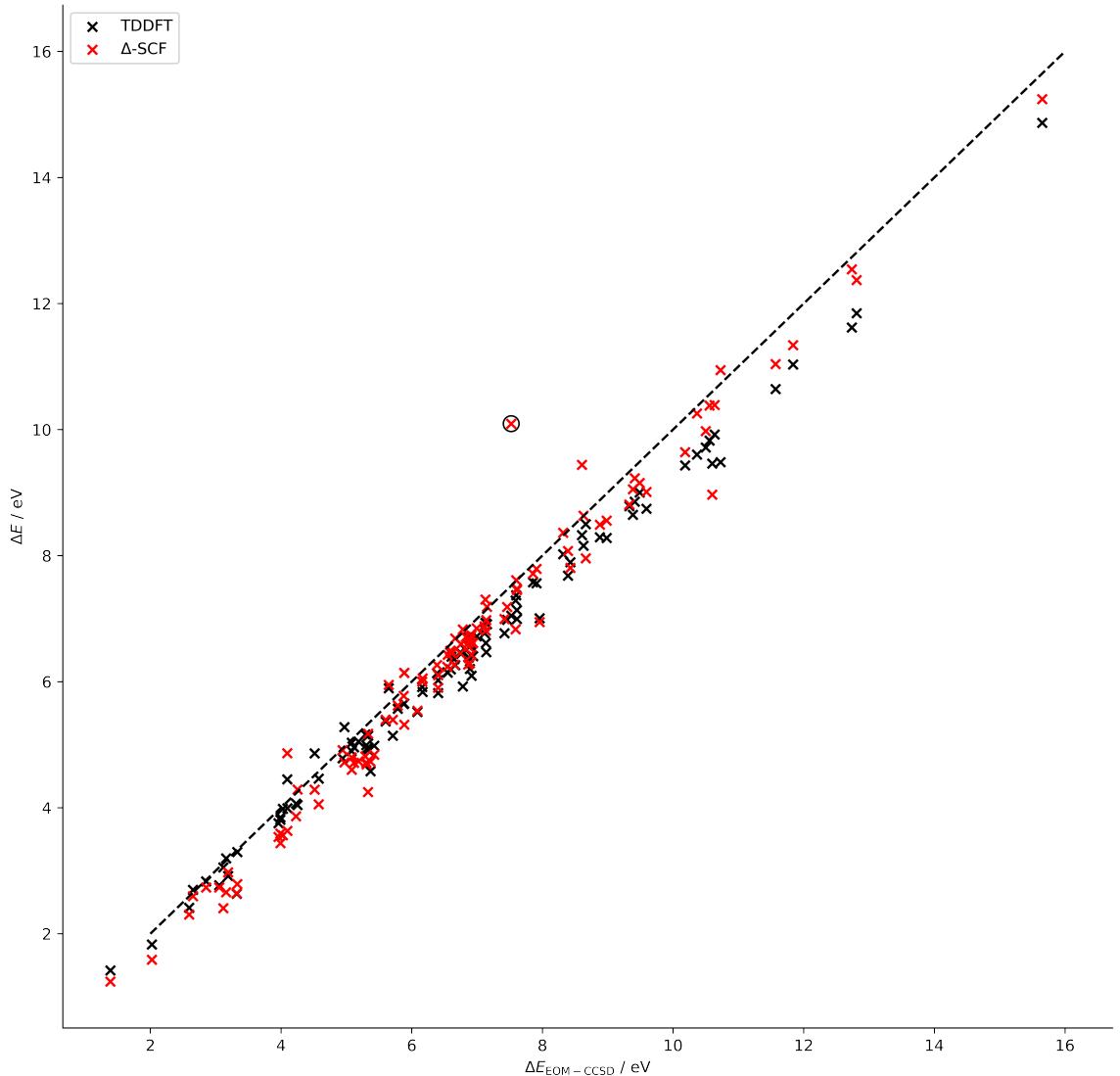


Figure 4.1: Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.

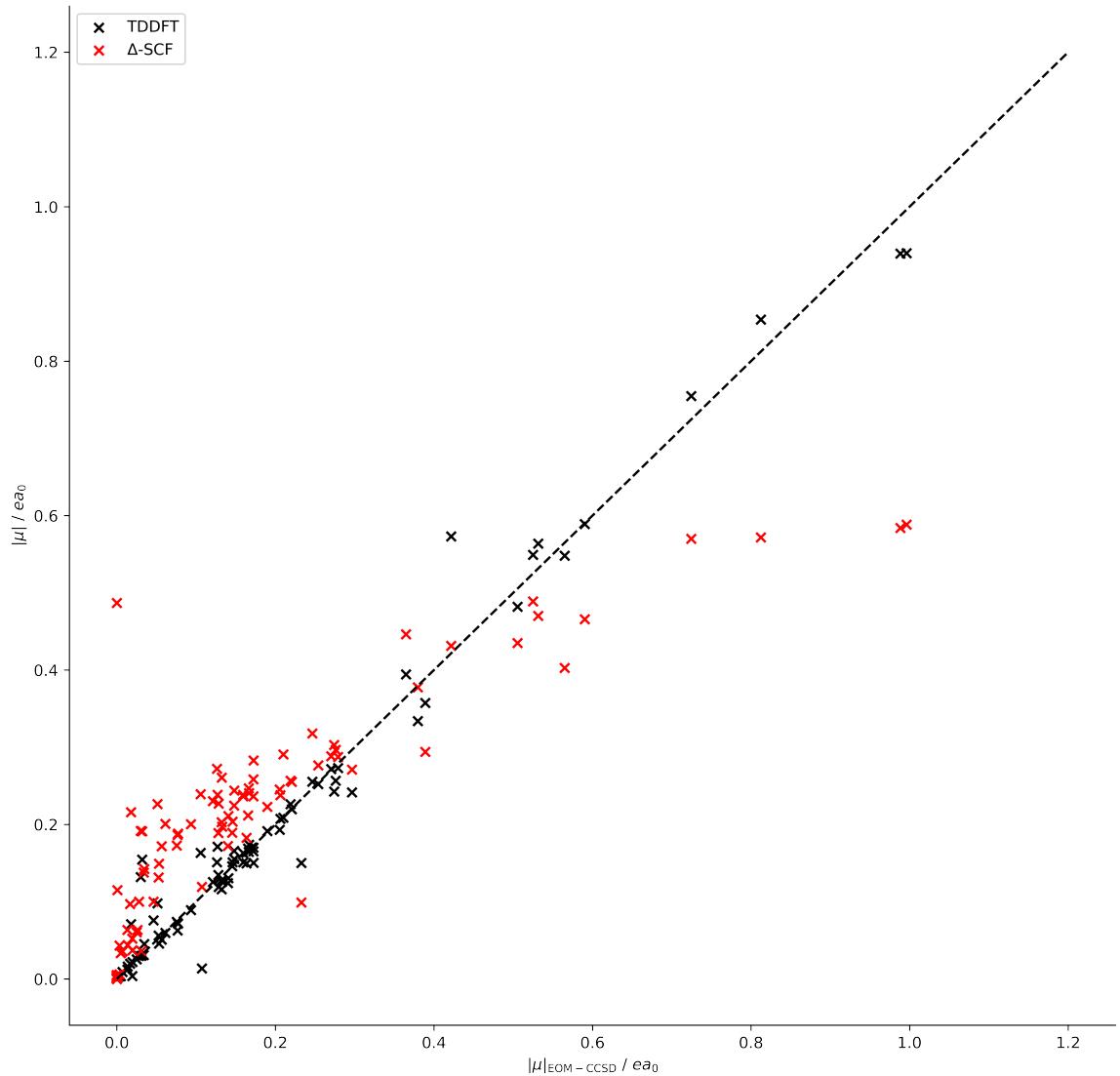


Figure 4.2: Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.

4.2.3 Non-orthogonality

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

$$(4.18) \quad S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle$$

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

$$(4.19) \quad q^{21} = \langle \Psi^2 | \Psi^2 \rangle$$

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

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

$$(4.20) \quad |\Psi_{\tilde{\nu}}\rangle = \sum_{\nu} |\Psi_{\nu}\rangle \left[\mathbf{S}^{-\frac{1}{2}} \right]_{\nu\tilde{\nu}}$$

where \mathbf{S} here is a block matrix

$$(4.21) \quad \mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

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

It was found that using this method for correcting the non-zero overlap of states, the origin-independence of the transition dipole moment was recovered (see figure 4.2.3). The transition dipole for each molecule in the test set systems was calculated for molecules translated by 100 Å in each of the x , y and z axes. This would induce an error for the non-orthogonalised states, which has been corrected for in the calculations with the symmetric orthonormalisation. 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 lead to nonphysical transition dipole

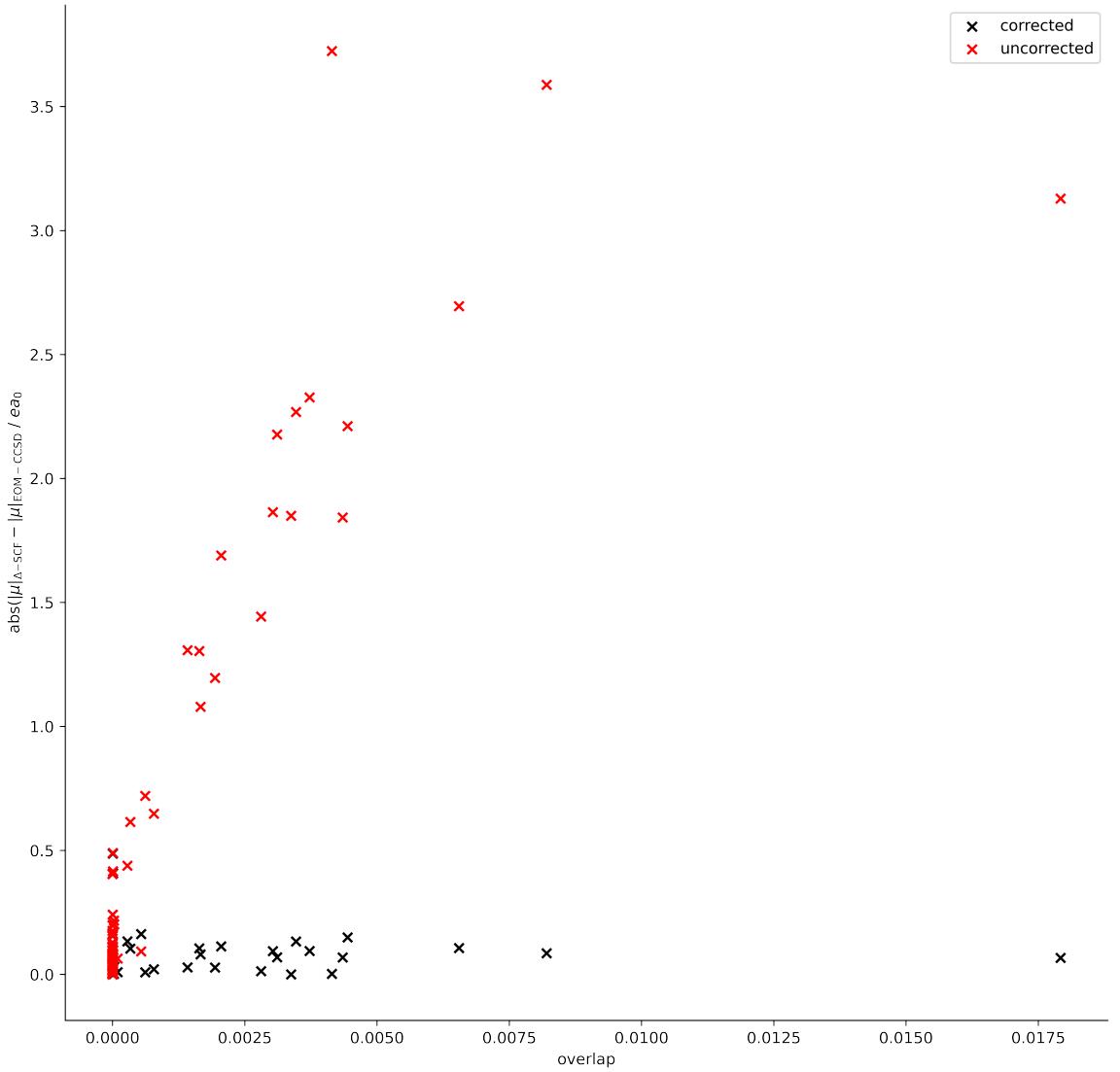


Figure 4.3: The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown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

magnitudes. In a large protein system, where chromophores can easily be 100 Å from the origin of the overall system, this would obviously present a much larger problem than for a vacuum-phase small molecule.

4.2.4 LHII Chlorophyll

Having demonstrated that TD-DFT is a good proxy for high-level methods, it is possible then to benchmark Δ -SCF for use in simulating a whole LHII system by looking at a set of chlorophyll geometries. This also has the additional benefit of testing Δ -SCF against more complex cases.

The PBE0 functional with Def2-SVP basis set [2, 28] was used to calculate both TD-DFT and Δ -SCF properties. This has been used previously for BChla structures, and has been shown to be a good balance between accuracy and cost [30].

The high degree of correlation between Δ -SCF and the TD-DFT results demonstrates it can detect changes in transition energies due to variation in geometries with an accuracy greater than the noise of random error from Δ -SCF. However the error in transition dipole magnitudes is larger than that of the small test set. This error is about 0.42 a.u. larger, but without EOM-CCSD or another high-level method, it's unclear whether this error might be from TD-DFT or Δ -SCF. Additionally, there is a clear correlation between the transition dipole magnitudes from TD-DFT and Δ -SCF, and so whilst quantitative statements cannot be made, qualitative assessments could be made from Δ -SCF data. For example, whilst the exact value of a transition dipole moment at a given geometry along a normal mode may not be accurate, the change in transition dipole moment from several geometries along the normal mode would be similar in variation to TD-DFT calculated properties.

An important point to take note of is that the excited state was not possible to calculate for all 27 chlorophyll geometries. One geometry repeatedly collapsed back down to the ground state. Methods such as increasing the amount of mixing of current and previous Fock matrices in Fock damping, using intermediate initial guesses like half-electron promotions, and alternative DIIS procedures, had little effect on improving this. 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, such as the initial ground state, rather than the previous SCF cycle, but implementing and testing these is outside the scope of this work.

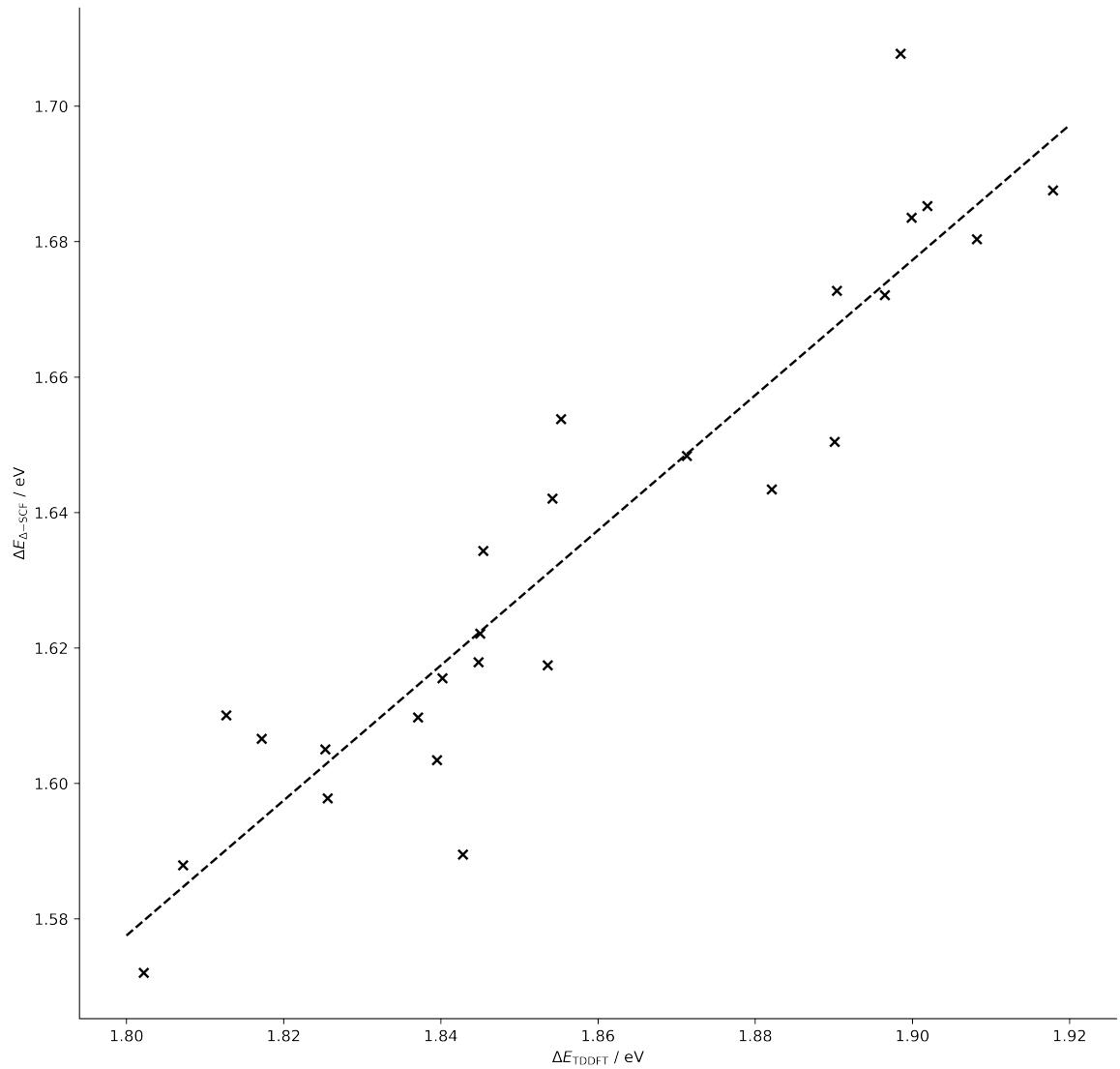


Figure 4.4: Transition energies from Δ -SCF for 26 chlorophyll geometries from the LHII protein of purple bacteria, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

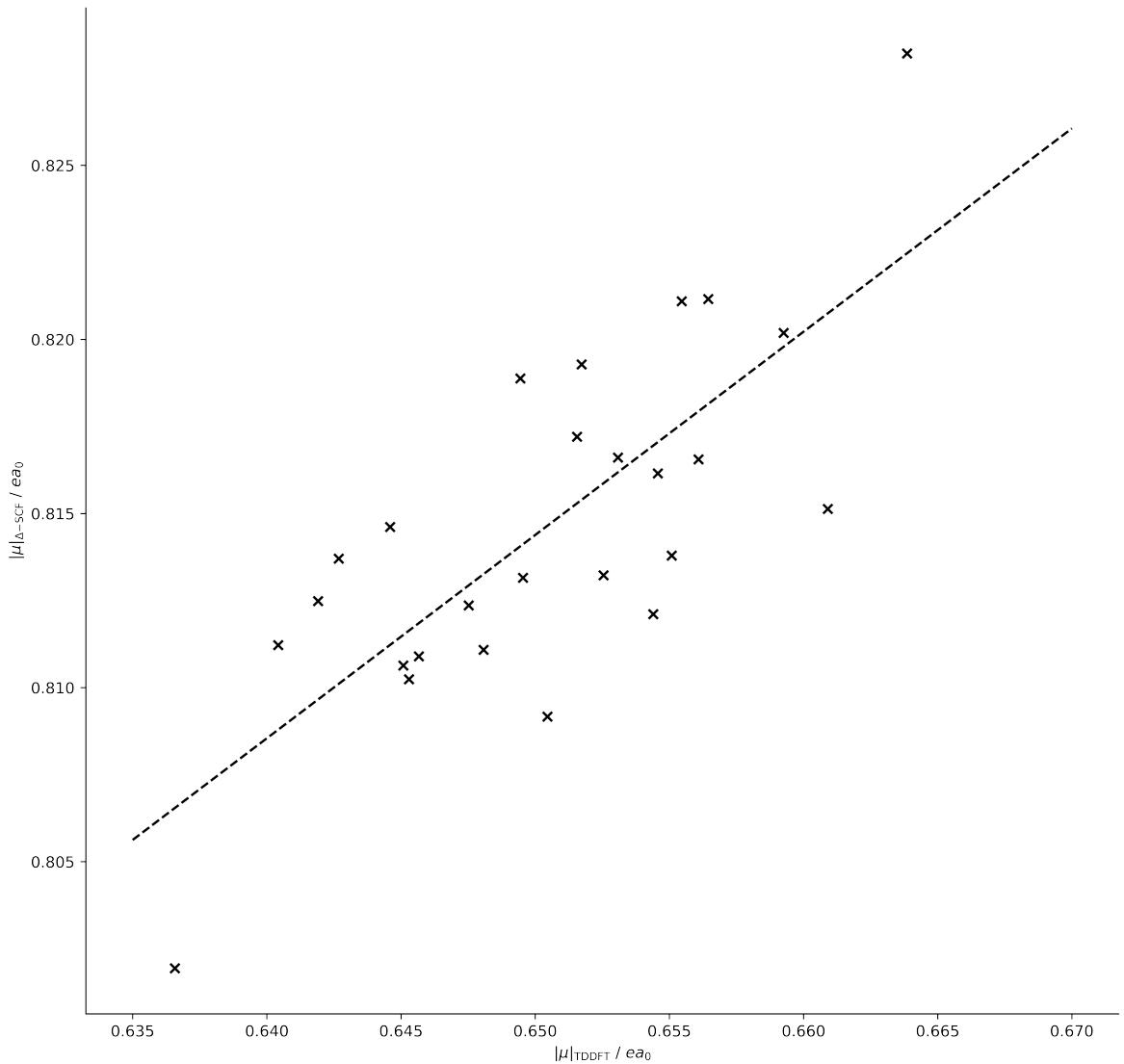


Figure 4.5: Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 4.2.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

4.2.5 xTB methods

Δ -xTB was tested on the same small molecule benchmark set. Its performance was compared to other low-cost methods, with a range of approximations for calculating transition properties, including the aforementioned sTDA-xTB method. To include this in the comparison, it was necessary to use spin-component-scaled second order coupled-cluster (SCS-CC2) [15, 16] reference data produced by Grimme et al. [12]. Concurrent to this work, the GFN0-xTB method was also implemented in QCORE. This method is similar to the GFN1-xTB method, but excludes any charge dependent terms in its Fock matrix so is not self-consistent. This method therefore only uses a single diagonalisation, with orbital solutions being the same for both ground and excited state. A transition energy from Δ -SCF with this method would functionally be the same as an eigenvalue difference, and this was included in the benchmarking to test whether it would be a possibility for predicting transition properties. Also included in this benchmarking are the eigenvalue differences from sTDA-xTB, given by the `xtb4stda` program [13], to observe how much the sTDA procedure can improve transition properties. `xTB4stda` is the first of two programs that runs a version of xTB, providing molecular orbital coefficients and energies for the sTDA program to then use to calculate the transition properties.

The full list of methods included in the comparison is:

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

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

was not possible. Instead, transition dipole orientations and plots of MOs were used as a proxy. This proved harder in the Δ -xTB case for two reasons. Firstly, the valence basis set for the Δ -xTB calculations are very different to those used in DFT. Second, this information was not available for the CC2 data, although the symmetry labels were. Additionally, inspecting the symmetry is very time-consuming and can not be automated. Every new method that was added to the benchmarking had to have every molecule individually inspected to make sure the symmetry of each transition was correct. Assigning symmetry to Δ -SCF results was investigated, but was not a straightforward implementation, as discussed in the following section.

4.2.5.1 Post-SCF Assignment of Symmetry

Symmetry is a common thread in many parts of electronic structure theory. It appears in normal mode analysis, wave-function analysis and assignment of electronic transitions. For this project, assigning symmetry to the transitions for Δ -SCF would require assigning symmetry to the MOs and overall wave-function of a molecule. Broadly speaking, 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 allows the symmetry to be given at any point in the SCF procedure, and allows the Hamiltonian to be organized into a block diagonal matrix. This can be useful when solving for a large basis set or large system as the matrix diagonalisation can be partitioned and parallelised over several cores or nodes on a cluster computer. However, this works best if the system is highly symmetric, which is often not the case when treating unoptimized systems, such as those from a molecular dynamics simulation, and is definitely not the case when looking at biological systems. The second approach, assigning symmetry after the SCF cycles, doesn't fix these drawbacks, but it does allow for codes which originally didn't have symmetry assignment to be extended without rewriting SCF code. The obvious drawback of doing assignment post-SCF is that symmetry can't be utilized during the SCF procedure. The second approach was opted for, as this was the easiest to implement in QCORE. The open source library libmsym [1] was used for point group assignment routines and finding the symmetry adapted linear combination of atomic orbitals. The steps for assigning orbital symmetry is as follows:

1. Determine the point group of the molecule, from the atomic positions
2. Set up the atomic orbitals in the libmsym representation.
3. Get the symmetry adapted linear combinations (SALCs) of atomic orbitals for each subspace.
These subspaces are the groups of symmetries that can be found in the point group of the molecule.
4. The SALCs can then be used to construct the transformation matrix T .

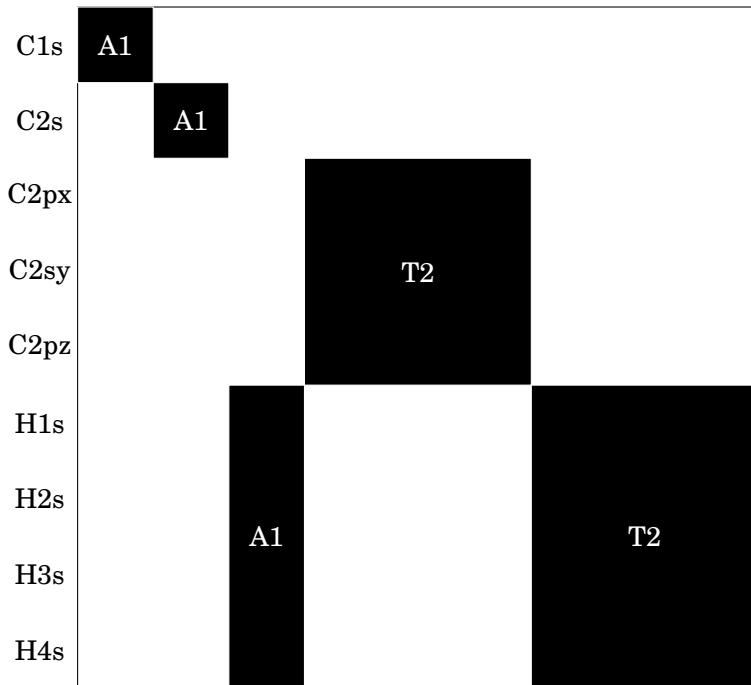


Figure 4.6: A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.

5. Assign the one electron molecular orbital (MO) for these subspace characters with the symmetry adapted linear combinations.
6. 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. It was found that this could accurately assign the MOs and overall wavefunction of methane the ground state.

To assign a label to each molecular orbital, the character of each orbital in all subspaces was looked at. This required transforming the molecular orbital coefficients \mathbf{C} into each subspace A by using the transformation matrix \mathbf{T} , defined by

$$(4.22) \quad \tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{S} \mathbf{C}$$

, and then summing the coefficients to obtain the character

$$(4.23) \quad P_A = \sum_v |\tilde{\mathbf{C}}_{A,\mu\nu}|$$

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

The MOs for an optimised methane geometry with an STO-3G basis set was correctly assigned - two occupied orbitals and one unoccupied orbital of A1 symmetry, and three occupied and unoccupied orbitals of T2 symmetry. The overall wavefunction symmetry can then be expressed as the product of all MO symmetries, reduced with the reduction formula

$$(4.24) \quad n = \frac{1}{h} \sum_R \xi_r(R) \xi_i(R)$$

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

However the assignment of MOs did not work well for excited states, due to the character from the subspace projection being unclear for many MOs. Furthermore, for non-abelian groups, where there are some degenerate E and T subspaces, this assignment procedure did not work. This is a similar problem to the assignment of symmetry for the TD-DFT and EOM-CCSD transitions in the earlier benchmarking. Often the reduction of ground state wavefunctions gave non-physical answers.

Extending the symmetry analysis to non-abelian point groups was beyond the scope of this work. Although it would be a useful feature for testing the benchmarking sets, chlorophyll molecules would be far from symmetric and so this type of assignment could not have been expected to have worked. In summary, transitions could not be confidently assigned for Δ -SCF with this method.

4.2.5.2 Δ -xTB Benchmarking Results

After considering this leading error, the assignment of symmetry was based on the previously used inspection of transition dipole orientations and transition density plots.

The distributions of absolute errors

$$(4.25) \quad \epsilon = \Delta E_{\text{method}} - \Delta E_{\text{SCS-CC2}}$$

for each of the benchmarking methods, as well as a generated distribution of sTDA-xTB results, are shown in fig-4.7. The means and standard deviations are reported in table 4.1.

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

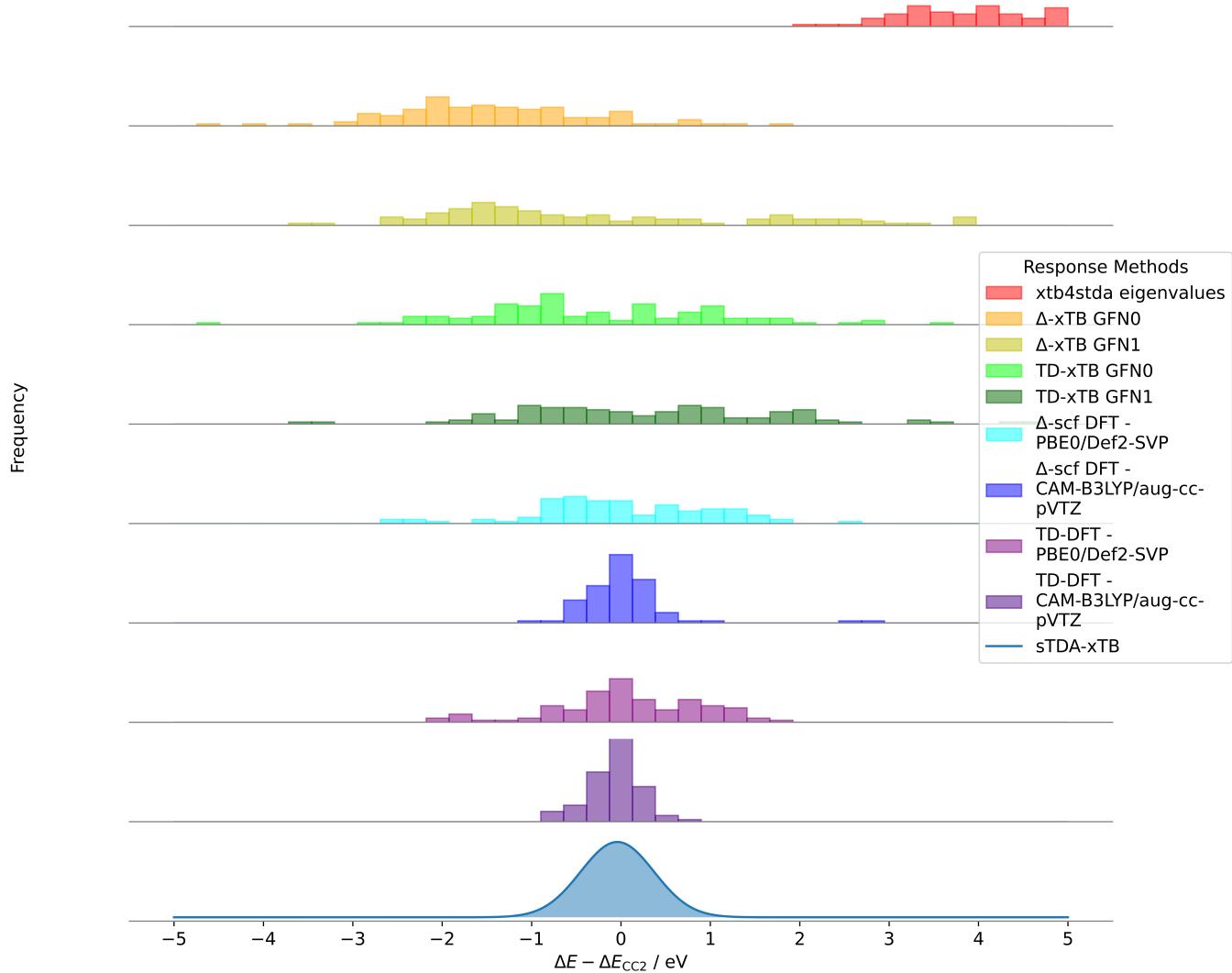


Figure 4.7: The distributions of errors compared to SCS-CC2 transition energies for the methods included in the $\Delta\text{-xTB}$ benchmarking.

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

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

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

The standard deviation should be well inside the value of the transition energy (1.8 eV for Q_y), so a 1 eV accuracy is used as the first test of usable accuracy. Next, in order to be accurate enough to predict variations that cannot be attributed to random noise, the standard deviation must be within the range of excitation energies for LHII chlorophylls, which from 4.2.4 can be seen to be around 0.2 eV. Any more than this and it's uncertain whether the difference between two excitation energies for different chlorophyll geometries are due to geometry reasons or due to random noise.

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

From these results, it's argued that decreasing the level of theory for the electronic structure dramatically decreases the accuracy of transition properties. The highest level electronic structure method has the best accuracy. CAM-B3LYP/aug-cc-pVTZ TD-DFT and Δ -SCF have mean errors of -0.18 eV and -0.14 eV and standard deviations of 0.34 and 0.28 eV respectively. Both methods are well within the accuracy needed to predict geometry-based variations. The outliers in the Δ -SCF results are known mixed transitions, as discussed earlier with the ethene dimer system. The lower level PBE0/Def2-SVP methods have a marked decrease in accuracy. On going from higher-level DFT to lower level, the standard deviation approximately doubles for both TD-DFT (0.34 eV to 0.79 eV) and Δ -SCF (0.28 eV to 0.50 eV). Again, the PBE0/Def2-SVP TD-DFT and Δ -SCF are comparable, with standard deviations of 0.79 eV and 0.59 eV, although the mean for Δ -SCF has a significant shift of -0.62 eV. However it's clear that the greater effect on accuracy is due to the change in response method rather than level of theory of the DFT calculation. Whilst the same conclusion is not as clear for the semi-empirical methods, the large standard deviations

of all the GFN based methods fail the first test of accuracy, and so can be immediately disqualified. They are all too inaccurate to justify a more detailed analysis. Comparing the DFT and GFN based methods, we can see the same trend that lowering the level of theory gives worse transition properties.

Overall, the most inaccurate method is the eigenvalue difference methods based orbital energies (eigenvalues of the Hamiltonian diagonalisation) from the `xtb4stda` method. The means and standard deviation was 4.39 eV and 1.26 eV respectively, a huge difference to the reported values (-0.04 eV and 0.41 eV respectively)[13]. However it does show that the full sTDA-xTB method can be as accurate as TD-DFT with a range-based density functional and triple zeta basis sets, and so arguably the sTDA method, and not the underlying xTB method, makes up a large part of the accuracy for predicting transition properties. Hence a similar response theory, based on a different xTB theory, could perform equally well if the xTB theory really does not factor into obtaining decent accuracy. This is investigated in the next chapter.

The result that GFN-xTB based methods are not accurate is not unexpected. As opposed other DFT method, which use *ab initio* or first principle parameters, the xTB methods were fit to target properties, and so could not be expected to be suitable for other properties outside the training data [3]. Whilst GFN-xTB is better for number and specificity of parameters than many other methods because of this "top-down" parameterisation approach, these parameters only extend the accuracy of predicted properties to different chemical systems (due to a lack of pair-wise parameters), and not to different properties altogether.

4.3 Conclusions

The transition properties of a test set of small molecules has been benchmarked with multiple Δ -SCF , TD-DFT and high-level methods. It has been shown that DFT based Δ -SCF and TD-DFT methods can reproduce the same transition energies and transition dipole magnitudes as EOM-CCSD to within reasonable levels of accuracy. For the set of small molecules, transition energies were predicted with a mean of less than 0.5 eV, and 0.07 a.u. for transition dipole magnitudes. Additionally, the issue of breaking the origin independence property of transition dipoles has been shown to be fixed by using a symmetric orthogonalisation of the two originally non-orthogonal states.

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

With all of the above benchmarking, reliably obtaining and assigning transitions predicted from Δ -SCF has proved to be an unsolved issue. Either Δ -SCF is formally unable to predict the correct character of transitions, as showcased in the ethene dimer mixed transition outlier, or it is unreliable in finding excited state solutions. This is best shown in the exclusion of a geometry of chlorophyll that could not be made to converge to the correct excited state, as well as the necessary use of Fock damping, altered DIIS procedures and intermediate initial guesses to investigate whether the convergence could be fixed.

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

Additionally, whilst DFT based Δ -SCF methods are shown to be accurate, GFN-xTB based methods were found to be inaccurate, to the point where it cannot be claimed to be a useful proxy to higher level methods. Due to the similarity in results for linear response and Δ -SCF methods over a range of electronic structure methods, from high level DFT to lower level GFN-xTB methods, this drop in accuracy is attributed to the different electronic structure theory used, rather than the response methods. This implies that altering the electronic structure method could lead to great improvements in the accuracy of a new response method.

The aim of this chapter was to determine whether Δ -SCF methods, which have a simple gradient theory and are less costly than TD-DFT, could provide a sufficiently accurate description of transition properties for use in an ab initio exciton framework. It's been shown that this is true with the condition that the underlying theory is sufficiently high. The issue of non-orthogonality has shown to have a solution that does not decrease the speed or accuracy of this method, and whilst the problem of symmetry assignment is still unsolved, this is not as much of an issue when only the Q_y transition of chlorophyll, which is well defined by its transition dipole, is needed.

However, the fact that higher level DFT is required for accurate transition properties is an issue for simulating a large volume of chlorophyll structures. The efficiency of these calculations is too low for an exciton framework, which would require in the region of 1e5 to 1e6 calculations for a full time series of LHII. Semi-empirical Δ -SCF methods, which would be efficient enough and have tractable gradients, prove inaccurate in their current form. However, it is demonstrated that a correct electronic structure and "top-down" parameterisation could make an accurate semi-empirical method. This is investigated in the following chapter.

CHLOROPHYLL SPECIFIC METHODS

This chapter reports the work done on designing and parameterising a novel method for response properties. The framework and theory for the method is outlined in section 5.1. Parameterisation details are given in section 5.2, including the reference data that compromised the training data, the objective function used and the algorithms that were used for optimisation. The accuracy of this new method is showcased in the final section 5.3.

5.1 Theory

This section covers the approximations to full linear response theory that leads to the novel chl-xTB method. Starting from the eigenvalue equation, that is derived more fully in the theory chapter but recapped here, a few key approximations lead to a framework that has tractable gradients and is less expensive than full linear response. This fulfills the requirements set out in the introduction of having an efficient and simple, but accurate, method of getting transition properties for LHII chlorophylls.

5.1.1 Full solutions TD-DFT equation

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

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

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

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

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

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

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

5.1.2 Approximations to Solutions

There are three approximations made in the chl-xTB method. First is the Tamm-Danoff approximation (TDA), which eliminates three blocks of the full linear response matrix. Second is a diagonal dominant approximation to eigenvalue solutions, where diagonal elements of a matrix are taken as an approximation to true eigenvalues. The third is a monopole approximation to two electron integrals which employs MNOK operators to accurately capture Coulomb and exchange functions. All these approximations reduce the amount of expensive electron integrals required to construct the full eigenvalue equation, whilst still capturing the important aspects of the electron density response.

5.1.2.1 Tamm-Danoff approximation and Diagonal Dominant A matrices

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

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

where the definitions of elements are the same as above, but it should be noted that the solutions \mathbf{X} will be not be same. This is formally the same as a CIS problem, and has been reported as a way to get back to CIS from full TD-DFT.

As an eigenvalue problem, solving for ω requires constructing and diagonalising the full \mathbf{A} and then diagonalising. The XC functional is dependent on ω values and so several iterations of this diagonalisation are needed to find stable, self-consistent solutions.

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

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

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

By removing the eigenvalue problem, the gradients of the excitation energies are more tractable. However the integral terms still require an iterative, self-consistent treatment. This can be simplified by using an approximation to these integrals that uses point charge interactions.

It should be noted that further cost is removed by not needing to calculate not only all of the off-diagonal elements but the diagonal elements for transitions that are not of interest. This however would only be in the case where transitions are not mixed. If transitions are mixed, then the coupling elements should be considered and approximating them to be zero would not be valid. This is considered in section 5.1.4.

5.1.3 Integrals approximations

The two electron integrals as written above are usually one of the most expensive part of an electronic structure calculation. These can be approximated with a monopole expansion, treating atomic sites as point charges, and using a charge-charge interaction for the energy.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.1.4 Q_y Transition

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

The Q_y transition is the one of the two transitions that make up the Q band in the absorption spectra of chlorophyll, the other being the Q_x transition. It is well known that the Q_y transition is important for electronic energy transfer, and predicting both transition energies and transition dipole magnitudes and orientations is important to construct frameworks for this transfer. The Q_y transition is mostly HOMO-LUMO in character (96%), with a small amount of HOMO-1 - LUMO+1 (remaining 4%). The analogous transition in the unsubstituted tetraphorphyrin ring has the transition dipole along the molecular axis defined by the N atoms, however due to the asymmetry introduced by substitutions and geometry deformations, this is usually not the case for BChla with a deviation from this axis of around 12°. The Q_y transition in chlorophyll has its dipole component lying along the N_A-N_C axis with Q_x lying orthogonal, along the N_B-N_D axis. Q_x has the reverse character to the Q_y transition, being mostly HOMO-1 - LUMO+1.

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

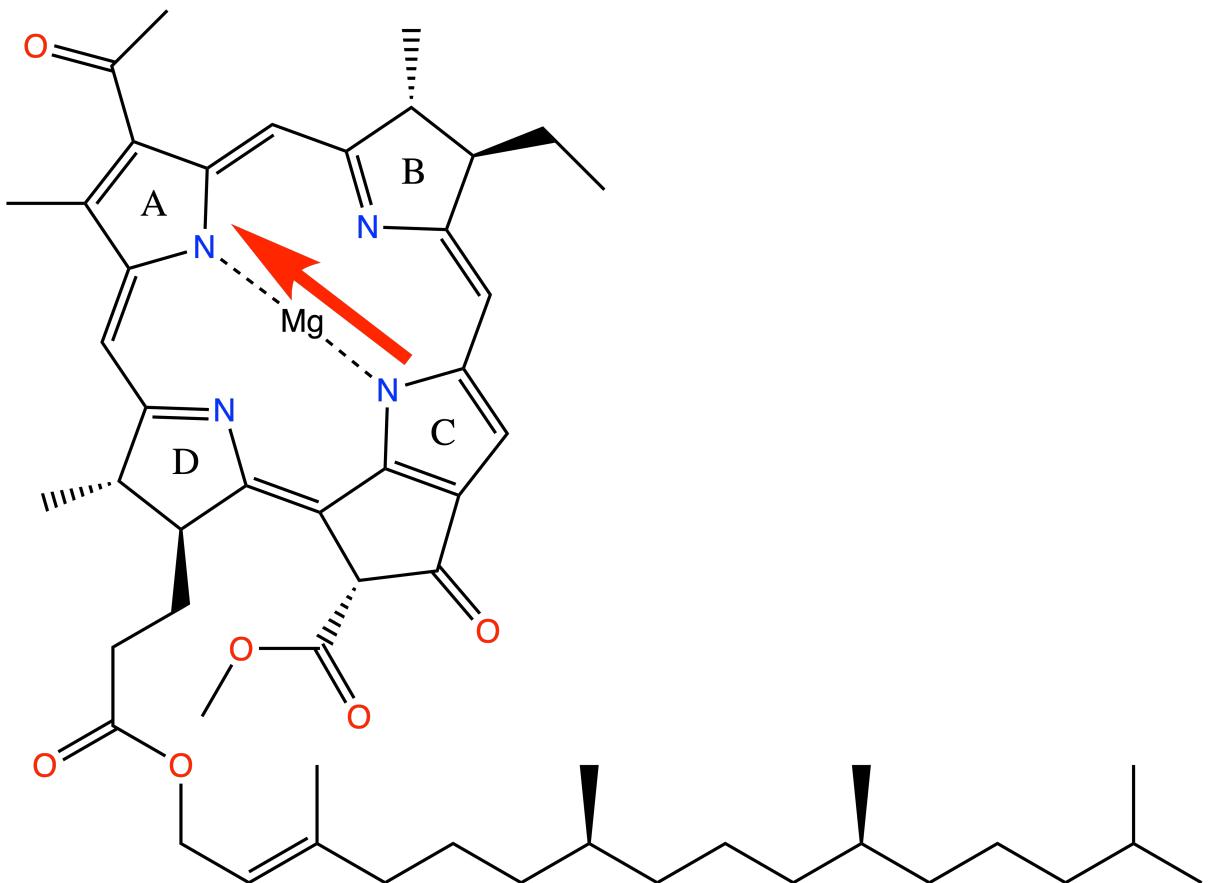


Figure 5.1: Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.

spectral density (discussed further in chapter ??).

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

The high amount of single-transition character in the Q_y transition makes it ideal for the approximations to transition energies set out earlier. The lack of coupling elements that would appear in the **A** matrix justifies both the use of the diagonal dominant approximation and TDA. The almost exclusive centering of electron density on main group elements make the use of MNOK integrals also valid.

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

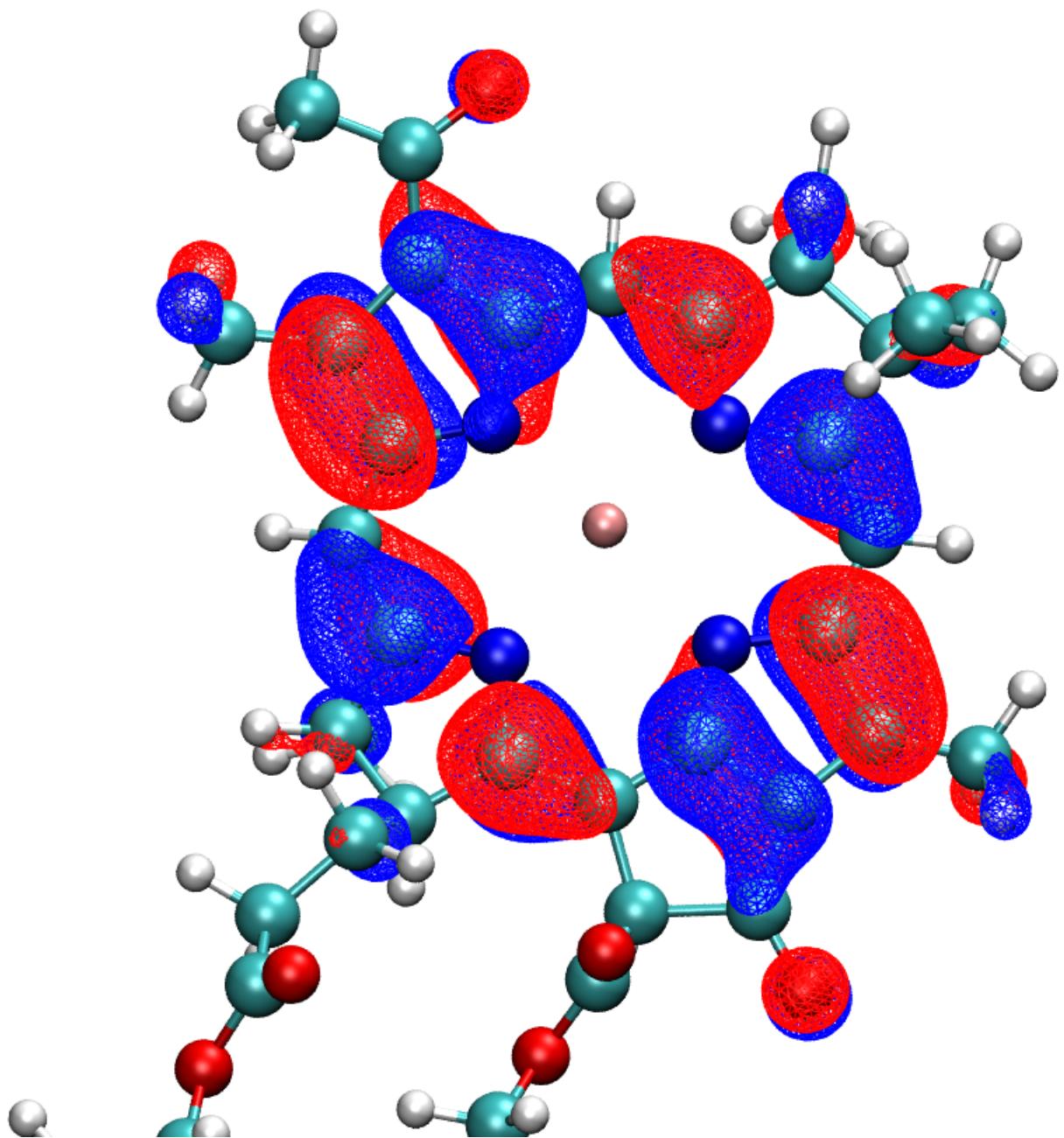


Figure 5.2: The HOMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.

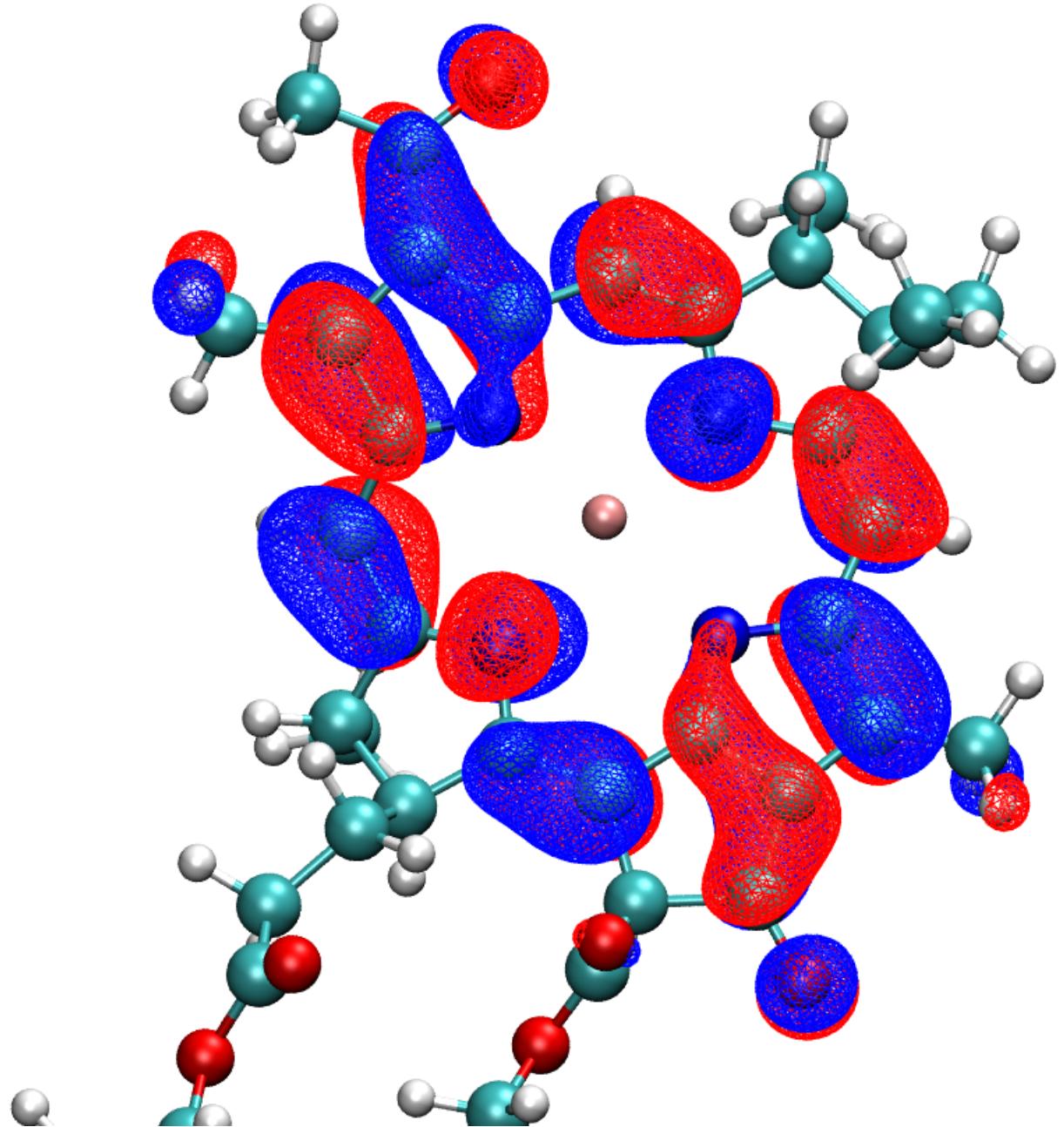


Figure 5.3: The LUMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.

The Q_y transition was then used to benchmark the novel response approximations. The optimisation and construction of reference and training set data is discussed later in section 5.2.

5.1.5 Changes to xTB parameters

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

As discussed in the theory chapter, the GFN1-xTB Fock matrix is made of both charge dependent and charge independent terms. The form of these terms, and definitions of parameters, is also given in the theory chapter. Only the charge dependent terms would have any effect on the partial and transition charges, thereby effecting the transition properties. The charge dependent terms are the first, second and third density fluctuation terms. The first order term is the leading term, and so parameters present in this term were chosen to be included in the optimisation procedure to prioritise parameters that would have the largest effect. Only some of these parameters are "free", whilst others are based on physical or *ab initio* values. These 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 incorrect 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 to transition properties is that they would lose their specificity to the GFN-xTB training set. However it is not in the scope of this work to find a semi-empirical method that would be able to calculate these properties for chlorophyll. chl-xTB is not used to calculate optimised geometries or hessians of chlorophyll as other methods could be expected to perform better. Also if this were the necessary, including more target properties in the parameter optimisation would decrease the accuracy to any one target, making the method worse overall. For example, the sTDA-xTB and GFN-xTB methods use different electronic structure parameters for this reason.

5.1.6 Transition and Excited State Density from the Ground State

Similar to Δ -SCF, the ground state orbital coefficients are taken to be a good approximation to the (Q_y) excited state coefficients. It was assumed then that the excited state could be calculated

without orbital relaxation, so that both the excited state and ground state could be constructed from the same set of molecular orbitals. The transition density is then calculated as

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

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

$$(5.14) \quad \mathbf{D}^0 = |\Psi^0\rangle\langle\Psi^0|$$

$$(5.15) \quad \mathbf{D}^1 = |\Psi^1\rangle\langle\Psi^1|$$

. 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, only one cycle of the SCC procedure would be necessary, which would eliminate the problem of convergence in the Δ -SCF excited state. It would also halve the computational time. Excited state properties, such as the molecular dipole and partial charges can also be calculated from the excited state density, which will be important for the exciton framework of the next chapter.

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

5.2 Parameterization

An objective function and algorithm is required to find minima in the parameter space, where the minima correspond to an optimised set of parameters. Reference data is also required as a target. All of these elements are important for minimising the amount of error in the final method, as well as determining how well chl-xTB performs for different problems and other properties.

5.2.1 Objective Function

At first, it would be argued that the root mean squared error (RMSE) to the reference transition energy should be the value that should be minimised, giving the objective function as

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

where \mathbf{x} is the set of parameters, ΔE_i , $\Delta E_{i,\text{ref.}}$ are the transition energies for system i from the chl-xTB method and reference method respectively, N is the number of systems used to calculate the objective function value (these 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, so the transition dipoles would be expected to have a large error. This can be fixed by including a metric for the error in transition dipoles. The other issue is that a low RMSE does not guarantee a high correlation. A measure of the correlation can be given by the coefficient of determination

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

where \hat{y} , y are the predicted and reference values, 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 mutually inclusive.

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

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

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

5.2.2 Minimisation Algorithms

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

5.2.2.1 Nelder-Mead

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2.2.2 Sequential Least-Squares Quadratic Programming

The SLSQP (Sequential Least-Squares Quadratic Programming) method is fundamentally different to the previous Nelder-Mead method, and follows a quasi-Newton procedure with additional factors to treat constraints.

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

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

however with an arbitrary amount of constraint functions c

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

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

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

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

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

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

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

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

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

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

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

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

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

A visualisation finding a parameter set \mathbf{x} corresponding to a local minimum of the Himmelblau function (a standard benchmark for optimisation algorithms) is shown in figures 5.4 and 5.5. 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 around the implementations of both methods make it a black-box that is hard to investigate further. The results of the optimisation is discussed in further detail in section 5.2.4.

5.2.3 Reference Data

The geometries for the training set used to optimise the chl-xTB method were taken from previously done molecular dynamics of the LHII protein [30]. The geometries of LHII were chosen from uncorrelated snapshots, although each of the 27 chlorophylls from each snapshot were included in the training data to account for the differences in binding pockets. The stochastic collection of LHII snapshots were chosen to cover a range of chlorophyll conformations to reduce the amount of artificial bias towards any particular geometries.

Training the chl-xTB parameters was done against PBE0 data. This data was chosen for the best accuracy-cost ratio, as well as having been previously used to investigate exciton properties for the LHII system [30]. Additionally, from the outset it was unknown how much training data would be necessary, and so keeping potential future costs of expanding the training data down was another factor in choosing this functional.

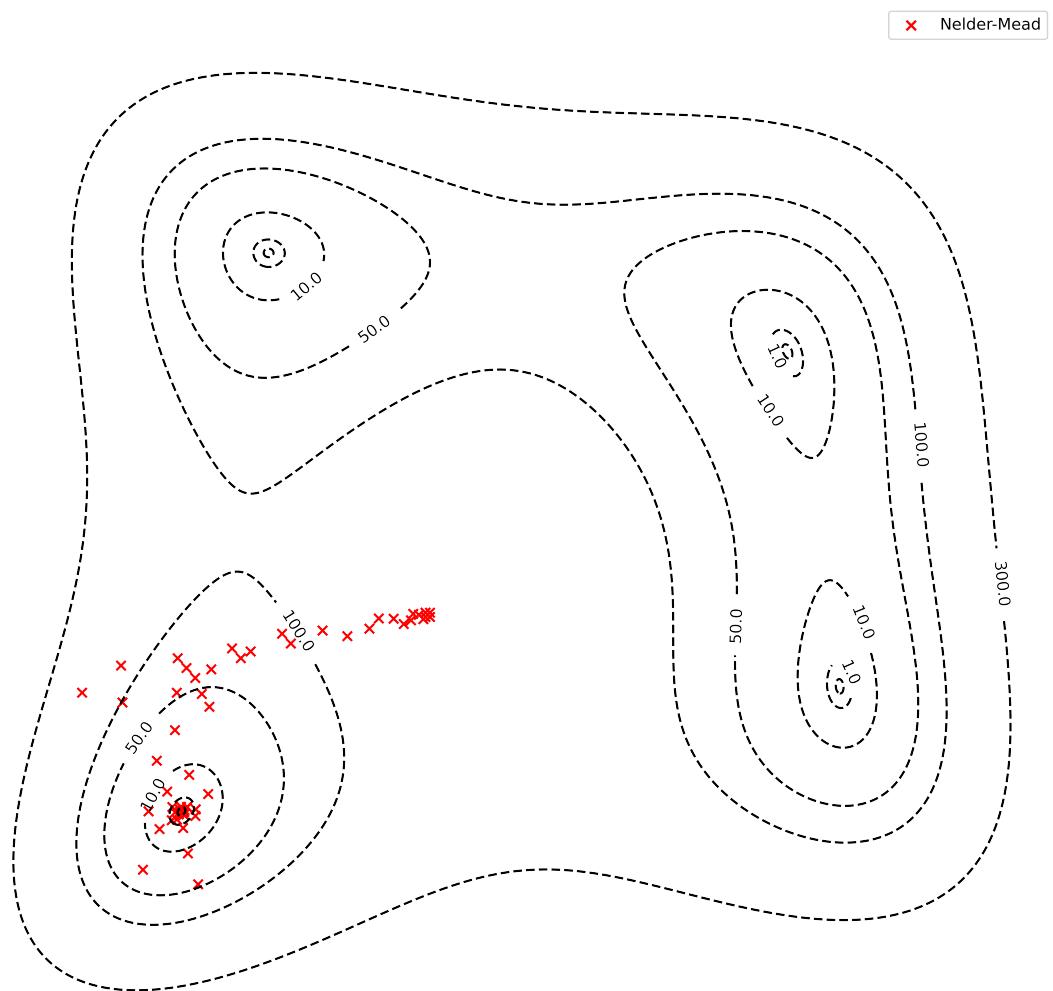


Figure 5.4: Positions of the Nelder-Mead simplex vertices for the HimmelBlau function.

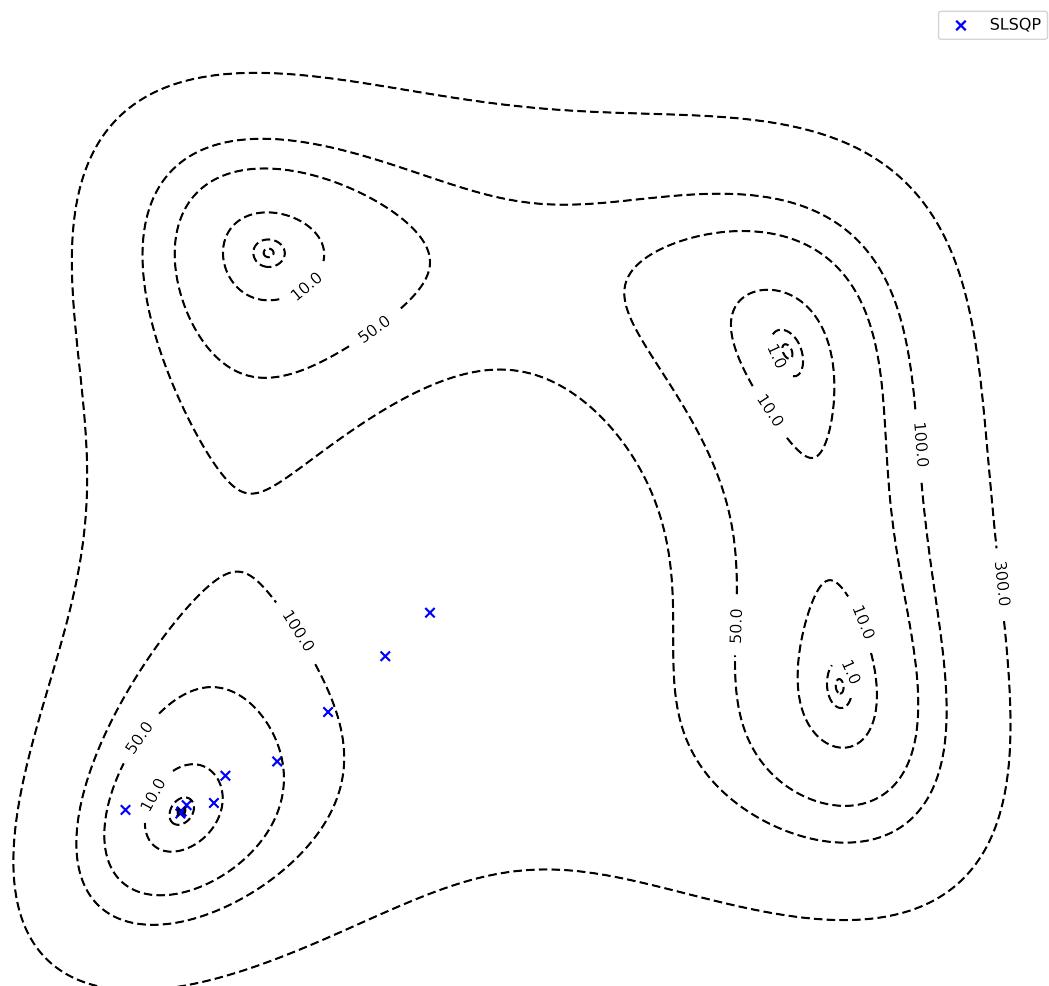


Figure 5.5: Positions of iterative solution vector \mathbf{x} from SLSQP for the HimmelBlau function.

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

Table 5.1: ref data vals

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

The errors and correlations of the reference data are shown in table 5.1. The methods included in the benchmarking were Δ -SCF, TD-DFT and eigenvalue difference all using the PBE0 functional and Def2-SVP basis set. Also included are CAM-B3LYP and BLYP functionals with Def2-SVP basis sets, as a higher and lower level reference respectively. The basis set was not changed as it has been found that the basis set has less importance on the accuracy than the functional [30].

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

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

The agreement of transition dipole magnitudes is much lower than excitation energies. The RMSE of Δ -SCF and eigenvalue difference methods is significantly higher (1.455 a.u. and 1.511 a.u. respectively) than the TD-DFT methods (0.172 a.u., 0.214 a.u. and 0.384 a.u. for CAM-B3LYP, ω -B97XD and BLYP respectively). The average magnitude of PBE0 transition dipoles is 2.751 a.u.,

with the average for Δ -SCF and eigenvalue difference being 4.287 a.u. and 4.342 a.u. respectively. This disparity is attributed to the lack of inclusion of Q_x transition character. As this dipole component of this transition is orthogonal to Q_y , it may reduce the transition dipole magnitude similar to the effect seen in the outliers in the previous chapter.

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

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

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

5.2.3.1 Training and testing set

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

	Hamiltonian	chl-xTB	GFN1/sTDA-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			
γ_K	2.147	2.000	
γ_J	4.012	4.000	
a_x	0.067	0.500	
D_{scale}	0.636	-	

Table 5.2: optimized parameters from SLSQP procedure.

5.2.4 Results

Overall, the method performs extremely well considering the limitations as set out above. Transition energies and dipole magnitudes are calculated well within acceptable RMSE and correlation limits.

The final parameters for the chl-xTB method are given in table 5.2. The best performing set of parameters had an RMSE of excitation energy of 0.014 eV with an R^2 value of 0.88, and an RMSE of transition dipole magnitude of 0.057 a.u. with an R^2 value of 0.40. Repeated optimisation runs gave parameter and objective function minima to similar values, and the difference in these values can be attributed to the complex solution space. These values for RMSE are well within the values for TD-DFT with various functionals, and the R^2 of transition energy is equally good. While the correlation is low, it is near to the expected correlation from Δ -SCF and eigenvalue difference. It can also be seen in figure 5.2.4 that the variation in chl-xTB transition dipole magnitude is much smaller than Δ -SCF, eigenvalue difference and ZINDO, as well as being close to the mean from PBE0. This is a better behaviour, similar to the statiscal method used before [30], than the other methods with low correlations.

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

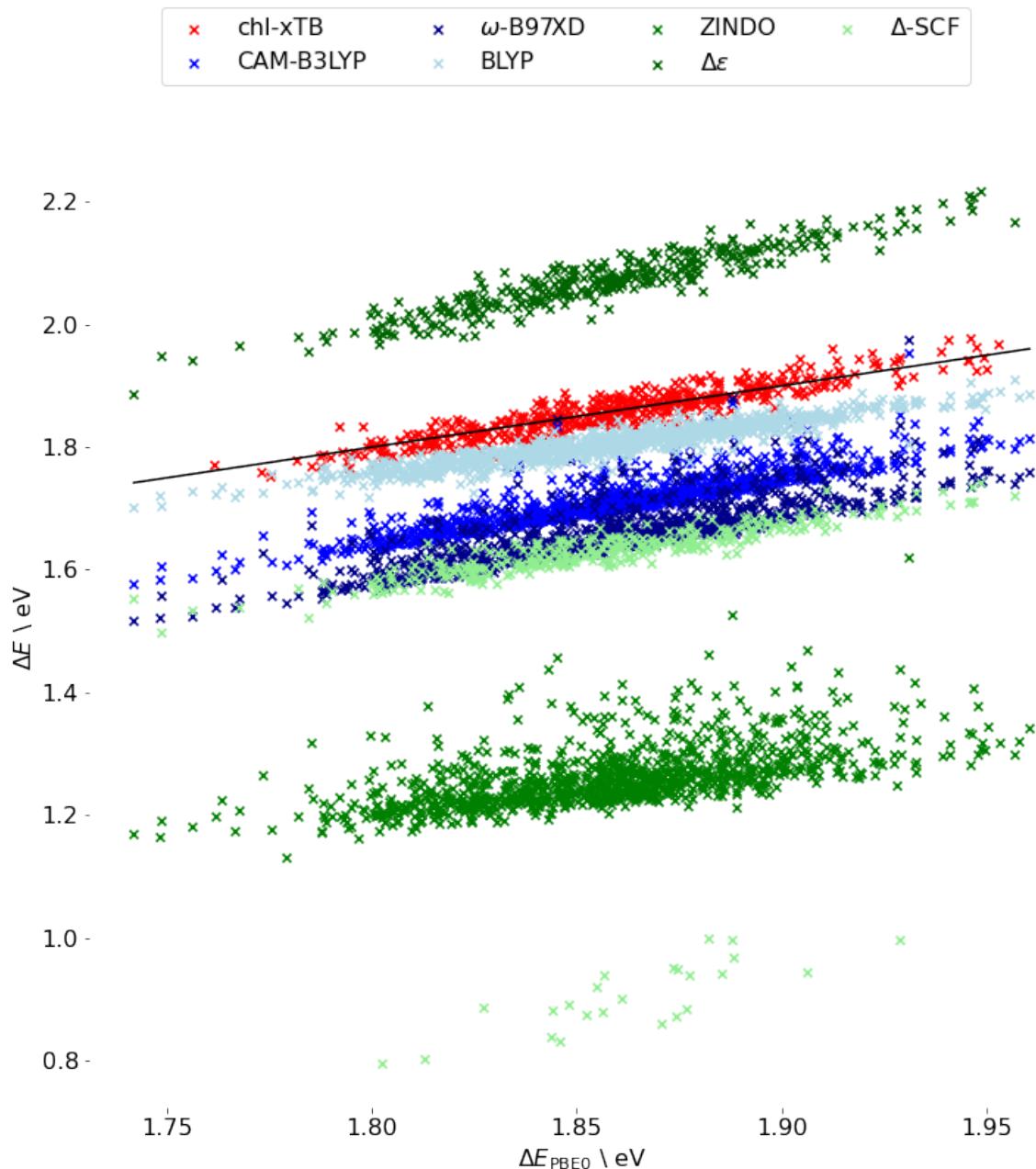


Figure 5.6: caption

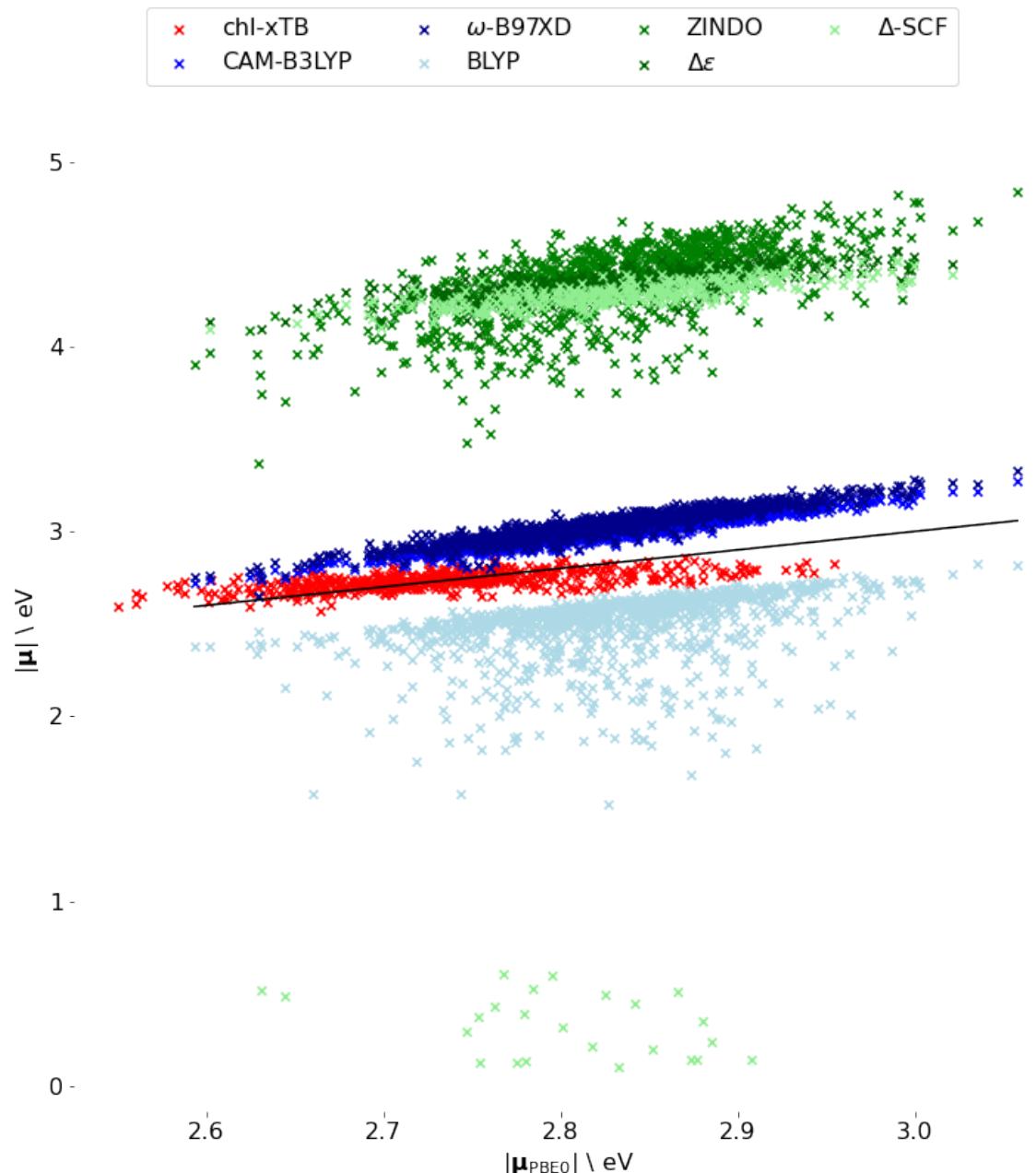


Figure 5.7: caption

The initial guess 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 5.2), with the exception of the α_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 [5].

5.3 Cross-validation

5.3.1 Vibrational Mode Coupling

Whilst the stochastic selection of BChla geometries should represent a large section of the conformational space in LHII, 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 correspond to geometries that vary by a single normal mode displacement at a time to reduce error cancellation, as well as to show how well chl-xTB predicts to the coupling of vibrational modes and transition properties.

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

Therefore the normal modes and transition properties were calculated for a truncated BChla with a hydrogen atom replacing the phytol tail, which made geometry optimisation possible, and also reduced the total number of vibrational modes.

Normal modes with the strongest coupling to the Q_y transition were chosen to most effectively scan the conformational space. These can be found by looking at modes which break the symmetry of the Q_y transition. In an ideal model, the magnesium and nitrogen centre have D_{4h} symmetry with the Q_y transition lying along the N_A, N_C axis. Vibrational modes with components along this axis will therefore couple to the transition. The movement of N_A, N_C atoms which induce $D_{4h}-D_{2h}$ and $D_{4h}-C_s$ symmetry breaking are shown in figure 5.8.

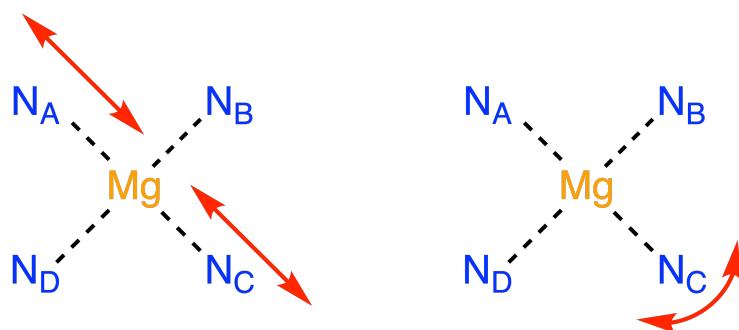


Figure 5.8: Normal modes of the nitrogen-magnesium centre of chlorophyll that have $D_{4h}-D_{2h}$ (left) and $D_{4h}-C_s$ (right) symmetry breaking components.

Normal modes that would also have this symmetry breaking component were indentified by how much the N_A - N_C displacement would change along each normal mode. A plot of these values, as well as the moving average, is shown in figure ???. A similar scan was made of the N_A - Mg - N_C angle, however less variance in this value was found and the peak positions did not match previously reported frequencies for strong coupling. Normal modes with the largest N_A - N_C were then chosen for the transition properties scan. It is interesting to note that the peaks in the moving average line approximate the peaks in the chlorophyll *a* absorption spectrum. The normal modes that were chosen had frequencies at 669.6, 701.7, 733.3, 745.5, 755.1, 1105.0, 1107.0, 1122.2, 1142.9, 1320.3, 1361.8 cm⁻¹, which roughly correspond to previously indentified normal modes with frequencies of 728 and 1156 cm⁻¹ [18].

For each selected normal mode, the geometry was propagated such that the sum of all atomic displacements from the optimised geometry was in units of 1 Å. This was done up to 3 Å as it was found for most normal modes the energy difference between the optimised geometry and 3 Å displaced geometry was greater than the thermal energy at 300 K. At each increment, the Q_y transntion energy and dipole was calculated using chl-xTB, as well as TD-DFT with a PBE0/Def2-SVP level of theory. Additionally, Δ -SCF and the eigenvalue difference method was used as a benchmark, along with CAM-B3LYP TD-DFT with the same basis set. A quadratic fit was also made for each of the methods and is shown in the plots. Some points for the Δ -SCF method are not shown as convergence for the excited state at these geometries were not possible for similar reasons discussed in the previous chapter.

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

It is argued that variations in chl-xTB transition properties, in addition to correctly correlated with geometry variations, are also correctly correlated to the coupling between normal modes and the Q_y transition. This is an important consideration for the later chapter on spectral density, where the assignment of spectral features are based on comparison to vibrational modes.

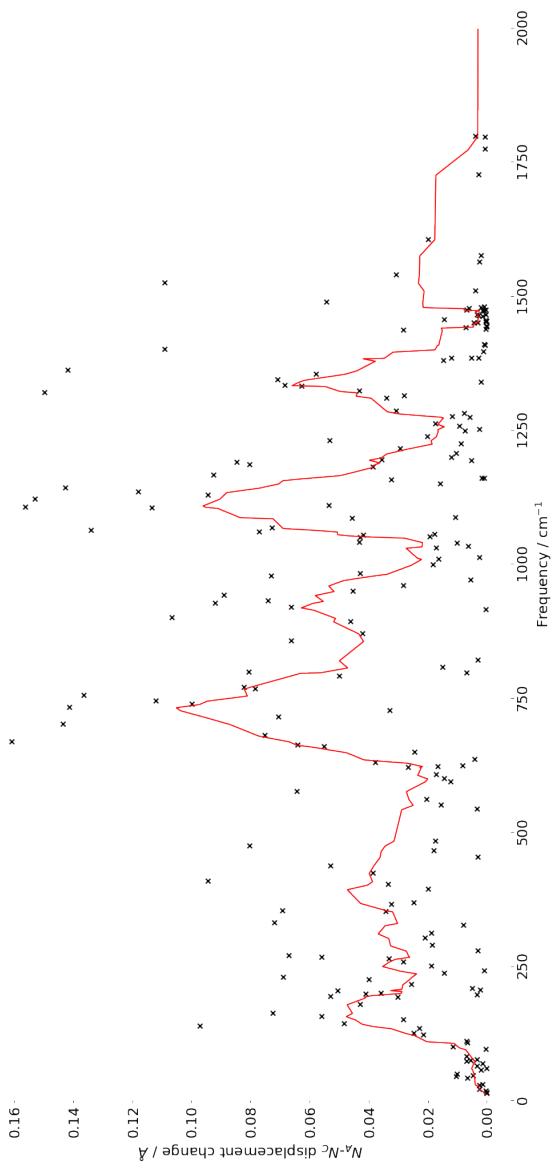


Figure 5.9: Change in the N_A - N_C displacements along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail.

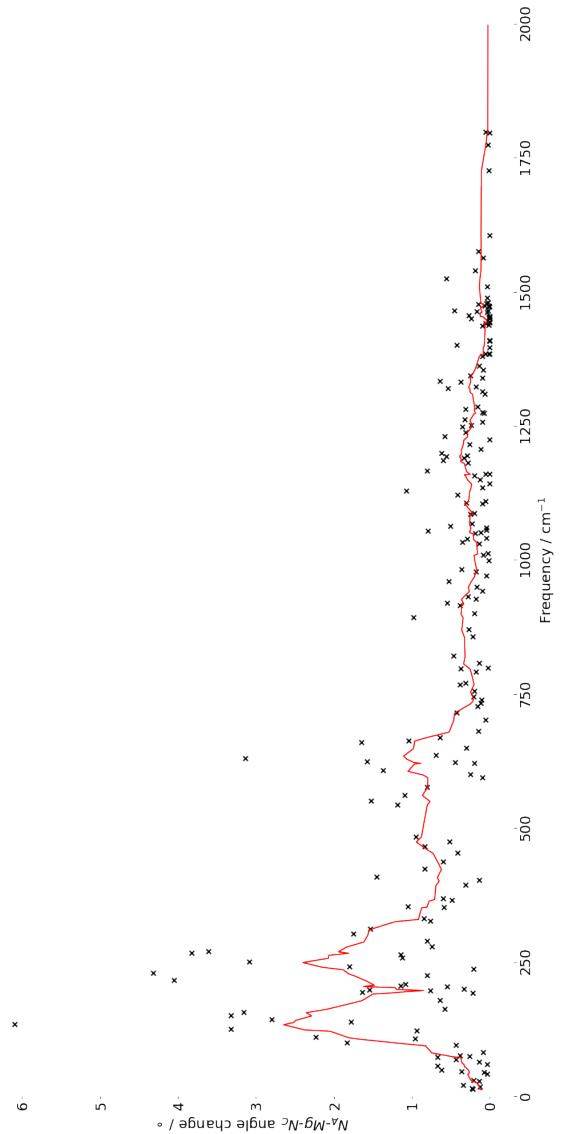
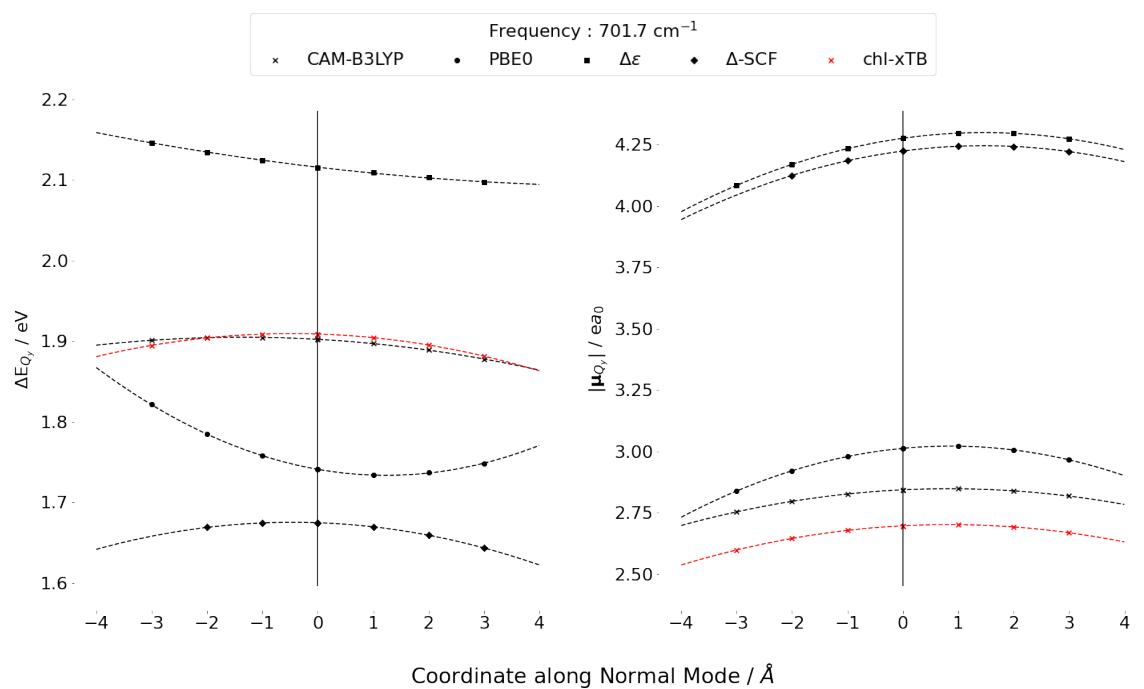
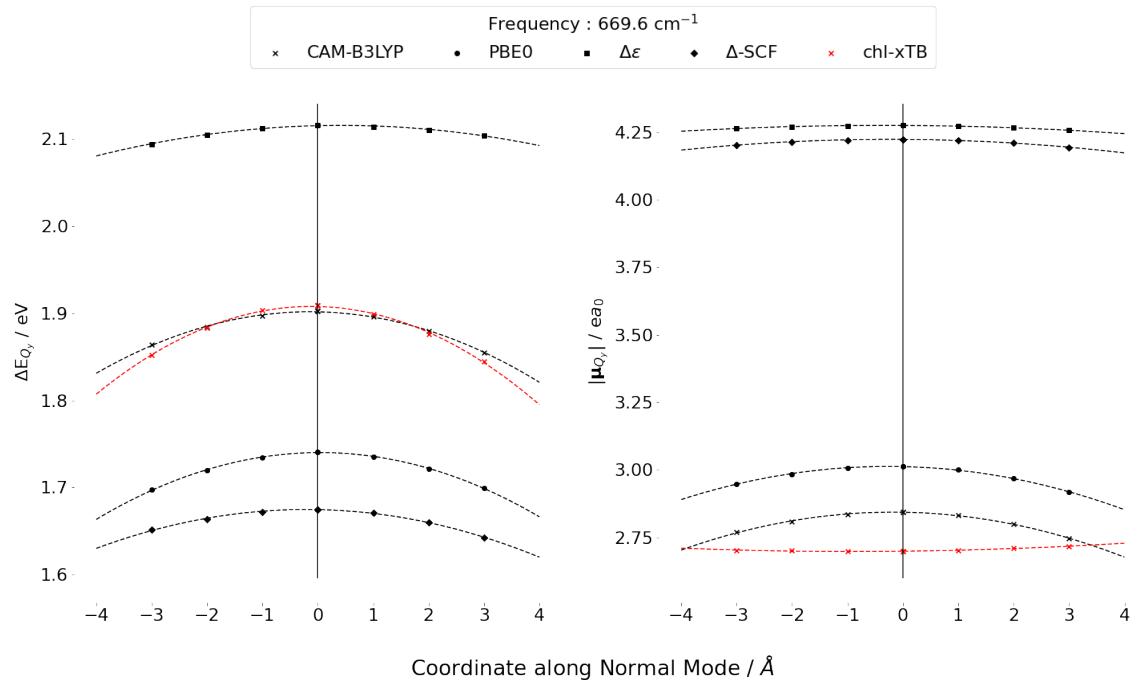
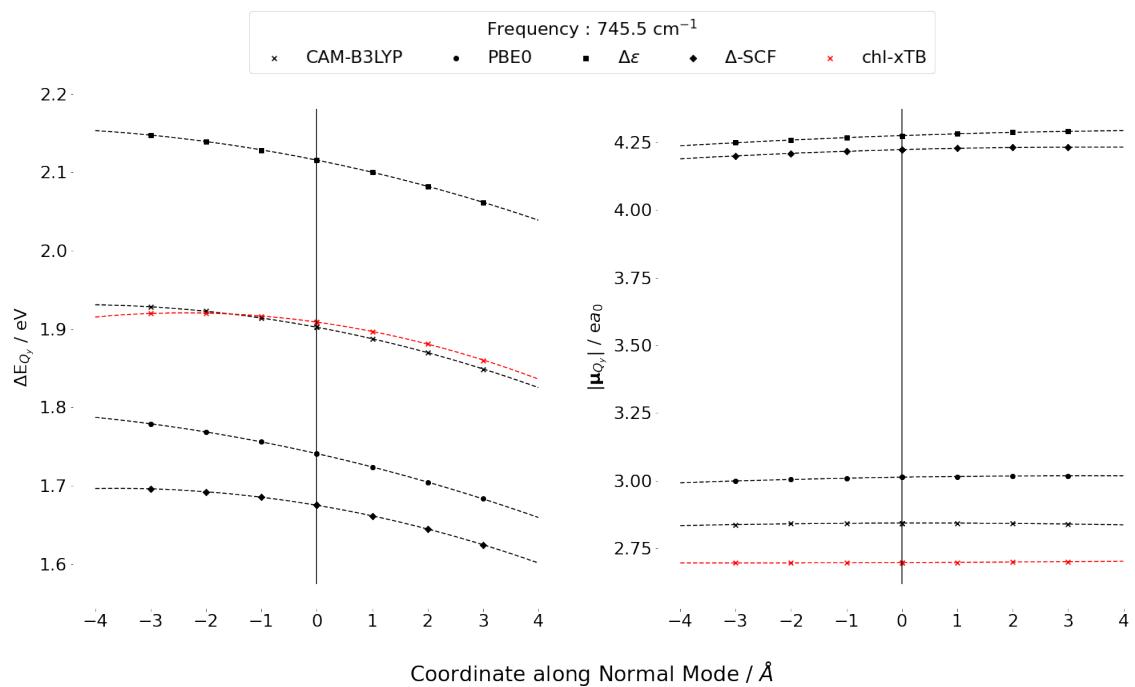
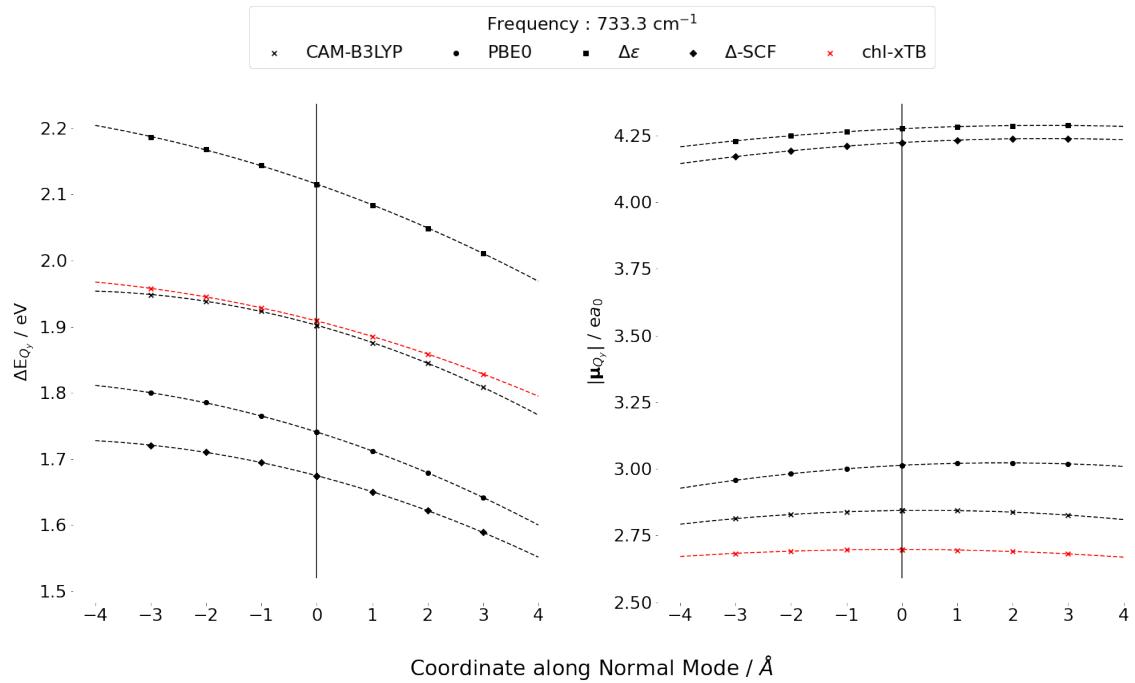
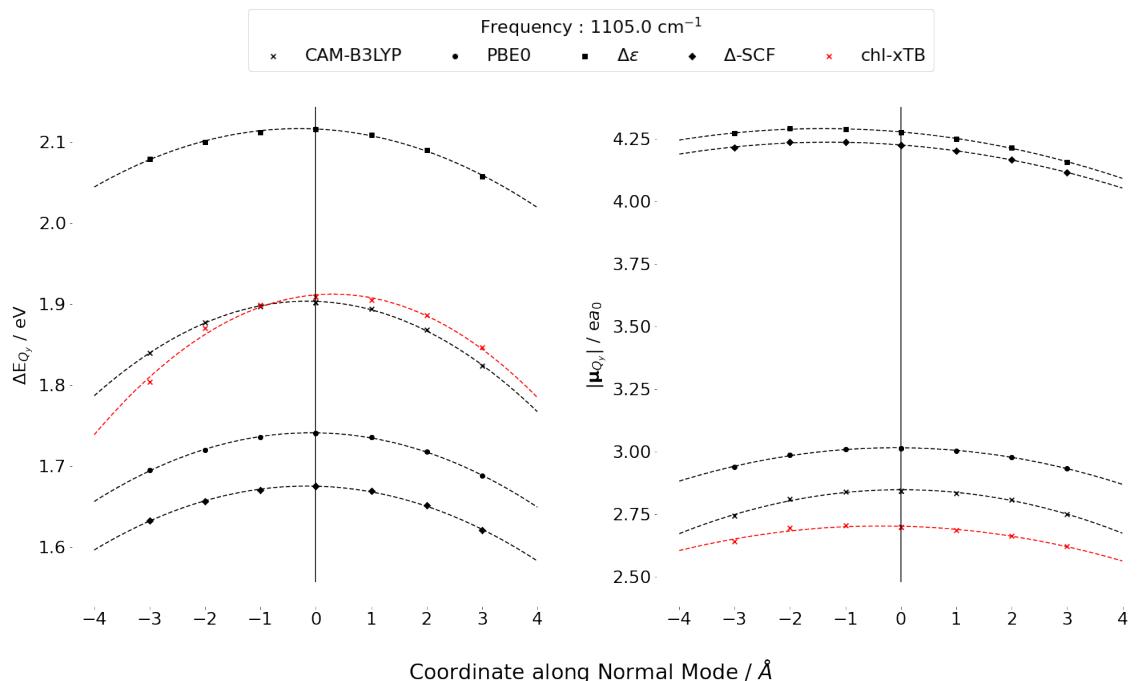
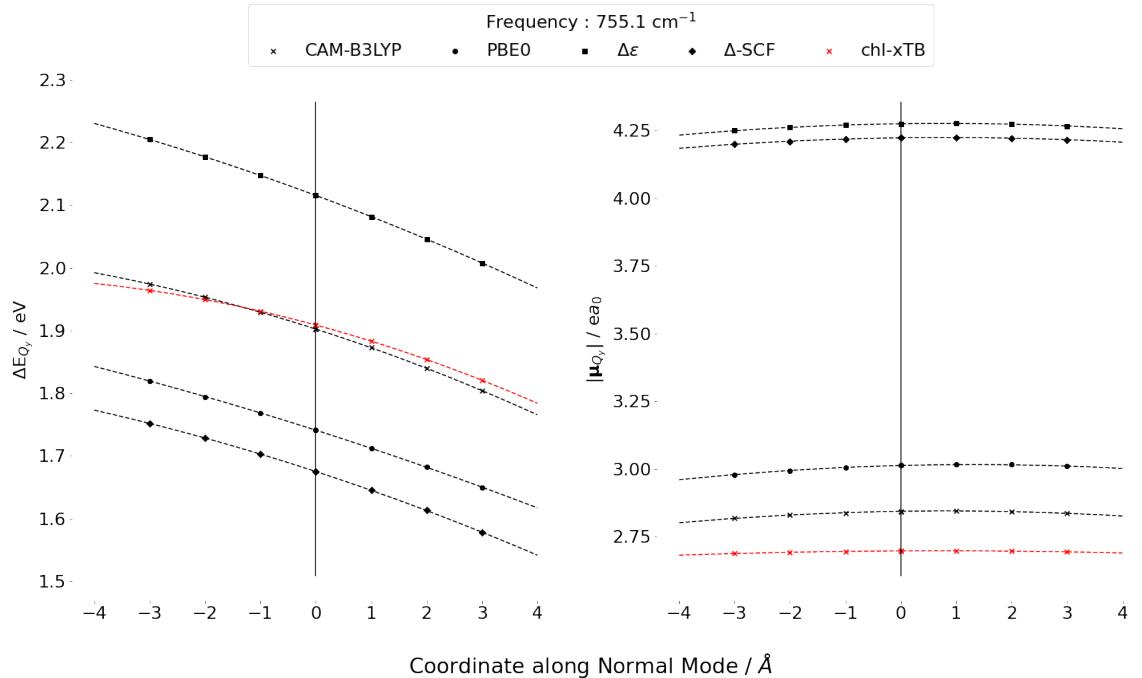


Figure 5.10: Change in the N_A -Mg- N_C angle along the set of GFN1-xTB normal modes for a chlorophyll molecule truncated at the phytol tail. The small smaller variance than in figure 5.3.1 led to this metric to not be used in normal mode choices.

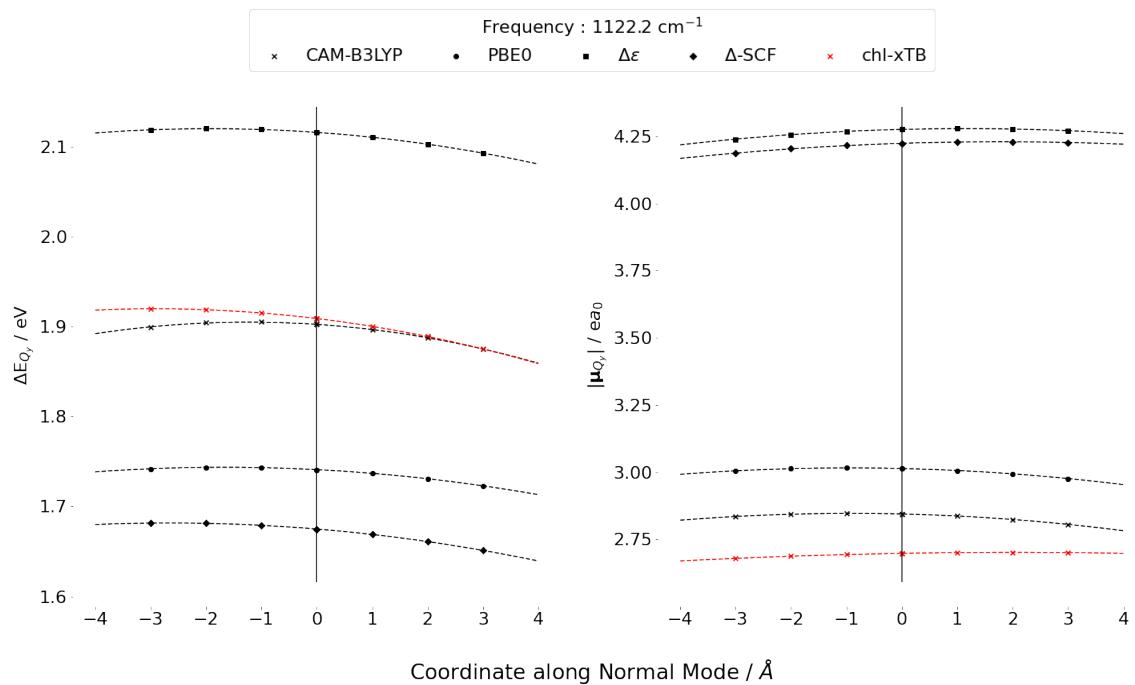
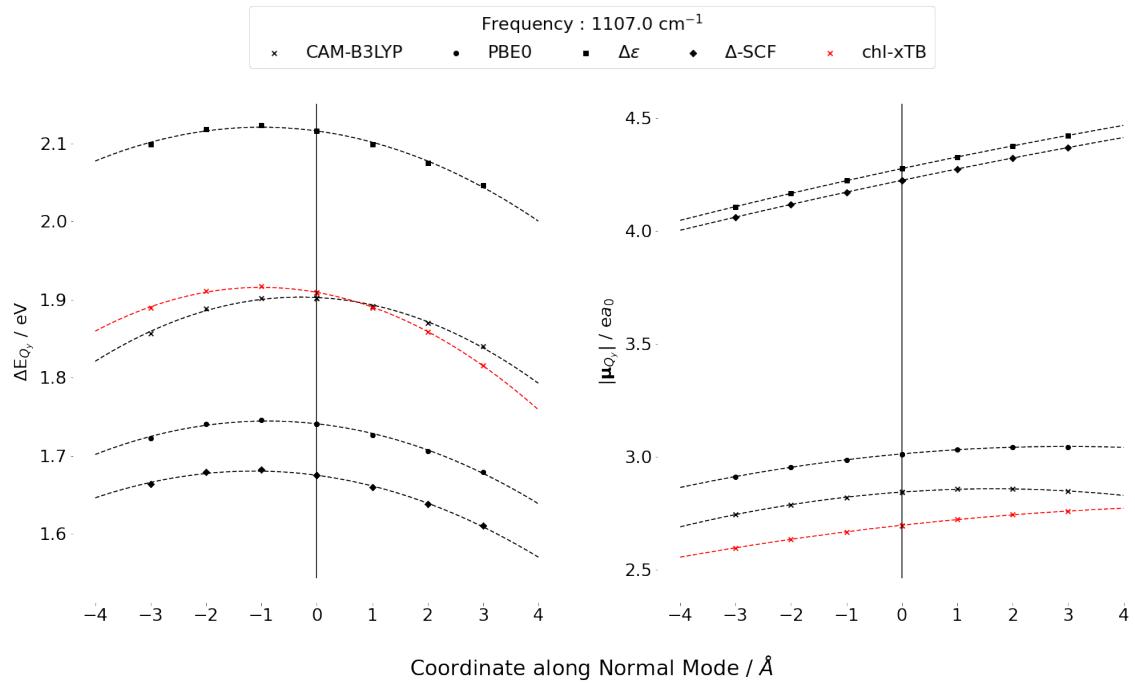


5.3. CROSS-VALIDATION





5.3. CROSS-VALIDATION



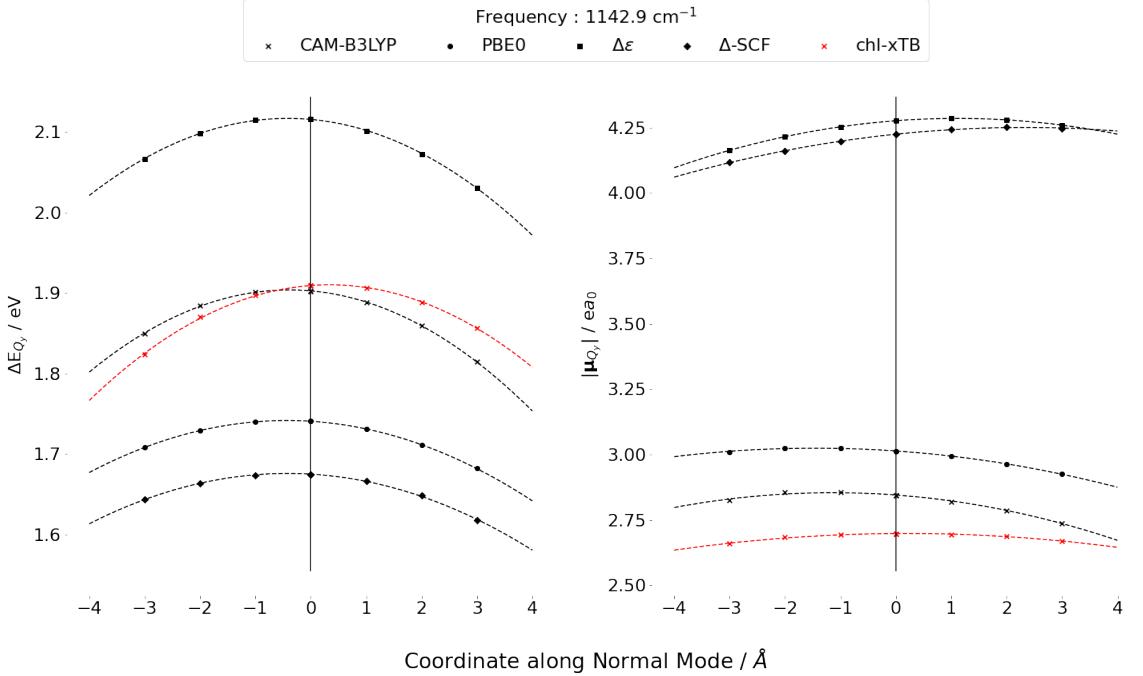


Figure 5.11: Transition energies and dipole magnitudes for the Q_y transition for geometries of truncated chlorophyll along normal modes, calculated with different response theories.

5.3.2 Absorption Spectra

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

The absorption spectra was calculated using frames from an MD trajectory of chlorophyll A in an explicit diethyl ether solvent. An explicit solvent was used to account for inhomogeneous broadening in the spectrum. The MD was performed with the OpenMM toolkit. Forcefield parameters for the chlorophyll were taken from a bespoke parameterisation for photosystem II [32], with the rest of the system using the OpenForceField. The structure for chlorophyll was taken from the same source as the bespoke forcefield, and packed with explicit solvent using the tools in the Mistral package. Equilibration and production steps were done with a Langevin integrator set to 300 K and a timestep of 0.5 femtoseconds. The system energy was minimised before running a

10ps equilibration. Frames were then taken from a 2ns simulation time, with structures taken every picosecond.

Transition properties for chlorophyll structures from every frame were calculated. This was done with the chl-xTB method, as well as PBE0 and CAM-B3LYP TD-DFT, both using the Def2-SVP basis set. For the TD-DFT methods, the lowest excited state, rather than always the Q_y state, was taken for each frame, so as to fully capture the Q band. Experimental data for the absorption spectra was also taken from Katz *et. al* [29]. Embedding effects were included in the chl-xTB Fock matrix with a particle mesh Ewald method. The real space term was calculated using QCORE, whereas the more complicated spline reciprocal space term was calculated with the HelPM library. The absorption spectrum for each method was calculated by a gaussian fit of a histogram of transition energies. The fit was normalised such that the area under all spectra matched that of the experimental spectrum. The absorption spectra, both with and without a single-parameter shift of excitation energies, can be seen in 5.12.

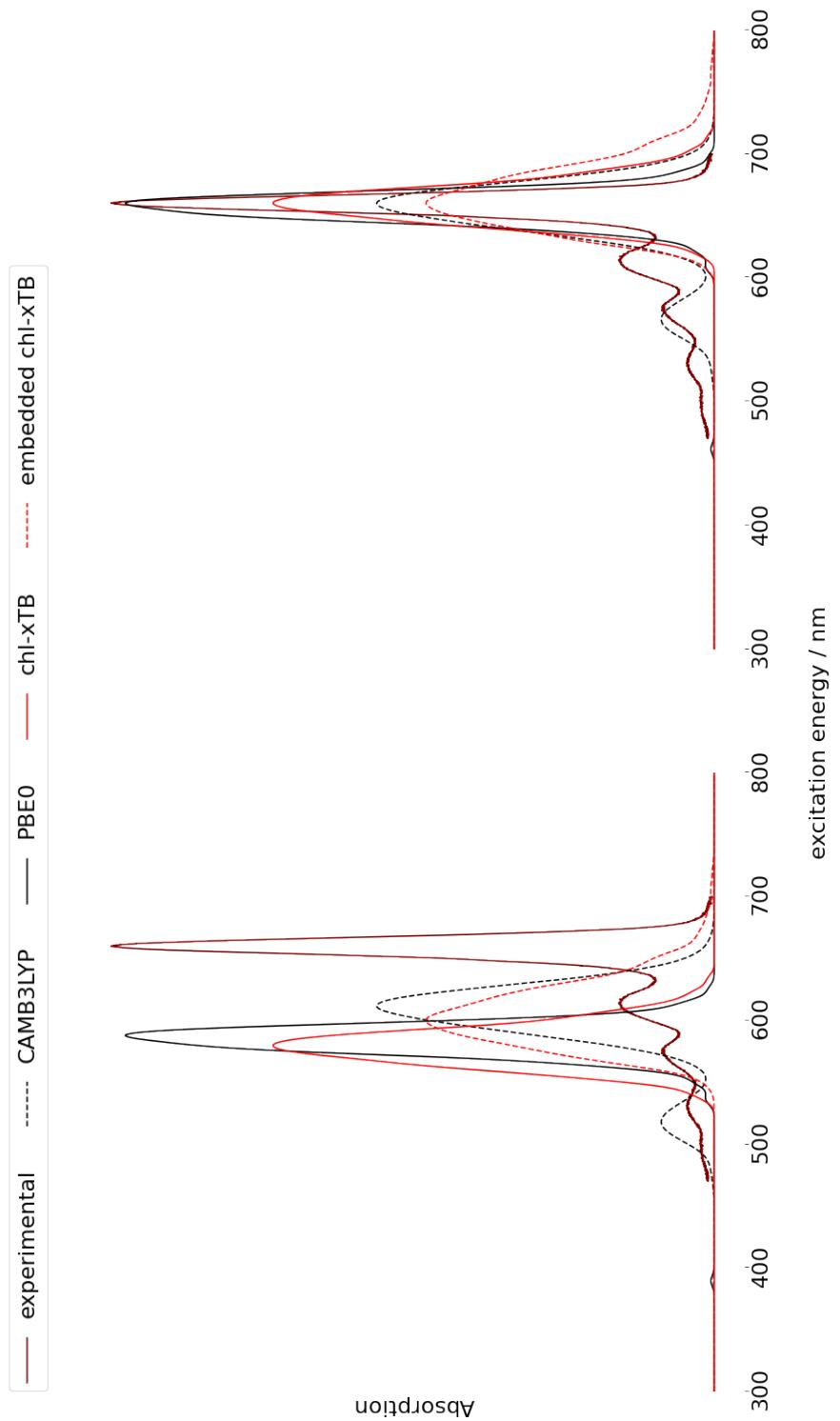
chl-xTB performs equally well as TD-DFT methods at simulating absorption spectra, although constrained by the limits of the method and the training data. It can be seen that the chl-xTB lineshape is similar in position and width to the PBE0 method. This is highly encouraging, as the functional groups on chlorophyll A can have a large effect on the Q_y transition, and so the good agreement here provides evidence that important chemical features are still captured in different systems. Although the chl-xTB lineshape is wider than the experimental spectrum, the CAM-B3LYP and PBE0 lines are also wider and so the chl-xTB is still within a reasonable expectation of accuracy. Also seen is the lack of other peaks from the Q band, which is present in the experimental spectrum. The CAM-B3LYP data shows one peak at 510 nm, with PBE0 having a very slight peak at around 380 nm. This is not present in the chl-xTB spectrum as other Q transitions are not calculated due to being outside of the training data. The embedded chl-xTB spectra is both red-shifted and broadened slightly, which is also observed in simulated spectra from other methods.

5.4 Conclusions

It has been shown that novel approximations to the full linear-response eigenvalue equation can give accurate predictions of transition properties from high-level methods. This has been shown by optimising a set of parameters for the response approximations, as well as for the underlying electronic structure.

Overall, chl-xTB performs as well as can be expected from the training data. Predictions of transition energies and transition dipoles are incredibly close to the PBE0 values, and well within the error between different high level DFT functionals. The variations of these transition properties can be attributed to geometry variations rather than random error.

From this it could be expected that improvements in the training data would yield better



5.4. CONCLUSIONS

accuracy. This might be done by extending the training data in either the conformational space, or by investigating other systems.

This was taken into consideration when investigating this method, for example using CAM-B3LYP as training data instead of PBE0. This approach would have two drawbacks. First would be that PBE0 has already been established for investigating LHII properties, and so a clearer comparison of results for the exciton system would be possible. Second is that a feature of this method is the reduced amount of computational cost required. If the amount of CAM-B3LYP data that may have been necessary to obtain increased too much then it would no longer be efficient to parameterise compared to using a different, lower level method. For other systems smaller than chlorophyll, this might not be the case.

CHAPTER



BEYOND MONOMER CHLOROPHYLL

This chapter investigates the applicability of chl-xTB for aggregate chlorophyll systems, beyond the monomers reported in the previous chapter. It is the properties of aggregate chlorophyll that make light harvesting systems so efficient. It is necessary to show that chl-xTB can be used to construct viable models for these systems.

6.1 Theory

6.1.1 Exciton States

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

To recap the exciton theory from the introduction, the exciton states $|\Psi\rangle$ can be constructed as a Hartree product of the states of individual sites (also referred to as monomers or chromophores)

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

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

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

where $|\phi_i\rangle^*$ is the excited state of monomer site i . States with more than one exciton are possible in this framework, however this is not necessary for light harvesting systems as stated in the

introduction. The Hamiltonian, including the "ground state" where there are no excitons in the system, is given in by

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

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

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

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

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

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

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

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

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

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

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

This Hamiltonian for exciton states is slightly different those reported in the literature, as the ground state where no excitons are present is not usually included. However, if the local excitations are sufficiently different in energy and the coupling between ground and excited states is weak, then the eigensolution corresponding to the ground state would have mostly ground state character. The other excited state eigensolutions would have very little ground state character, and so the block matrix of just excited state contributions can be taken which would return the Hamiltonian used in many other studies. Calculating the ground state however is necessary if all the states for the exciton system are required, which is the case for the spectral density investigation of the next chapter.

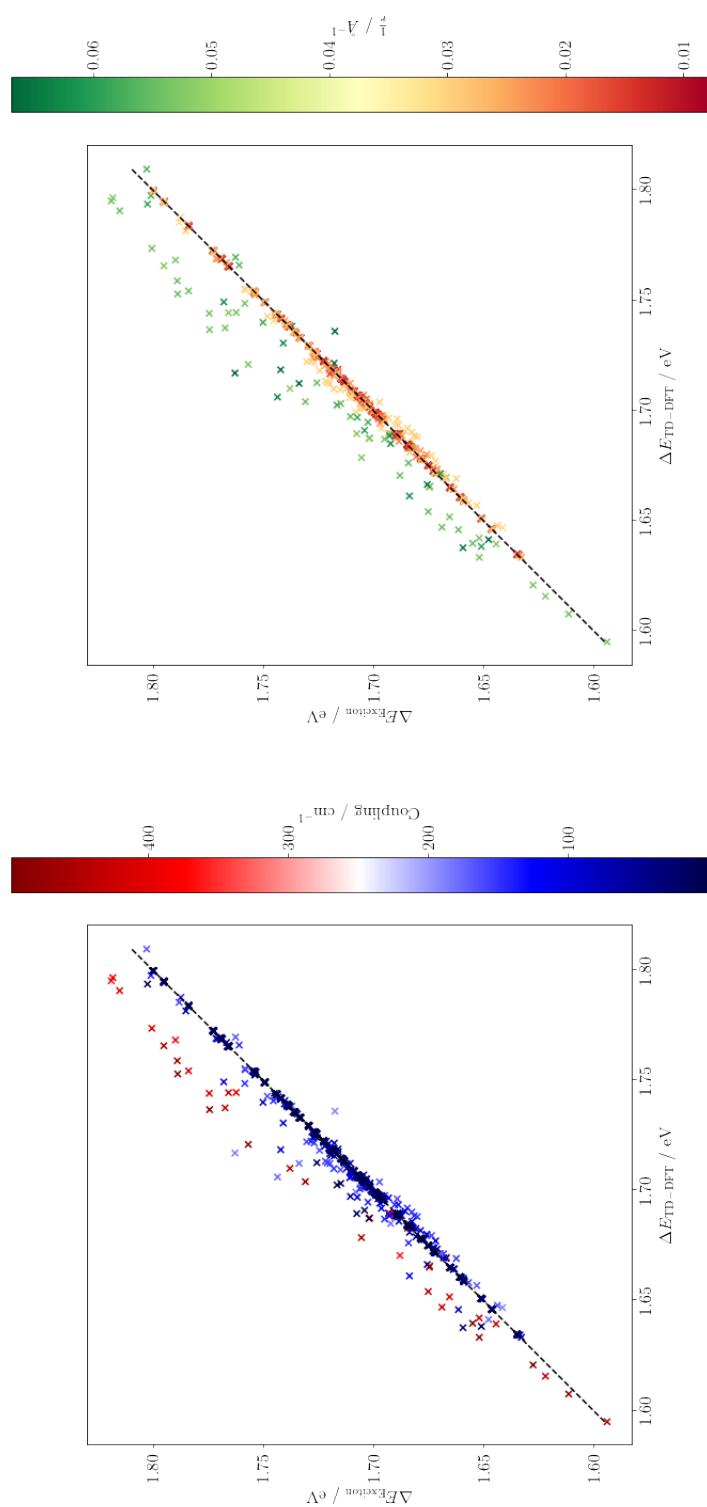
6.2 Same-functional Benchmarking

It has been well established in many other studies that the Frenkel exciton Hamiltonian with the framework set out above can predict dimer response properties fairly well, as has been discussed in more detail in the introduction. This type of study was repeated here to set a benchmark for the later comparison with chl-xTB. Well known causes of error between the exciton model and full TD-DFT, such as inter-chromophore distance and treatment of coupling of excited states, were explored. This was done with two studies: first comparing the transition energies predicting by the exciton model against full TD-DFT for a set of chlorophylls from LH2; second with a conformational scan of chlorophyll dimers along rotational axes, to pinpoint if errors may be due to any specific interactions.

6.2.1 LH2 Dimers

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

A clear relationship can be seen between the full dimer calculations and the exciton model. The systems in which the error is significant can be seen to be correlated to the distance between the monomers, with a lower distance giving a higher error. There is a smaller amount of correlation with the coupling value from the exciton Hamiltonian (taken as the coupling between the excited states, as generally the coupling value to the ground state is lower), implying that there are effects that are not included in the exciton model rather than an improper model of the coupling. These effects could be from not including higher energy transitions, such as the Q_x transition, or



charge transfer transitions. However for the large majority of cases the exciton model predicts full dimer TD-DFT with decent accuracy, within the range of previously reported exciton models.

6.2.2 Rotation

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

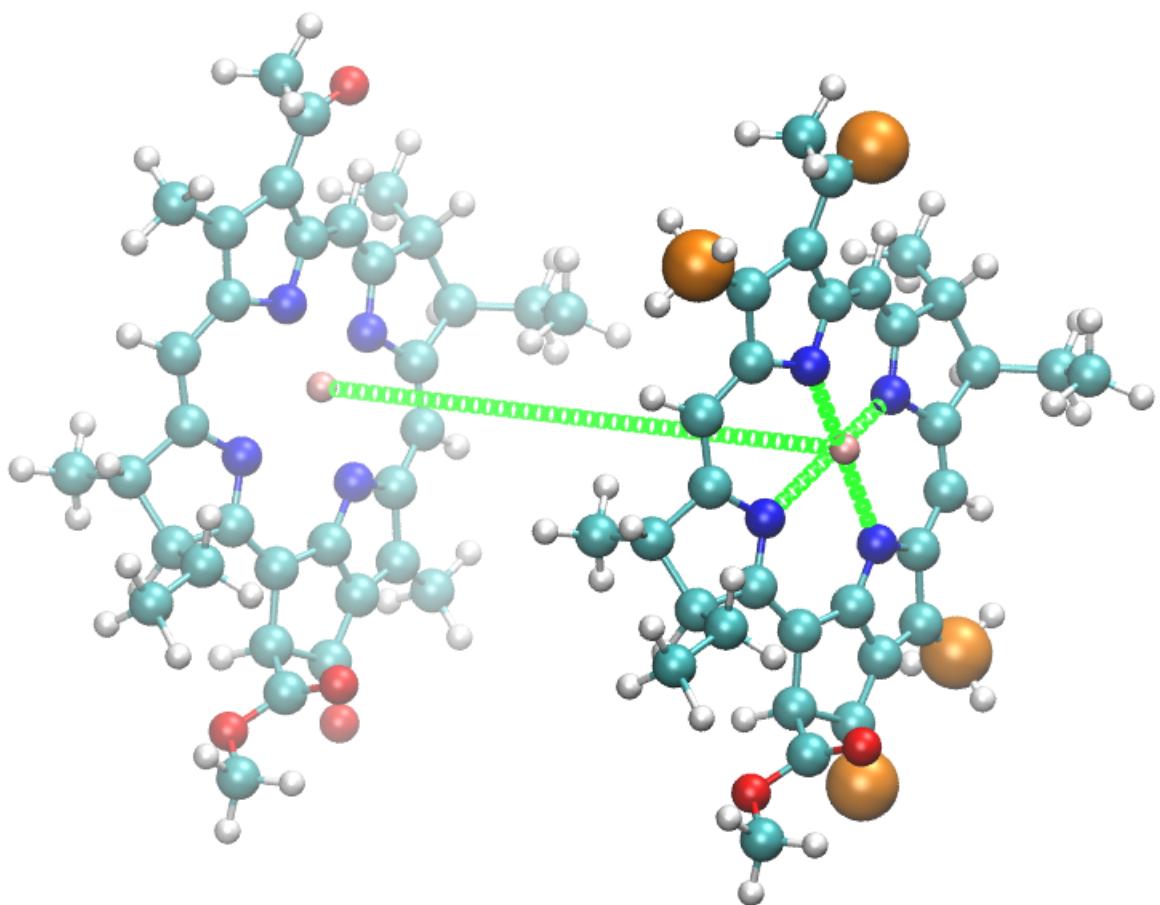
$$(6.8) \quad v_{mn} = \frac{\vec{\mu}_m \cdot \vec{\mu}_n}{R_{mn}}$$

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

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

In order to make assignment of exciton states easier, the geometries of the two truncated chlorophyll monomers were altered such that there would be a distinct gap in their transition energies. This gap was not wider than the variation in transition energies found in LHII 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. From these plots it is clear to see the same trends in 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 profile, where the exciton model predicts the higher excited state at a little above the TD-DFT profile. A similar effect is seen in the LH2 dimers, where the errors in transition energies are almost all overestimates, and is attributed to a higher coupling value in the exciton model than is present in TD-DFT. This higher coupling could be due to the asymptotic behaviour of a bare point charge



6.2. SAME-FUNCTIONAL BENCHMARKING

interaction. Using a short range damping operator might correct this, but this investigation 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 as these dimers are further apart they would be expected to have less error, as seen in fig.

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. However as similar models to the one benchmarked here have been shown to be useful in previous studies, it is out of the scope of this work to include these effects.

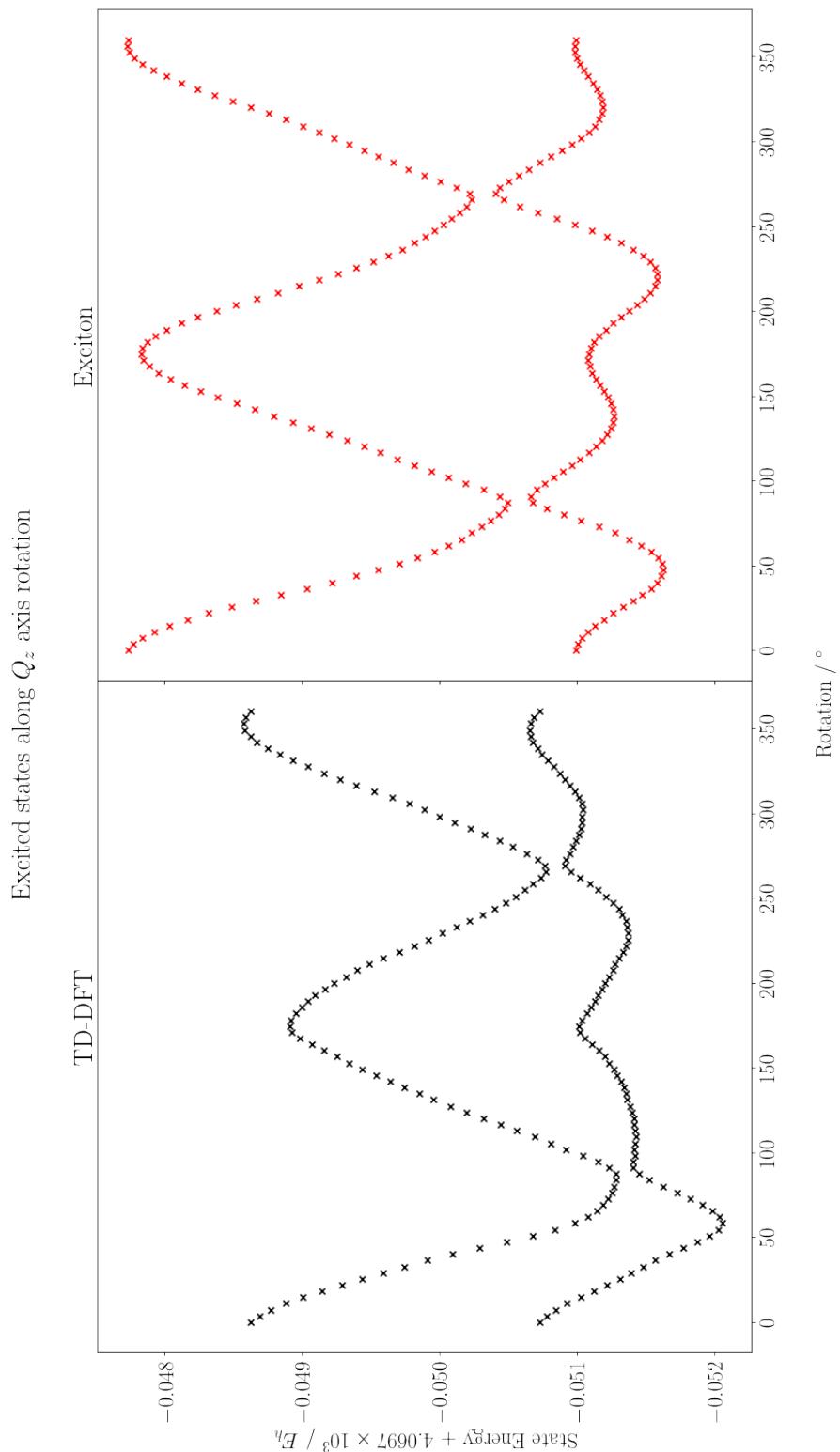


Figure 6.1: caption

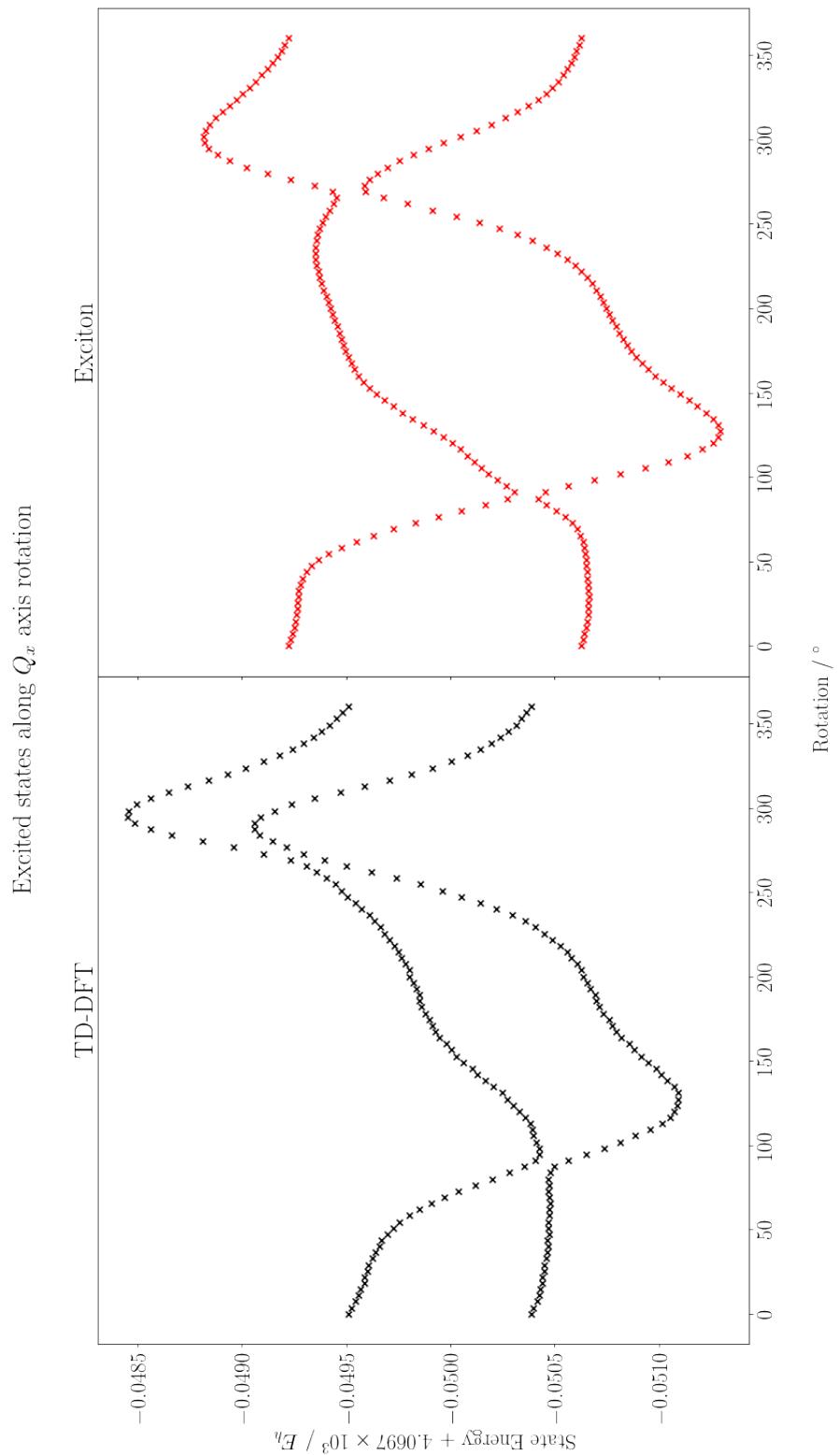


Figure 6.2: caption

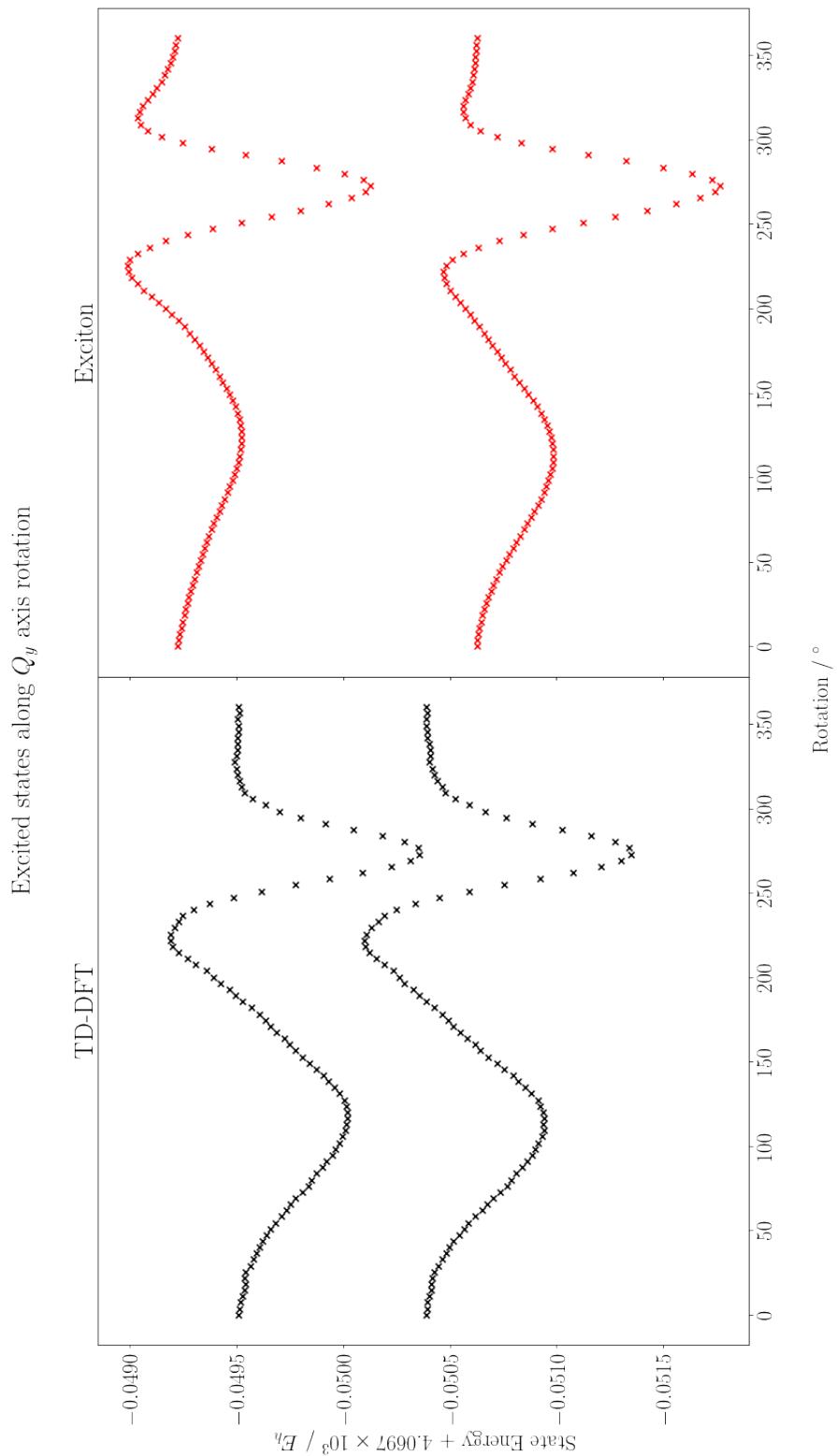


Figure 6.3: caption

6.3 chl-xTB Excitons

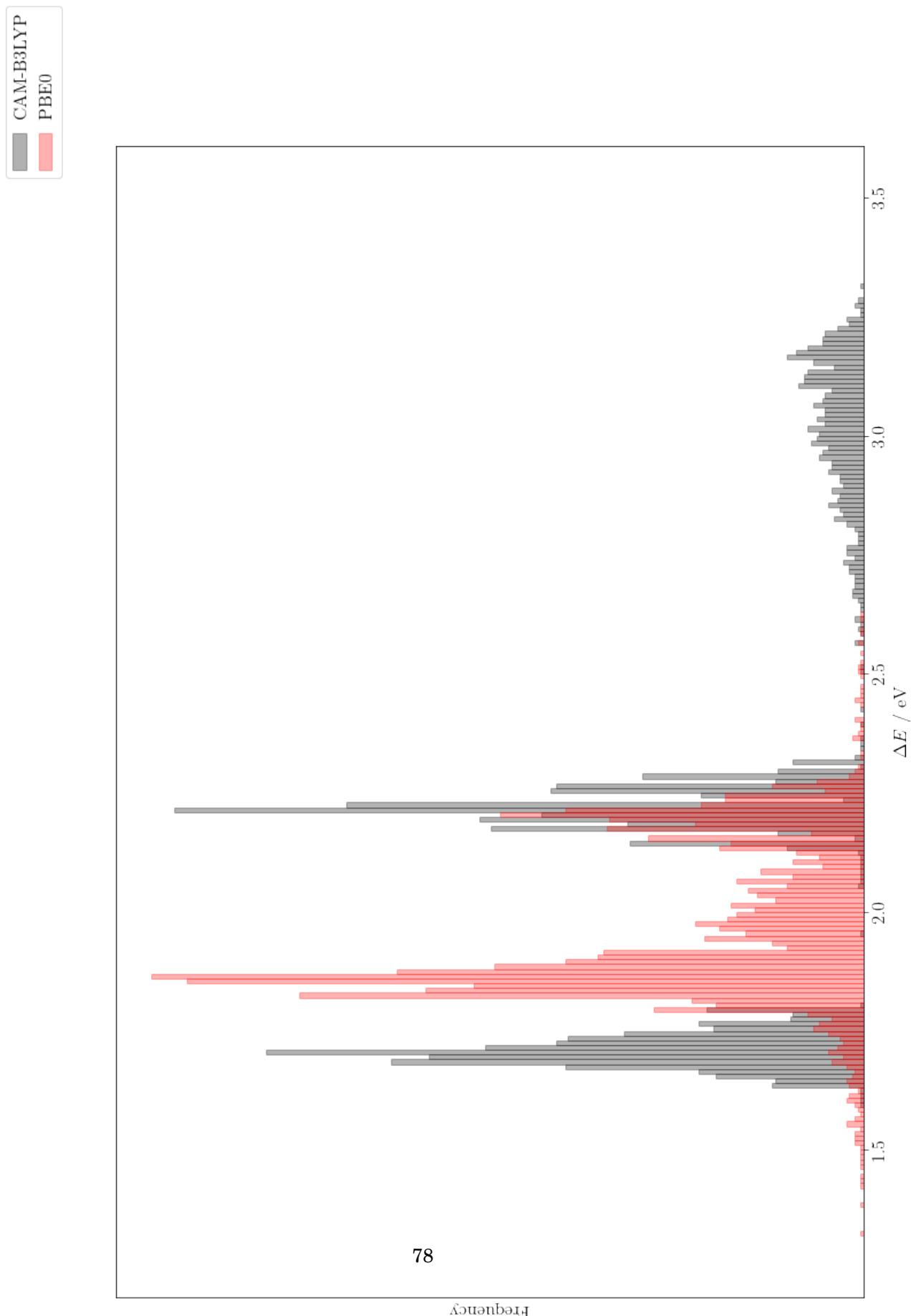
6.3.1 Full Dimer comparison

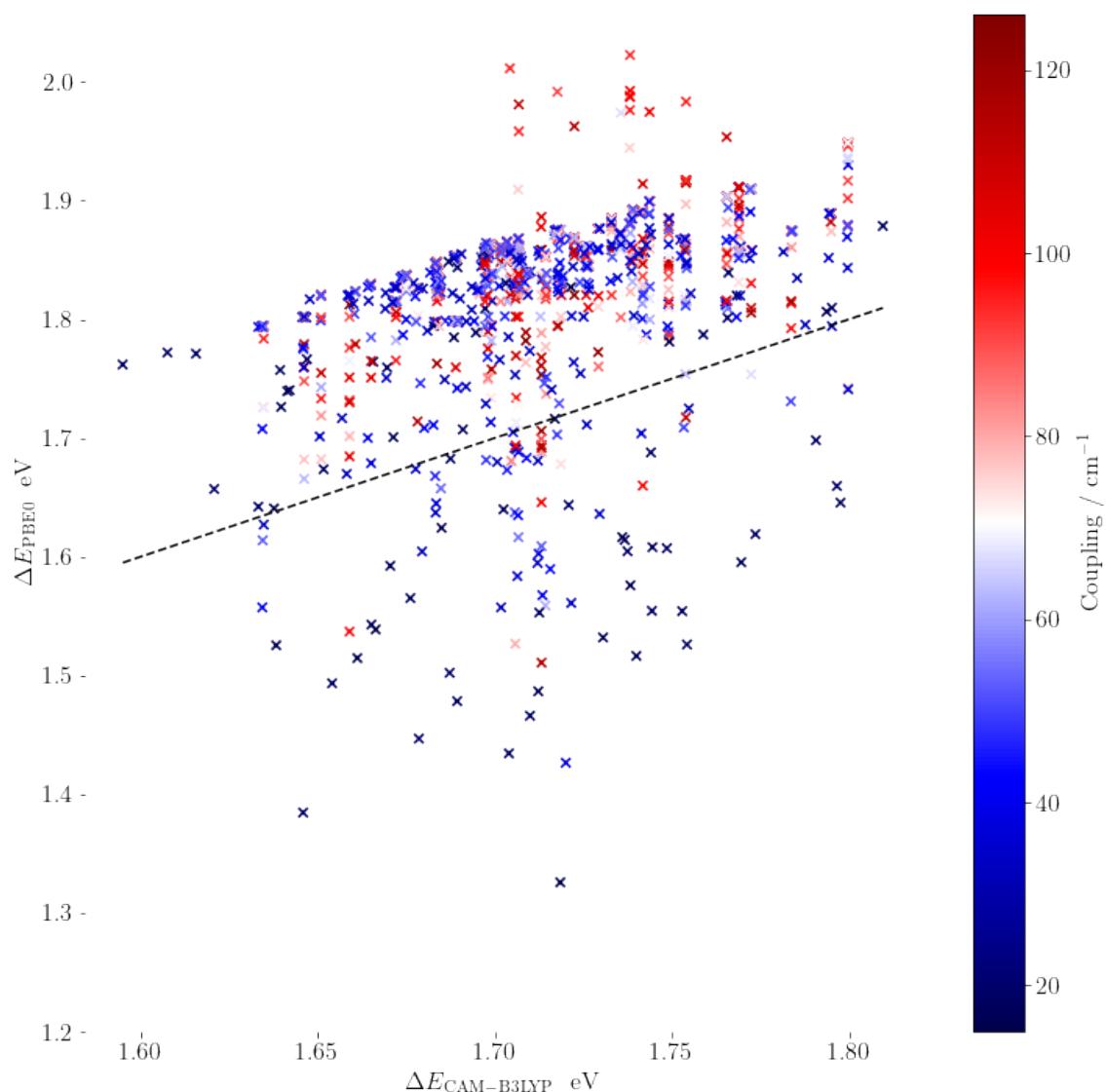
6.3.1.1 Discarding PBE0

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

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

Here the individual Q_y and higher energy transitions, that are well resolved by CAM-B3LYP, are mixed. This leads these transitions to be uncorrelated to CAM-B3LYP transitions, as seen in the first panel of figure. Hence the PBE0 dimer data would not make a good comparison for either the PBE0 exciton or the chl-xTB exciton models, and so was discarded from benchmarking. For the remaining benchmarking, the exciton model is compared to CAM-B3LYP TD-DFT.





6.3.1.2 Assignment of States

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

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

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

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

For the Frenkel exciton result, the location was taken as the monomer which had the most amount of character in the eigenvector solution. For example, if the diagonal elements of the Hamiltonian were in order of ground state, transition on A, transition on B, and the eigenvector solution of the lower energy exciton state had mostly "transition on A" character, the exciton was assigned as localised on monomer A.

6.3.1.3 Comparison

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

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

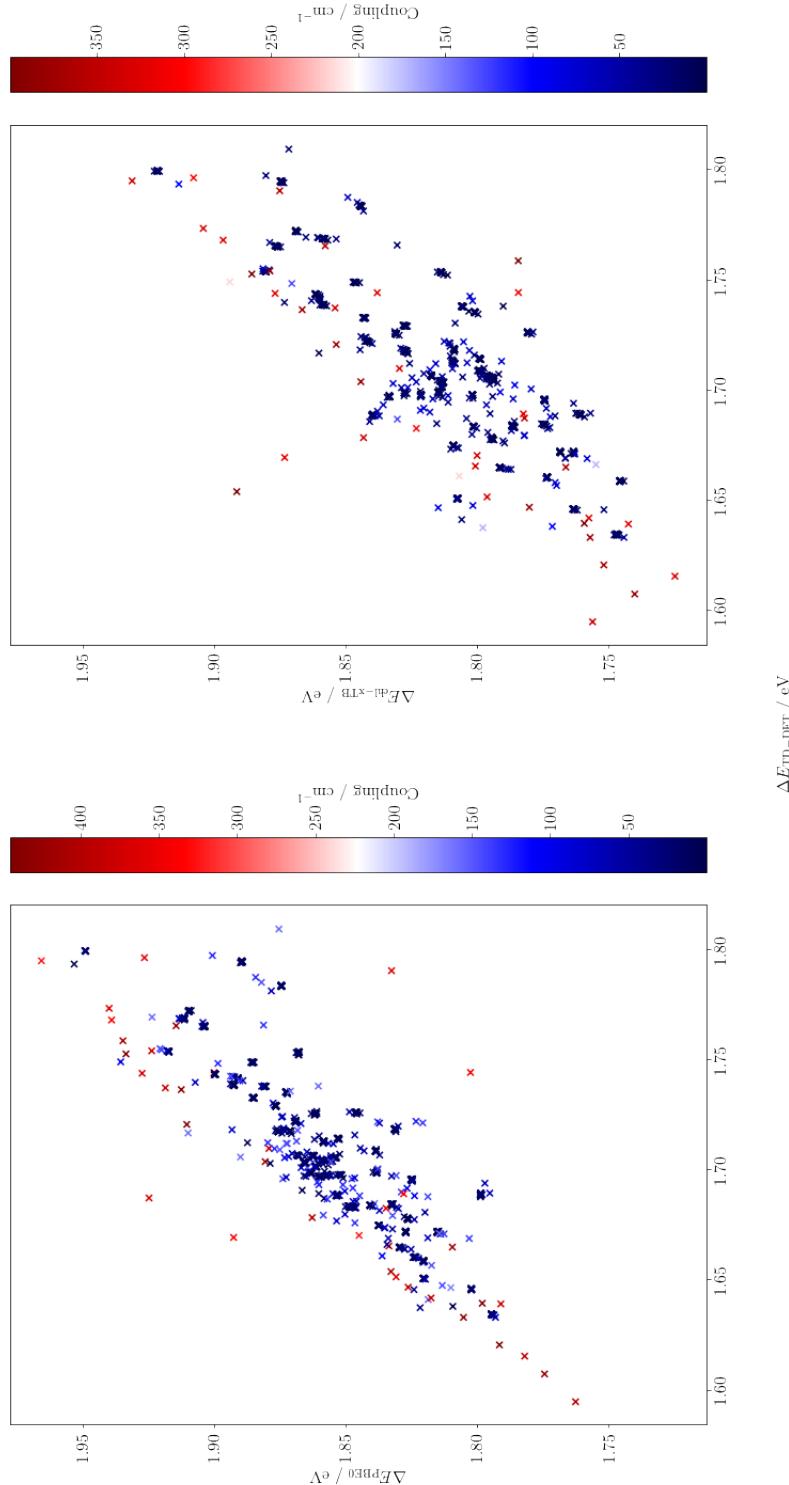
to overestimation of transition energy implied that the exciton model was the leading cause of error. This implies that at the opposite end, where there is large separation and low coupling between chromophores, the exciton model would have little effect and so any error here would be due to the monomer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

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

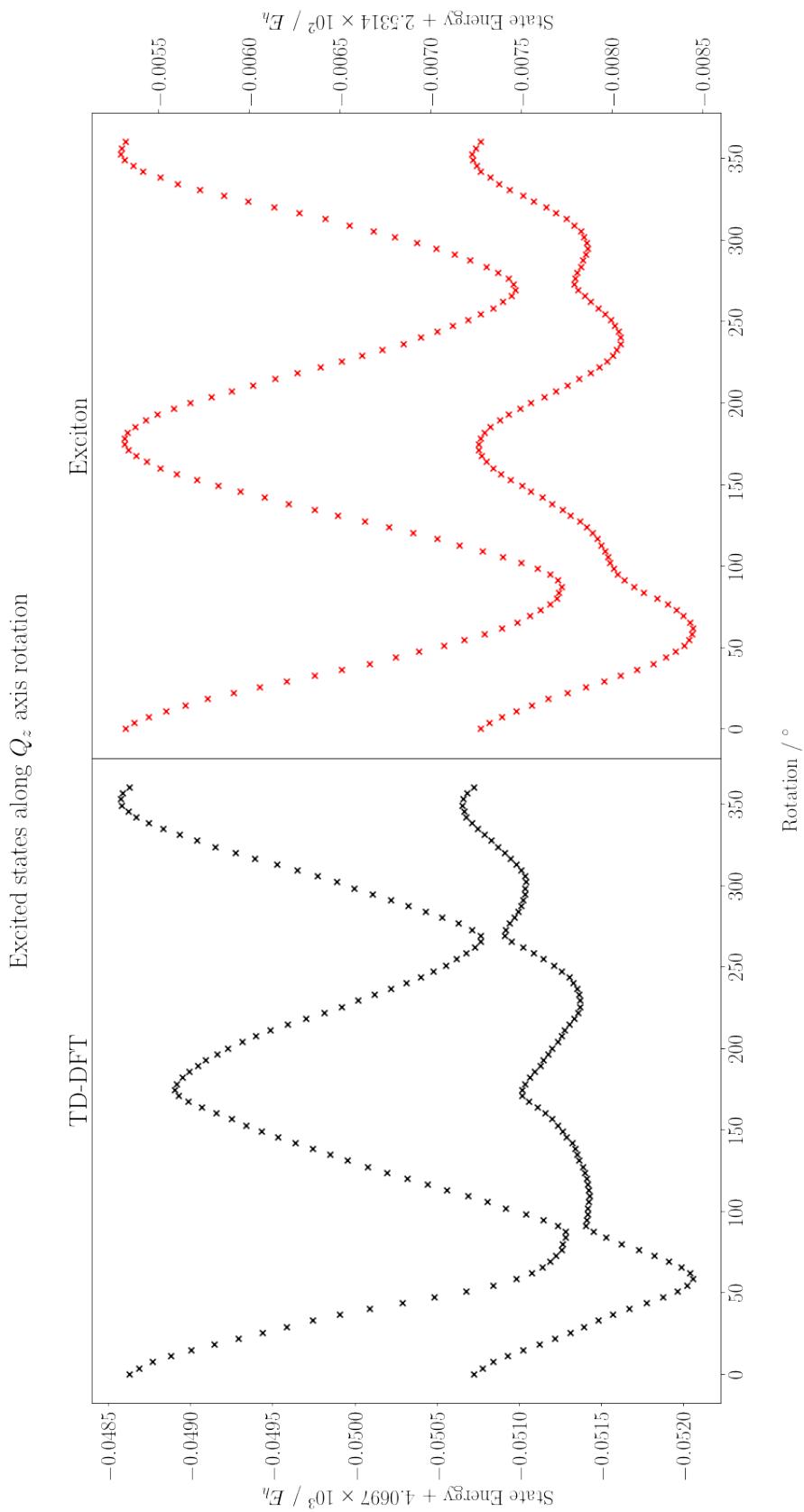
6.3.1.4 chl-xTB Scans

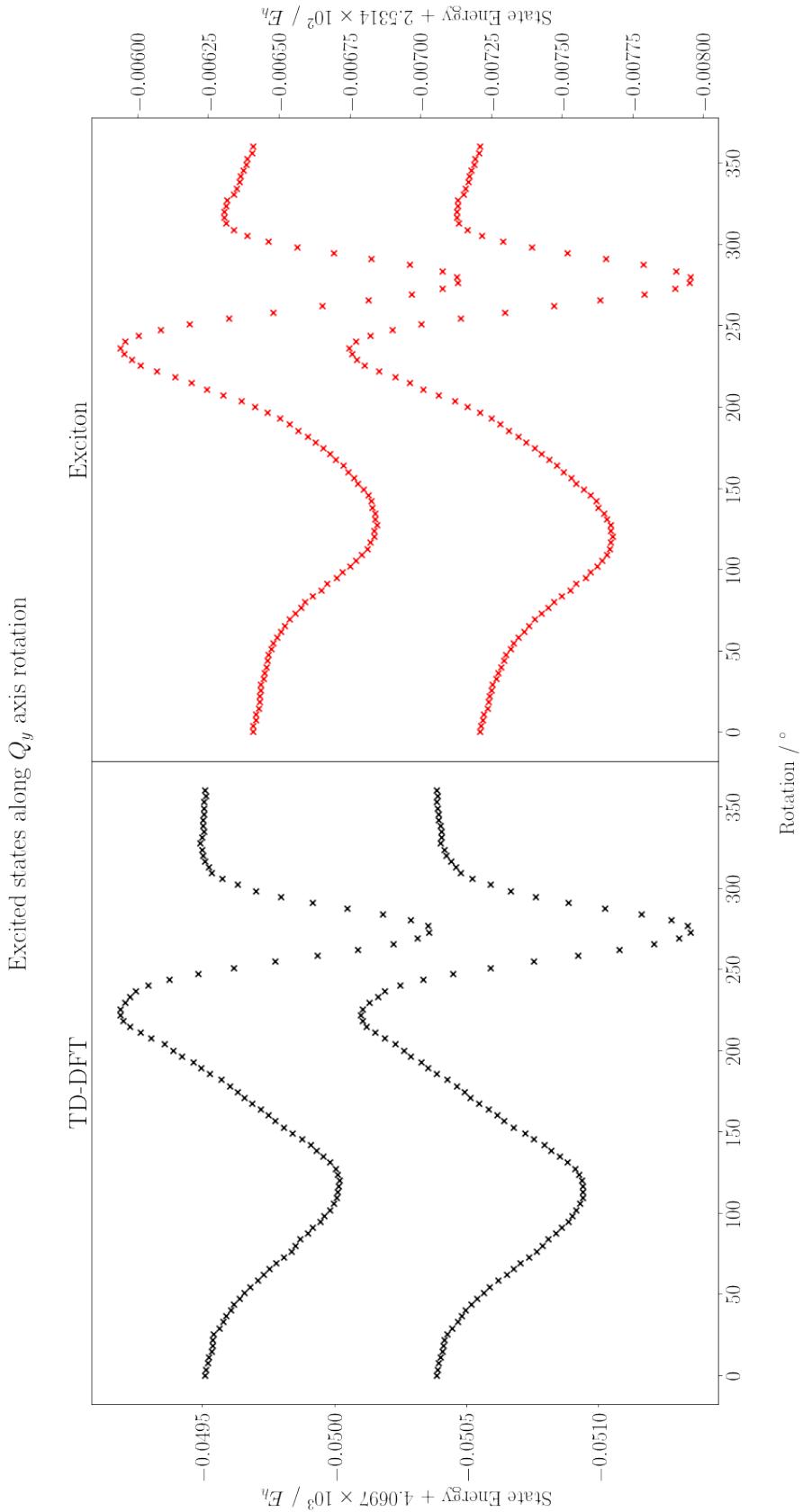
Similar to the CAM-B3LYP scans above, the profile of excited dimer states was a useful test in detailing the accuracy of chl-xTB excitons. For the same reasons as the LH2 dimers, this could not be benchmarked against chl-xTB or PBE0 dimer properties, and so was compared to CAM-B3LYP again. Similar qualitative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

In this case where there is little random error, the exciton model rather than chl-xTB properties are 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 inter-chromophore geometry, such as angle and distance, is for the most part very similar for all dimers, especially nearest neighbours. In contrast, in this study the intra-chromophore geometry is constant, and the intra-chromophore geometry is changed. This removes the compound issue between chl-xTB and CAM-B3LYP monomer

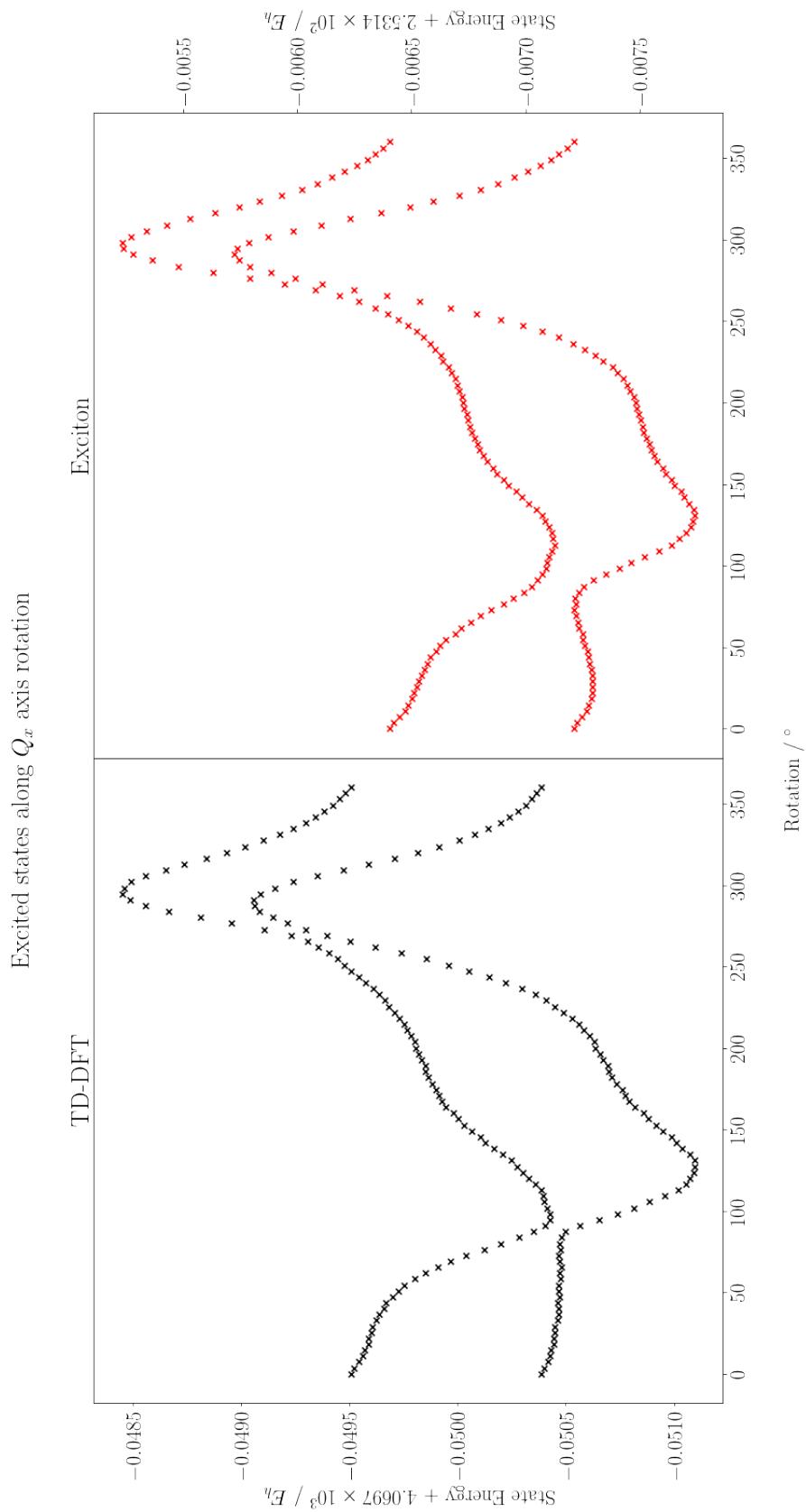


6.3. CHL-XTB EXCITONS





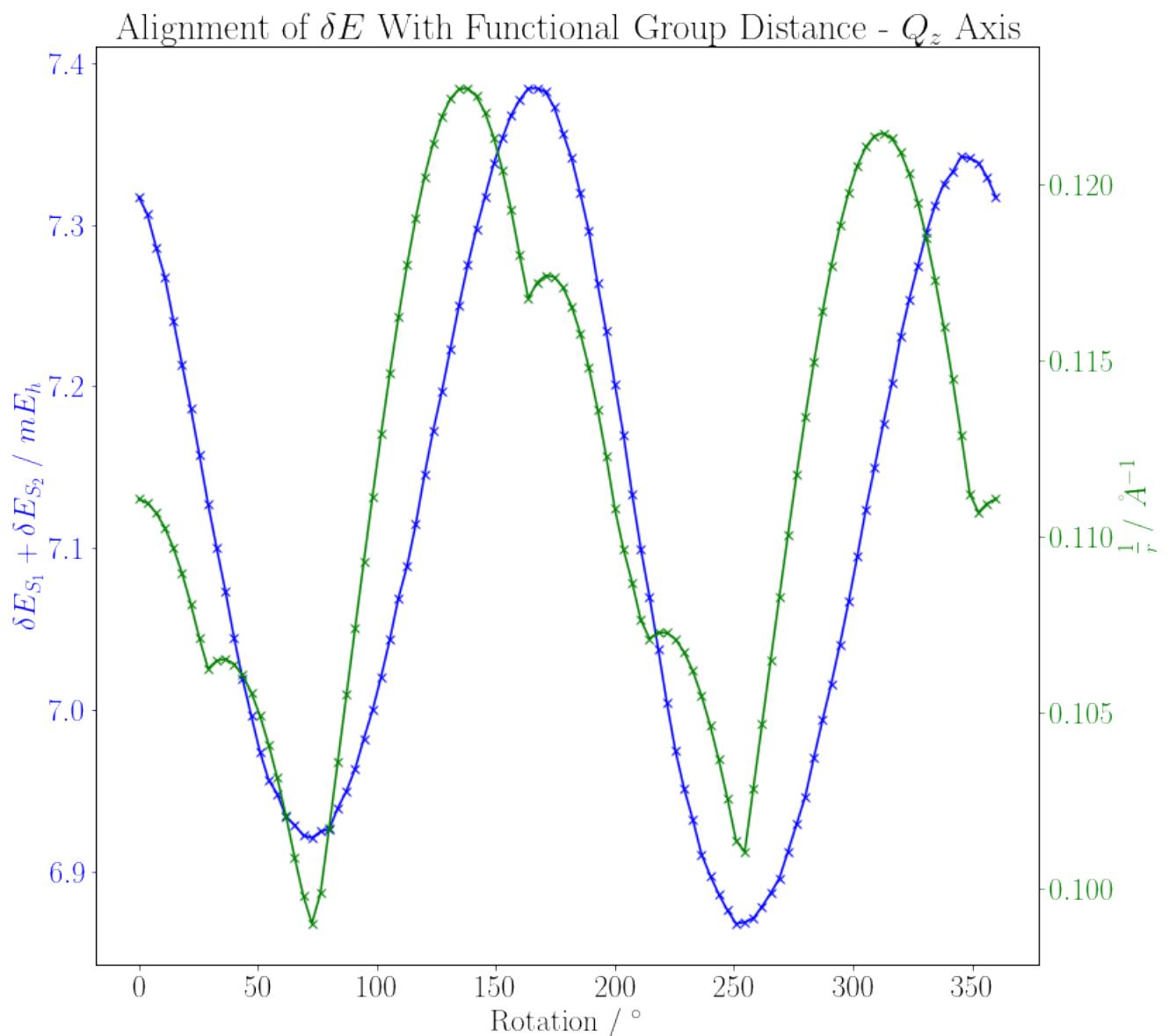
6.3. CHL-XTB EXCITONS

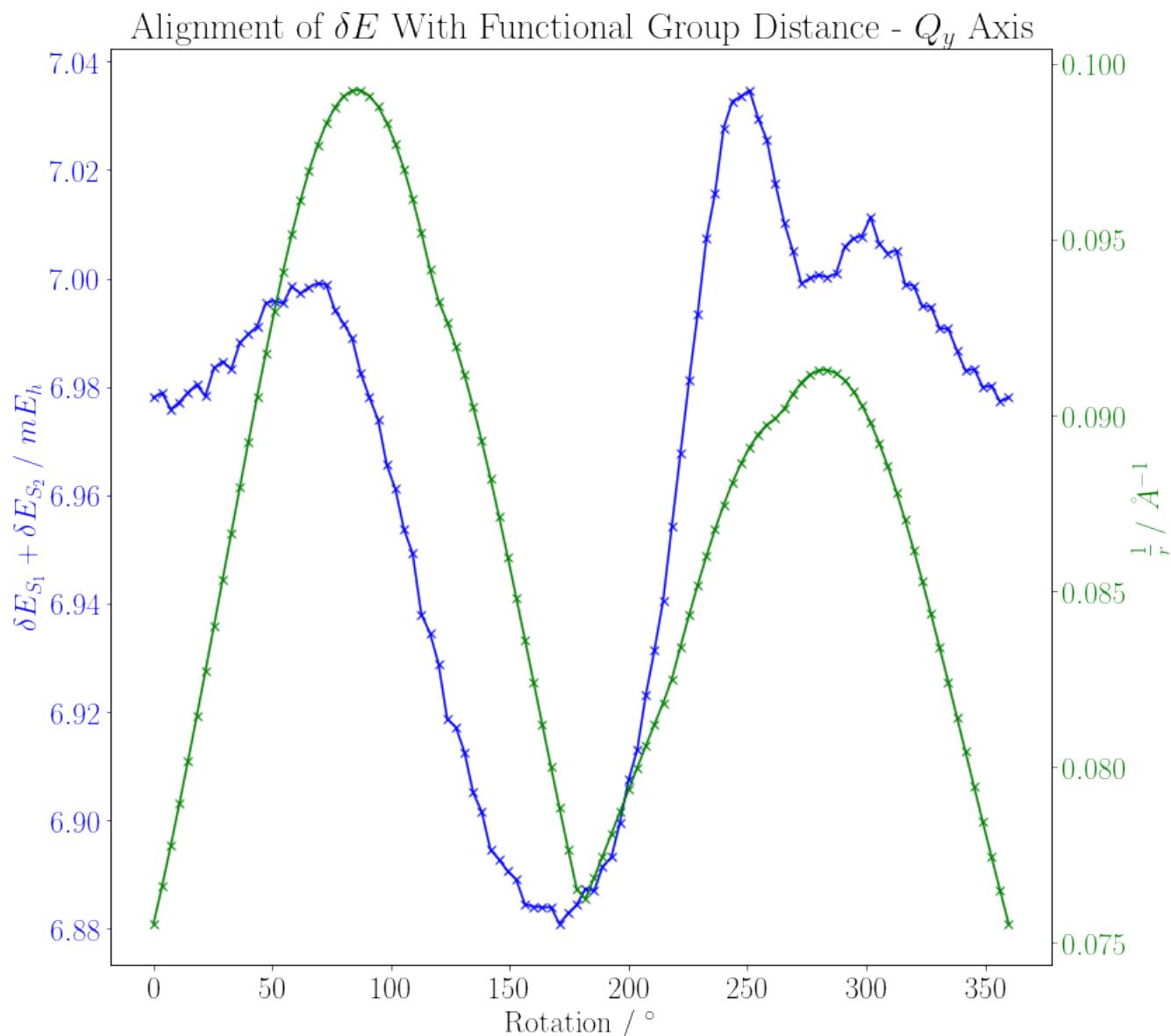


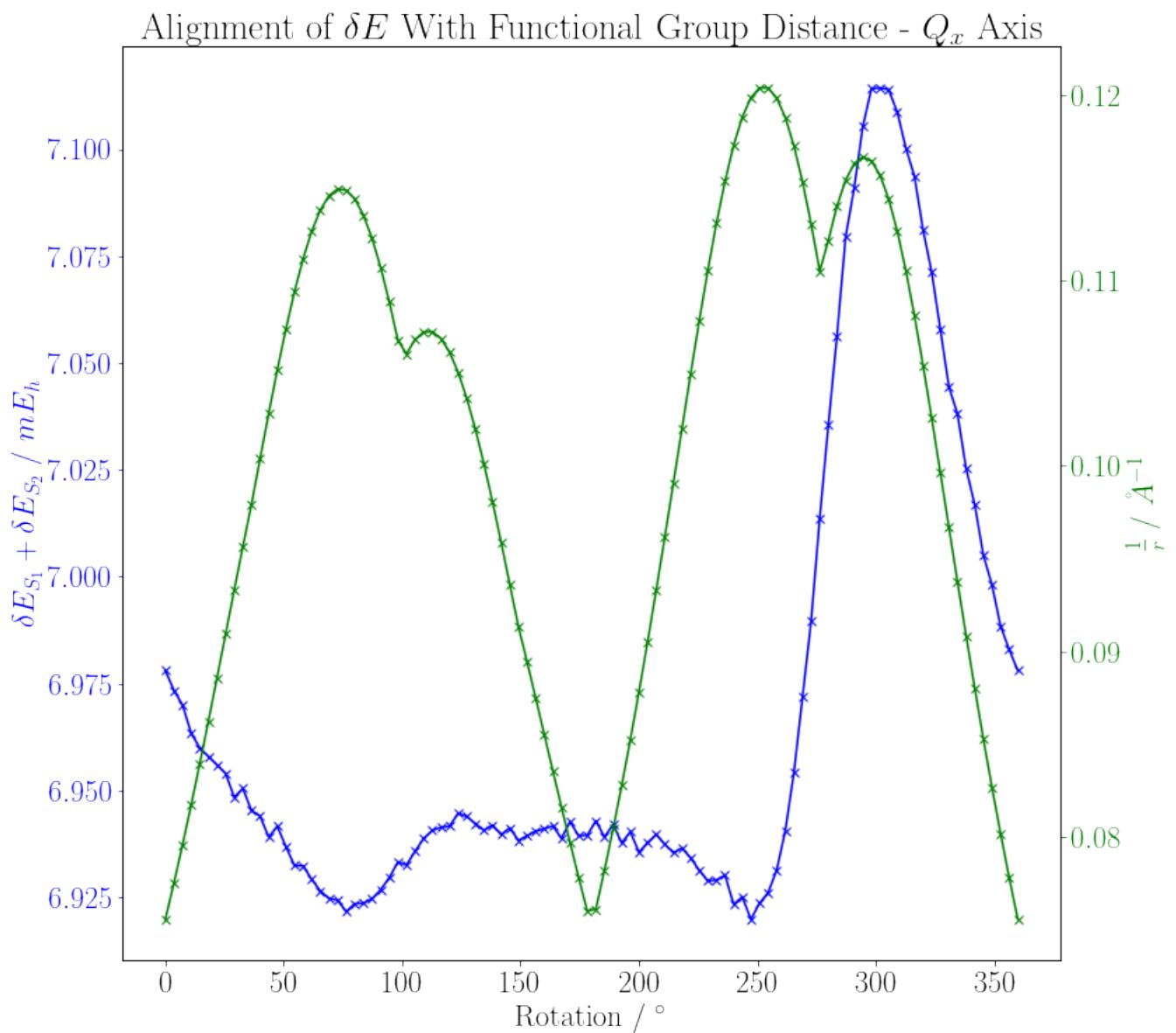
properties being different for intra-chromophore geometries, and just focuses on whether a non-varying chl-xTB transition can supply decent properties for a varying dimer system. One observation that showcased this difference was that the coupling values predicted by earlier version of chl-xTB were much larger than CAM-B3LYP. This was attributed to the much larger transition dipole magnitude that is observed in Δ -SCF and the eigenvalue difference methods of the previous chapter, and led to the inclusion of the transition density matrix scaling factor. Without removing the variation due to intra-chromophore geometry, this effect may not have been clear.

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

As the energy value of the excited states is very different between CAM-B3LYP and chl-xTB, a better comparison would be in the error of transition energies from the ground exciton state to the excited states, taken as the difference between the states. The profile of the sums of the absolute values of the error in these two transitions along the angle of rotation, alongside the reciprocal distance of the closest functional groups on the porphyrin ring is shown in fig. It can be clearly seen that in the cases of peaks in error in transition energy, there is a corresponding peak in the proximity of the functional groups. It would be expected that a better approximation of electron interaction at these distances, 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.







6.4 Concentration Quenching

An ideal use case for the workflow as discussed so far would be for a system where either the level of theory or volume of calculation would make both full dimer calculations as well as individual monomers too expensive. An example of this would be reconstructing potential energy surfaces (PESs) from distributions of energies where a systematic scan of a reaction coordinate is not possible. This is the case for calculating the rate constants of conversion between the excited state (ES) and charge separation (CS) state of a chlorophyll dimer, which is important for understanding the quenching mechanism observed in some chlorophyll systems. This rate constant is dependent on the free energy change ΔA and reorganisation energy λ

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

where H_{AB} is the coupling value between the ES and CS states, k_B is the usual Boltzmann factor and T is the temperature. ΔA and λ are calculated from the minima of the ES and CS PESs. However as these surfaces are dependent on both the internal geometry of the dimers and the solvent system around it, it is not possible to reduce this space to simple reaction coordinates and perform systematic scans. Instead, the PES can be reconstructed using the distribution of energy values that it defines.

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

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

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

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

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

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

The k_{CS} rate was calculated for a series of chlorophyll dimers at a constant separation, solvated in diethyl ether. The rate was also calculated for a pair of chlorophylls embedded in the LHII protein. MD simulations were used to generate the thermally distributed geometries of the dimer systems. The diethyl ether systems were prepared using Packmol into 100 Å boxes, with forcefield parameters taken from OpenForceField version 1.3.0. For the solvent systems, an additional constraint was used to keep the dimer at a constant distance in order to properly

inspect the effect of chlorophyll distance on charge separation. These constraints were set to 8, 10, 12, 14 and 20 Å. Each system was equilibrated for 10 ps at 300 K with a Langevin Integrator, with a production simulation of 500 ps performed afterwards. Geometries were taken every 1 ps for 500 total geometries. The LH2 geometries were calculated using the same MD method but using the LH2 structure and forcefield developed by Cardoso Ramos *et al.*.

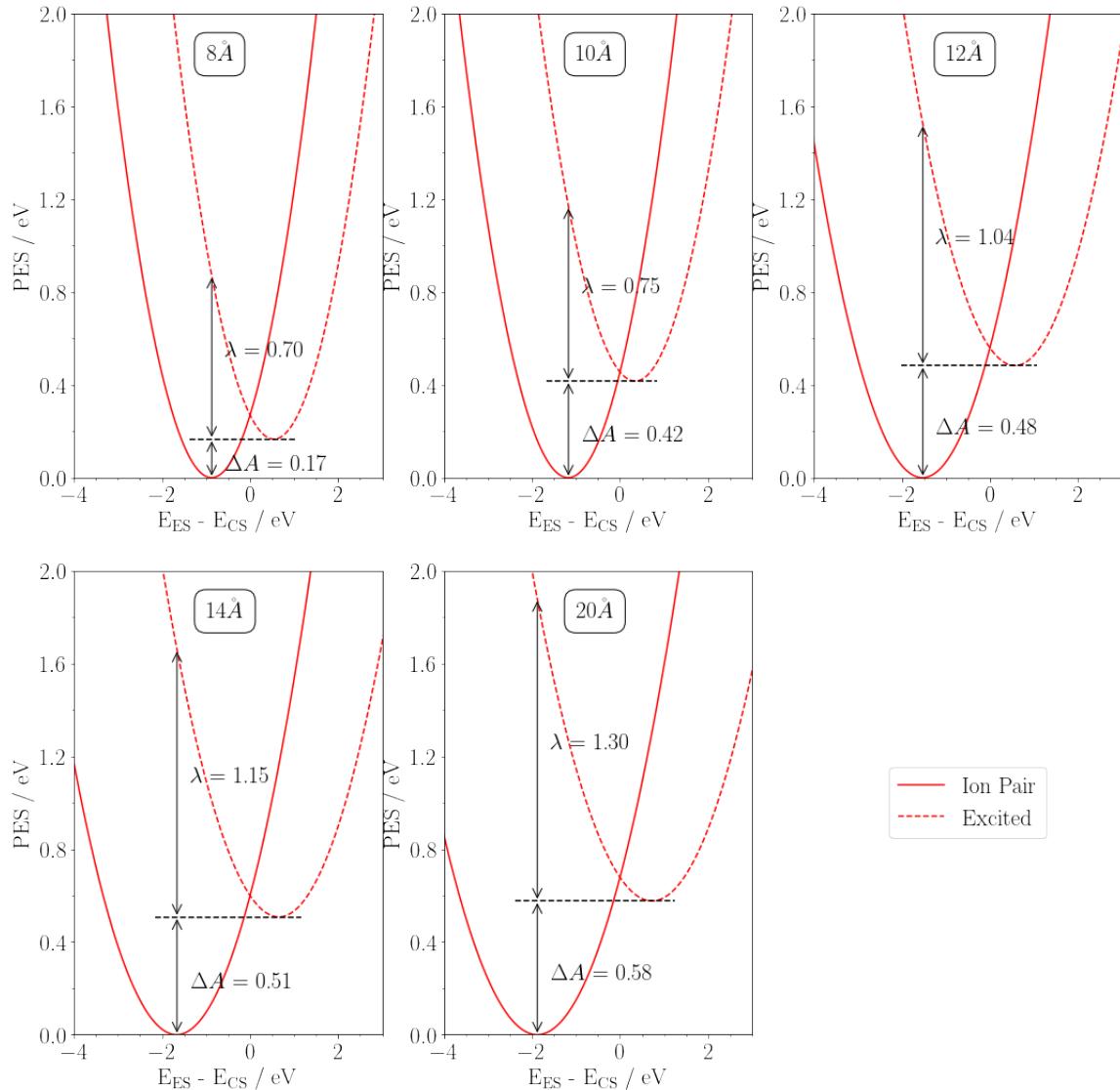
Dimer excited states were calculated for all of the geometries, using the Frenkel exciton Hamiltonian and chl-xTB. The lowest energy exciton states were taken for the distribution to reconstruct the PES. To account for environmental effects, the solvent and other chlorophyll(s) were included as point charges (with charge values taken from the forcefields used) to polarize the chl-xTB calculations. Fitting normal distribution values μ and σ was done with the `scipy.stats` module. Good fits were achieved with this method, justifying the assumption that the PESs follow a quadratic function.

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

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

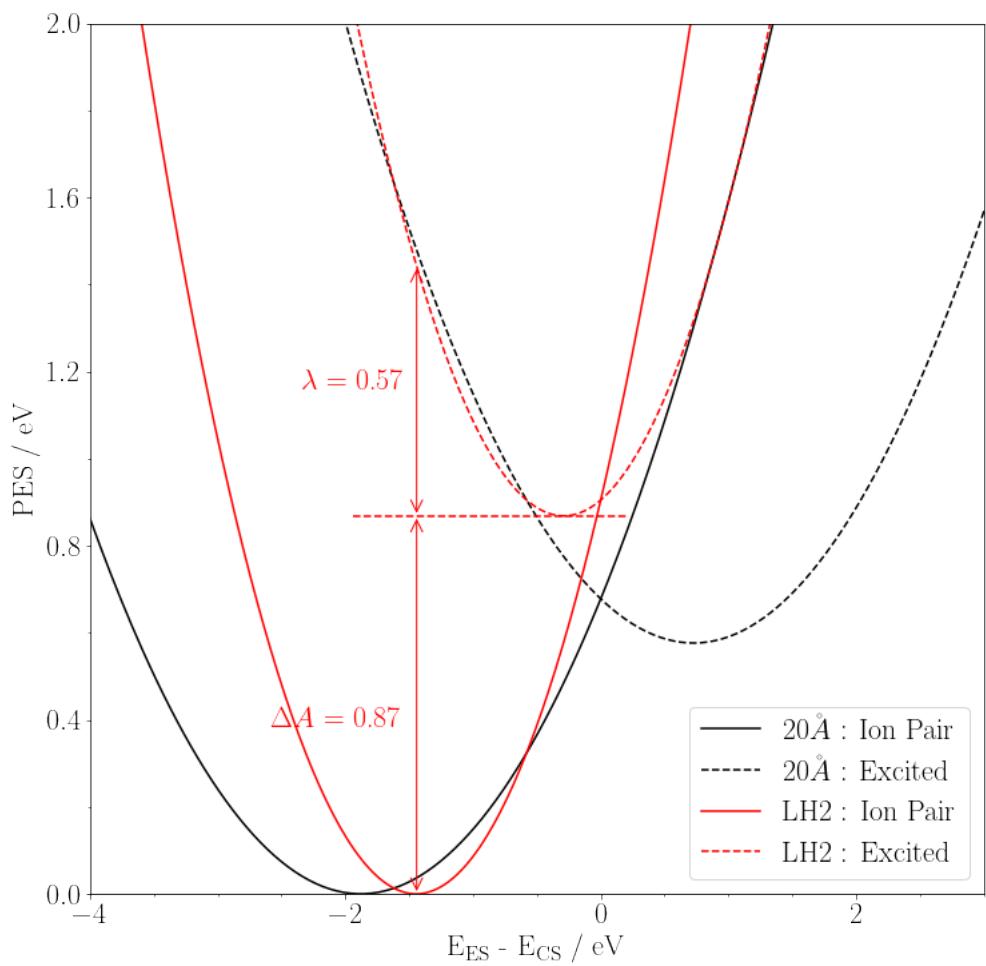
The rate constant also decreases with the distance between the chlorophyll dimers, as can be seen in the λ and ΔA values for 8 Å, 10 Å, 12 Å, 14 Å and 20 Å. As the systems get more separated both ΔA and λ get greater, exponentially reducing the rate constant. This is in contrary to the previously held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

Comparing the constrained systems to the LH2 system reveals how the protein structure biases the rate constant to be even lower. The average separation of chlorophylls in the LH2 system is around 9 Å, which is in the region where transition to the CS state is most competitive. This is in contrast to the "better designed" FMO complex, where the average separation of 12 Å would drastically reduce the quenching pathway. However, as can be seen in fig., the free energy change is much higher for LH2 than for even the 20 Å, attributed to the destabilisation of



the CS state. Additionally, the crossing point for the electron transfer is to the right of the CS state minima, implying that any transition from ES to CS would quickly reverse. This illustrates how the LH2 protein actively reduces the rate constant for quenching which would otherwise be more present at such close separations.

6.5 Conclusions



LIGHT HARVESTING COMPLEXES

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

7.1 LH2

In the introduction it was stated that spectral densities of individual sites can be calculated with an explicit electronic treatment for each frame of a time series.

For larger systems, such as the 27-site exciton system in LH2, the increased volume of calculation necessary mean that approximations have to be made that averages over intra-site properties, reducing the level of detail. Due to the efficiency of the chl-xTB exciton method, it was possible to calculate the transition properties for each site for each frame of a time series, and construct and diagonalise the Frenkel exciton Hamiltonian.

This chapter tries to answer two questions about environmental frequency coupling to the exciton states. The first is whether the presence of low frequency modes in the spectral density, which would be indicative of a large scale motion of the protein scaffold. It can be seen on AFM images of the LH2 protein that there are distinct deformations of the circular symmetry, which would indicate a ring- breathing mode in the chlorophyll site positions. It is not fully understood whether this would have an effect on the exciton states. Second would be by how much is the exciton state spectral density different to that of the sites. This would be an indicator of whether the protein structure promotes some special interaction between sites, or if the protein

only affects intra-chromophore geometry. Whilst these two questions only probe vague details of the protein structure, there are other conclusions possible that would be useful for further studies. The amount of intra- verses inter-site difference would inform how much detail is needed for future studies on LH2. It would also help in the design of artificial systems, which is discussed later.

7.1.1 Spectral Density Method

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

There are several definitions of spectral density. This chapter uses the definitions laid out by Mallus *et al.*. This describes the spectral density as the real part of the Fourier transform of the classical autocorrelation function. The autocorrelation here is the deviation of transition energies from the mean correlated with itself

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

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

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

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

In practice, the spectral density is usually calculated from a discrete series of a property calculated from the frames of an MD simulation. Hence the autocorrelation and Fourier transforms are also discrete. This places a limit on the frequencies that the spectral density can cover, proportional to both the length of the MD for the low frequency limit and the interval between frames for the high frequency limit. The fastest vibrations would be due to hydrogen atoms with a timescale of about, giving the timestep interval to be about , which is in line with other similar studies. The lower frequency limit is an open ended limit, dependent on how many frames it is possible to calculate the exciton model for. This is discussed in more detail later.

7.2 Calculating LH2 Excitons

The time series of exciton states was calculated from a series of structures of LH2 generated by molecular dynamics. These simulations were done with the OpenMM package. The forcefield and geometry was the same as used by *et al.*. For each simulation, the system was equilibrated for 60

ps. The production workflow length was 300 ps, with frames taken every 2 fs similar to the work done by *et al.*. A Langevin integrator set at 300 K was used with a timestep of 2 fs. Non-bonded interactions were treated with a PME method.

Exciton states were calculated with the same method as discussed in the previous chapter, extending beyond the dimer system to include all chlorophyll sites. The elements in exciton Hamiltonian were calculated using chl-xTB properties. A highly parallelised version of the code used to calculate chl-xTB properties was implemented in order to achieve good scaling. As machines with 20 cores were readily accessible, it was found that running sites as a serial chl-xTB calculation whilst inside a parent program that setup and collected the chl-xTB results, gave the best performance.

The main bottleneck in calculating these states was the volume of chl-xTB properties needed, as it was found construction and diagonalisation of the exciton Hamiltonian was negligible. The 300 ps MD generated 150,000 individual frames, requiring 4,050,000 individual chl-xTB runs. The time for each chl-xTB calculation is \approx 1 second, so this presented about 46 days of CPU walltime. This was significantly reduced by using the highly parallelised program. 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 still be more efficient use of resources.

The coupling values obtained from the chl-xTB exciton framework match those of previous studies well.

7.2.1 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.

Point charge embedding was used in the previous chapter for LH2 dimers, with the non-periodic implementation utilising the open source `helpME` library. The CPU time of performing non-periodic point charge embedding was profiled using the LH2 MD frames. It was found that due to implementation problems with re-using splines and grid positions, the time for each chl-xTB calculation was increased from \approx 1 second to \approx 40 seconds. Additionally the PME routines required multiple cores in order to get the best possible performance, which competed with the site parallelisation of serial 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 as 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 *et al.*, where any point-charge interaction has a prefactor screening term that is dependent on the

distance between point charges. For example for a coupling term between two exciton states, a single element in the sum of chromophore-chromophore interactions would be given by

$$(7.3) \quad V_{mn} = \frac{f}{4\pi\epsilon_0} \sum_{i,j} \frac{\mathbf{q}_i^T \mathbf{q}_j^T}{R_{ij}}$$

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

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

where A , B and f_0 are constants.

With this screening factor implemented, it was found that the spectral density was not appreciably different, and the only effect was to decouple the B800 and B850 exciton states. This is best demonstrated in the simulated absorption spectra for LH2 with and without the screening factor, as well as the breakdown of exciton states into the density contributions from each site.

7.2.1.1 Absorption Spectra

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

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

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

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

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

where j, m 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

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

. As this intensity is dependent on the orientation of the unit vector ϵ , the average intensity can be found by either calculating the intensities for a large distribution of randomised unit vectors

and taking the average, or by taking the analytic spherical average. Both these methods were tested and give the same values, as shown in fig.

The simulated LH2 spectra with and without screening factors, alongside the experimental spectrum, are shown in fig. 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, and the best fit is achieved by using a more idealised method to calculate exciton transition energies.

7.2.1.2 Site Contributions to Exciton Density

The decoupling of ring structures is also seen in the density contributions of chlorophyll sites to overall exciton states. The eigenvector solutions contain the amount of site character in each state, with the square of this being equal to the exciton density on a specific site. For example, the second element in each eigenvector is the corresponding amount of character from an excitation on the first chlorophyll site (the first element corresponding to the ground state, with no excitation on any chlorophylls). The average of density values is shown in fig for both exciton states calculated with and without the screening factor, as well as the difference between them.

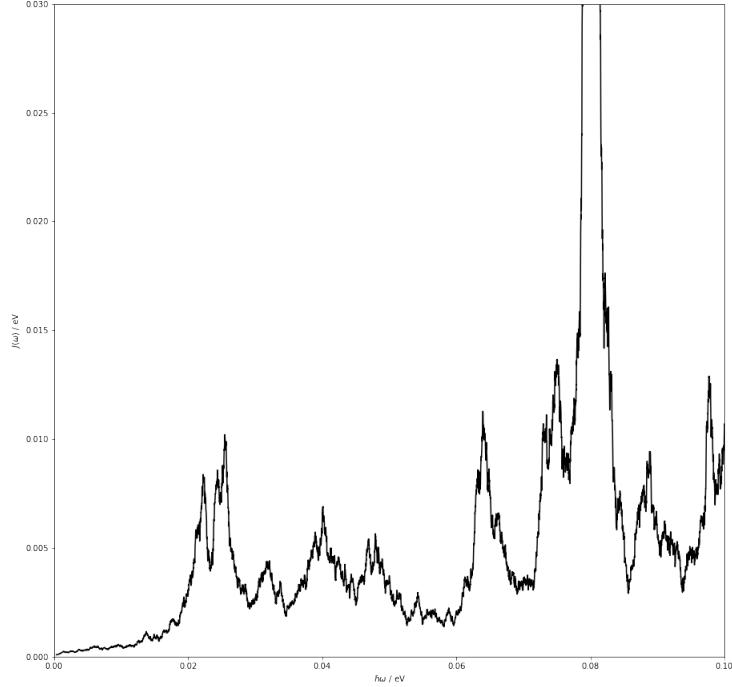
The main difference is the change in the amount of density shared between states localised on B800 and B850 sites. The plot of the density difference shows how density is localised more on B800 sites for exciton states calculated with the screening factor. This is also true for the B850 sites. However it does not reduce the delocalisation of intra-ring sites - the amount of density shared between B850 sites with B850 sites stays effectively the same. Even without the screening factor, the contribution of B800 and B850 sites to the same state was relatively low, and so the exclusion of the screening factor is argued to not be detrimental to the model.

Overall, whilst it was possible to include embedding effects into the chl-xTB exciton model, it may not change the overall qualitative behaviour of the exciton states. Calculating the chlorophyll system in a vacuum would still be a valid decision to make. The vacuum model has been effective in the past, and captures many of the important effects. This should also be true for the spectral density as the leading cause of exciton state energy variation would be in effect of intra-chlorophyll geometry on the coupling and transition energy.

7.3 Spectra

7.3.1 Sites

Similar to other studies, it was possible to produce a spectral density for Q_y transitions at individual sites. This is shown in figure , with the same scale as the spectral density by *et. al.* Similar features such as the peaks at eV are present in both. The discrepancy between both can be explained by the different forcefields and response methods used. It could then be expected



that the features found in the spectral density of chl-xTB exciton states to be reasonable as the individual sites show the same features as those reported before in the literature.

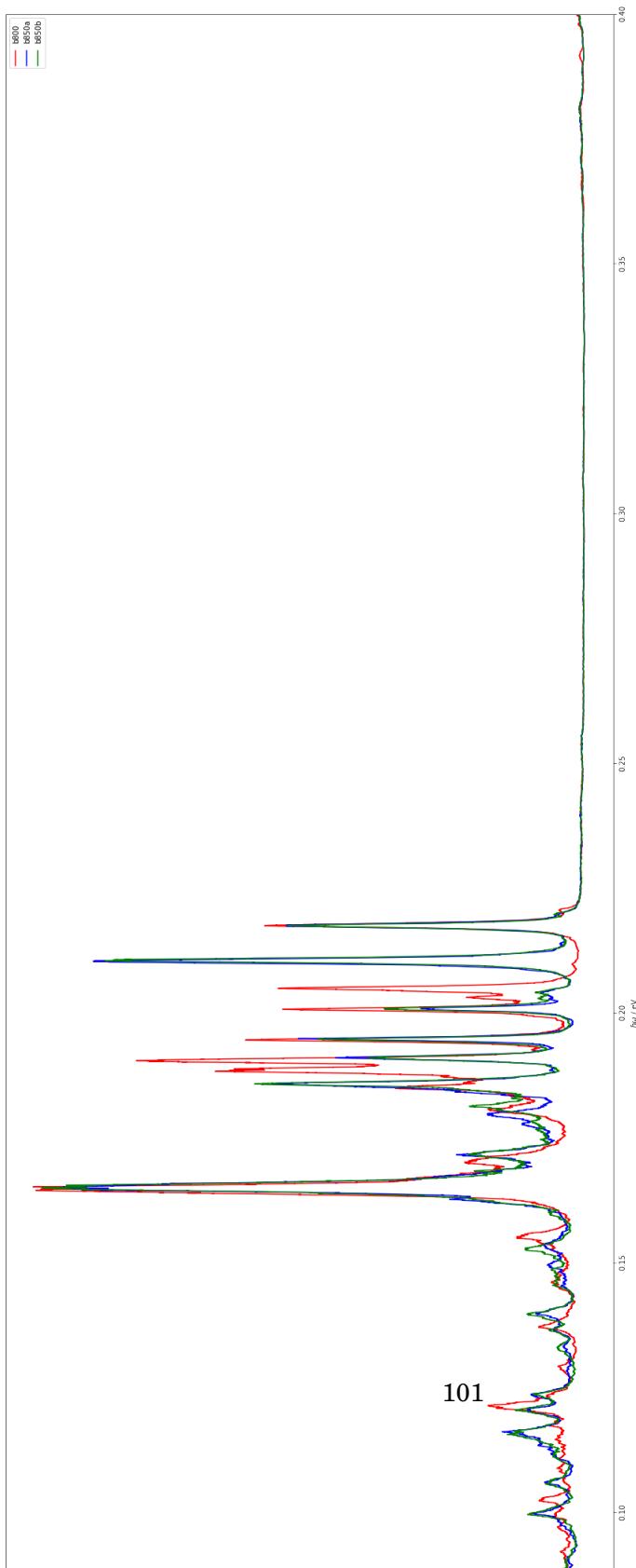
Expanding the frequency domain and averaging the spectra by the ring type clearly shows the effect of the protein scaffold at different sites. In line with previous studies, the region around 0.2 eV show the greatest features, and is attributed to intra-chromophore normal modes [26]. The difference between B800 and B850 lines also indicates different binding site environments, in line with other studies.

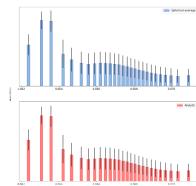
7.3.2 Exciton states

The exciton spectral density was calculated for the time series of the exciton transition energies (calculated as the difference between a state and the exciton ground state) of each state bar the ground state. The absorption probability was calculated for each exciton states at all MD frames. The time series of these probabilities is shown in fig. It can be seen that the 2nd and 3rd lowest energy exciton states have the highest probability of absorption.

It was then possible to plot the spectral densities of all states, weighted by their average absorption probability. The relative absorption probability for a state was found not to vary much, justifying taking the average over a large time span. This plot is shown in fig.

7.3. SPECTRA





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 and so mostly staying close to a mean value. This would reduce the fluctuation of exciton transition energies, which in turn would decrease the magnitude of any peak in the spectral density.

It can also be seen that major features are common in both the exciton and site spectra. This implies that the protein environment effect on the transition at a single chlorophyll site is much greater than any effect on the coupling between chlorophyll sites. Whilst not wholly surprising based on the lack of correlation between atomic positions in chlorophylls, it is interesting to note that there is not as much constraint on the coupling as there is for the different ring sites. Many of the exciton states are delocalised across several states, and so it would be imagined that the coupling that decides this delocalisation would be important to control. What is found here is that there is little consideration for the variations in coupling outside of variations due to inter-chlorophyll angle and distance, which stay relatively constant throughout the MD trajectory.

The absorption probability weightings offer some more notable observations. For the most part, exciton states with lower probabilities correspond to lower probabilities. This does give some credit to the idea that the environment does influence some states in order to achieve some difference properties. However as these are not outside of the features seen in the site energies it is hard to say this is borne from effects on coupling. Additionally, it is clear to see a lack of B800 features in the state spectra. This probably has common cause with the suppression of the B800 peak in simulated absorption spectra. The most notable phenomena is the outlying exciton state which, across the entire spectrum domain, has a much larger density than all other states. This would imply one exciton state in particular has much larger fluctuations in transition energy than others. This state is probably localised in the B850 ring by the absence of any B800 features at the highest frequency end of the spectrum.

There is also a lack of low frequency modes in the exciton state spectra, outside of the range of intra-chlorophyll motions. These frequencies would correspond with large scale movements of the protein scaffold, causing changes in inter-chlorophyll separations or angles. This would have a pronounced effect on the coupling in the exciton state. However the lack of these modes in the spectral density imply that either these motions are not present in LH2, or are on a timescale that is too long for the MD simulation. Reasons for the lack of appearance of these modes is discussed later.

7.4 Assigning Specific Motions

Reporting the results so far has been limited due to the lack of assignment of which specific motions are coupling to the environment. This section reports on some of the tests used to assign features in the spectral densities. This includes comparing the LH2 spectral densities to a spectral density of monomer chlorophyll embedded in diethyl ether, as well as a spectrum constructed from Huang-Rhys factors calculated with chl-xTB response properties and a GFN1-xTB hessian. An explanation of the lack of low frequency features is also given by an estimation of the force constant in the interchromophore separation.

7.4.1 Ether system

While environmental coupling differences for chlorophyll sites just in LH2 can be seen from the results so far, it is not clear how other environments couple to the Q_y transition. The spectral density of chlorophyll in diethyl ether was calculated to give some light on this question. The two extremes for possible answers would be that there is almost no variation in the spectral density, implying that the features in the spectral density are intrinsic to the chlorophyll geometry, or there is a large variation and so there is a large berth for different environmental coupling.

Geometries were taken from an MD simulation of a diethyl-ether embedded chlorophyll, using a solvent box made with the `packmol` program with a single chlorophyll molecule in a 64 Å box

with 1054 diethyl-ether molecules. The simulation was run with OpenMM using parameters for chlorophyll taken from the LH2 forcefield, and parameters for diethyl ether taken from the General Amber ForceField (GAFF). The system was equilibrated for 60 ps, with a production workflow of 300 ps run afterwards. Structures were taken every 2 fs. A Langevin integrator set at 300 K was used, with a timestep of 2 fs.

The Q_y transition was calculated for every frame of the MD trajectory. The spectral density of these transition energies was calculated with equations . A plot of this can be seen in fig.

It can be seen that the features in the spectral density for diethyl-ether embedded chlorophyll are similar to the features from the LH2 single chlorophyll and exciton states. This implies that the coupling of environment to energy fluctuations is due to mostly intrinsic properties of the chlorophyll geometry. With this in mind, a more detailed view of the normal modes in chlorophyll might help assignment of spectral density features. This is the focus of the next section.

7.4.2 Huang Rhys Factors

In the sections discussing LH2 and diethyl-ether environmental coupling, it is argued that features in the spectral density are mainly due to intrinsic properties of the chlorophyll geometry. These properties would most likely be the coupling of internal motions of chlorophyll to the Q_y transition. A larger coupling of motion would imply greater variation in the Q_y transition energy, which would cause larger features in the spectral density. The coupling of normal modes to electron transition can be calculated with Huang-Rhys factors, defined by the difference between minima in the excited and ground state energy surfaces along normal mode coordinates. The proposed idea was that modes with high Huang-Rhys factors may correspond to features in the spectral density, and this correspondence might be observed by comparing a spectrum constructed from Huang-Rhys factors. This would require calculating these factors for all normal modes.

The normal modes used to calculate Huang-Rhys factors were calculated with a hessian calculation on an optimised single chlorophyll structure. In trying to obtain an optimised geometry for chlorophyll, it was found that rotations in the phytol tail were too unconstrained for a converged geometry to be achieved. Instead a truncated chlorophyll with a hydrogen atom replacing the phytol group was used again. The optimised geometry was then used to calculate normal modes with GFN1-xTB.

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

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

where ω_i is the angular frequency of the normal mode i and x_i^m is the displacement vector in mass weighted coordinates. The Q_y transition energy was calculated for a series of structures

with successive values of q_i . Fits of the ground state and excited state energies were made with quadratic functions, from which it was possible to make estimates of the q value for a minimum ground state energy (q_{ground}) and excited state energy (q_{excited}).

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

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

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

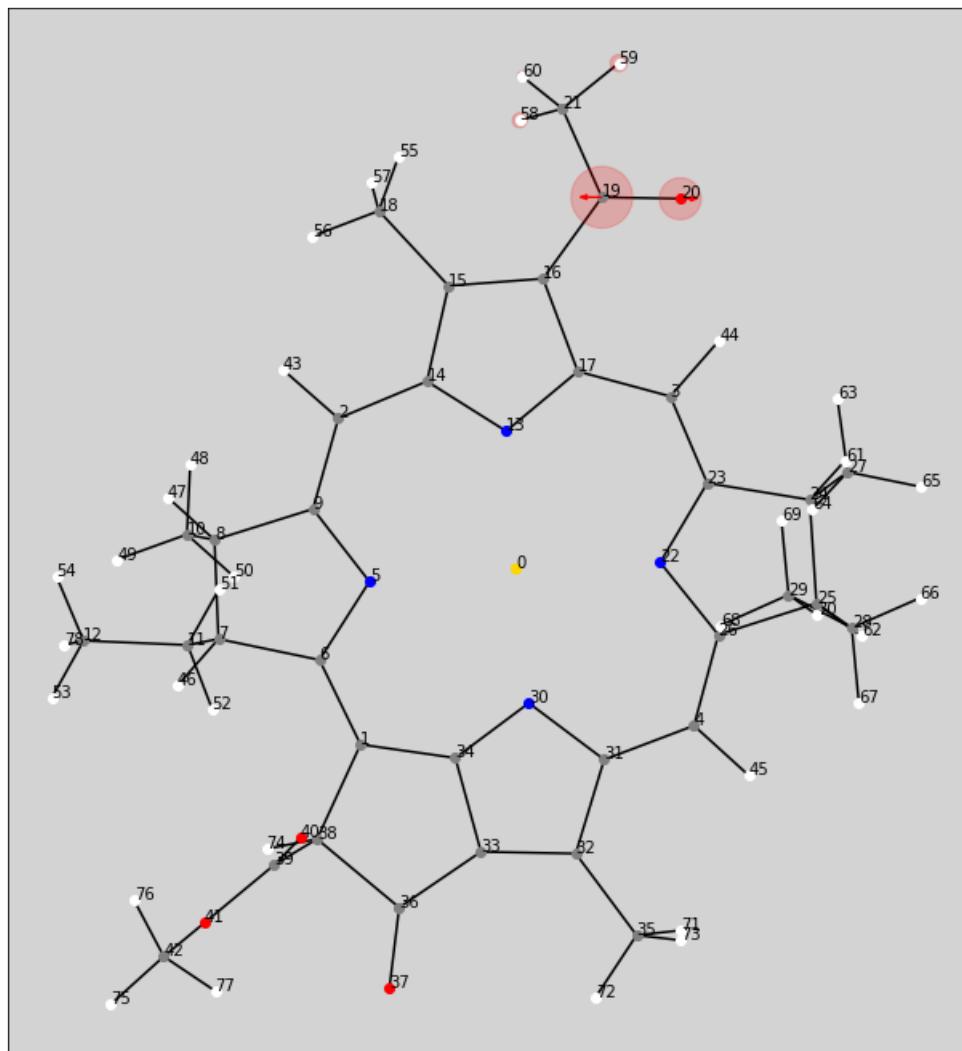
At the low frequency region it is fairly clear that these modes are suppressed in LH2 and ether. It could be argued that this is due to the high force constants, and so would not be expected to be observable motions. Without further investigation it is hard to say for what reasons these normal modes are not present but it is clear that they do not explain any features in the spectral densities.

The high frequency normal modes correspondence to spectral density features is less clear. Whilst two peaks in the Huang-Rhys spectrum are similar to the full spectral density, these may just be coincidence. It is argued that it is more likely that whole sections of the normal modes are shifted from the frequencies observed in the spectral densities, due to the differences in the forcefield and GFN1-xTB methods. Whilst full assignment of the spectral density features are outside the scope of this work, it can be clearly seen that with a method of producing hessians from forcefield parameters it would be possible to make a more compelling comparison of Huang-Rhys factors and the spectral density.

7.4.2.1 Estimating Force Constants for Low Frequency Modes

While there is evidence for the protein scaffold affecting chlorophyll motion and subsequent Q_y transition and coupling, there is little evidence for any effects outside of intra-chlorophyll range.

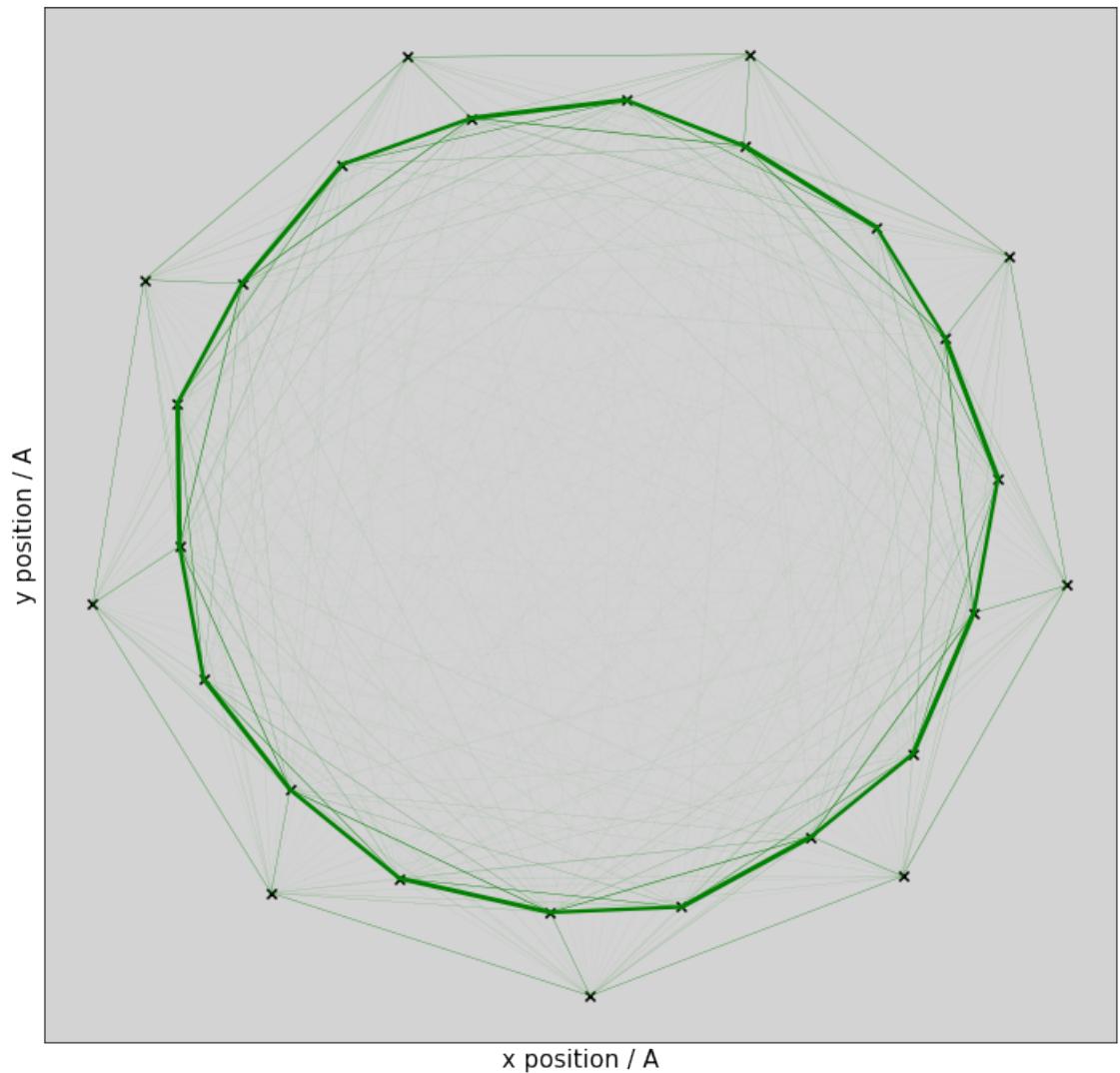
Treating the chlorophyll sites as nodes on a circular spring, it was thought that the ring-breathing vibrational modes of this circular spring would cause features in the spectral density of the exciton sites. This would be due to the change in separation of chlorophyll sites, which would have a large effect on the coupling values. The lack of these ring-breathing modes could be due to the setup of the experiment, and that with a different study these features might be present. The first concern in the setup is that the crystal structure used for LH2 is a low energy structure, and so there would be little potential energy in the ring-breathing modes to start any observable motions. To counter-act this, ring-breathing motion could be promoted by additional force constants, but this would introduce artificial features into the spectral density. Additionally it could be that heating the system could induce these motions more organically, however this would also affect the spectral density as the whole thermal environment would be changed.



This first explanation assumes that the ring breathing modes are not accessible from the starting geometry, which may not have been the case. Alternatively, it could be that the timescale of these motions is much larger than the resolution possible by the length of the MD used here. A rough estimate of this could be made by looking at the force-constant of separating two chlorophylls. The force constant is proportional to the gradient of the energy surface, which can be constructed from the energy of a dimer pair at various separations. This was done, and a quadratic fit was used to give estimates of the gradient at specific points. The plot in fig shows the dimer energies and the quadratic fitting.

With this model the force constant was calculated to be around , which gives a oscillation period of . This is far beyond what could have been observed in the MD trajectory used in this chapter. However there are several issues with this argument. First is that it does not include the protein scaffold, which would have a much larger effect on interchromophore separation than the potential energy of the chlorophylls. Second is that from the positions of chlorophyll separation on the PES in fig, it can be seen that the separations in LH2 are well outside of the region where a quadratic fit is reasonable.

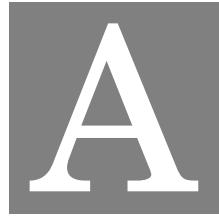
7.5 conclusions



DISCUSSION

Preamble

8.1 Transition Property Approximations**8.2 Further Investigations into LHII****8.3 Coherence**



APPENDIX A

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

A.1 Electronic Structure Codes

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

A.2 Computational Hardware

BIBLIOGRAPHY

- [1] *mcodev31/libmsym: molecular point group symmetry lib.*
- [2] C. ADAMO AND V. BARONE, *Towards reliable density functional methods without adjustable parameters: The PBE0 model*, J. Chem. Phys1, 110 (1999), p. 6158.
- [3] C. BANNWARTH, C. EIKE, S. EHLERT, A. HANSEN, P. PRACHT, J. SEIBERT, S. SPICHER, AND S. GRIMME, *Extended tight-binding quantum chemistry methods*, WIREs Comput. Mol. Sci., 11 (2020), p. 1493.
- [4] N. A. BESLEY, A. T. B. GILBERT, AND P. M. W. GILL, *Self-consistent-field calculations of core excited states*, J. Chem. Phys, 130 (2009), p. 124308.
- [5] Y. CHO, S. J. BINTRIM, AND T. C. BERKELBACH, *A simplified GW/BSE approach for charged and neutral excitation energies of large molecules and nanomaterials*, (2021).
- [6] 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.
- [7] J. GASTEIGER AND M. MARSILI, *A new model for calculating atomic charges in molecules*, Tetrahedron Lett., 19 (1978), pp. 3181–3184.
- [8] J. GAVNHOLT, T. OLSEN, M. ENGELUND, AND J. SCHIØTZ, *self-consistent field method to obtain potential energy surfaces of excited molecules on surfaces*, Phys. Rev. B, 78 (2008), p. 075441.
- [9] A. T. B. GILBERT, N. A. BESLEY, AND P. M. W. GILL, *Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM) †*, J. Phys. Chem., 112 (2008), p. 13164.
- [10] T. GIMON, A. IPATOV, A. HESSELMANN, AND A. GÖRLING, *Qualitatively Correct Charge-Transfer Excitation Energies in HeH+ by Time-Dependent Density-Functional Theory Due to Exact Exchange Kohn-Sham Eigenvalue Differences*, J. Chem. Theory Comput, 5 (2009), p. 27.

BIBLIOGRAPHY

- [11] S. GRIMME, *A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules*, J. Chem. Phys., 138 (2013), p. 244104.
- [12] S. GRIMME AND C. BANNWARTH, *Supporting information for: Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)*, tech. rep.
- [13] ———, *Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)*, J. Chem. Phys., 145 (2016), p. 54103.
- [14] S. GRIMME, C. BANNWARTH, AND P. 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 = 1-86)*, J. Chem. Theory Comput., 13 (2017), pp. 1989–2009.
- [15] C. HÄTTIG AND F. WEIGEND, *CC2 excitation energy calculations on large molecules using the resolution of the identity approximation*, J. Chem. Phys., 113 (2000), p. 5154.
- [16] A. HELLWEG, S. A. G. GRUÑN, AND C. H. HÄTTIG, *Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states*, Phys. Chem. Chem. Phys., 10 (2008), pp. 4119–4127.
- [17] S. W. J. HUNT AND W. A. GODDARD, *Excited States of H₂O using improved virtual orbitals*, Chem. Phys. Lett., 3 (1969), pp. 414–418.
- [18] J. KIM, J. JEON, T. HYUN YOON, AND M. CHO, *Two-dimensional electronic spectroscopy of bacteriochlorophyll a with synchronized dual mode-locked lasers*, Nat. Commun., 11 (2020), p. 6029.
- [19] G. KLOPMAN, V. 86, J. LINEVSKY, K. S. SESHADEVI, AND D. WHITE, *A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules*, J. Am. Chem. Soc., 86 (1964), pp. 4550–4557.
- [20] T. KOWALCZYK, S. R. YOST, AND T. VAN VOORHIS, *Assessment of the D-SCF density functional theory approach for electronic excitations in organic dyes*, J. Chem. Phys., 134 (2011), p. 164101.
- [21] T. LIU, W.-G. HAN, F. HIMO, . G. M. ULLMANN, D. BASHFORD, A. TOUTCHKINE, K. M. HAHN, AND L. NOODLEMAN, *Density Functional Vertical Self-Consistent Reaction Field Theory for Solvatochromism Studies of Solvent-Sensitive Dyes*, J. Phys. Chem. A, 108 (2004), pp. 3545–3555.

- [22] P. O. LÖWDIN, *Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction*, Phys. Rev., 97 (1955), pp. 1474–1489.
- [23] A. V. MARENICH, S. V. JEROME, C. J. CRAMER, AND D. G. TRUHLAR, *Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases*, (2012).
- [24] K. NISHIMOTO AND N. MATAGA, *Electronic Structure and Spectra of Some Nitrogen Heterocycles*, Zeitshrift fur Phys. Chemie Neue Folge, 12 (1957), pp. 335–338.
- [25] K. OHNO, *Some Remarks on the Pariser-Parr-Pople Method*, Theor. chim. Acta, 2 (1964), pp. 219–227.
- [26] C. OLBRICH AND U. KLEINEKATHÖFER, *Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II*, J. Phys. Chem. B, 114 (2010), p. 12427.
- [27] C. C. J. ROOTHAAN, *New Developments in Molecular Orbital Theory*, Rev. Mod. Phys., 23 (1951), p. 69.
- [28] A. SCHÄFER, H. HORN, AND R. AHLRICHS, *Fully optimized contracted Gaussian basis sets for atoms Li to Kr*, J. Chem. Phys, 97 (1992), p. 2571.
- [29] H. STRAIN, M. THOMAS, AND J. KATZ, *Spectral Absorption Properties of Ordinary and Fully Deuteriated Chlorophylls a and b*, Biochem. Biophys. Acta., 75 (1963), pp. 306–311.
- [30] C. STROSS, M. W. VAN DER KAMP, T. A. OLIVER, J. N. HARVEY, N. LINDEN, AND F. R. MANBY, *How static disorder mimics decoherence in anisotropy pump-probe experiments on purple-bacteria light harvesting complexes*, J. Phys. Chem. B, 120 (2016), pp. 11449–11463.
- [31] S. B. WORSTER, O. FEIGHAN, AND F. R. MANBY, *Reliable transition properties from excited-state mean-field calculations*, J. Chem. Phys, 154 (2021), p. 124106.
- [32] L. ZHANG, D.-A. SILVA, Y. YAN, AND X. HUANG, *Force Field Development for Cofactors in the Photosystem II*, J. Comput. Chem., 33 (2012), pp. 1969–1980.
- [33] T. ZIEGLER, A. RANK, AND E. J. BAERENDS, *On the Calculation of Multiplet Energies by the Hartree-Fock-Slater Method*, Acta, 43 (1977), pp. 261–271.

