
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 Introduction	1
1.1 Quantum Exploits in Light Harvesting Systems	1
1.1.1 Electronic Energy Transfer	1
1.1.2 Coherence	1
1.2 Light-Matter Response	1
1.3 Electronic Structure for Large Systems	1
1.4 The Aim	1
2 Mean-Field excited states	3
2.1 Theory	4
2.1.1 Δ -SCF	4
2.1.2 Eigenvalue Difference	5
2.1.3 Transition Density and Dipole Moments	5
2.1.4 Semi-empirical extensions	6
2.2 Benchmarking	8
2.2.1 Reference Data and test set	9
2.2.2 Small Systems	9
2.2.3 Non-orthogonality	12
2.2.4 LHII Chlorophyll	14
2.2.5 xTB methods	17
2.3 Conclusions	23
3 Chlorophyll specific methods	25
3.1 Theory	25
3.1.1 Approximations to Solutions	26
3.1.2 Integrals approximations	26

TABLE OF CONTENTS

3.1.3	Q_y Transition	28
3.1.4	chl-xTB method	28
3.2	Parameterization	31
3.2.1	Reference Data	31
3.2.2	Objective Function	32
3.2.3	Minimisation Algorithms	33
3.2.4	Results	39
3.3	Benchmarking and Cross-validation	39
3.3.1	Transition properties	39
3.3.2	Potential Energy Surfaces	39
3.3.3	Absorption Spectra	41
4	Exciton Method	55
4.1	Theory	55
4.1.1	Exciton States	55
4.1.2	Embedding	55
4.2	Truncated Chlorophylls	55
4.2.1	Rotation	55
4.2.2	Distance	55
4.3	LHII Pairs	55
4.3.1	Assignment of States	55
4.3.2	Comparison	55
5	Atomistic Modelling of Light Harvesting Complexes	61
5.1	LHII	69
5.1.1	Spectral Density Method	69
5.1.2	Molecular Dynamics Method	69
5.2	Approximating Spectral Densities	69
5.2.1	Hessians	69
5.2.2	Huang Rhys Factors	69
5.2.3	Chlorophyll distances	69
5.3	Environmental Effects	69
5.3.1	Screening	69
5.3.2	Embedding	69
5.4	Sites, states and couplings	69
5.4.1	Sites	69
5.4.2	Exciton states	69
5.4.3	Coupling	69
5.5	Excitation Energies	69

TABLE OF CONTENTS

6 Discussion	71
6.1 Transition Property Approximations	71
6.2 Further Investigations into LHII	71
6.3 Coherence	71
A Appendix A	73
A.1 Electronic Structure Codes	73
A.2 Computational Hardware	73
Bibliography	75

LIST OF TABLES

TABLE	Page
2.1 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.	22
3.1 optimized parameters from SLSQP procedure.	39

LIST OF FIGURES

FIGURE	Page
2.1 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.	10
2.2 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.	11
2.3 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.	13
2.4 The transition energies from Δ -SCF for the 26 chlorophyll geometries, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line.	15
2.5 The transition dipole magnitudes from Δ -SCF for the 26 chlorophyll geometries, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line.	16
2.6 A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.	19
2.7 The distributions of errors to CC2 transition energies for the methods included in the Δ -xTB benchmarking.	21
3.1 correlations of energies	31
3.2 correlations of transition dipole moments	32
3.3 Nelder-mead	37
3.4 SLSQP	38
3.5 fig:training scatter	40
4.1 caption	56
4.2 caption	57

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.

1.1 Quantum Exploits in Light Harvesting Systems

Begins a section.

1.1.1 Electronic Energy Transfer

Begins a subsection.

1.1.2 Coherence

1.2 Light-Matter Response

1.3 Electronic Structure for Large Systems

1.4 The Aim

MEAN-FIELD EXCITED STATES

Previous Published Work

The work presented in this chapter is based on joint work with Dr Susannah Bourne-Worster, and was published in March 2021[26]. The account given in this chapter includes research that was also reported in this publication, namely the parts of section 2.2 from 2.2.1 up to but excluding 2.2.5.

This chapter collates all of the work done on investigating and benchmarking Δ -SCF methods, at 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 were also investigated for the mean-field Δ -SCF method, as well as assignment of transitions based on symmetry. It was found that while the DFT methods give reasonable results, the semi-empirical Δ -SCF methods are not as accurate. This result is unexpected in the context of the benchmarking, however has reasonable explanations, and will guide the work of the later chapter on developing a new semi-empirical method.

2.1 Theory

2.1.1 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:

$$(2.1) \quad \Delta E = E_2 - E_1$$

Finding the excited state correctly can be an issue, and for ease is usually assumed to be similar to the ground state. In its simplest form, the Δ -SCF method calculates the ground state with normal DFT or other mean-field methods, and then calculates the excited state by rerunning the same method with the excited state occupation numbers. A full description of both the ground as excited state is given as the orbital coefficients output from the two calculations.

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[15]. However, it was argued that this procedure would exacerbate errors from finding the ground state, and that the excited state was not a proper SCF solution[8]. 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. Each iteration in an SCF procedure produces new molecular orbital coefficients by solving the Roothaan-Hall equations[23], generally given as an eigenvalue problem:

$$(2.2) \quad \mathbf{FC}^{\text{new}} = \mathbf{SC}^{\text{new}}\epsilon$$

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

$$(2.3) \quad \mathbf{F} = f(\mathbf{C}^{\text{old}})$$

The amount of similarity of orbitals can be estimated from their overlap:

$$(2.4) \quad \mathbf{O} = (\mathbf{C}^{\text{old}})^{\dagger} \mathbf{SC}^{\text{new}}$$

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

$$(2.5) \quad p_j = \sum_i^n O_{ij} = \sum_v^N \left[\sum_\mu^N \left(\sum_i^n C_{i\mu}^{\text{old}} \right) S_{\mu v} \right] C_{vj}^{\text{new}}$$

where μ, ν are orbital indices. The population can then be given by the set of orbitals with the highest projection p_j . 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[8]. The method of using this orbital overlap is called the maximum overlap method (MOM)[8].

Δ -SCF has been shown to be cheap alternative to TD-DFT and other higher level methods[4, 7, 18], without considerable losses of accuracy in certain cases, especially for HOMO-LUMO transitions[17]. 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, and so would be better for any dynamic simulation[7].

2.1.2 Eigenvalue Difference

Another approximation to full response theory would be to 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 are set to zero. This gives the difference between the excited state energy and the ground state energy, the transition energy, as just the difference of the orbital energies between the orbital an electron has been excited (ϵ_e) to and the orbital has been excited from (ϵ_g):

$$(2.6) \quad \Delta E = \epsilon_e - \epsilon_g$$

Additionally, transition properties can be calculated by construction transition density matrices from only the ground state solution, needing only a single SCF optimization. Generally, eigenvalue difference methods are not seen as accurate response methods, but can offer a quick and easy initial value[9].

2.1.3 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 **D**²¹ can be written as:

$$(2.7) \quad \mathbf{D}^{21} = |\Psi_1\rangle\langle\Psi_2|$$

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 given by

$$(2.8) \quad \mathbf{D}^{21} = \mathbf{C}^{(2)} \text{adj}(\mathbf{S}^{21}) \mathbf{C}^{(1)\dagger}$$

where \mathbf{S}^{mn} is the overlap of the two states. The adjunct of the overlap correspond to Löwdin's normal rules for non-orthogonal determinants [19]. For transition dipole elements, this is:

$$(2.9) \quad \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle = \sum_{jk} \mu_{jk}^{21} \text{adj}(\mathbf{S}^{21})_{jk}$$

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 this overlap, the orbital inner products, can be defined as the inner product of the two states:

$$(2.10) \quad |\mathbf{S}^{21}| = \langle \Psi_2 | \Psi_1 \rangle$$

The general definition of the transition dipole

$$(2.11) \quad \mu^{1 \rightarrow 2} = \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle$$

can be expressed with this transition density matrix as:

$$(2.12) \quad \begin{aligned} \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle &= \text{tr}(\hat{\mu} | \Psi_1 \rangle \langle \Psi_2 |) \\ &= \text{tr}(\hat{\mu} \mathbf{D}^{21}) \end{aligned}$$

2.1.4 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 [12]. 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 acronym for "Geometries, Frequencies, Non-Covalent - eXtended Tight Binding". This method was chosen for two reasons. The first being that the GFN-xTB method was already implemented in the QCORE package. This significantly reduced the amount of effort required for this project. Additionally, there would be other users and developers who would help with implementation of this new method. Second is that a similar method has already been published that calculates transition properties, the precursor to the GFN-xTB methods. This is the sTDA-xTB method [11].

2.1.4.1 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. The accuracy in calculating transition energies with this method is very good, with the error to high-level results being around 0.3 - 0.5 eV.

Similar to other xTB methods, the sTDA-xTB method is a tight-binding method that uses empirically fit 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 in 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.

The two basis sets are used to construct formally similar Fock matrix elements, however 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:

$$(2.13) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left(k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left(h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

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 is the overlap, 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 CM5 [20] 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 [6], 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 the previous work in the Grimme group on the simplified Tamm-Danoff Approximation [10]. There are several approximations made between full linear response theory and the sTDA method. First is the Tamm-Dannoff approximation, where the **B** matrix is ignored. The second approximation is to use monopole approximations with Mataga-Nishimoto-Ohno-Klopman operators instead of explicit 2 electron integrals to calculate matrix elements, as well as neglecting the density functional term.

Transition charges are used to calculate these MNOK integrals. The charge q_{nm}^A , the charge centred on atom A given by transition density for the transition of $n \rightarrow m$, are computed using a Löwdin population analysis:

$$(2.14) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m}$$

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

$$(2.15) \quad \mathbf{C}' = \mathbf{S}^{\frac{1}{2}} \mathbf{C}$$

and μ is an index that runs over the atomic orbitals (AO).

The operator is the MNOK[21][22][16] damped coloumb function, with different exponents y_K and y_J for exchange and coulomb integral respectively. The α_x parameter is included to recover the amount of Fock exchange mixing in the original matrix element equation, and is a free parameter. 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 **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 **A** matrix, albeit with a tight-binding method for molecular orbital coefficients, the gradient of the transition properties would still be difficult to calculate.

2.2 Benchmarking

The first step in testing whether a GFN-xTB based Δ -SCF method (which we name Δ -xTB) could predict TD-DFT transition properties with decent accuracy was to test a set of small molecules, as well as bacterial chlorophylls from the LHII protein against high-level results.

2.2.1 Reference Data and test set

The test set comprised 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 relative simplicity of this test set was chosen to minimise any other factors that could cause complication in analysing the results. For example choosing a consistently correct basis set and functional for the Δ -SCF results would have been challenging if it was necessary to include transition metal complexes, as well considering CPU walltime constraints. Additionally, the size of these molecules made symmetries much easier to inspect by hand.

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 [5].

2.2.2 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, however 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, as this is a superposition of both spin-same $\alpha \rightarrow a, \alpha$ and spin-flipping $\alpha \rightarrow a, \beta$ excitations. The spin-purification formula:

$$(2.16) \quad \Delta E_S = 2\Delta E^{i,\alpha \rightarrow a,\alpha} - \Delta E^{i,\alpha \rightarrow a,\beta}$$

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

The results of comparing transition energies and transition dipole magnitudes are shown in figures 2.1 and 2.2.

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 2.1 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

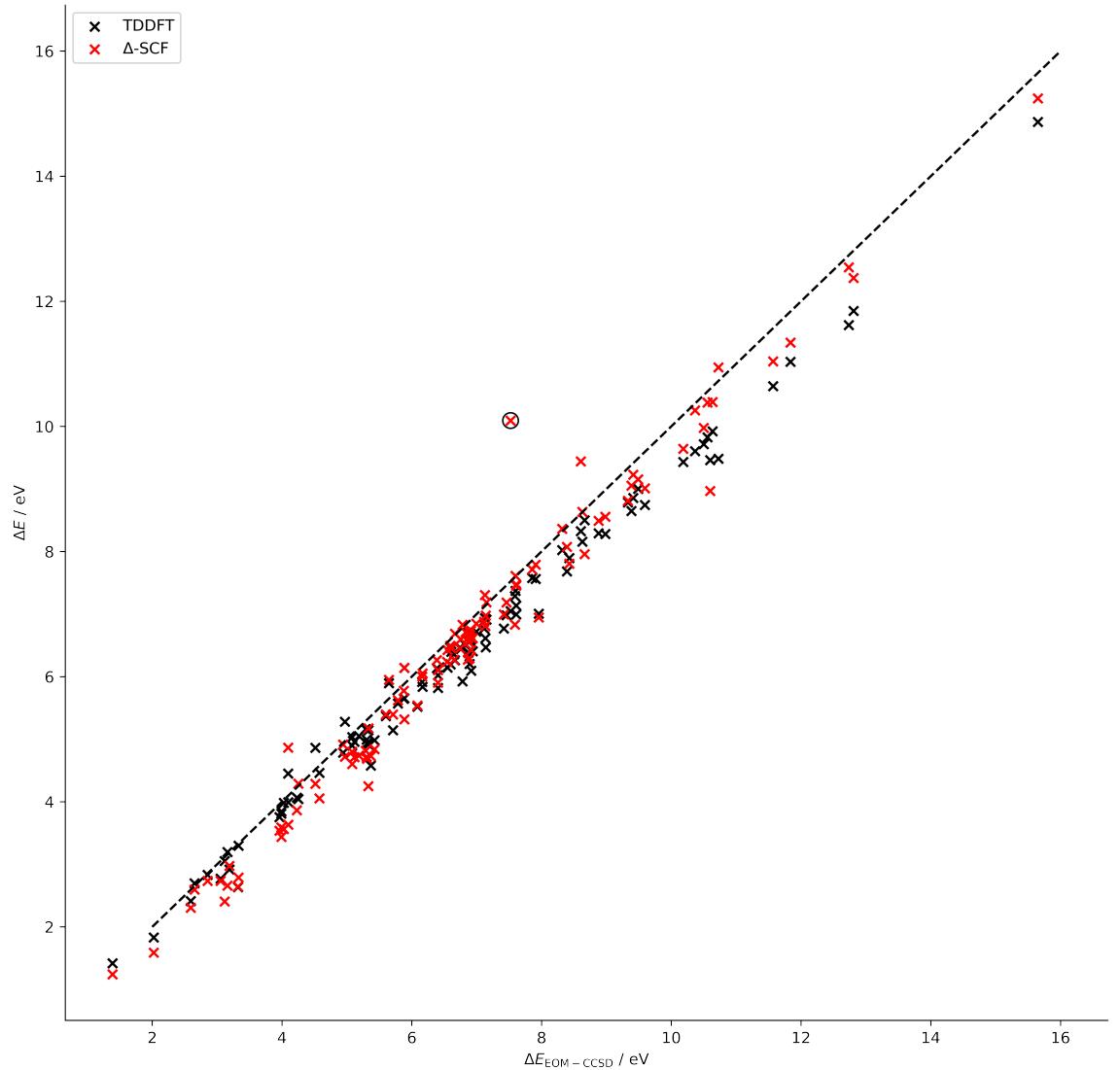


Figure 2.1: Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.

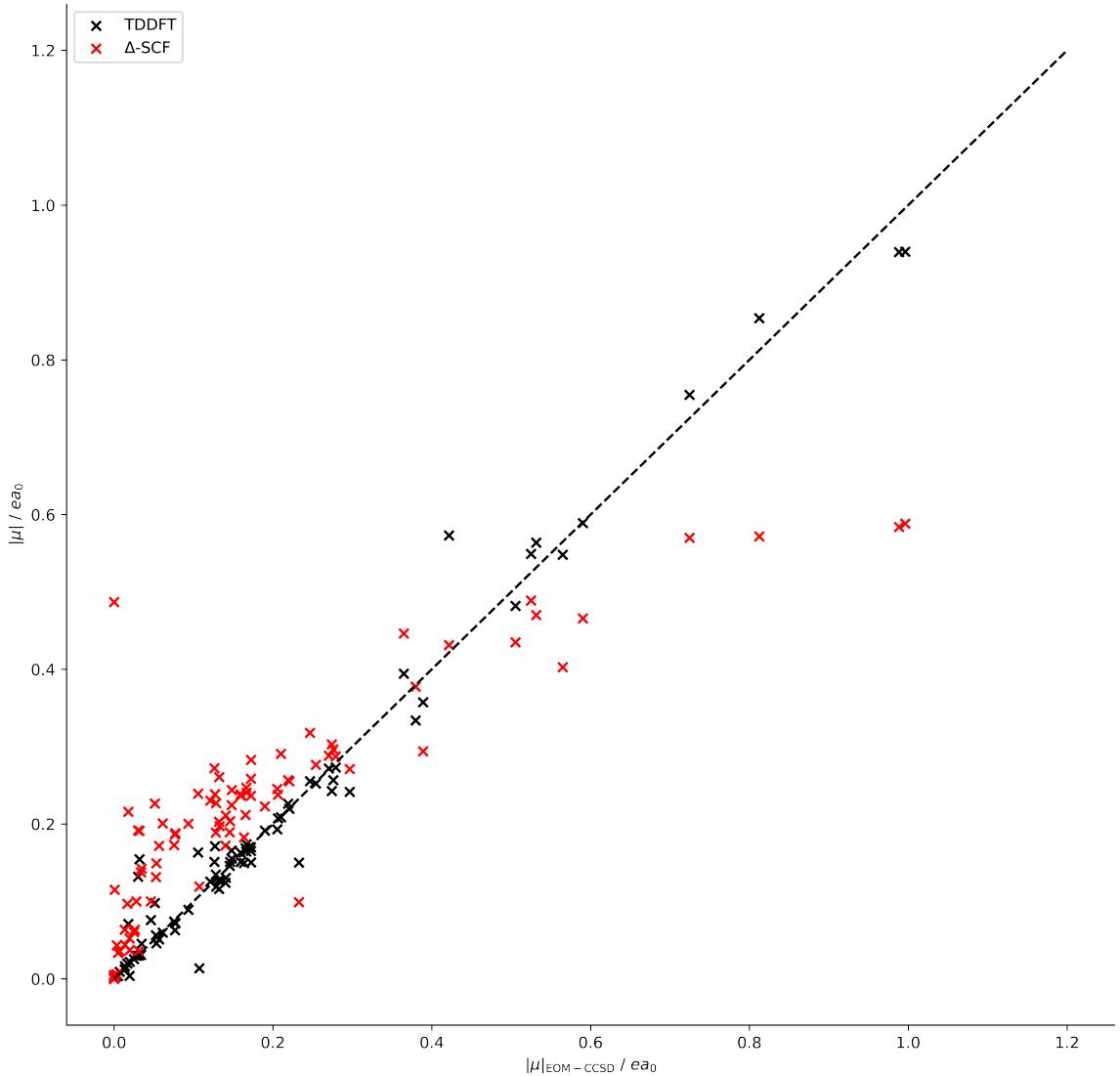


Figure 2.2: Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.

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.

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.

2.2.3 Non-orthogonality

There is, however, 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 the set of orbitals $\{|\phi_j^{(n)}\rangle\}$ that will be orthogonal within the same state, however the overall states will not be orthogonal to each other, such that the inner product:

$$(2.17) \quad S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle$$

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

$$(2.18) \quad q^{21} = \langle \Psi^2 | \Psi^2 \rangle$$

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, a 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:

$$(2.19) \quad |\Psi_{\tilde{\nu}}\rangle = \sum_{\nu} |\Psi_{\nu}\rangle \left[\mathbf{S}^{-\frac{1}{2}} \right]_{\nu\tilde{\nu}}$$

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

$$(2.20) \quad \mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

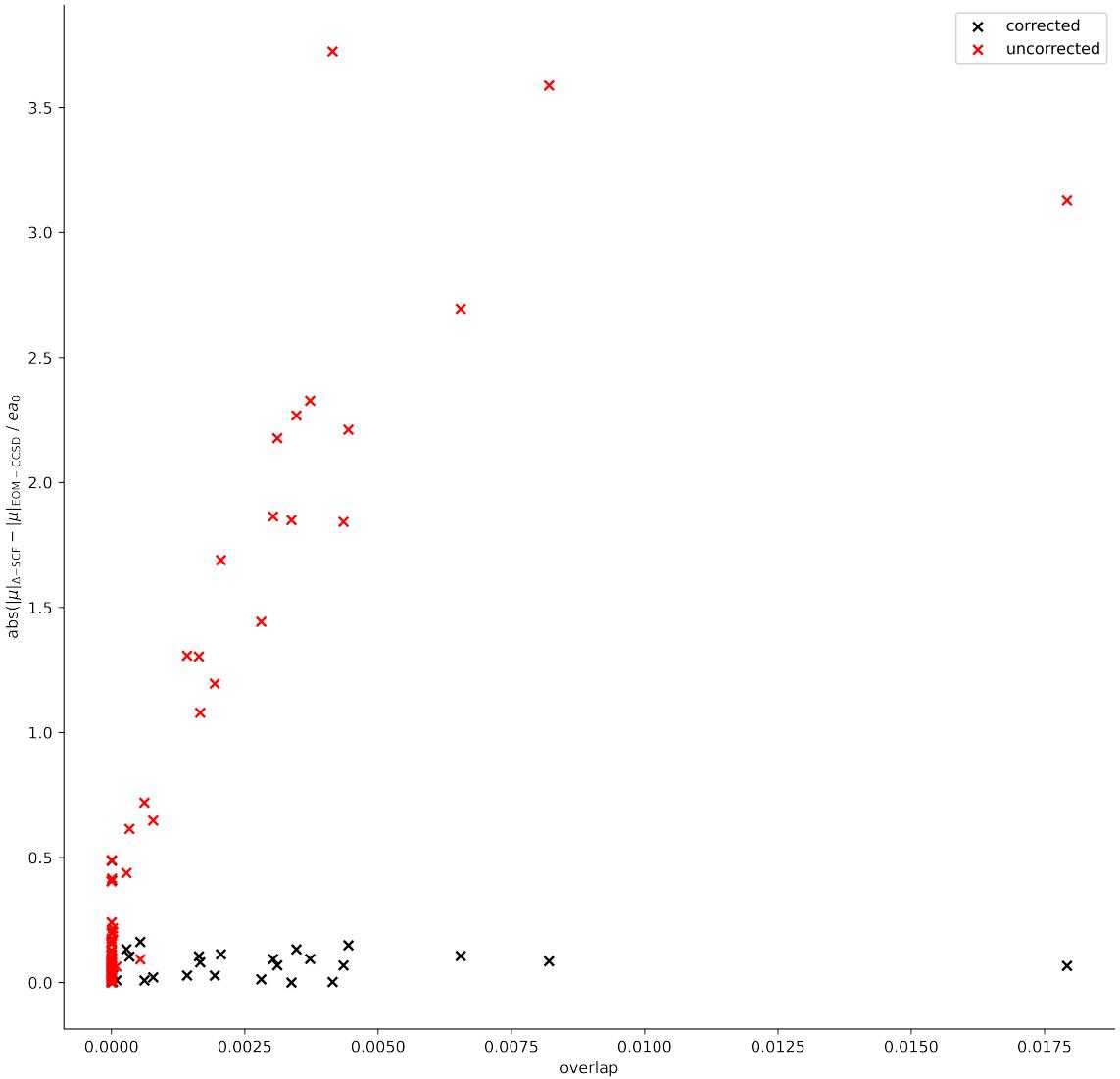


Figure 2.3: The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown in red, whilst those with the symmetric orthogonalisation correction are shown in black.

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 2.2.3). 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 magnitudes. In a large protein system, where these distances can be on this scale, this would obviously present a much larger problem than for a vacuum phase small molecule.

2.2.4 LHII Chlorophyll

The accuracy of Δ -SCF was also tested on a set of bacterial chlorophyll A (BChla) molecules. These are a much larger and more complex system, and so present a tougher challenge for the Δ -SCF method. Additionally, it is a more relevant test for the goal of simulating a whole LHII complex.

As each BChla is 140 atoms, EOM-CCSD could not be used as the reference method due to resource constraints, so TD-DFT at a PBE0/Def2-SVP level of theory was used instead [2, 24]. This has been used previously for BChla structures, and has been shown to be a good balance between accuracy and cost [25]. The Δ -SCF was run with the same density functional and basis set.

Δ -SCF can be seen to give transition energies to within the accuracy of TD-DFT, such that the variations in transition energy can be said to be due to physical intra-molecular reasons rather than random error from the Δ -SCF method. Δ -SCF could therefore be used to accurately predict geometry dependent properties.

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 can be confidently made from Δ -SCF data.

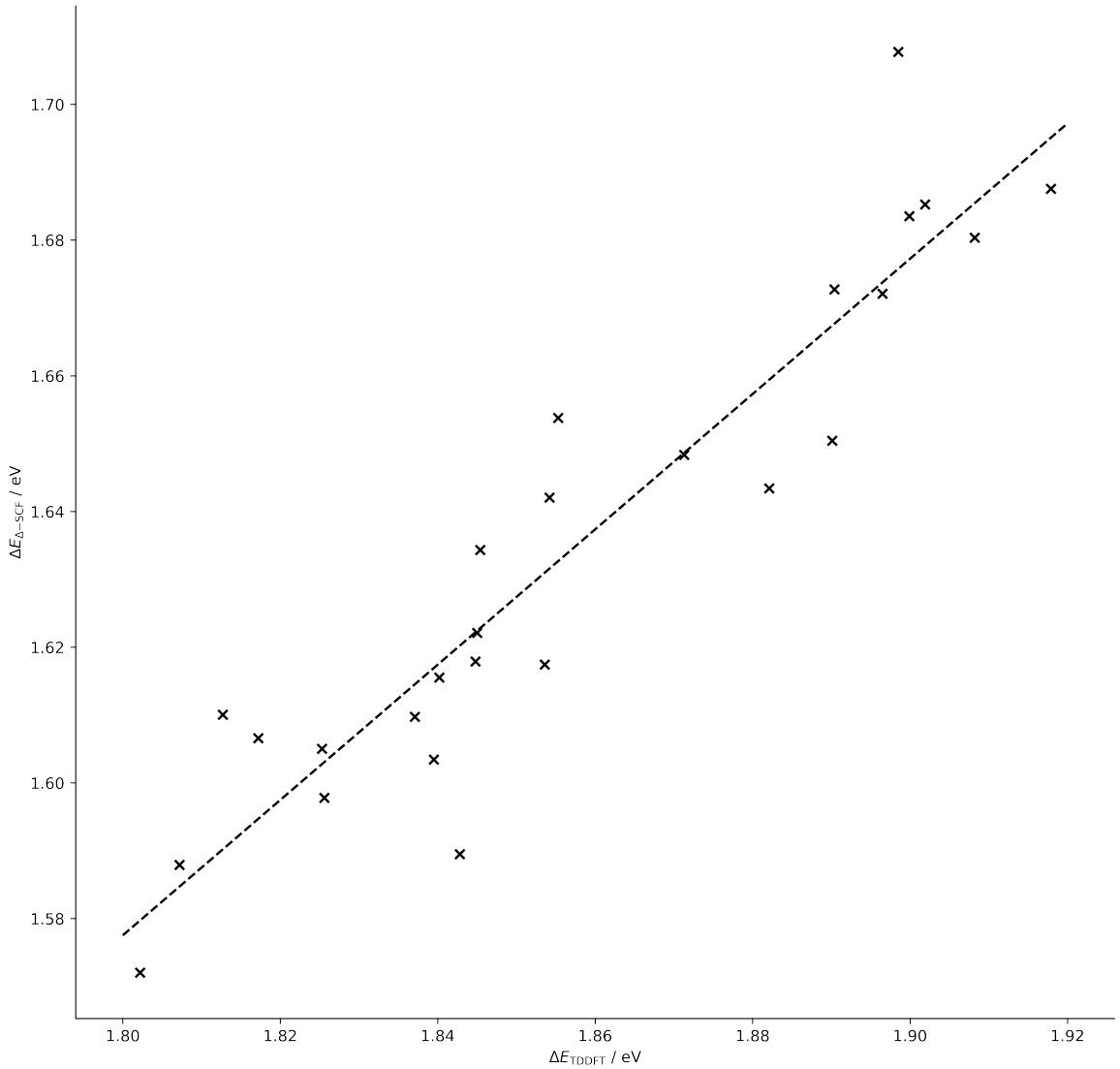


Figure 2.4: The transition energies from Δ -SCF for the 26 chlorophyll geometries, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line.

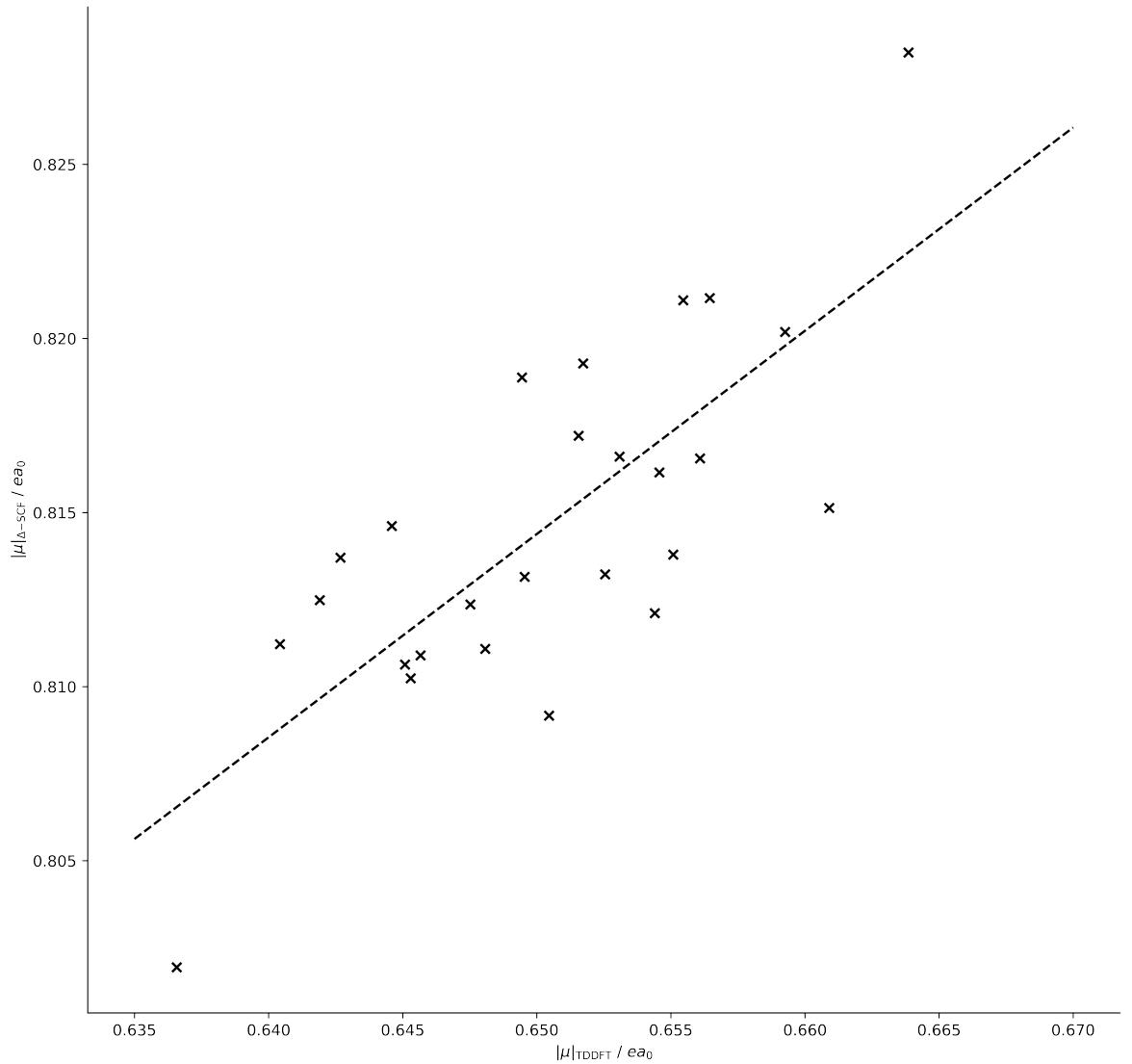


Figure 2.5: The transition dipole magnitudes from Δ -SCF for the 26 chlorophyll geometries, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line.

2.2.5 xTB methods

The same benchmarking set was used with the Δ -xTB method. The number of methods in the comparison was also expanded to cover a range of approximations in calculating transition properties, including the aforementioned sTDA-xTB method. To include this in the comparison, it was necessary to use the spin-component-scaled second order coupled-cluster (SCS-CC2) [13, 14] reference data produced by the Grimme group. 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 is using the eigenvalue differences from sTDA-xTB, given by the `xtb4stda` program, to observe how much improvement to transition properties is caused by the sTDA procedure. `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.

All methods included were:

- 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, for additional reference.
- 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, as if Δ -SCF excited state has converged to different transition than TD-DFT and CC2 it would have different symmetries 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. First is the valence basis set for the Δ -xTB calculations are very different to those used in DFT. Second is that this information was not available for the CC2 data, however 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.

2.2.5.1 Post-SCF Assignment of Symmetry

Considering 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. Setup 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 .

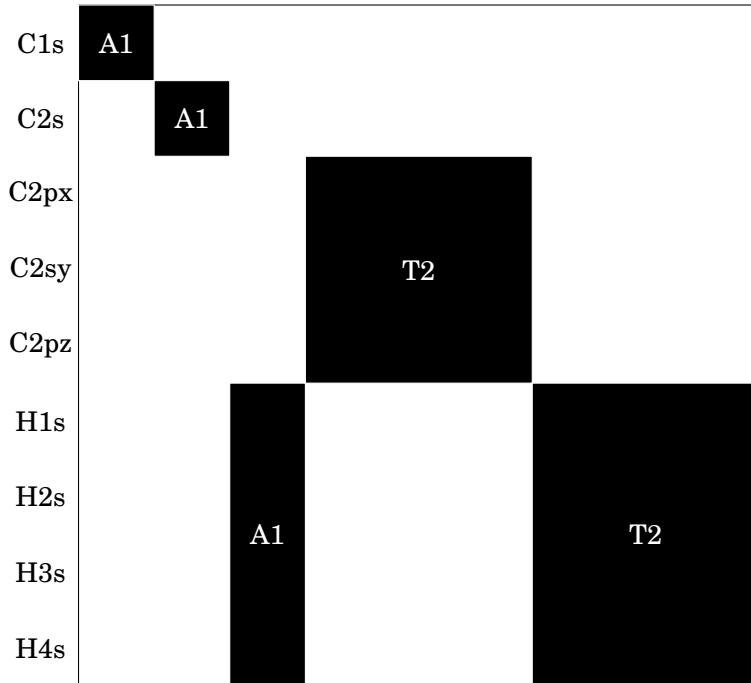


Figure 2.6: A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.

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 the each subspace A by using the transformation matrix \mathbf{T} :

$$(2.21) \quad \tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{S} \mathbf{C}$$

and then summing the coefficients to obtain the character P_A :

$$(2.22) \quad P_A = \sum_v |\tilde{\mathbf{C}}_{A,\mu v}|$$

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 so 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 were 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:

$$(2.23) \quad n = \frac{1}{h} \sum_R \xi_r(R) \xi_i(R)$$

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. Additionally was that for non-abelian groups, where there are some degenerate E and T subspaces, this assignment 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.

Using symmetry decomposition for the non-abelian point groups and analysis based on the antisymmetric component of the direct products was discussed, however this would create a more open-ended project. Additionally, while this may 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 total transitions could not be confidently assigned for Δ -SCF with this method.

2.2.5.2 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, however the difficulty of doing this for the entire range of methods should be noted.

The distributions of errors for each of the benchmarking methods, as well as a generated distribution of sTDA-xTB results, are shown in fig-2.7. The means and standard deviations are reported in table 2.1.

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

2.2. BENCHMARKING

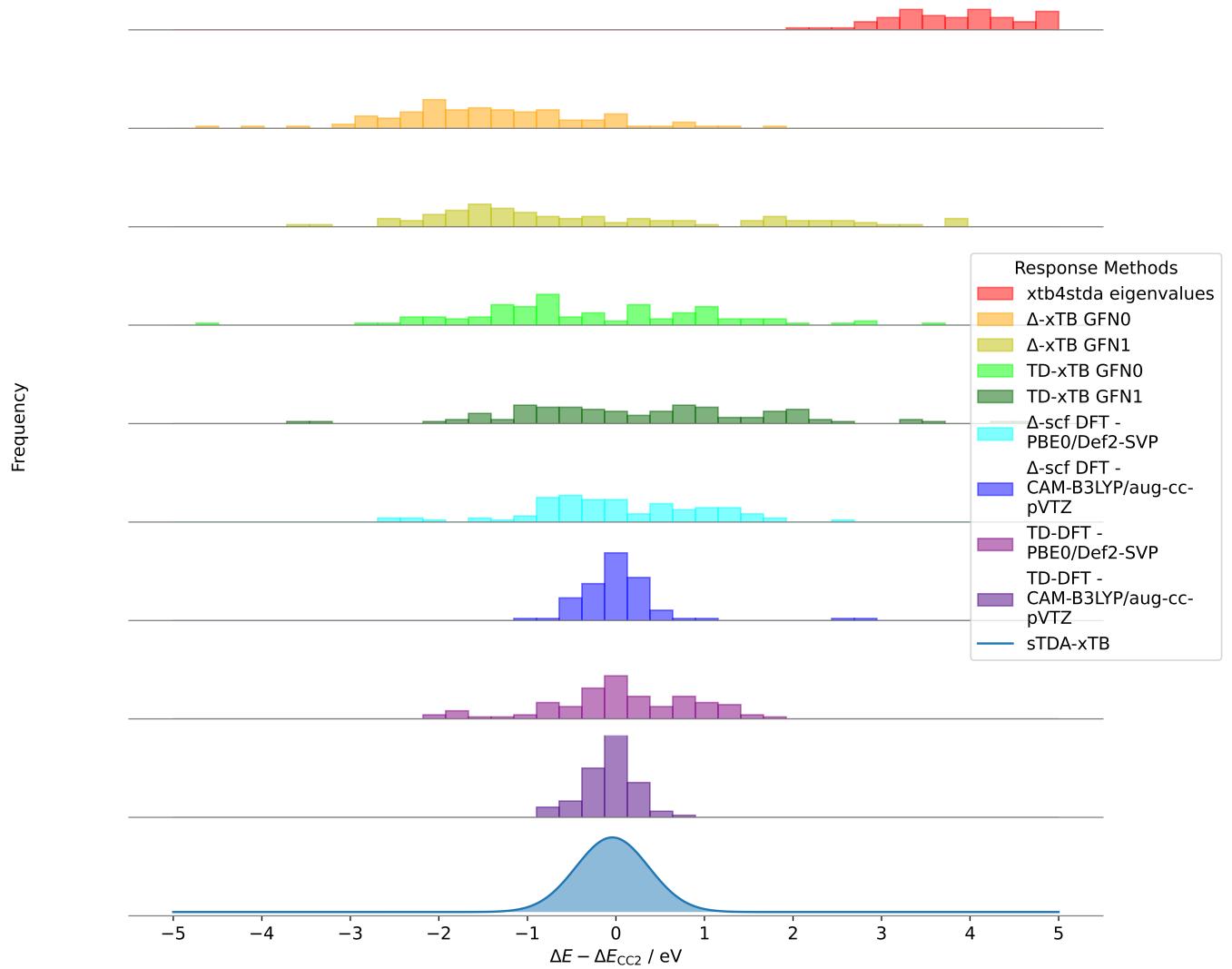


Figure 2.7: The distributions of errors to CC2 transition energies for the methods included in the Δ -xTB benchmarking.

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 2.1: 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 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.

The sTDA-xTB line is generated from the mean and standard deviation of the errors as reported (-0.04 eV and 0.41 eV respectively)[11]. This was included as a reference line to show how the full sTDA-xTB method is arguably as accurate as TD-DFT with a range-based density functional and triple zeta basis sets. Again CAM-B3LYP/aug-cc-pVTZ methods, both the linear-response and Δ -SCF, are accurate to within this range as well. The outliers in the Δ -SCF results are known mixed transitions, as discussed earlier with the ethene dimer system. Overall, these results are inline with the previous set of benchmarking. The 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). Additionally, Both linear-response and Δ -SCF methods perform similarly when compared to each other, and so the leading cause of error is attributed to be the electronic structure method, and not the response method.

It can be seen then that the GFN-xTB based methods performed much worse than the DFT methods. There is a relatively small difference between linear-response GFN-xTB and Δ -xTB results, again implying that poor electronic structure theory gives poor transition properties. Both the GFN0- and GFN1-xTB based methods proved inaccurate, and there is a marked drop when using GFN0. The systematic shift, shown in the mean error of -1.50 eV for response method based on out-of-the-box GFN0-xTB would not be viable. the GFN0- Δ -xTB method is especially bad. It's concluded from this that a

Overall, the most inaccurate method is the eigenvalue difference methods based orbital

energies (eigenvalues of the Hamiltonian diagonalisation) from the `xtb4stda` method. Arguably the sTDA method then makes up a large part of predicting transition properties accurately.

The result that GFN-xTB based methods are not accurate is not unexpected. As opposed to using *ab initio* or first principle parameters, the xTB methods fit to target properties, and so could not be expected to be suitable for every case [3]. Whilst the GFN-xTB methods are better than many other methods in this parameterisation, using a few pair-wise parameters as possible to improve extensibility to systems outside the training set, these parameters are trained on ground state properties and not excited state properties. These parameters are found to be unsuitable for predicting transition properties.

2.3 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 would 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 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.

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 3.1. Parameterisation details are given in section 3.2, 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 3.3.

3.1 Theory

Excitations from linear-response TD-DFT are given by solving the non-Hermitian eigenvalue equation:

$$(3.1) \quad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

where \mathbf{A} , \mathbf{B} are matrices whose elements describe the perturbation response of the electron density. \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 elements respectively, and in TD-DFT, with a global hybrid density functional, are given by:

$$(3.2) \quad A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2(i|a|j|b) - \alpha_x(i|j|a|b) + (1 - \alpha_x)(i|f_{XC}|j|b)$$

$$(3.3) \quad B_{ia,jb} = 2(ia|bj) - a_x(ib|aj) + (1 - a_x)(ia|f_{XC}|bj)$$

where indices a, b and i, j refer to virtual and occupied orbitals respectively, a_x is the value of non-local Fock exchange in the XC functional f_{XC} . The integral (here in Mulliken notation) can be seen to be of Coulomb type for the \mathbf{A} matrix and exchange for the \mathbf{B} matrix.

3.1.1 Approximations to Solutions

3.1.1.1 Tamm-Dancoff Approximation and Diagonal Dominant A matrices

One of the earliest approximations applied to full TD-DFT was the Tamm-Dancoff approximation, where only virtual-occupied effects were taken into account. This sets all of the elements of \mathbf{B} matrix to zero, and reduces the full eigenvalue equation to

$$(3.4) \quad \mathbf{AX} = \omega \mathbf{X}$$

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, the way to solve for ω is to construct the full \mathbf{A} and then diagonalise. Additionally, this is required as the XC functional is dependent on ω values and so several iterations of this diagonalisation are needed to find stable, self-consistent solutions.

However, the diagonal elements of the \mathbf{A} matrix may be an approximation to the eigenvalue solutions. This would be in the limit where the coupling elements, the off-diagonal elements, go to zero, which would be the case for an excitation that is mostly made up of a single transition. Excitation energies therefore would be given by:

$$(3.5) \quad \omega_{ia} = A_{ia,ia} = (\epsilon_a - \epsilon_i) + 2(ia|ia) - a_x(ii|aa) + (1 - a_x)(ia|f_{XC}|ia)$$

If the integrals are assumed to also be small compared to the first term difference of ϵ_a and ϵ_i , then the excitation energy would then be equal to just this eigenvalue difference, hence its use as a proxy for higher-level TD-DFT.

3.1.2 Integrals approximations

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

be used to obtain the atomic (centered on atom A) charges densities q_{nn}^A and transition charge densities q_{nm}^A :

$$(3.6) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m}$$

where μ are indices of AOs centered on atom A , and \mathbf{C}' are the orthogonalised MO coefficients, following the same orthogonalisation scheme as previously used:

$$(3.7) \quad \mathbf{C}'_n = \mathbf{S}^{\frac{1}{2}} \mathbf{C}_n$$

where \mathbf{S} are the AO overlaps, and \mathbf{C} the unorthogonalised orbitals.

The two electron integrals can be approximated by using MNOK integrals. For both coloumb and exchange type integrals, this is done with a short-range damped MNOK operators. The integral is approximated by:

$$(3.8) \quad (pq|rs) = \sum_A^N \sum_B^N q_{nm}^A q_{nm}^B \Gamma_{AB}$$

where N is the total number of atoms in the system, p, q, r and s electron indices. For Coloumb type integrals, the operator Γ_{AB} is given by:

$$(3.9) \quad \Gamma_{AB}^J = \left(\frac{1}{(R_{AB})^{y_J} + (\alpha_x \eta)^{-y_J}} \right)^{\frac{1}{y_J}}$$

where R_{AB} is the interatomic distance of atoms A, B , η is the average of the chemical hardness of atoms A and B . These values are defined as:

$$(3.10) \quad \eta(A) = \frac{\delta^2 E(A)}{\delta^2 N^2}$$

but in practice are precomputed and used as parameters. y_J is a global parameter and usually set to integer values. α_x is the previously defined amount of Fock-exchange mixing.

For exchange type, the operator is:

$$(3.11) \quad \Gamma_{AB}^K = \left(\frac{1}{(R_{AB})^{y_K} + \eta^{-y_K}} \right)^{\frac{1}{y_K}}$$

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

As the α_x parameter can "mop up" many of the exchange effects, the density functional in equation 3.5 is also neglected to further reduce computational cost.

The final form of the expression used to calculate excitation energies is then given by:

$$(3.12) \quad A_{ia,ia} = (\epsilon_a - \epsilon_i) + \sum_{A,B}^N \left(2q_{ia}^A \Gamma_{AB}^K q_{ia}^B - q^A \Gamma_{AB}^J q^B \right)$$

where the exchange term uses transition charges q_{ia} , and the coloumb term uses normal charges.

The approximations outlined so far have been based on both the sTDA [10] and sTDA-xTB [11] methods, which show great success at both reducing the amount of computational cost as well as accurately predicting transition properties for small molecules and larger systems.

3.1.3 Q_y Transition

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, as 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 by around 12 °.

Plots of the electron density of the HOMO, LUMO and transition density show how this transition is delocalised over large sections of the porphyrin ring, with approximate C_2 symmetry along the molecular axes.

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 chlorophyll data in the previous chapter.

3.1.4 chl-xTB method

This section will set out the novel procedure for calculating transition properties for chlorophyll, which is named chlorophyll-xTB, abbreviated to chl-xTB.

This method, in the form reported here, is specific to the Q_y transition for bacterial chlorophyll A systems. This transition constitutes the entire training set, and the validity of this method outside of this range is not tested. This is due to the aim that this method could predict variations in values for transition properties that would be due to atomic geometry reasons, which is a high level of specificity. Initially it was investigated whether this method could be applicable to a broad range of systems and transitions. 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 correctly predict geometry variations. Additionally, 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. It is argued that this is an improvement on a broader method, that would have been much harder and more expensive to train.

The first part of the chl-xTB method was to adjust the electronic structure method for a better starting ground state density. As shown in the previous chapter, this was an issue with the Δ -xTB methods. The GFN1-xTB Fock matrix is made of both charge dependent and charge independent terms, and only the charge dependent terms would have any effect on the transition properties as these would effect the partial and transition charges. The charge dependent terms are the first, second and third density fluctuation terms. The first order term is the leading term, and so these parameters were chosen to be included in the optimisation procedure. Of these, only certain parameters are "free", as in based on the top-down method of parameterisation, whilst others are based on physical or *ab initio* values. One set of free parameters which were altered for this method are the Hückel parameters k_l , where l is the angular momentum of the orbital, and the global scaling parameters. In the original GFN1-xTB method there are also free parameters called global scaling parameters, which were used to adjust for some pairs of elements were the general procedure led to incorrect bond lengths. Similar parameters were also included, but only for the Mg and N atoms. Whilst other atoms types could have been part of this global scaling scheme, it was found that only these two were necessary to achieve a decent the accuracy. Using this small amount of parameters would reduce the risk of overfitting to the training set. An obvious drawback of altering these parameters to fit to transition properties is that they would lose their specificity to geometry optimisation and normal-frequencies, as well as to the non-covalent interactions although should be less effected. However it is not in the scope of this work to find a semi-empirical method to do be able to calculate these properties for chlorophyll. The 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.

The second part of this method calculate the transition density based on a Δ -SCF like approach to the excited state density, but without orbital relaxation. The ground state orbital coefficients are taken to be a good approximation to the (Q_y) excited state coefficients. The transition density is then calculated as:

$$(3.13) \quad D_{Q_y}$$

. This is then used to calculate the transition charges with the Mulliken scheme, which is in

turn used to calculate the MNOK integrals of the form eq. Here the free parameters a_x, y_J, y_K are optimised specific to the Q_y transitions, whilst η is calculated using the GFN1-xTB chemical hardnesses for consistency, as well as to avoid overfitting. Similar to $\Delta\text{-SCF}$, the transition dipoles can be calculated as the trace of the dipole operator with the transition 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 $\Delta\text{-SCF}$ excited state. It would also halve the computational time. Excited state properties, such as the molecular dipole and partial charges can also be calculated as the excited state density can be constructed in a similar fashion, which will be important for the exciton framework of the next chapter.

It's clear that this method is heavily inspired by sTDA-xTB. However there are some key differences in this theory that constitutes novel work. The most obvious is the ground state method. sTDA-xTB uses a bespoke version of the xTB formalism, which whilst similar has some key differences. These include the use of a geometry dependent basis set, whereas the chl-xTB method uses a fixed basis set. The Fock matrix terms for chl-xTB are also different, and for the most part the same as GFN1-xTB, with the notable exception of the extended Hückel term which uses novel chl-xTB specific parameters. Third is the SCC procedure, which is replaced in sTDA-xTB for single diagonalizations. First is that the sTDA methods still solve the eigenvalue problem, and so require constructing the entire \mathbf{A} matrix. As stated earlier, it is assumed for transitions were the corresponding sections of the \mathbf{A} matrix are diagonal dominant this is unnecessary. Additionally, instead of the Löwdin scheme for atomic charges, the Mulliken scheme was used. This gives the charges centered on atom A as:

$$(3.14) \quad q^A = Z_A - \sum_{\mu \in A} D_{\mu\nu} S_{\mu\nu}$$

where \mathbf{D} is the reduce one-electron density, and Z_A is the atomic number. For transition charges, this density can be switch with the transition density. The choice to use Mulliken charges over Löwdin charges was made due to the known instability of Löwdin charges with respect to orbital rotation, however it is known that Mulliken charges are highly dependent on the basis set chosen.

Additionally, a scaling factor for the transition density was employed to attempt to recover some of the effects of neglecting off diagonal elements from the \mathbf{A} matrix.

Excitation Energies correlation				
	PBE0	dscf	eigdiff	camb3lyp
PBE0	1	0.915415	0.930766	0.952334
dscf	0.915415	1	0.846607	0.863979
eigdiff	0.930766	0.846607	1	0.925034
camb3lyp	0.952334	0.863979	0.925034	1
BLYP	0.730346	0.637576	0.606743	0.637641

Figure 3.1: correlations of energies

3.2 Parameterization

3.2.1 Reference Data

The geometries for the training set used to optimise the chl-xTB method were taken from previously done classical MD of the LHII protein. These geometries were chosen as the stochastic variations of the BChla molecules would represent a range of configurations, which would train chl-xTB to accurately predict variations of transition properties on vibrational modes.

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 three levels of TD-DFT.

The actual training data used was the PBE0 data. This was chosen for the best accuracy-cost ratio, as well as having been previously used to investigate exciton properties for the same system. 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.

Transition Dipole Magnitudes correlations

	PBE0	dscf	eigdiff	cam
PBE0	1	0.5891	0.566434	0.566434
dscf	0.5891	1	0.977046	0.977046
eigdiff	0.566434	0.977046	1	0.977046
cam	0.762861	0.800859	0.773252	0.773252
BLYP	0.0832043	-0.0344407	-0.0098512	0.0098512

Figure 3.2: correlations of transition dipole moments

3.2.2 Objective Function

The design of the objective function, the function that is minimised in the optimisation procedure, is also important to find the best parameters.

At first, it would be argued that the root mean squared error (RMSE) to the PBE0 transition energy should be the value that should be minimised, giving the objective function as

$$(3.15) \quad f_{\text{RMSE}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_i^N (\Delta E_i - \Delta E_{i,PBE0})^2}$$

however this 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:

$$(3.16) \quad R^2 = 1 - \frac{\sum_i^N (\hat{y}_i - y_i)^2}{\sum_i^N (\hat{y}_i - \bar{y})^2}$$

The correlation is a better metric for determining if chl-xTB has a small enough random error to have predict variations of transition properties from different geometries. A low RMSE and high coefficient of determination (R^2) are not mutually inclusive.

The full objective function used was:

$$(3.17) \quad f_{\text{full}}(\mathbf{x}) = \lambda_1 \text{RMSE}(\Delta E) + \lambda_2 \text{RMSE}(|\mu|) + \lambda_3 R^2(\Delta E)$$

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.

3.2.3 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 nonlinear problems. The optimisation was performed using SciPy's `minimize` function in the `optimize` module. The first method tested was the "Nelder-Mead" option, the algorithm for which is described below.

3.2.3.1 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 minimum of the function $f(\mathbf{x})$:

$$(3.18) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

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

$$(3.19) \quad f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$$

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:

$$(3.20) \quad \mathbf{x}_r = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1})$$

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:

$$(3.21) \quad \mathbf{x}_e = \mathbf{x}_0 + \gamma(\mathbf{x}_r - \mathbf{x}_0).$$

. 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:

$$(3.22) \quad f(\mathbf{x}) = \begin{cases} \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_r - \mathbf{x}_0) & \text{if } f(\mathbf{x}_r) < f(\mathbf{x}_{n+1}) \\ \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_{n+1} - \mathbf{x}_0) & \text{otherwise } f(\mathbf{x}_r) \geq f(\mathbf{x}_{n+1}) \end{cases}$$

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

$$(3.23) \quad \mathbf{x}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$$

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

One 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 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.

3.2.3.2 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:

$$(3.24) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

however with the added constraints:

$$(3.25) \quad c_i(\mathbf{x}) = 0$$

$$(3.26) \quad c_j(\mathbf{x}) \leq 0$$

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:

$$(3.27) \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

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

$$(3.28) \quad \mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_n^m \lambda_n g_n(\mathbf{x})$$

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

$$(3.29) \quad \min_d f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 \mathcal{L}(\mathbf{x}_k, \lambda_k) d$$

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:

$$(3.30) \quad \min_d \| (\mathbf{D}_k)^{\frac{1}{2}} (\mathbf{L}_k)^T d + (\mathbf{D}_k)^{-\frac{1}{2}} (\mathbf{L}_k)^{-1} \nabla f(\mathbf{x}_k) \|$$

where the matrices \mathbf{L} , \mathbf{D} are from a diagonal decomposition of \mathbf{B} :

$$(3.31) \quad \mathbf{L}_k \mathbf{D}_k (\mathbf{L}_k)^T = \mathbf{B}_k$$

. 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.

Both these methods were used to find optimal parameters, and it was found that the SLSQP method performed much better, 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 values out from the SLSQP optimisation are physically reasonable, and within the range of similar parameters in the xTB methods.

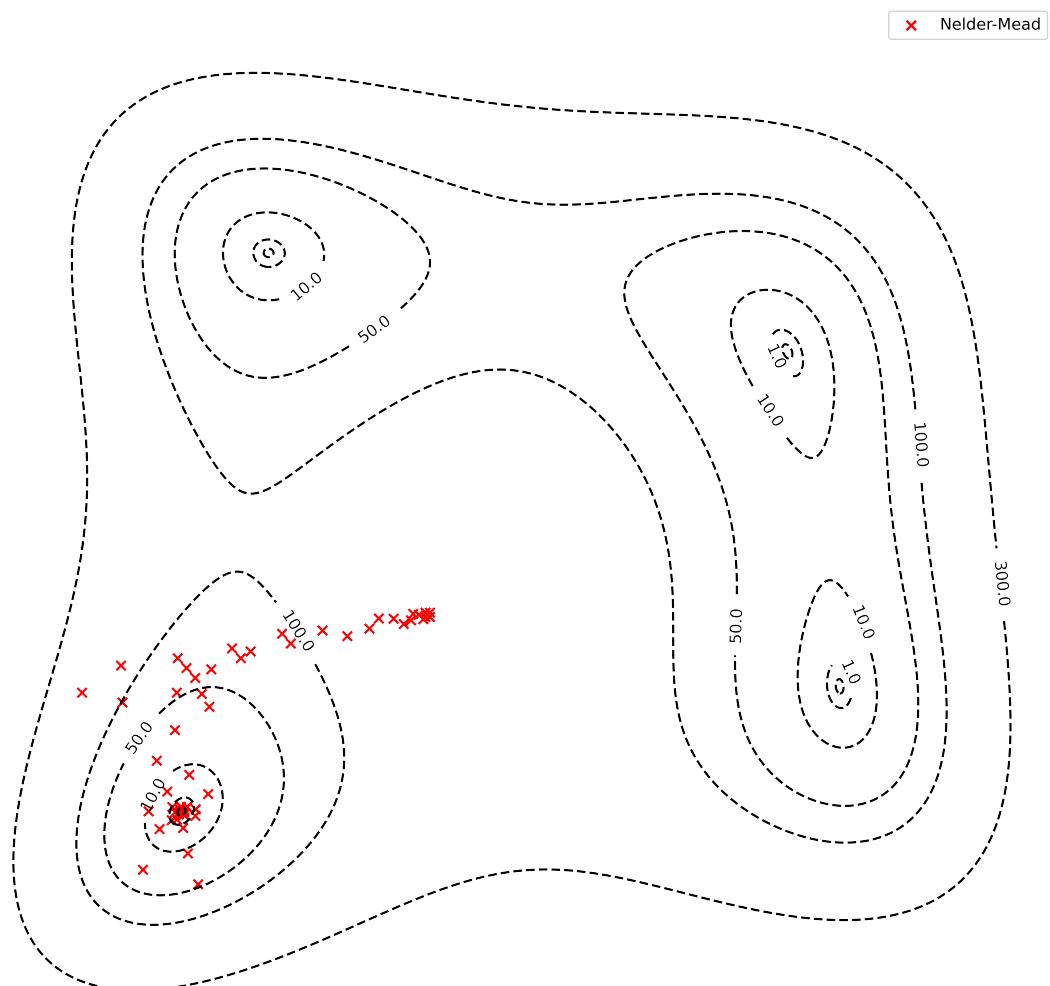


Figure 3.3: Nelder-mead

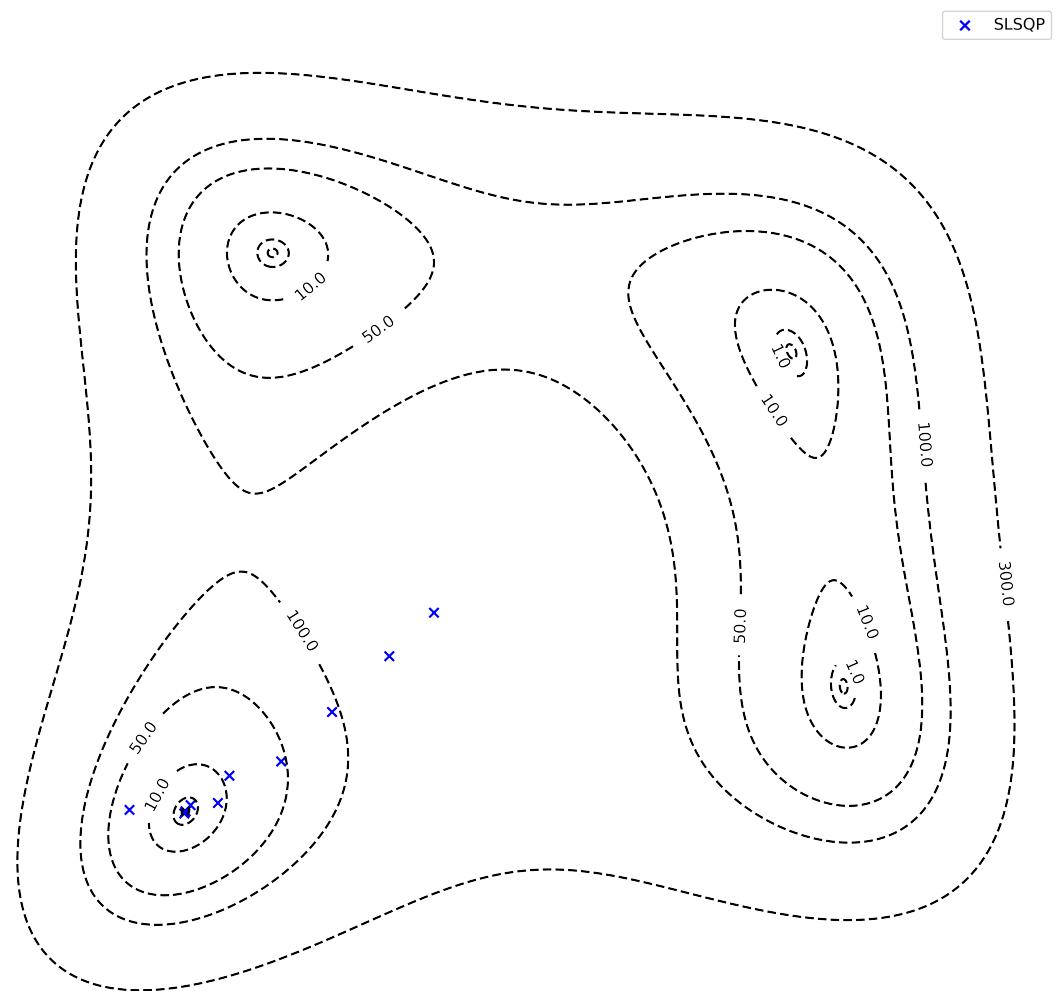


Figure 3.4: SLSQP

y_K	1.0
y_J	1.0
a_x	1.0

Table 3.1: optimized parameters from SLSQP procedure.

3.2.4 Results

The final parameters for the chl-xTB method are given in table 3.1. The best performing set of parameters had an RMSE of excitation energy of, RMSE of transition dipole magnitude of , and a R^2 value of . 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.

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. It was also found that 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.

3.3 Benchmarking and Cross-validation

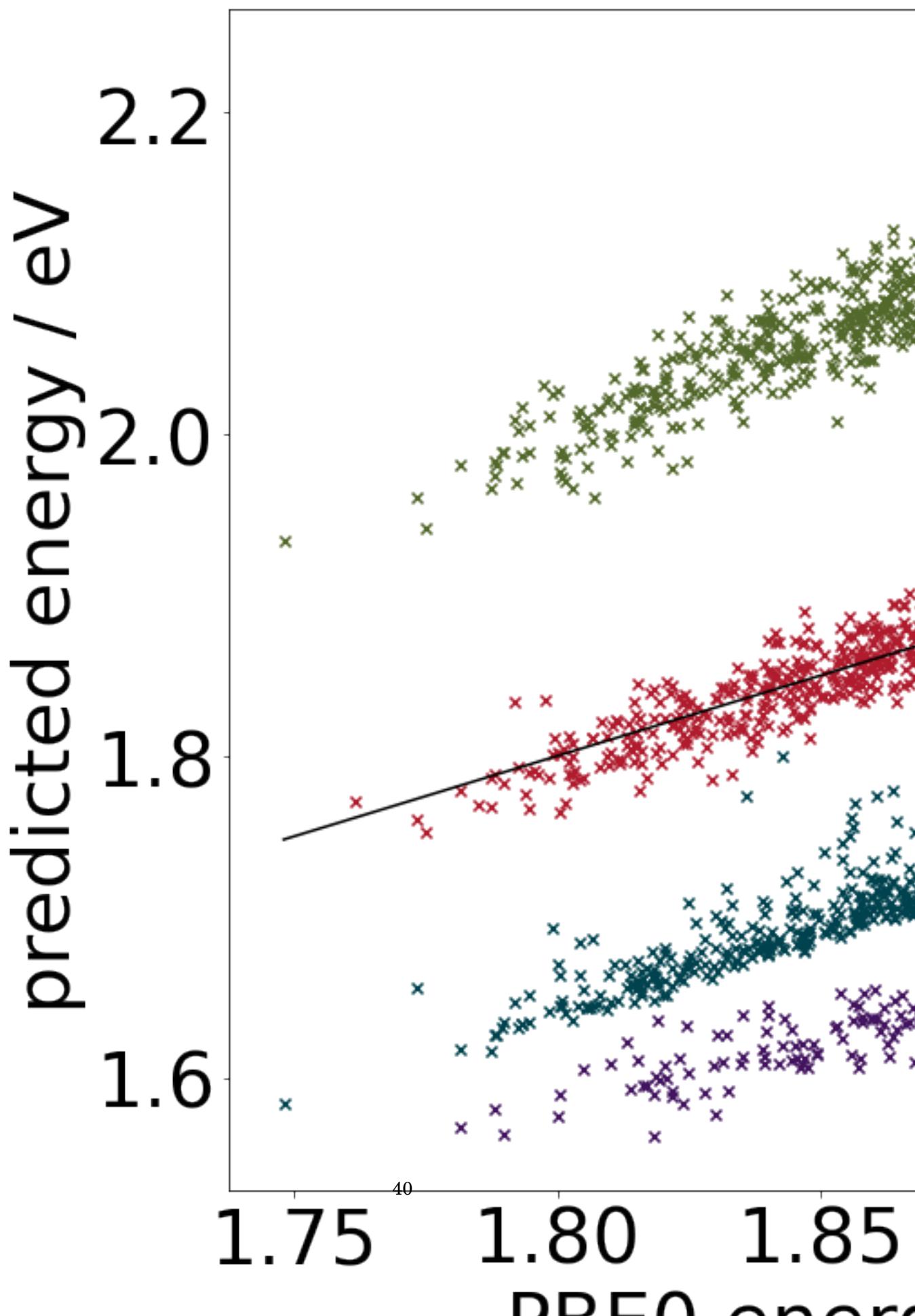
3.3.1 Transition properties

The predicted values for Q_y transition energies are shown against other methods in figure ???. It can be seen that the chl-xTB energies are well within the range of other DFT functionals, and on a similar level of accuracy and correlation, which is a huge improvement over the lower level methods.

3.3.2 Potential Energy Surfaces

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 all conformations. The errors in transition properties between PBE0, benchmarked against other DFT methods, and chl-xTB was then calculated for geometries along multiple normal modes.

The geometries for this test were not taken to be BChla for two reasons. There are 140 of atoms in BChla, which means the number of normal modes is 414, and with anywhere between 10-20 coordinates along each normal mode, this represents a large number of calculations to do 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

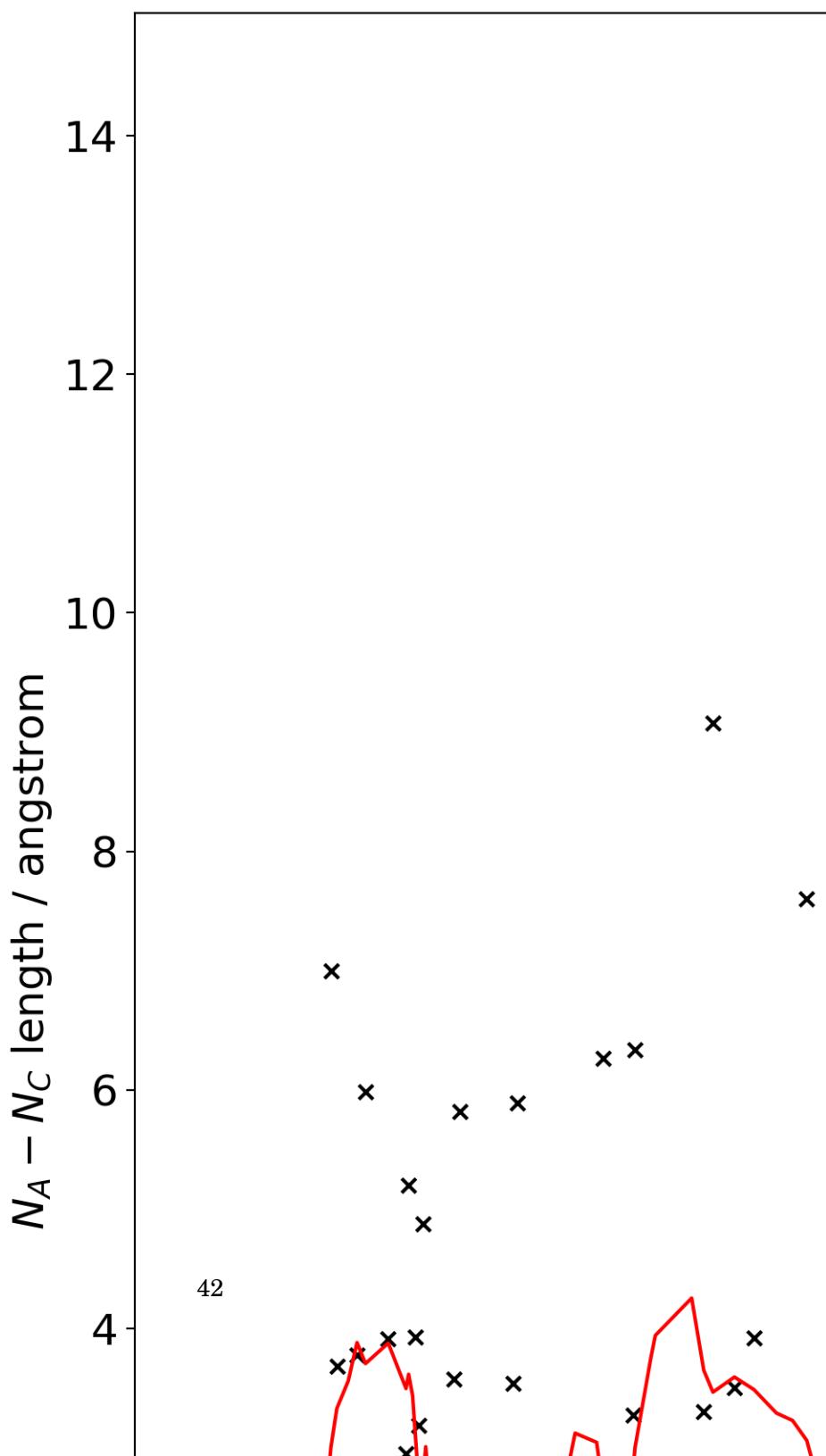


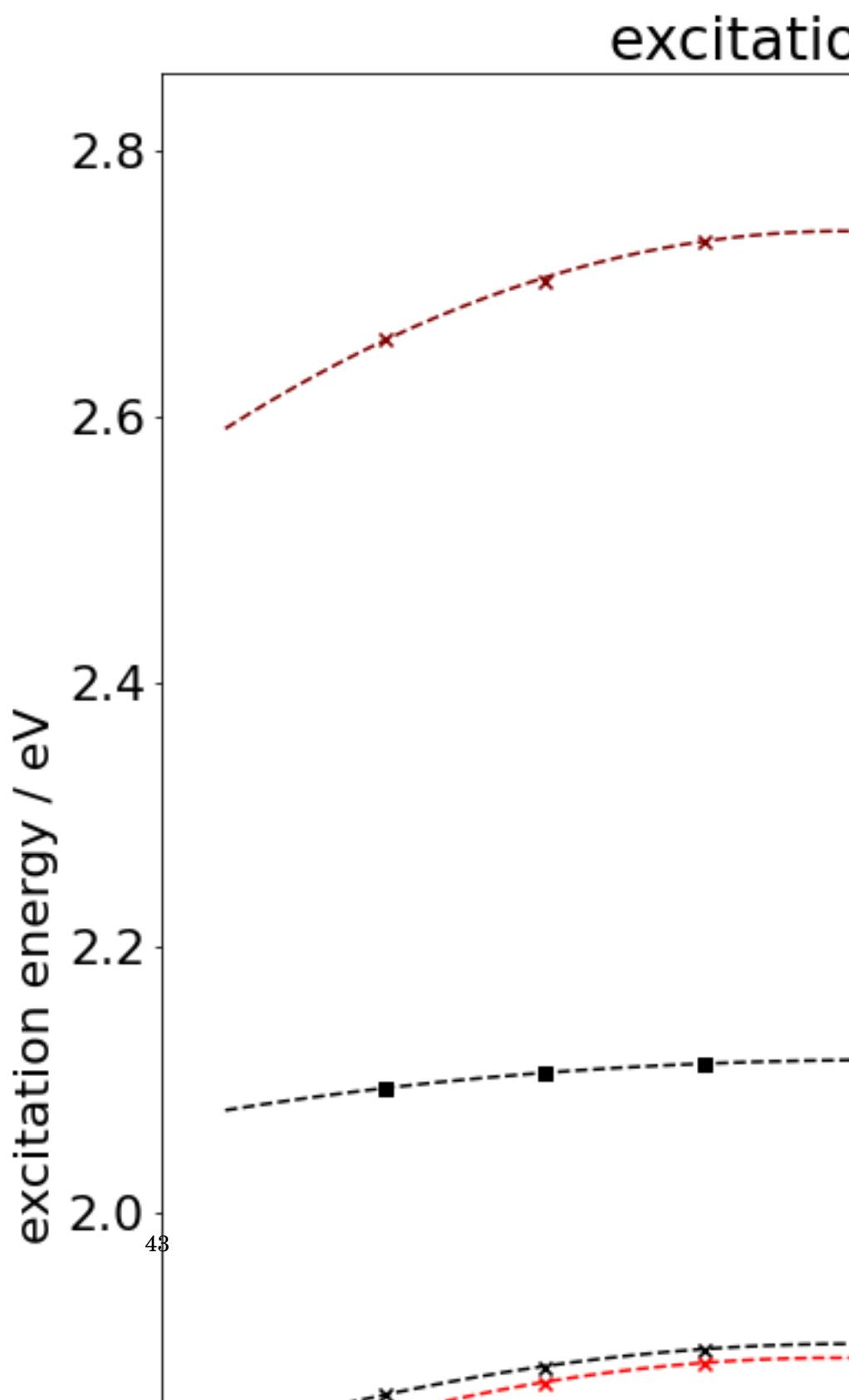
the carbon chain. Therefore the normal modes and transition properties were calculated for a truncated BChla, with a hydrogen atom replacing the phytol tail, instead of a full BChla molecule.

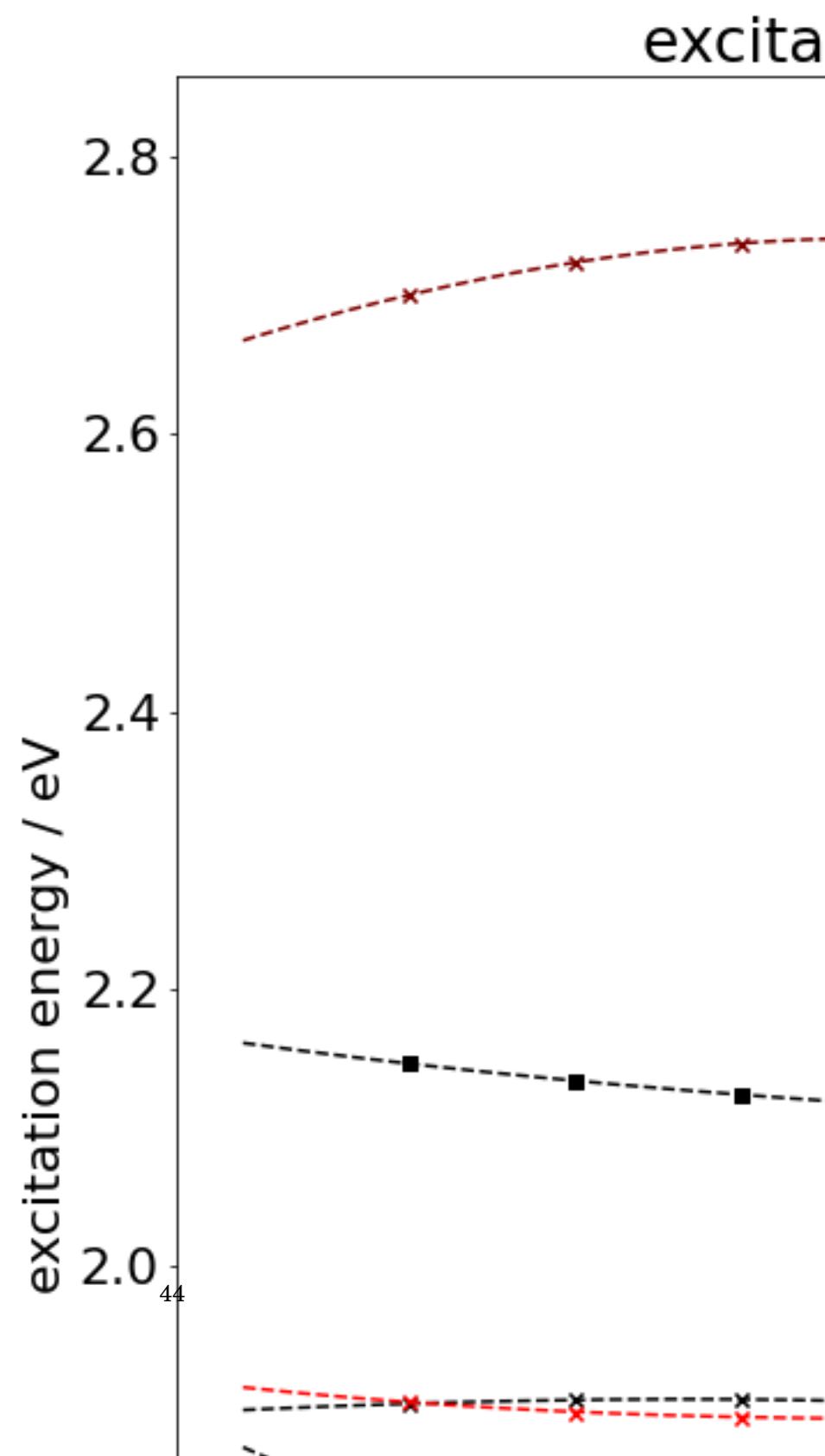
To further reduce the number of calculations needed, only normal modes where strong coupling to the Q_y transition were taken. It's known that only certain symmetry breaking normal modes will couple to the Q_y transition. These can be proxied by looking at the difference in N_A , N_C displacements as the normal mode is scanned. A plot of these values, as well as the moving average, is shown in figure ?? . It is interesting to note that the peaks in the moving average line closely approximates the peaks in the BChla absorption spectrum.

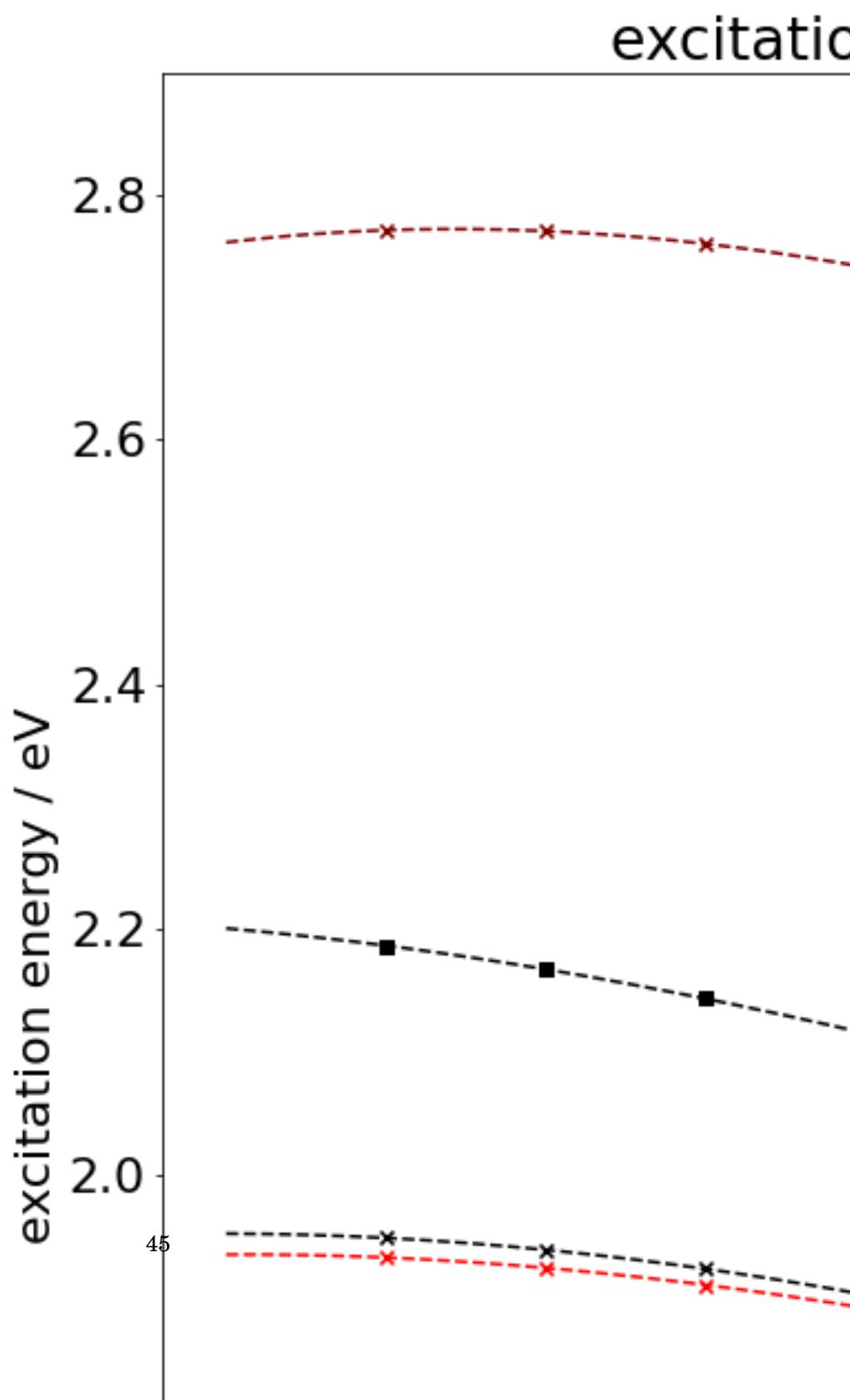
The modes with highest displacement were then chosen for scans, and the results are shown in the set of figures.

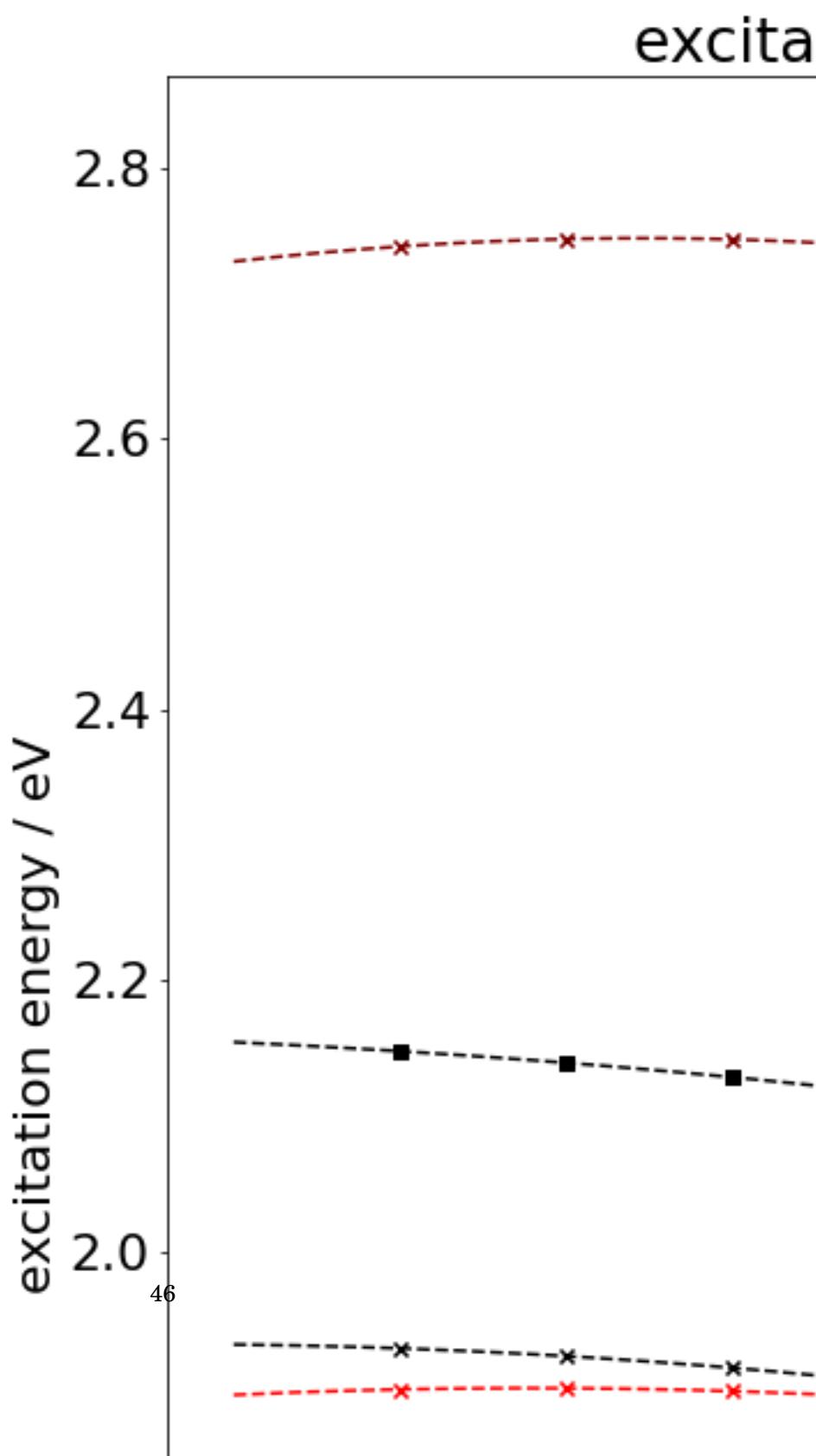
3.3.3 Absorption Spectra

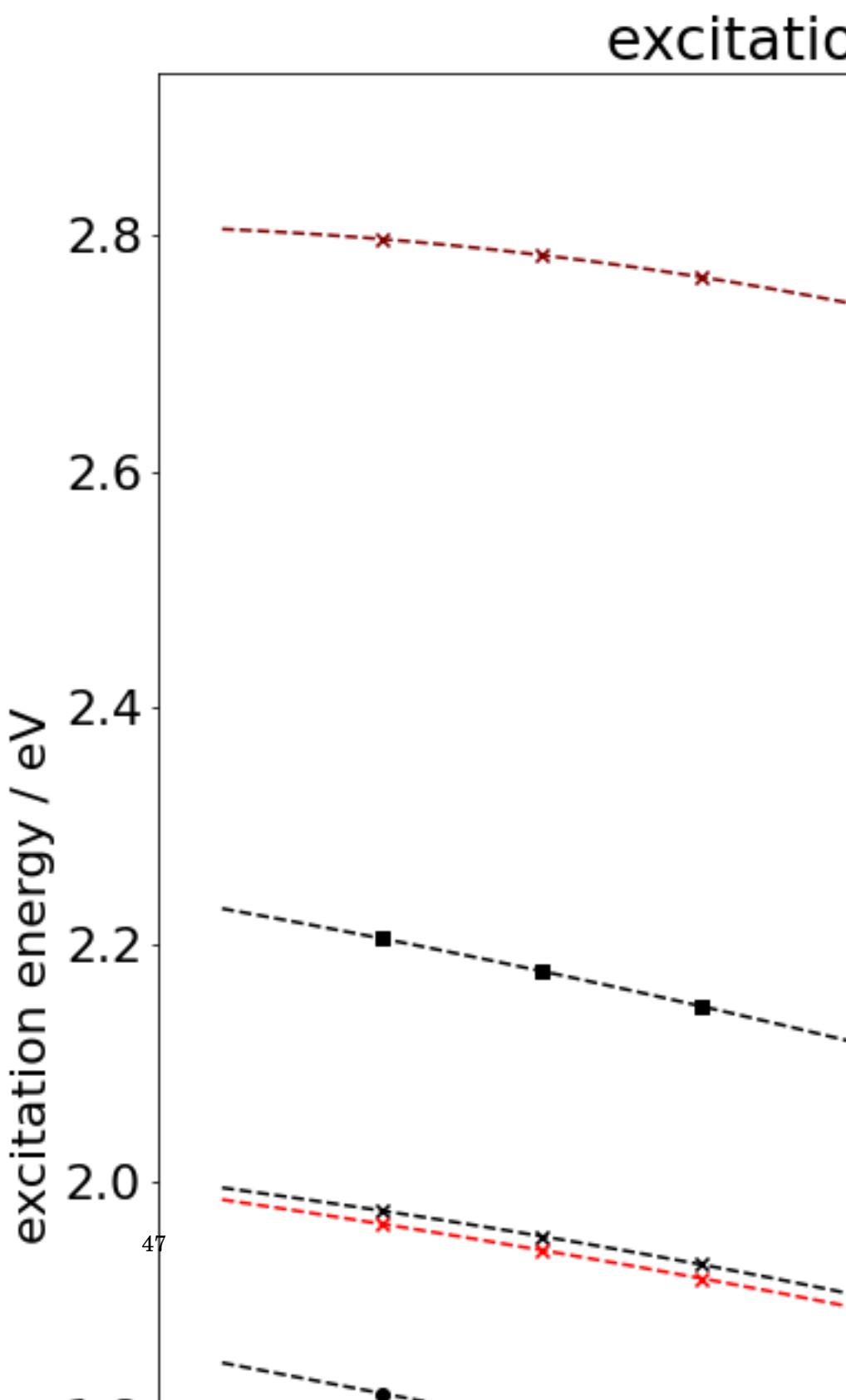


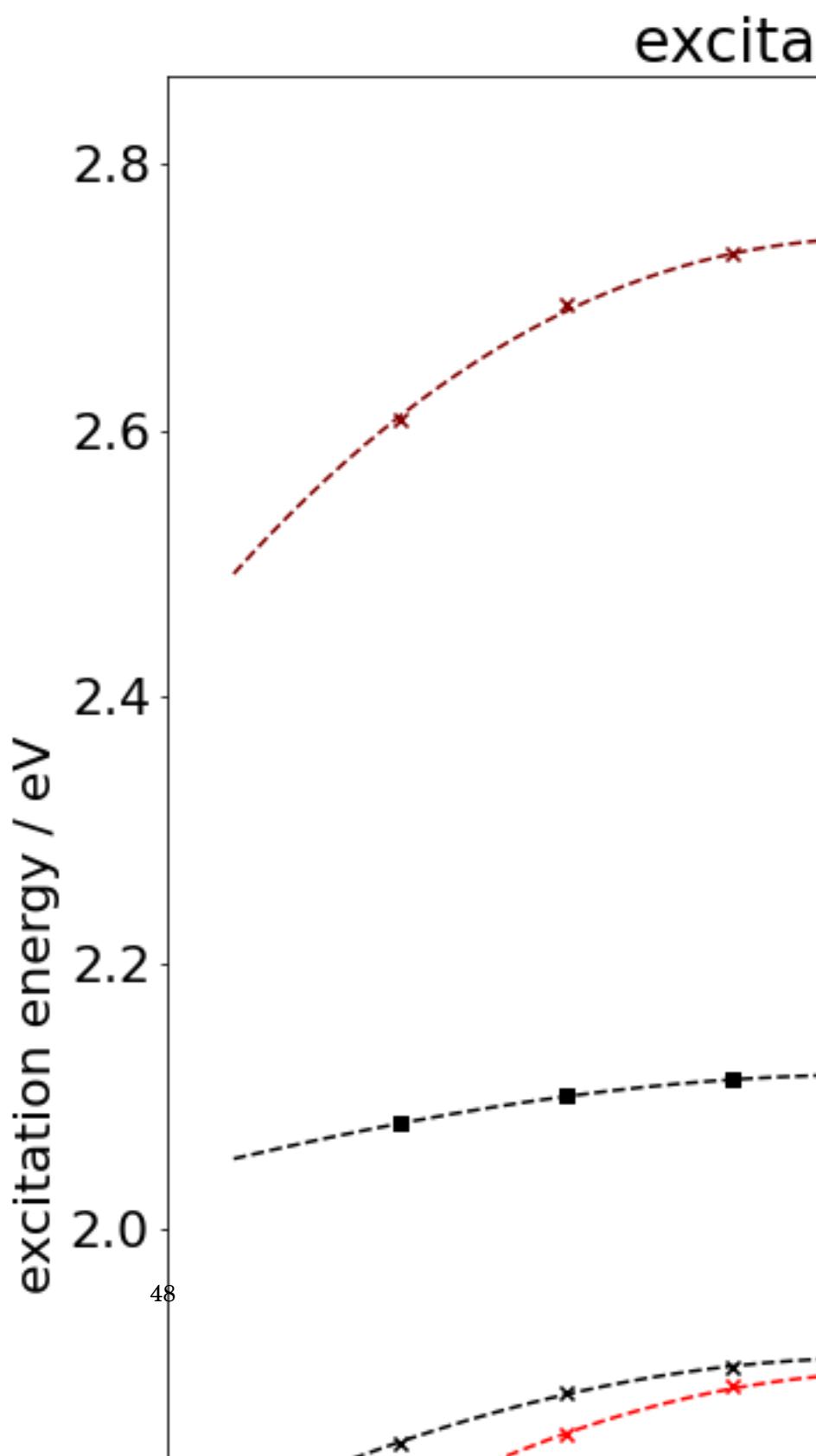


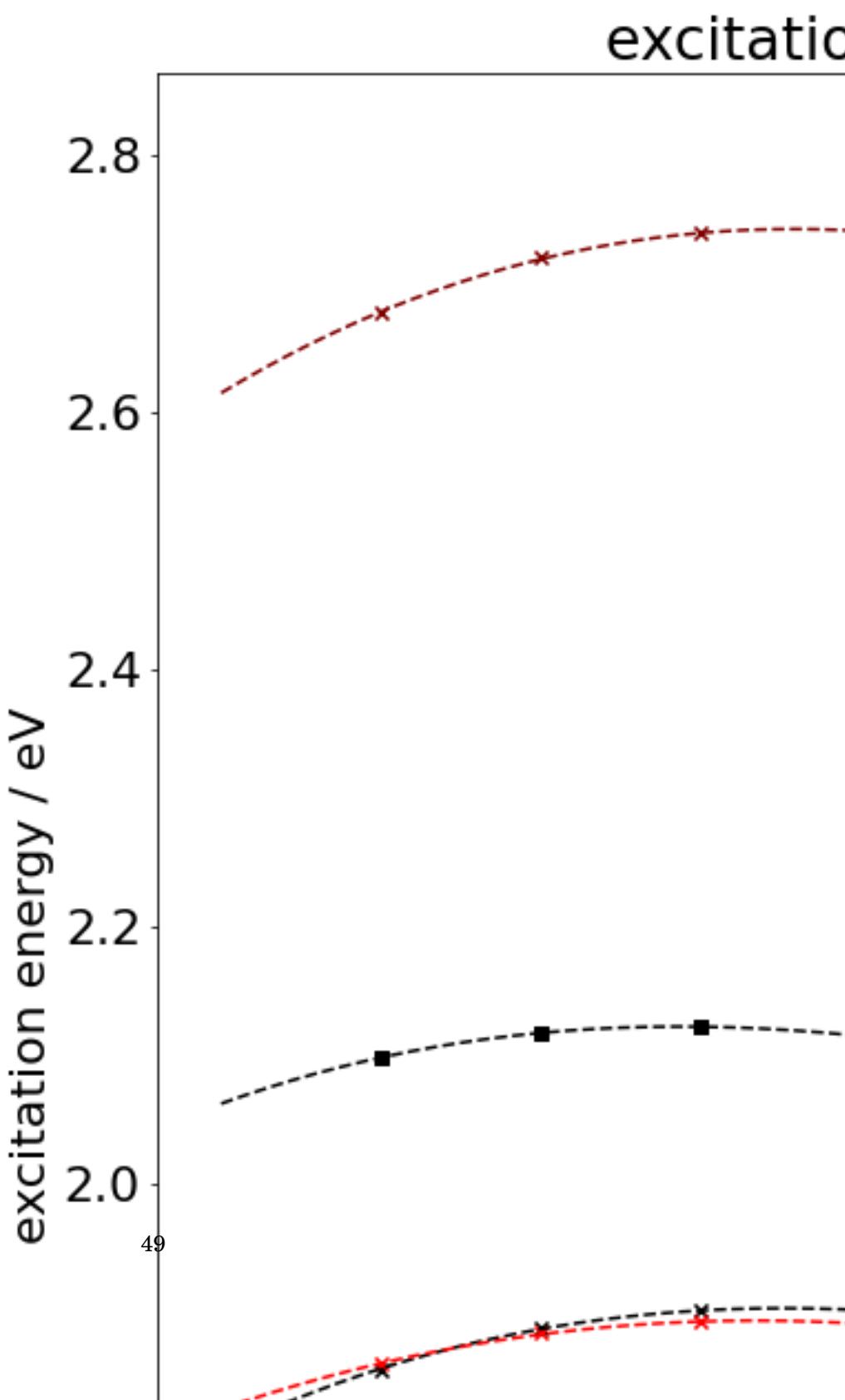


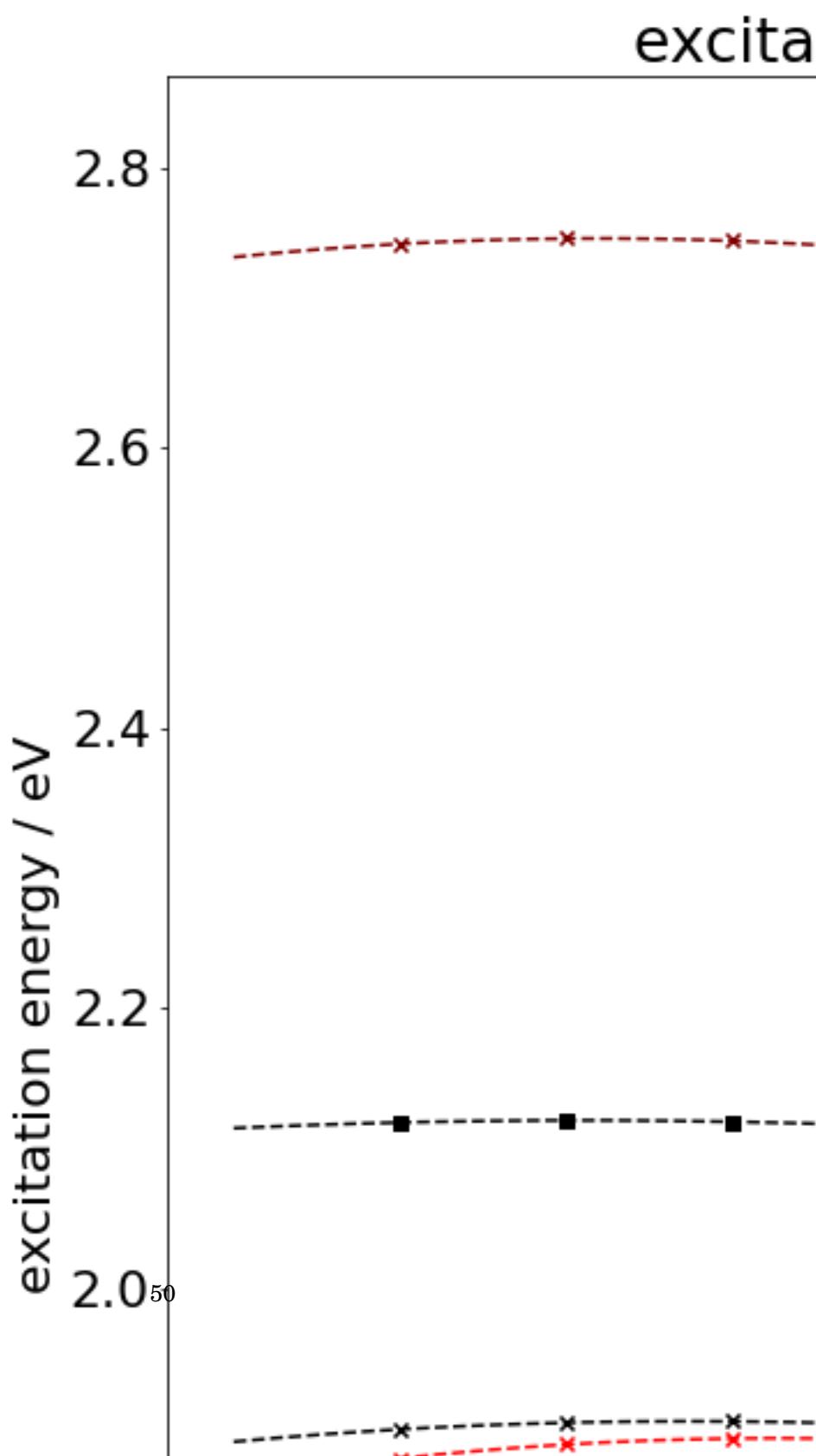


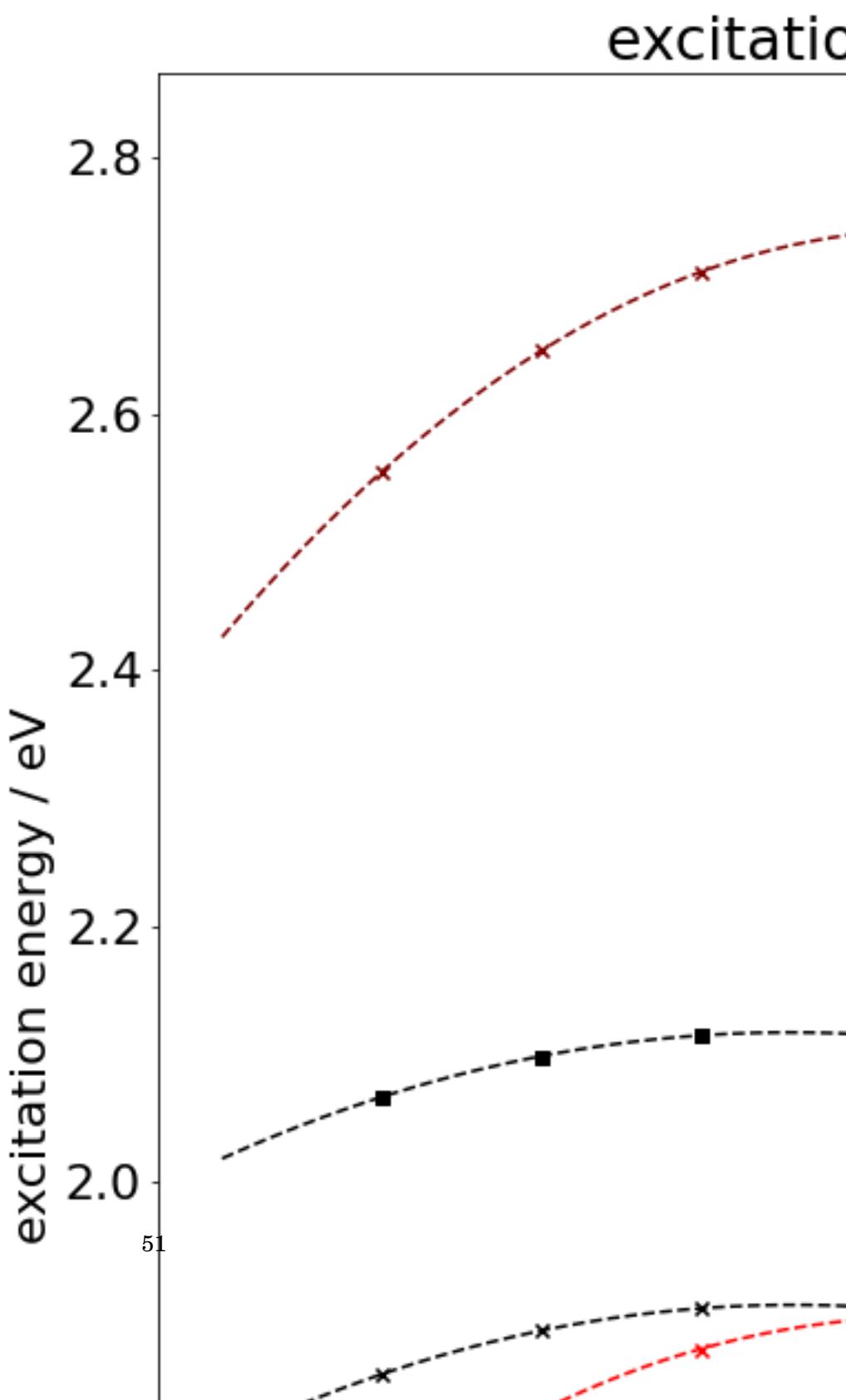


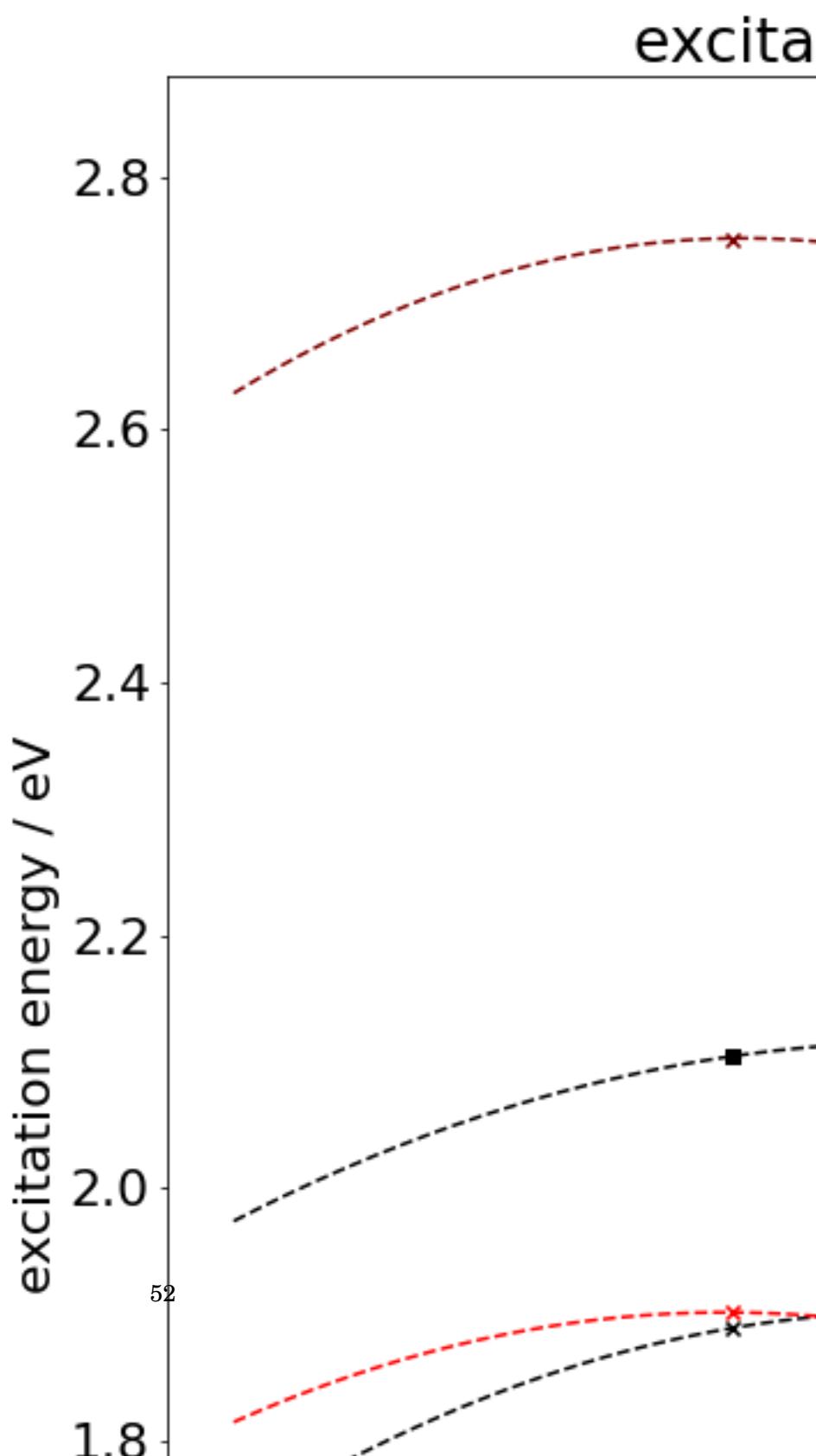


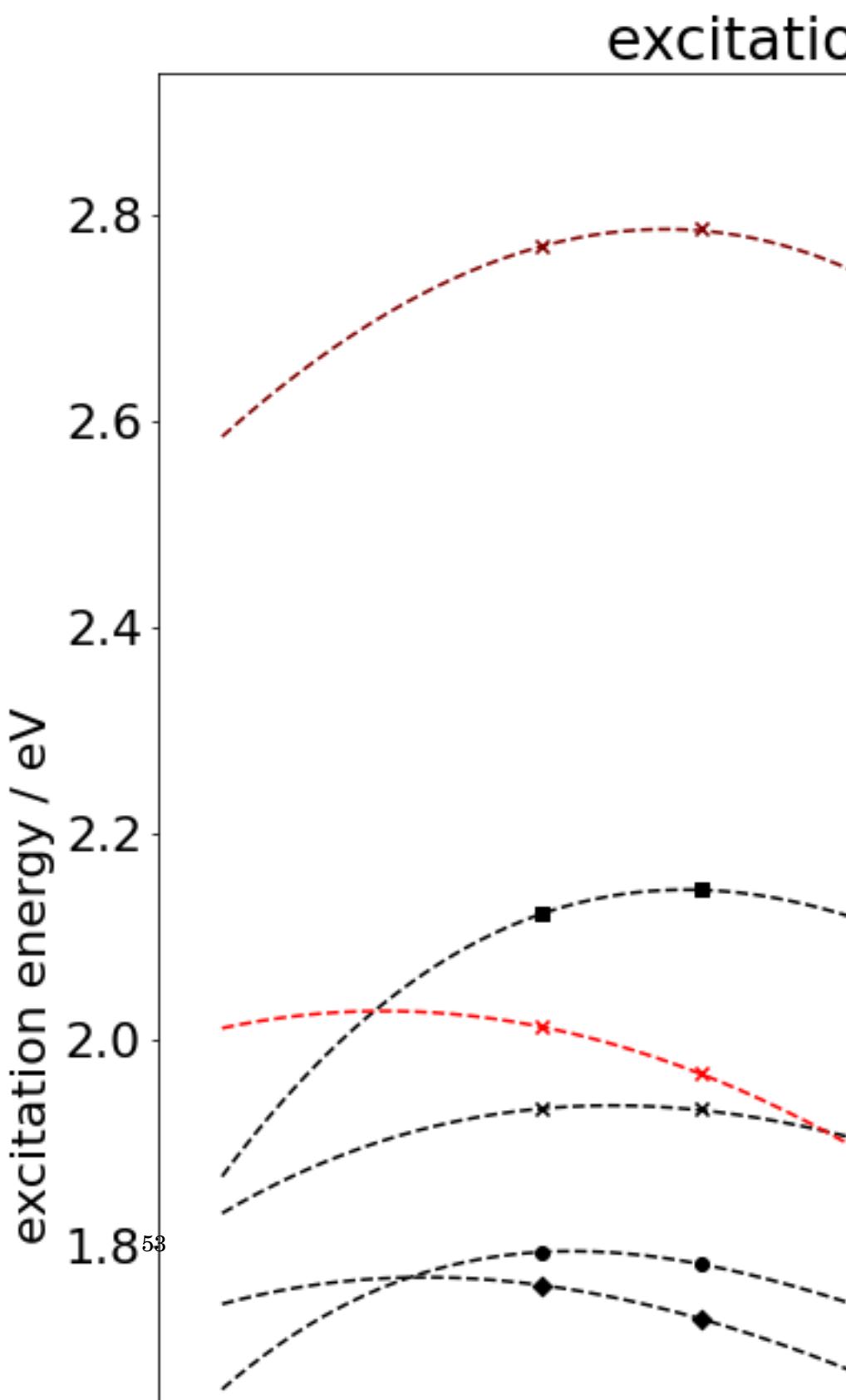












C H A P T E R



EXCITON METHOD

P^{reamble}

4.1 Theory

4.1.1 Exciton States

4.1.2 Embedding

4.2 Truncated Chlorophylls

4.2.1 Rotation

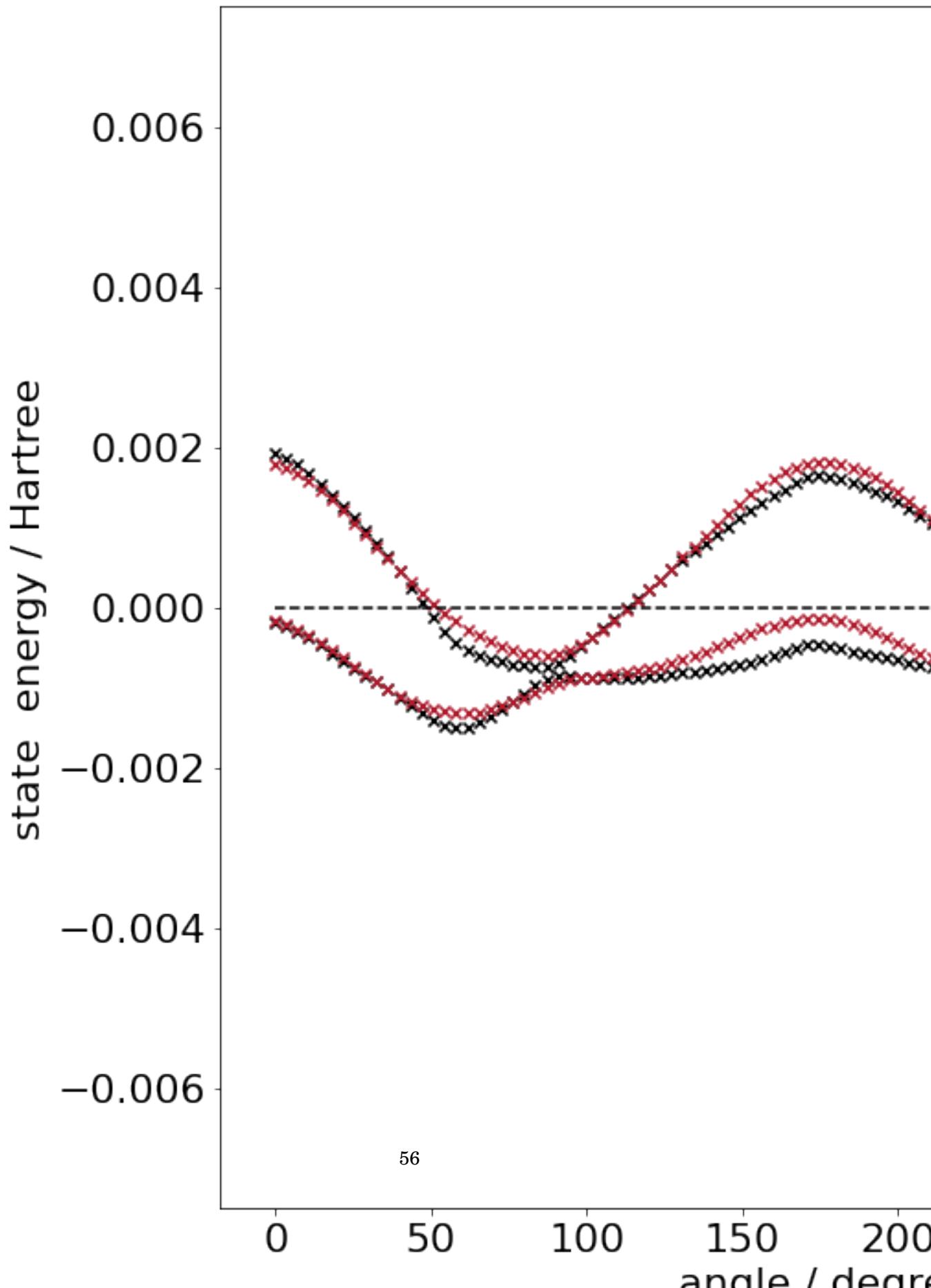
4.2.2 Distance

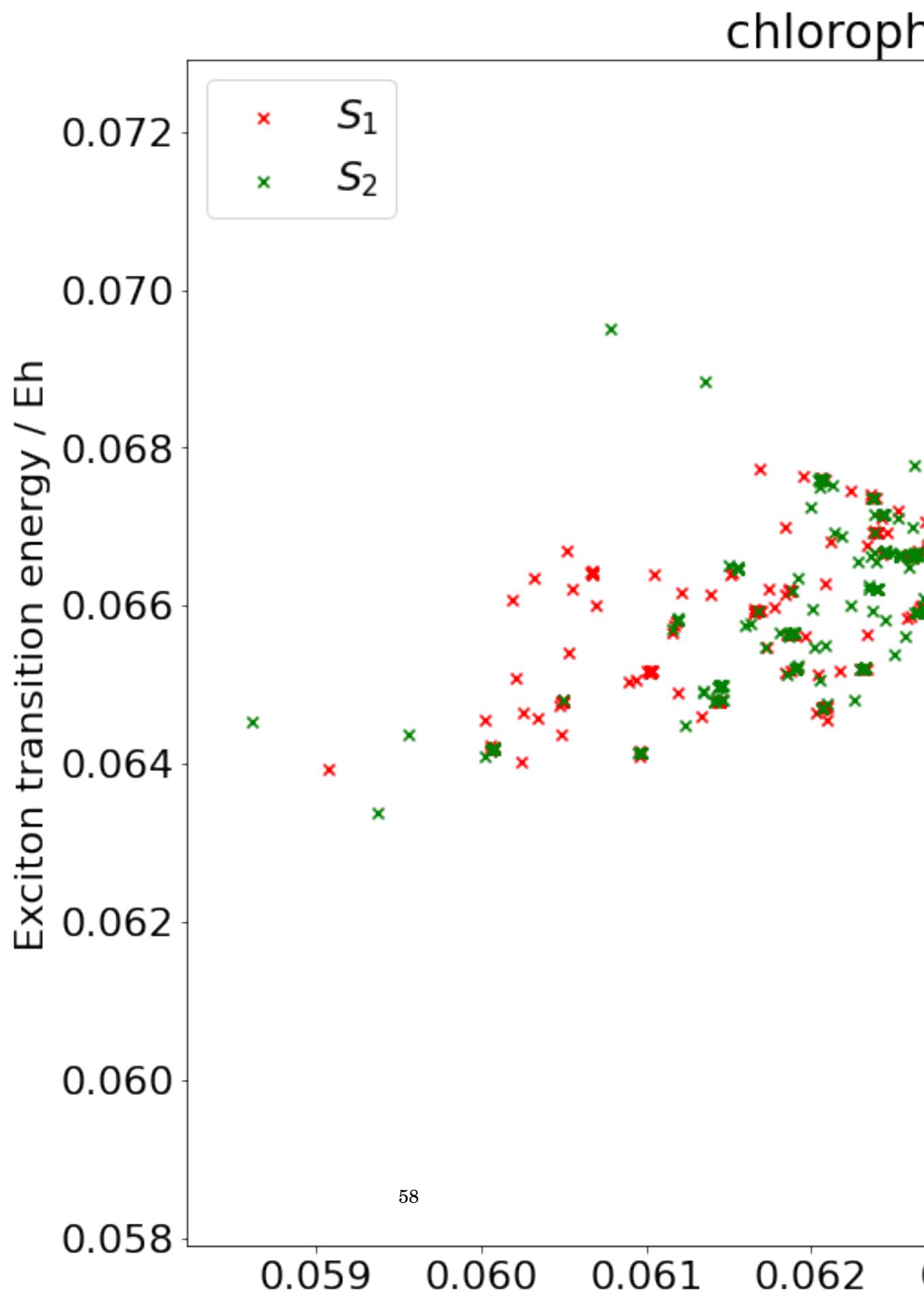
4.3 LHII Pairs

4.3.1 Assignment of States

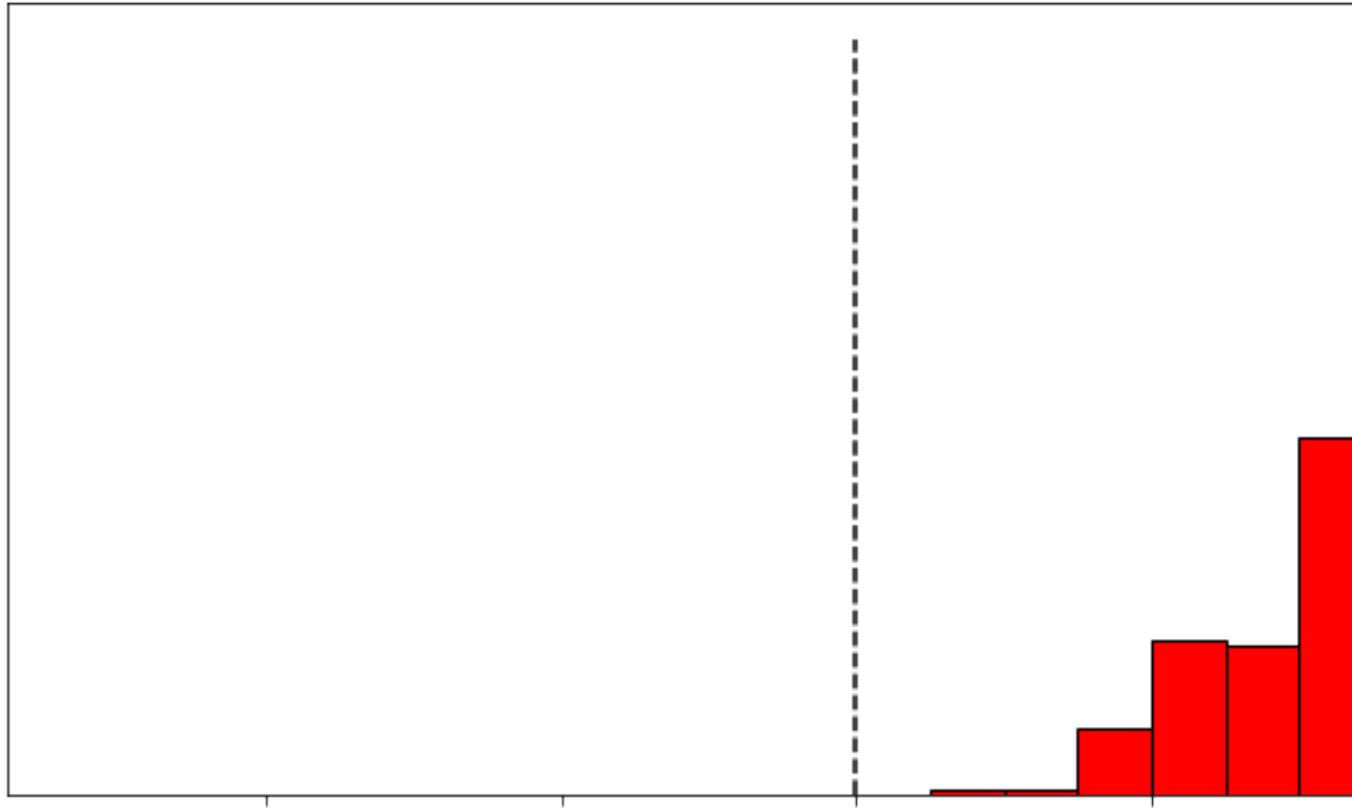
4.3.2 Comparison

Excited states of a ro

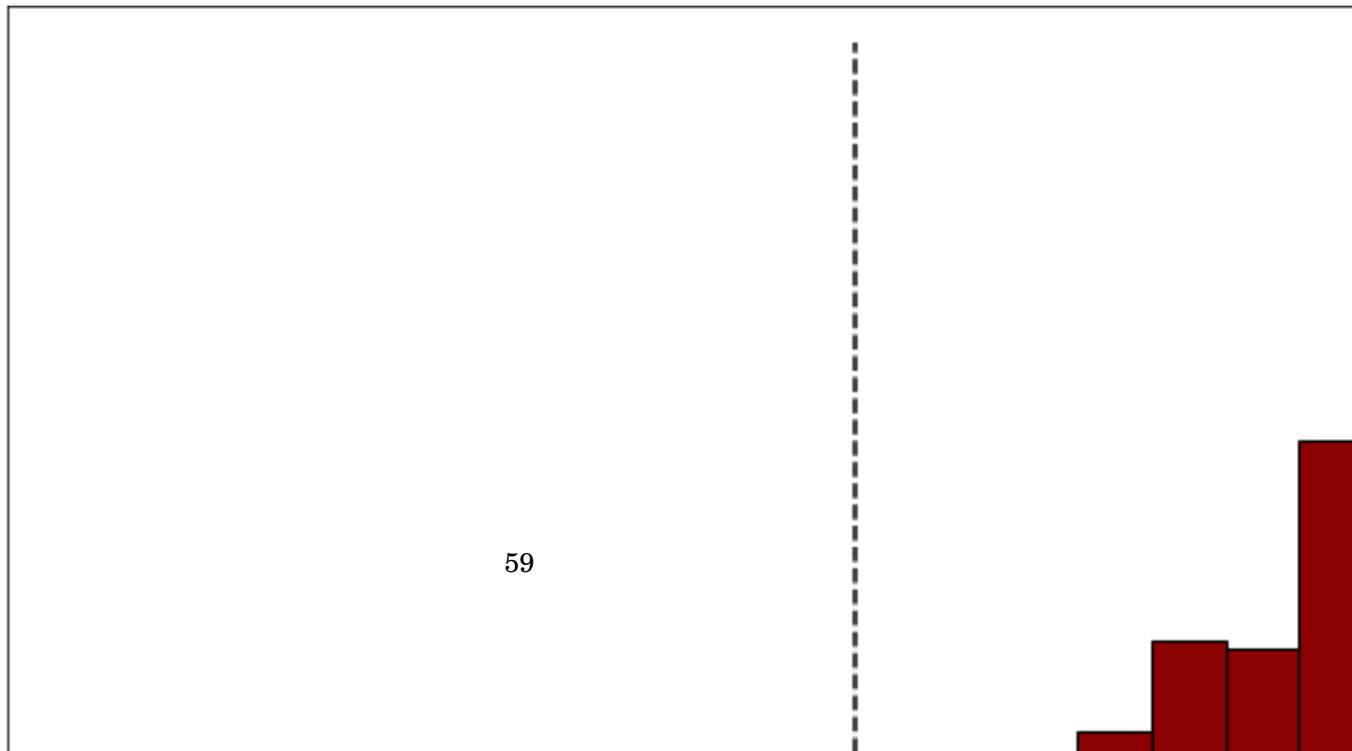




energy ordered



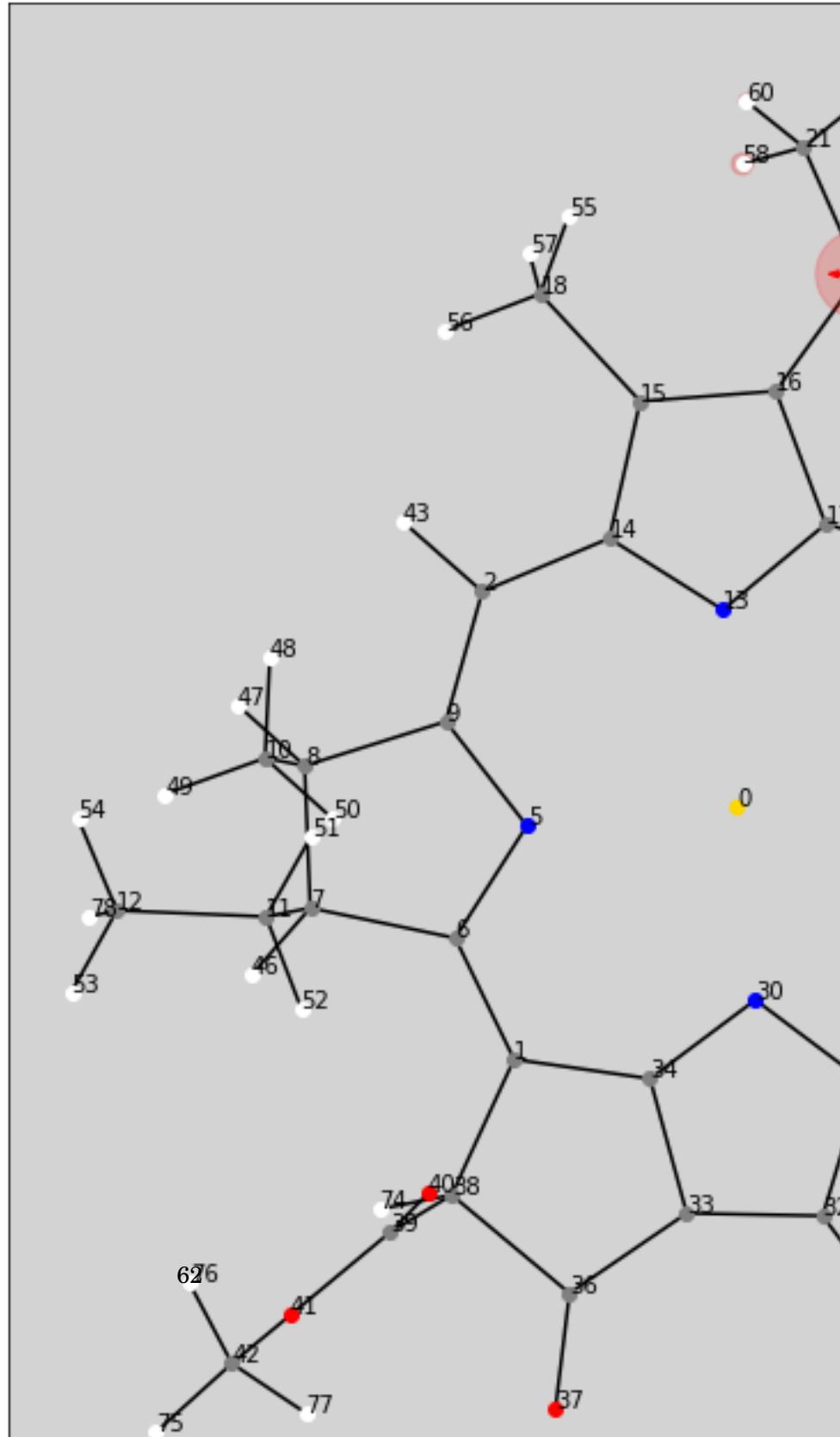
centre ordered

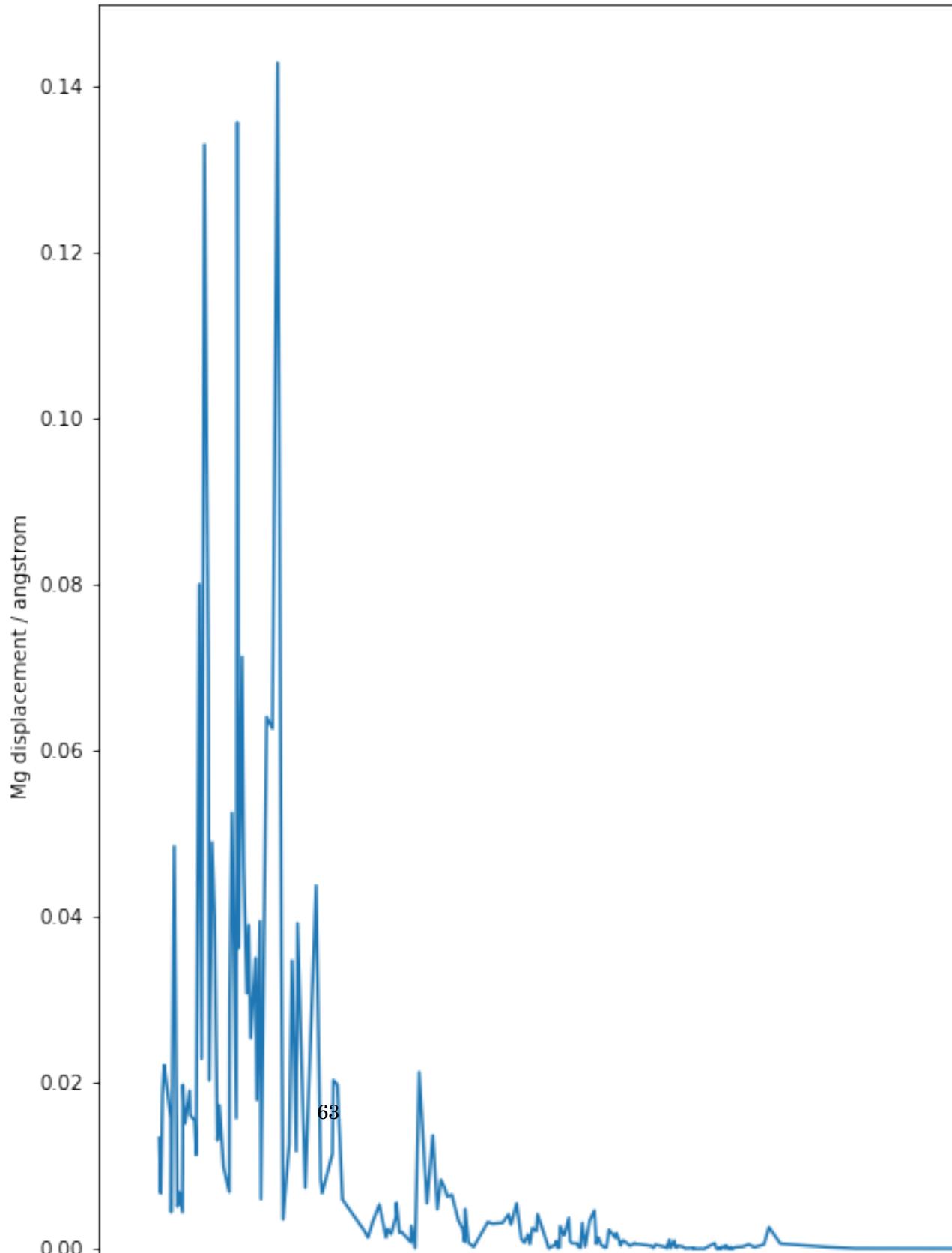


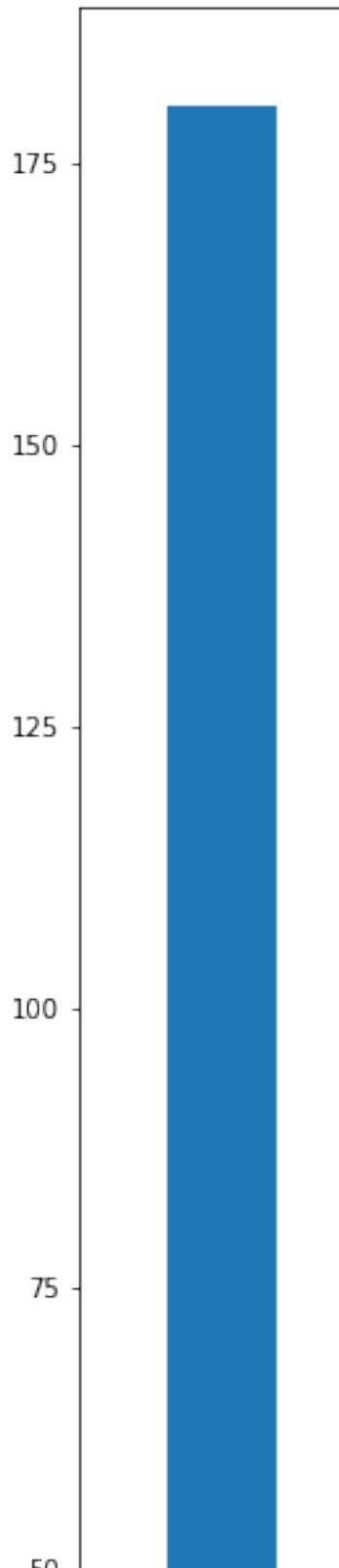
C H A P T E R

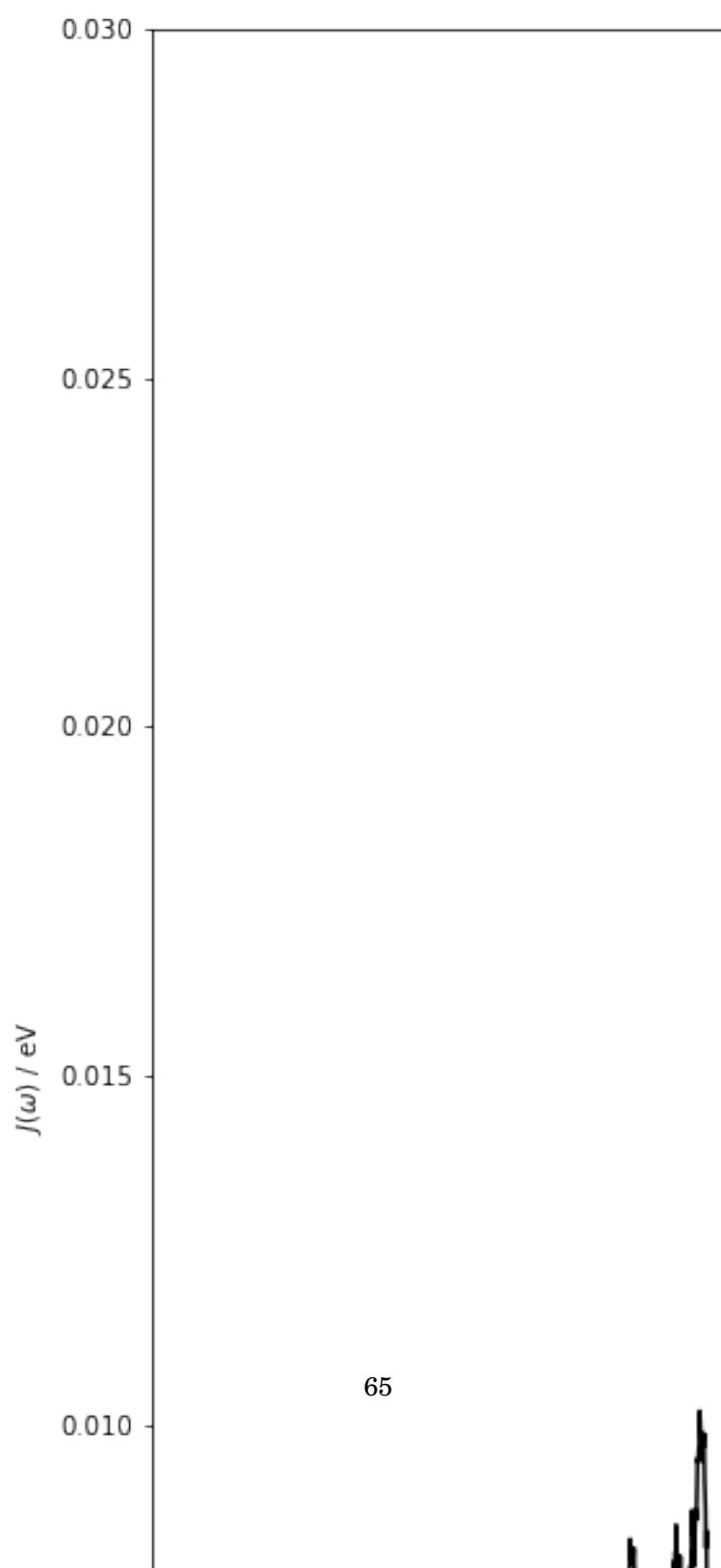


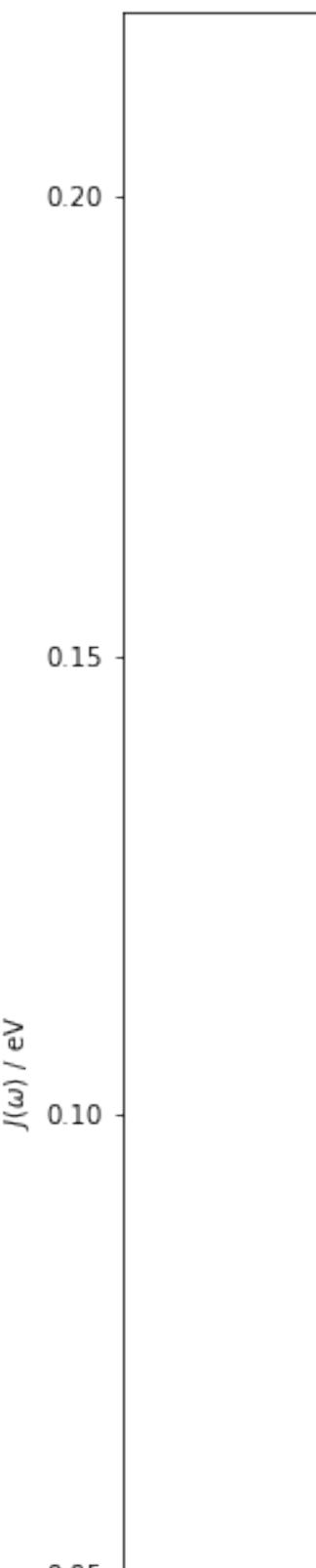
ATOMISTIC MODELLING OF LIGHT HARVESTING COMPLEXES

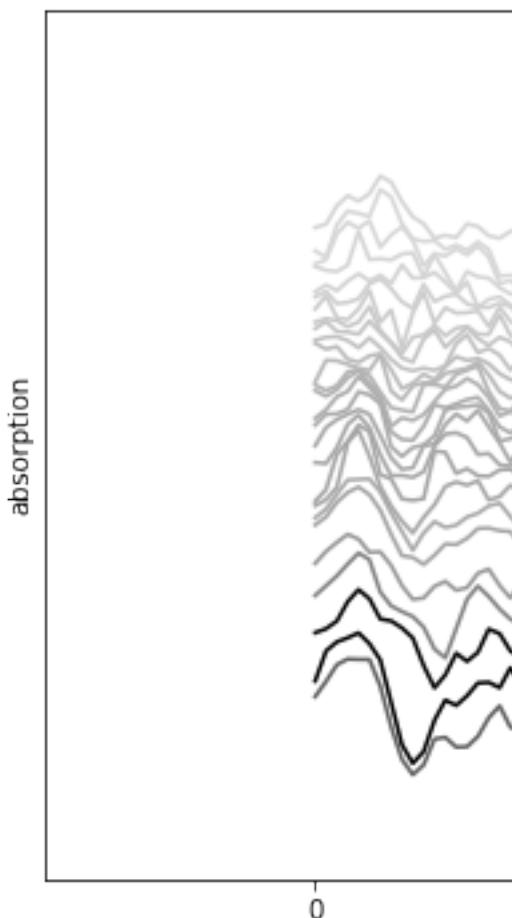


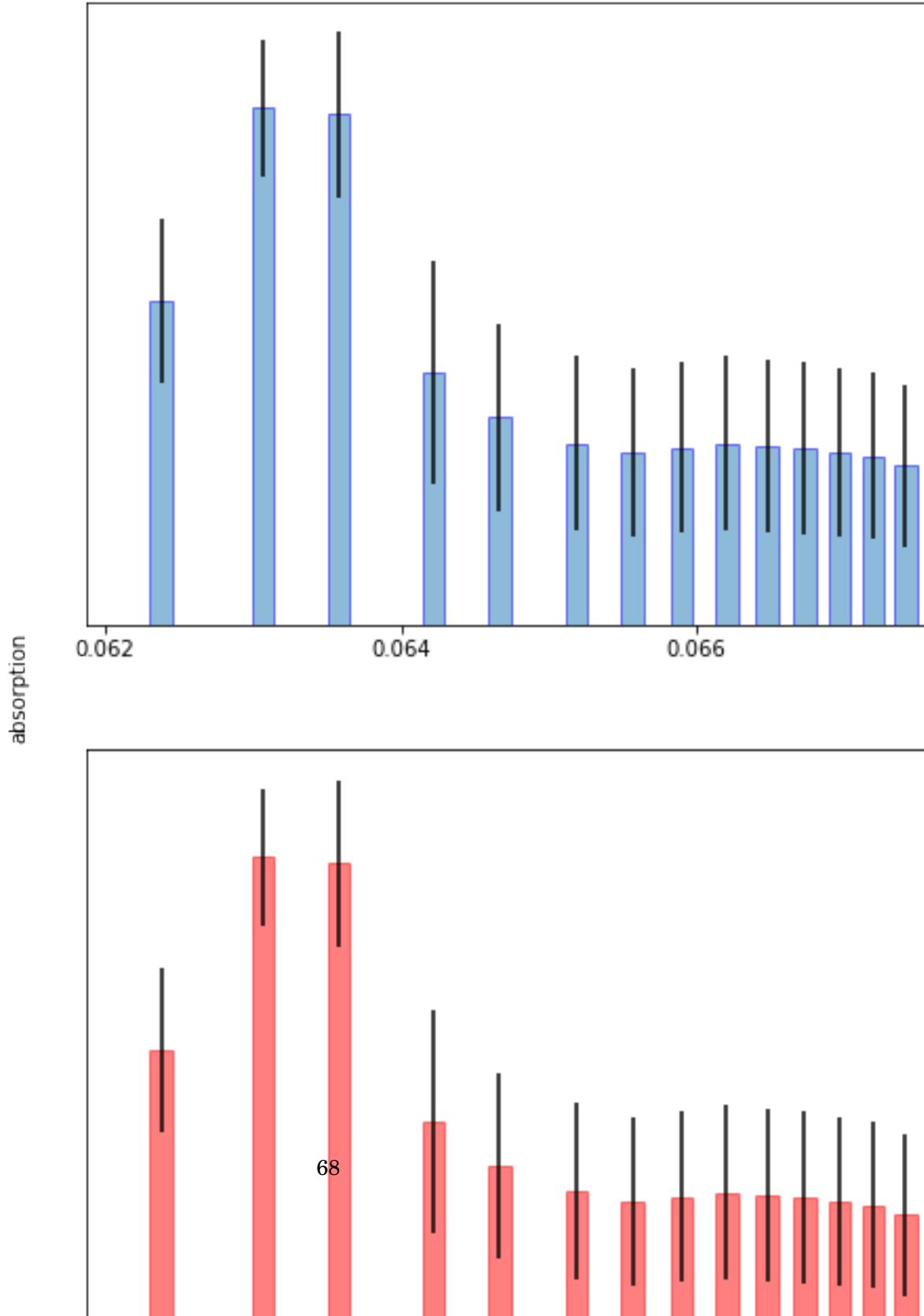












5.1 LHII

5.1.1 Spectral Density Method

5.1.2 Molecular Dynamics Method

5.2 Approximating Spectral Densities

5.2.1 Hessians

5.2.2 Huang Rhys Factors

5.2.3 Chlorophyll distances

5.3 Environmental Effects

5.3.1 Screening

5.3.2 Embedding

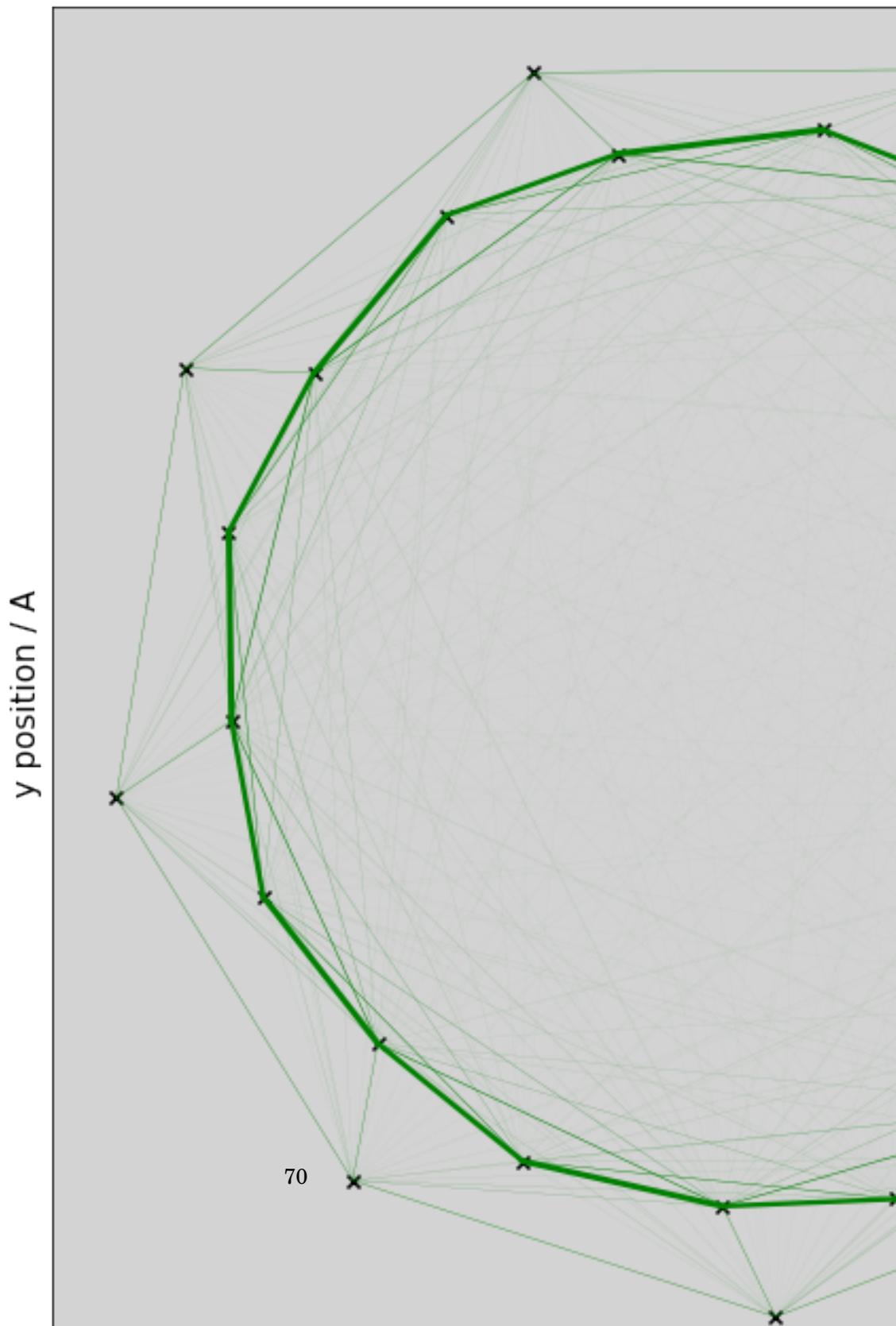
5.4 Sites, states and couplings

5.4.1 Sites

5.4.2 Exciton states

5.4.3 Coupling

5.5 Excitation Energies



C H A P T E R



DISCUSSION

Preamble

6.1 Transition Property Approximations

6.2 Further Investigations into LHII

6.3 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] 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.
- [6] J. GASTEIGER AND M. MARSILI, *A new model for calculating atomic charges in molecules*, Tetrahedron Lett., 19 (1978), pp. 3181–3184.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

BIBLIOGRAPHY

- [11] S. GRIMME AND C. BANNWARTH, *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.
- [12] 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.
- [13] 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.
- [14] A. HELLWEG, S. A. G. GRU“N, AND C. H. HA“TTIG, *Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states*, Phys. Chem. Chem. Phys., 10 (2008), pp. 4119–4127.
- [15] S. W. J. HUNT AND W. A. GODDARD, *Excited States of H₂O using improved virtual orbitals*, Chem. Phys. Lett., 3 (1969), pp. 414–418.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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).
- [21] K. NISHIMOTO AND N. MATAGA, *Electronic Structure and Spectra of Some Nitrogen Heterocycles*, Zeitshrift fur Phys. Chemie Neue Folge, 12 (1957), pp. 335–338.

BIBLIOGRAPHY

- [22] K. OHNO, *Some Remarks on the Pariser-Parr-Pople Method*, Theor. chim. Acta, 2 (1964), pp. 219–227.
- [23] C. C. J. ROOTHAAN, *New Developments in Molecular Orbital Theory*, Rev. Mod. Phys., 23 (1951), p. 69.
- [24] 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.
- [25] 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.
- [26] 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.

