
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	6
2.1.4 Semi-empirical extensions	7
2.2 Benchmarking	9
2.2.1 Reference Data and test set	9
2.2.2 Small Systems	9
2.2.3 Non-orthogonality	13
2.2.4 LHII Chlorophyll	15
2.2.5 xTB methods	18
2.3 Conclusions	24
3 Chlorophyll specific methods	27
3.1 Theory	27
3.1.1 Full solutions TD-DFT equation	27
3.1.2 Approximations to Solutions	28

TABLE OF CONTENTS

3.1.3	Integrals approximations	29
3.1.4	Q_y Transition	31
3.1.5	Changes to xTB parameters	32
3.1.6	Transition and Excited State Density from the Ground State	35
3.2	Parameterization	37
3.2.1	Objective Function	37
3.2.2	Minimisation Algorithms	38
3.2.3	Reference Data	41
3.2.4	Results	45
3.3	Benchmarking and Cross-validation	46
3.3.1	Transition properties	46
3.3.2	Potential Energy Surfaces	48
3.3.3	Absorption Spectra	61
3.4	Conclusions	63
4	Exciton Method	65
4.1	Theory	65
4.1.1	Exciton States	65
4.1.2	Embedding	65
4.2	Truncated Chlorophylls	65
4.2.1	Rotation	65
4.2.2	Distance	65
4.3	LHII Pairs	65
4.3.1	Assignment of States	65
4.3.2	Comparison	65
5	Atomistic Modelling of Light Harvesting Complexes	71
5.1	LHII	79
5.1.1	Spectral Density Method	79
5.1.2	Molecular Dynamics Method	79
5.2	Approximating Spectral Densities	79
5.2.1	Hessians	79
5.2.2	Huang Rhys Factors	79
5.2.3	Chlorophyll distances	79
5.3	Environmental Effects	79
5.3.1	Screening	79
5.3.2	Embedding	79
5.4	Sites, states and couplings	79
5.4.1	Sites	79

TABLE OF CONTENTS

5.4.2	Exciton states	79
5.4.3	Coupling	79
5.5	Excitation Energies	79
6	Discussion	81
6.1	Transition Property Approximations	81
6.2	Further Investigations into LHII	81
6.3	Coherence	81
A	Appendix A	83
A.1	Electronic Structure Codes	83
A.2	Computational Hardware	83
	Bibliography	85

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.	23
3.1 ref data vals	44
3.2 optimized parameters from SLSQP procedure.	45

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.	11
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.	12
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.	14
2.4 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.	16
2.5 Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 2.2.4, 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.	17
2.6 A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.	20
2.7 The distributions of errors compared to SCS-CC2 transition energies for the methods included in the Δ -xTB benchmarking.	22
3.1 Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.	32
3.2 The HOMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.	33
3.3 The LUMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.	34
3.4 Positions of the Nelder-Mead simplex vertices for the HimmelBlau function.	42
3.5 Positions of iterative solution vector \mathbf{x} from SLSQP for the HimmelBlau function.	43
3.6 fig:training scatter	47
3.7 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.	48

LIST OF FIGURES

3.8 Chlorophyll in diethyl ether.	62
4.1 caption	66
4.2 caption	67

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

CHAPTER



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[28]. 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 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

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:

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

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 [25], generally given as an eigenvalue problem:

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

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,

$$(2.3) \quad \mathbf{F} = f(\mathbf{C}^{n-1})$$

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

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

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

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

where μ, 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, 20], without considerable losses of accuracy in certain cases, especially for HOMO-LUMO transitions [19]. 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].

2.1.2 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),

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

. 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].

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 \mathbf{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

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

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

$$(2.9) \quad S_{jk}^{21} = \langle\phi_j^2|\phi_k^1\rangle$$

. The dependence on the adjunct of the overlap can be understood using Löwdin's normal rules for non-orthogonal determinants [21]. In the same way, the transition dipole moment is given by

$$(2.10) \quad \langle\Psi_2|\hat{\mu}|\Psi_1\rangle = \sum_{jk} \mu_{jk}^{21} \text{adj}(\mathbf{S}^{21})_{jk}$$

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

$$(2.11) \quad |\mathbf{S}^{21}| = \langle\Psi_2|\Psi_1\rangle$$

The general definition of the transition dipole

$$(2.12) \quad \mu^{1\rightarrow 2} = \langle\Psi_2|\hat{\mu}|\Psi_1\rangle$$

can be expressed with this transition density matrix as:

$$(2.13) \quad \begin{aligned} \langle\Psi_2|\hat{\mu}|\Psi_1\rangle &= \text{tr}(\hat{\mu}|\Psi_1\rangle\langle\Psi_2|) \\ &= \text{tr}(\hat{\mu}\mathbf{D}^{21}) \end{aligned}$$

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 [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.

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 [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,

$$(2.14) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left(k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left(h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

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) [22] 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:

$$(2.15) \quad q_{nm}^A = \sum_{\mu \in A} C'_{\mu n} C'_{\mu m}$$

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

$$(2.16) \quad \mathbf{C}' = \mathbf{S}^{\frac{1}{2}} \mathbf{C}$$

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 2.2.

Approximations to full 2 electron integrals are given by charge-charge interaction damped by the MNO[23][24][18] 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.

2.2 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.

2.2.1 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].

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, 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:

$$(2.17) \quad \Delta E_S = 2\Delta E^{i,\alpha \rightarrow a, \alpha} - \Delta E^{i,\alpha \rightarrow a, \beta}$$

was used to correct for the true singlet excitation energy ΔE_S [30].

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 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 2.2 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.

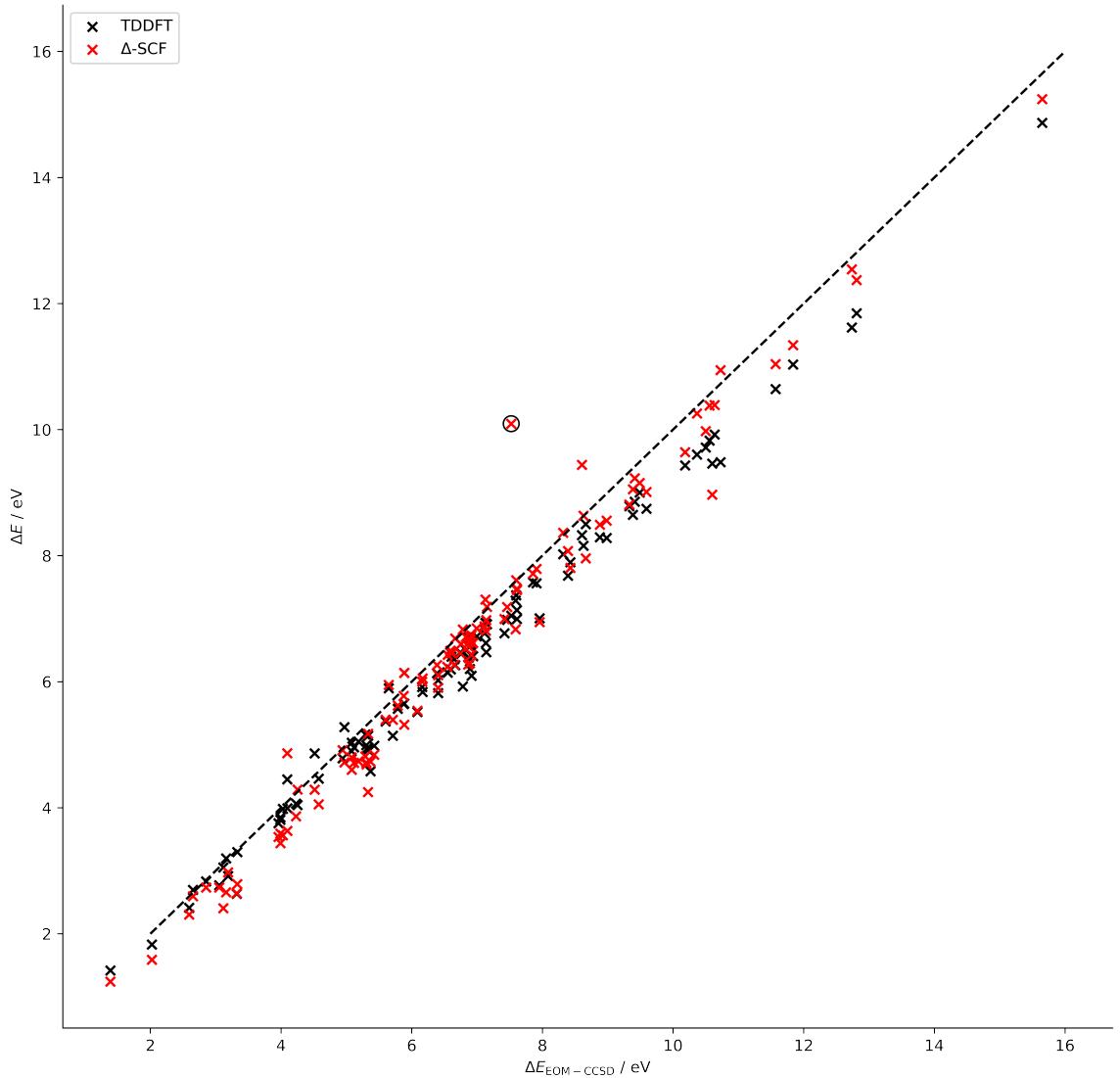


Figure 2.1: Transition energies ΔE from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD energies, with the line $y = x$ (dashed) for reference. The ethene dimer outlier has been circled.

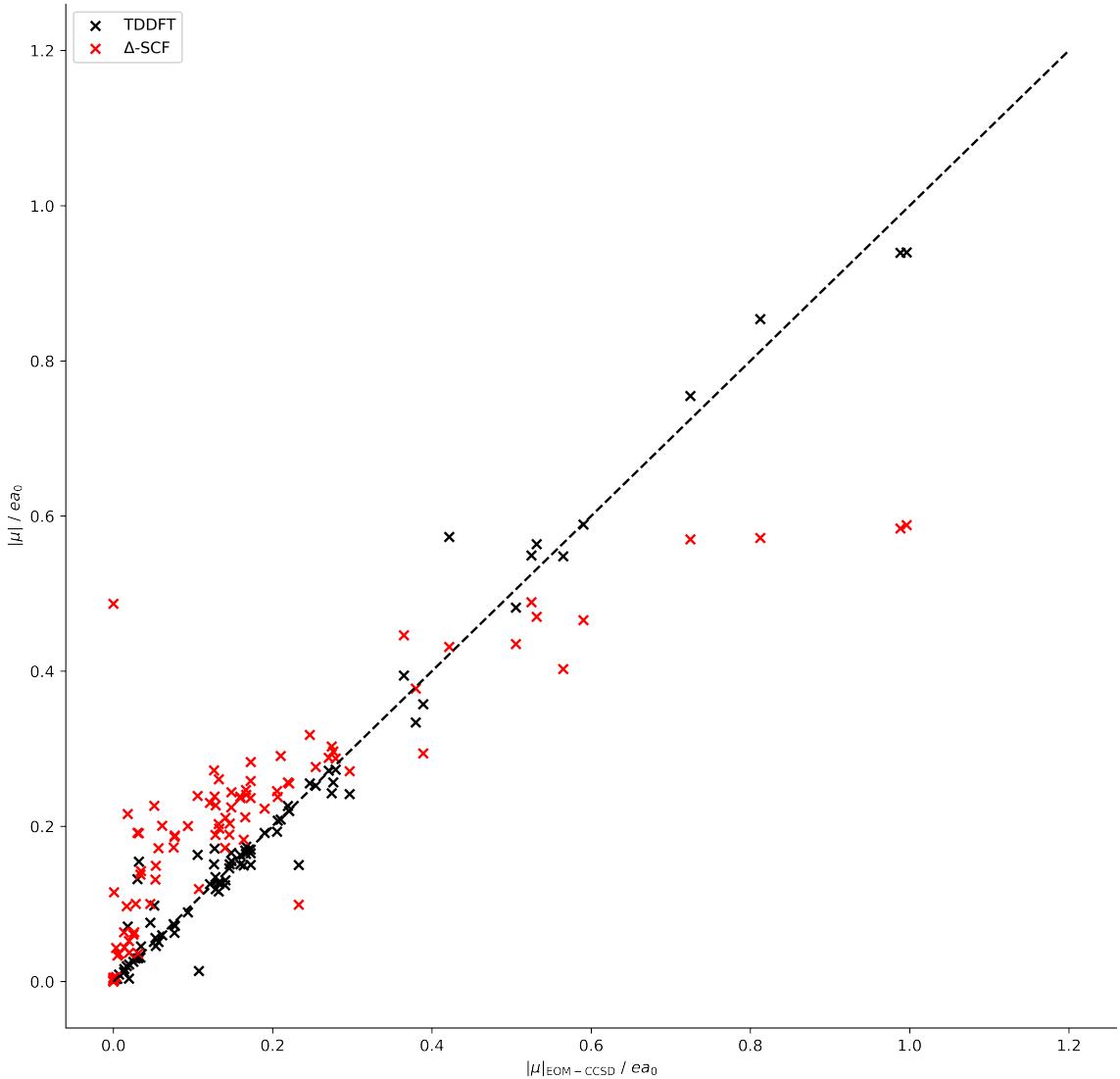


Figure 2.2: Transition dipole magnitudes from TD-DFT (black) and Δ -SCF (red) plotted against EOM-CCSD transition dipole magnitudes, with the line $y = x$ (dashed) for reference.

2.2.3 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,

$$(2.18) \quad S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle$$

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

$$(2.19) \quad q^{21} = \langle \Psi^2 | \Psi^2 \rangle$$

, 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

$$(2.20) \quad |\Psi_{\tilde{\nu}}\rangle = \sum_{\nu} |\Psi_{\nu}\rangle \left[\mathbf{S}^{-\frac{1}{2}} \right]_{\nu\tilde{\nu}}$$

where \mathbf{S} here is a block matrix

$$(2.21) \quad \mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

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 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 orthogonalisation. It should be noted that whilst this effect is dependent on how large the overlap may be, and it could be argued that with a small overlap this effect may not be large, having any large translation of the molecule (on the order of hundreds of angstroms) can lead to nonphysical transition dipole

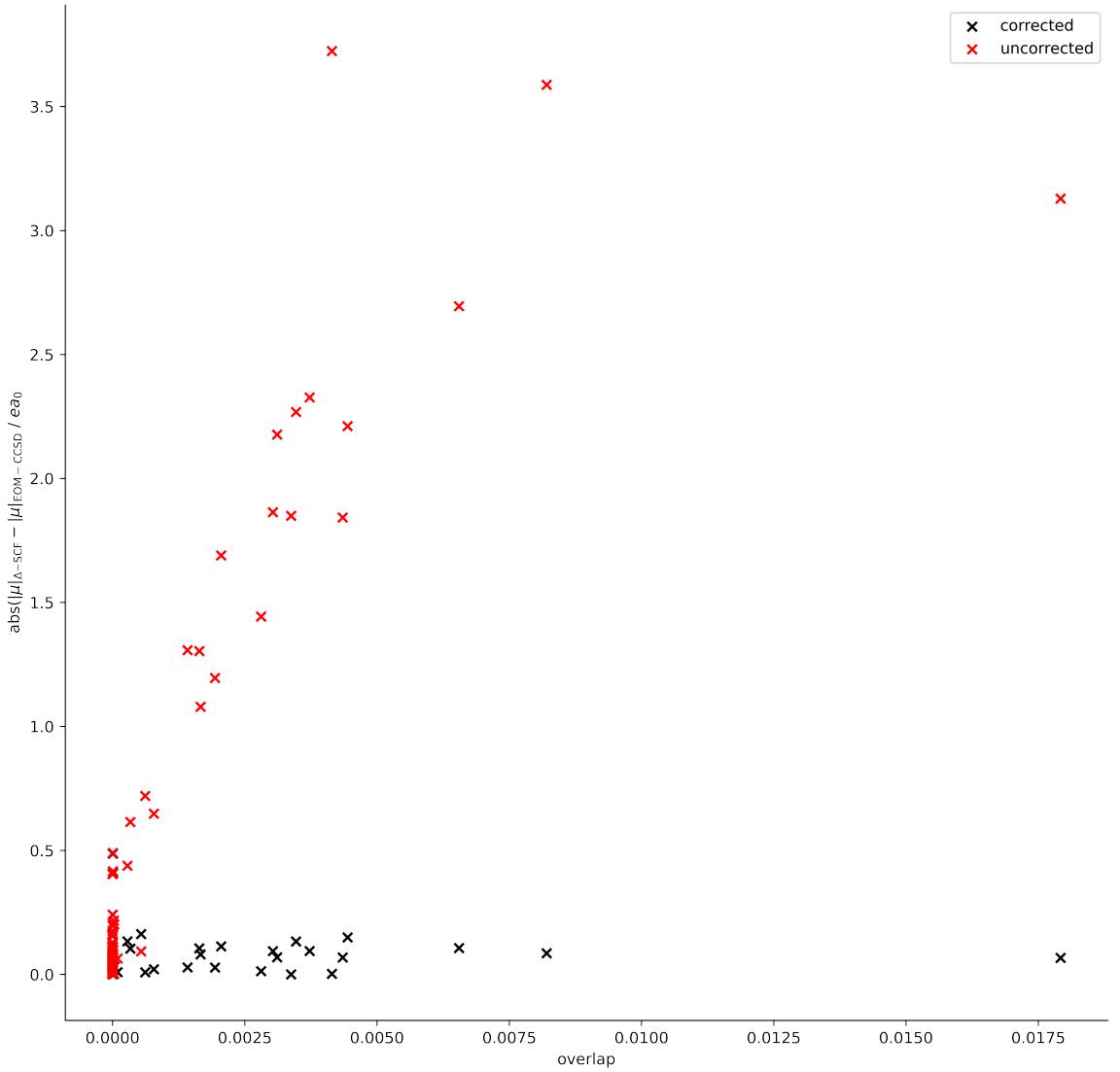


Figure 2.3: The absolute value of the error in transition dipole magnitude between Δ -SCF and EOM-CCSD, plotted against the Δ -SCF overlap of the ground and excited state. All systems were translated by 100 Å in all cartesian axes. Transition dipole magnitudes calculated without any correction are shown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

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.

2.2.4 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, 26] 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 [27].

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.

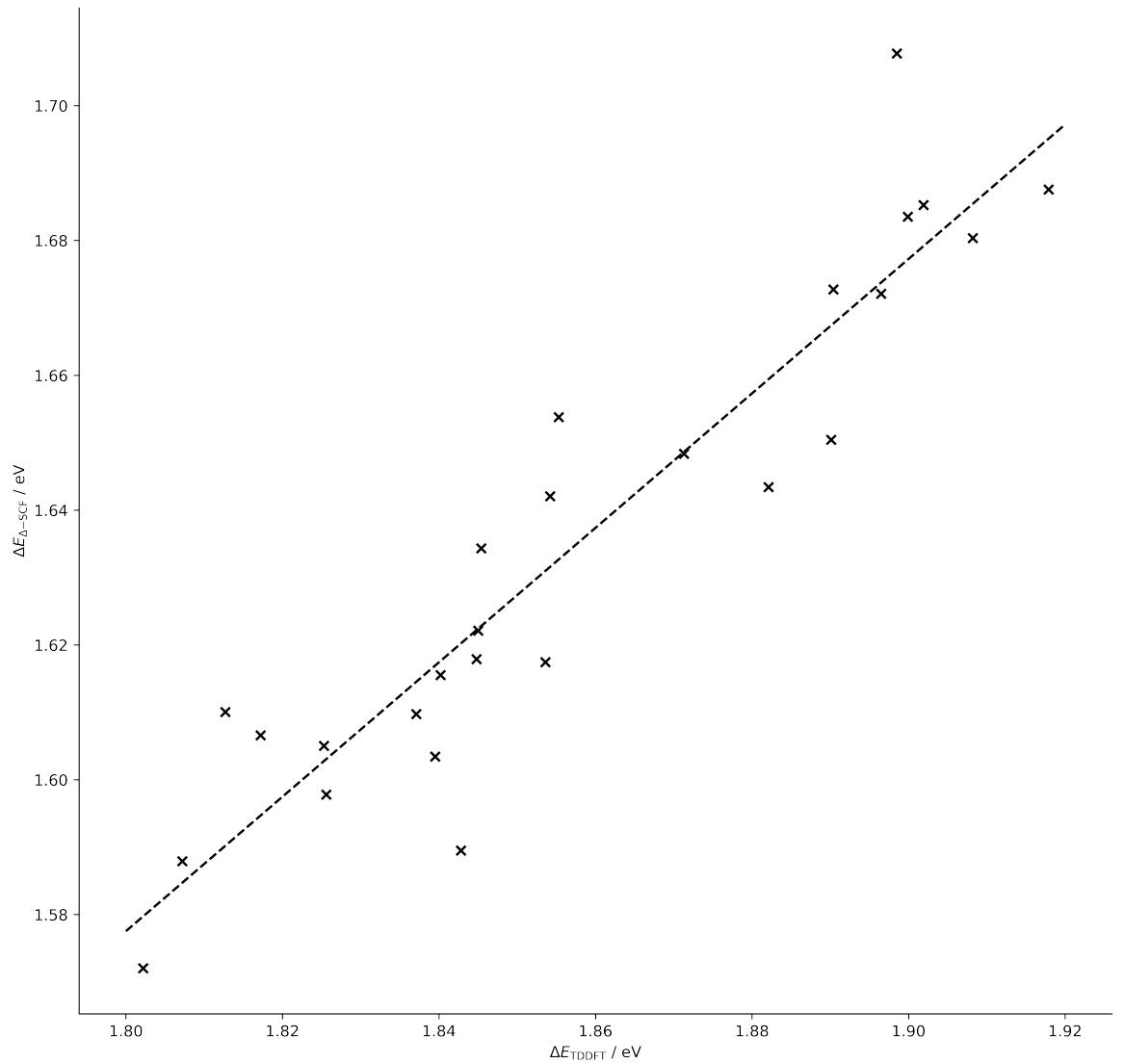


Figure 2.4: Transition energies from Δ -SCF for 26 chlorophyll geometries from the LHII protein of purple bacteria, plotted against energies from TD-DFT. The line of best fit ($R^2 = 0.87$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

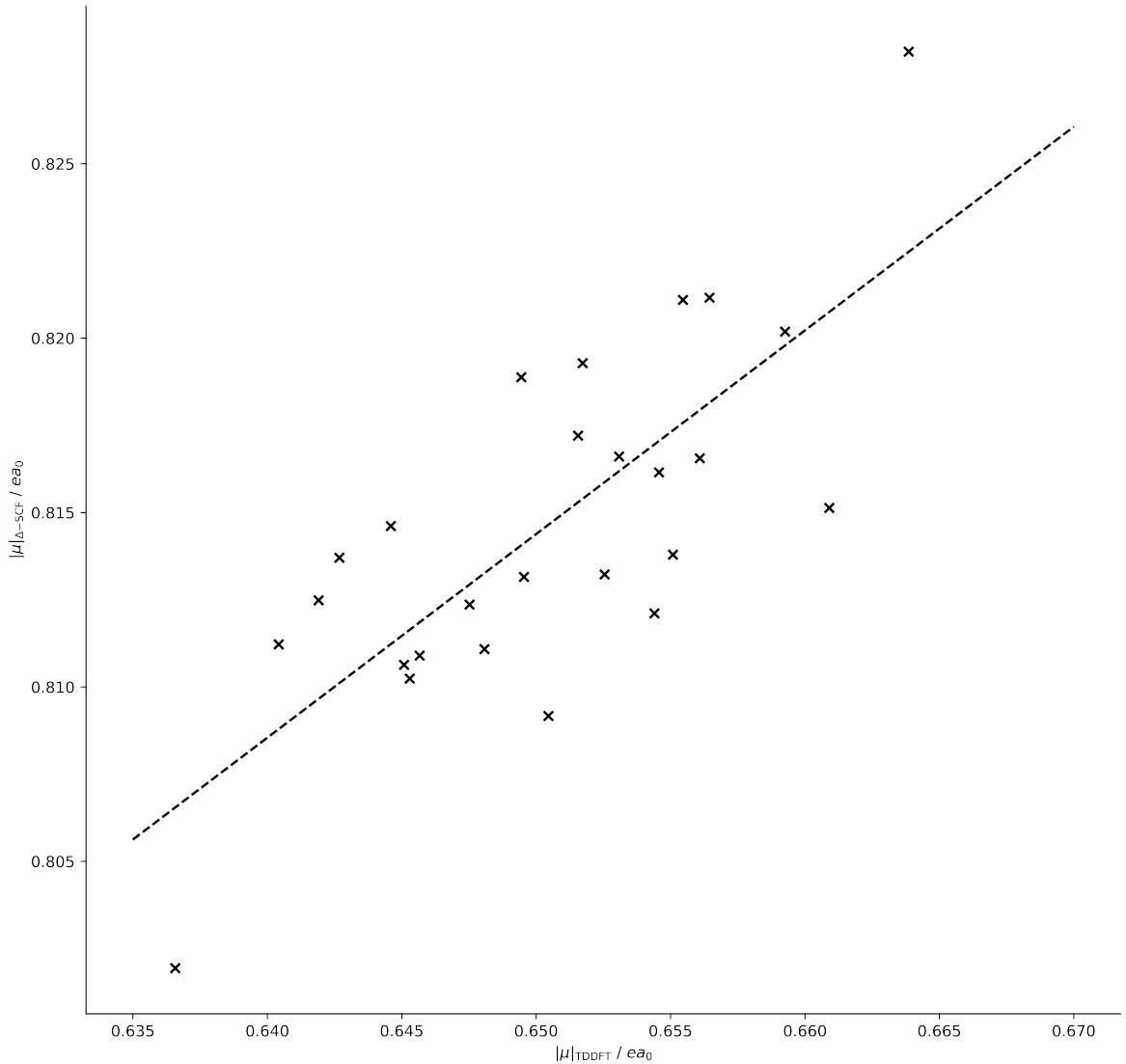


Figure 2.5: Transition dipole magnitudes from Δ -SCF for the same 26 chlorophyll geometries as 2.2.4, plotted against dipole magnitudes from TD-DFT. The line of best fit ($R^2 = 0.57$) is shown as the dashed line. Both methods used a PBE0/Def2-SVP level of theory.

2.2.5 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.

2.2.5.1 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 .

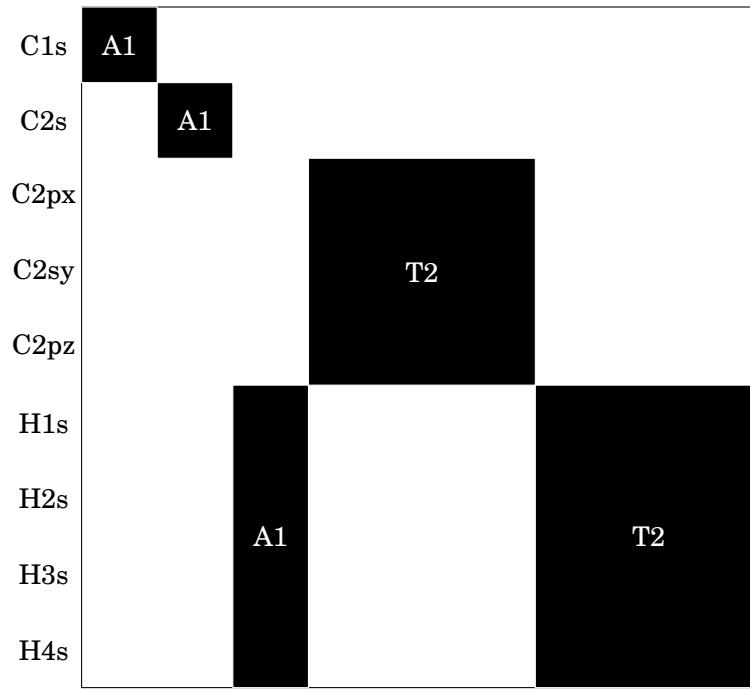


Figure 2.6: A breakdown of the symmetry orbitals in STO-3G methane into the subspaces present in the T_d point group.

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

$$(2.22) \quad \tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{S} \mathbf{C}$$

, and then summing the coefficients to obtain the character

$$(2.23) \quad P_A = \sum_v |\tilde{\mathbf{C}}_{A,\mu\nu}|$$

, 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

$$(2.24) \quad n = \frac{1}{h} \sum_R \xi_r(R) \xi_i(R)$$

, 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.

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.

The distributions of absolute errors

$$(2.25) \quad \epsilon = \Delta E_{\text{method}} - \Delta E_{\text{SCS-CC2}}$$

for each of the benchmarking methods, as well as a generated distribution of sTDA-xTB results, are shown in fig-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

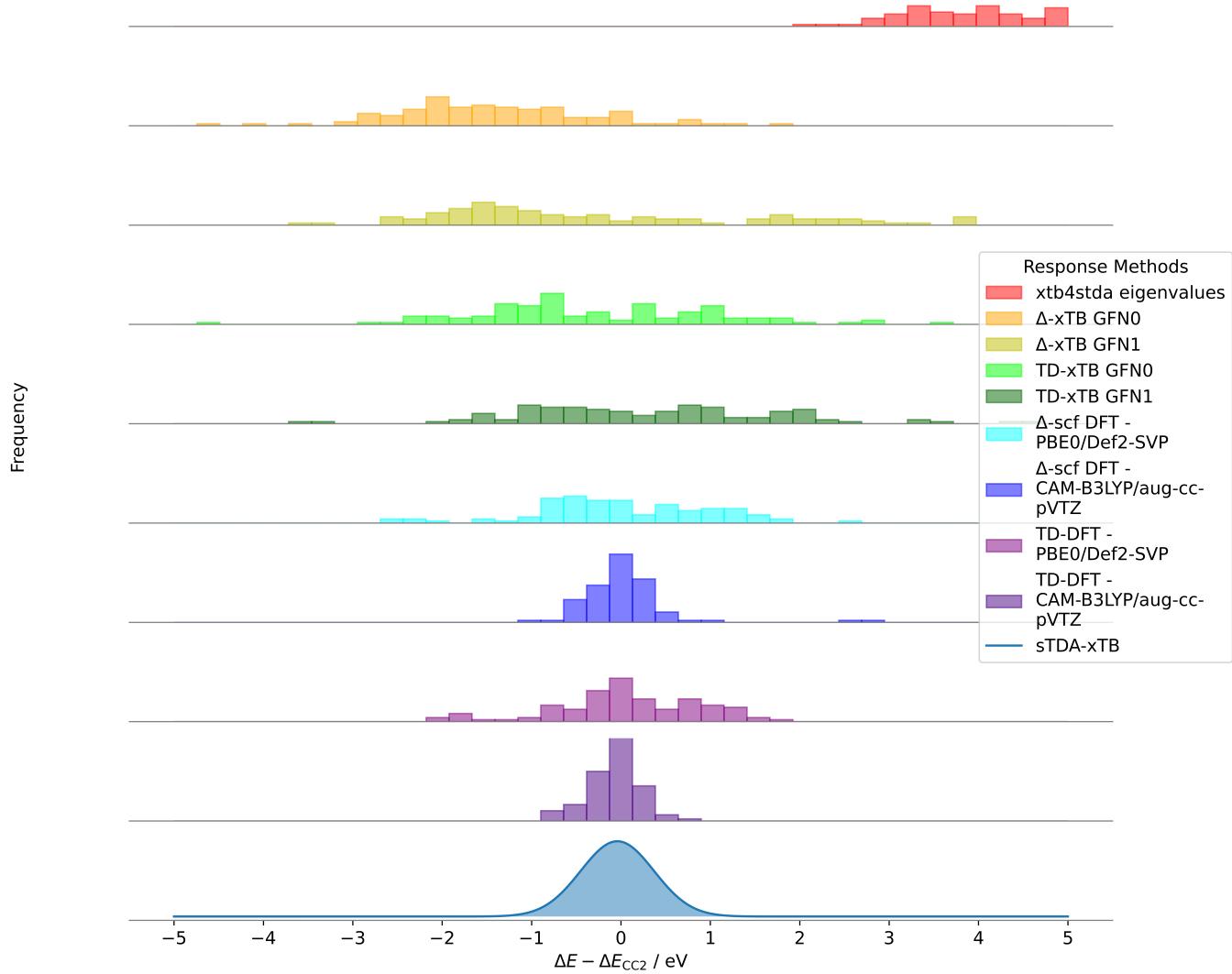


Figure 2.7: The distributions of errors compared to SCS-CC2 transition energies for the methods included in the Δ -xTB benchmarking.

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 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 2.2.4 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.

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 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 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

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 forefills the requirements set out in the introduction of having an efficient and simple, but accurate, method of getting transition properties for LHII.

3.1.1 Full solutions TD-DFT equation

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

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

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

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

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

where indices a, b and i, j refer to virtual and occupied orbitals respectively, 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 non-zero for diagonal elements only.

3.1.2 Approximations to Solutions

There are three approximations made in the chl-xTB method. First is a diagonal dominant approximation to eigenvalue solutions, where eigenvalues are equal to the diagonal elements. Second is the Tamm-Dancoff approximation (TDA), which eliminates three blocks of the full linear response matrix. 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.

3.1.2.1 Tamm-Dancoff approximation and Diagonal Dominant \mathbf{A} matrices

One of the earliest approximations applied to full linear response was the Tamm-Dancoff approximation, where only virtual-occupied contributions are calculated. This sets all of the elements of \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, 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.

However, 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:

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

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.

Additionally, 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 would be large and by approximating them to be zero would not be valid. This is considered in the 3.1.4.

3.1.3 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 point charges, and using a charge-charge interaction for the energy.

Atom centered transition charges can be given by summing the reduced one-electron transition density \mathbf{D}^{mm} multiplied the orbital overlap \mathbf{S} for all orbitals centred on atom A :

$$(3.6) \quad q_A^{mn} = \sum_{p \in A} D_{pq}^{mn} S_{pq}$$

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

$$(3.7) \quad q_A = Z_A - \sum_{p \in A} D_{pq} S_{pq}$$

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:

$$(3.8) \quad (pq|rs) \approx \frac{1}{2} \sum_A^N \sum_B^N q_A^{mm} q_B^{mn} \Gamma_{AB}$$

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:

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

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:

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

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:

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

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

As the a_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 calculated excitation energies is then given by:

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

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 3.2.3). 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.

3.1.4 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 °.

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.

The high amount of single-transition character in the Q_y transition makes it ideal for the approximations to transition energies set out above. The lack of coupling elements that would

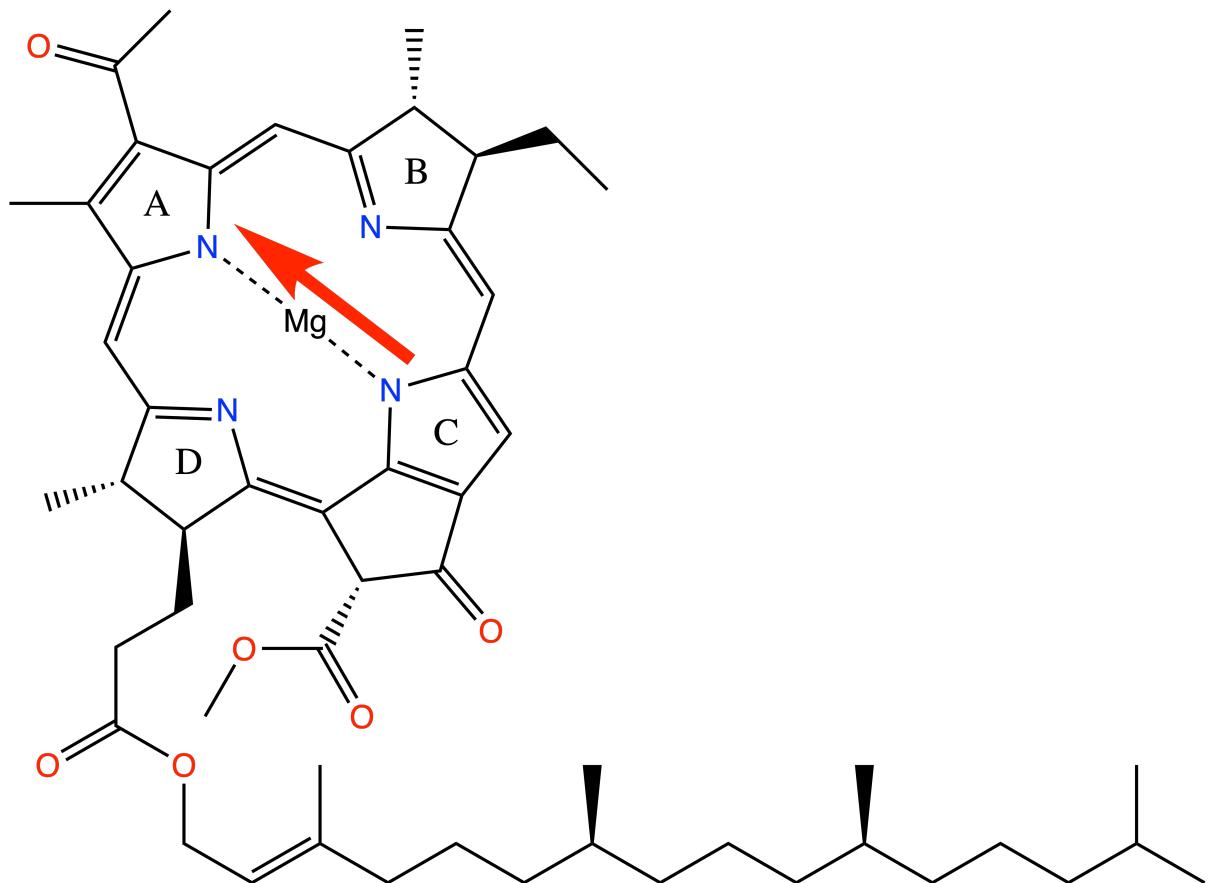


Figure 3.1: Bacterial Chlorophyll a (BChla) with a model Q_y transition dipole.

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 proxy for the accuracy of transition dipole moments. As a single value, this is ideal for optimisation both for use in an objective function as well as for discarding incorrect transitions.

The entire training set was constructed from Q_y transitions from bacterial chlorophyll A systems. The geometries and linear response methods used to construct this is discussed in 3.2.3.

3.1.5 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 tranistion properties. Whilst DFT methods could have been used as for mulliken partial

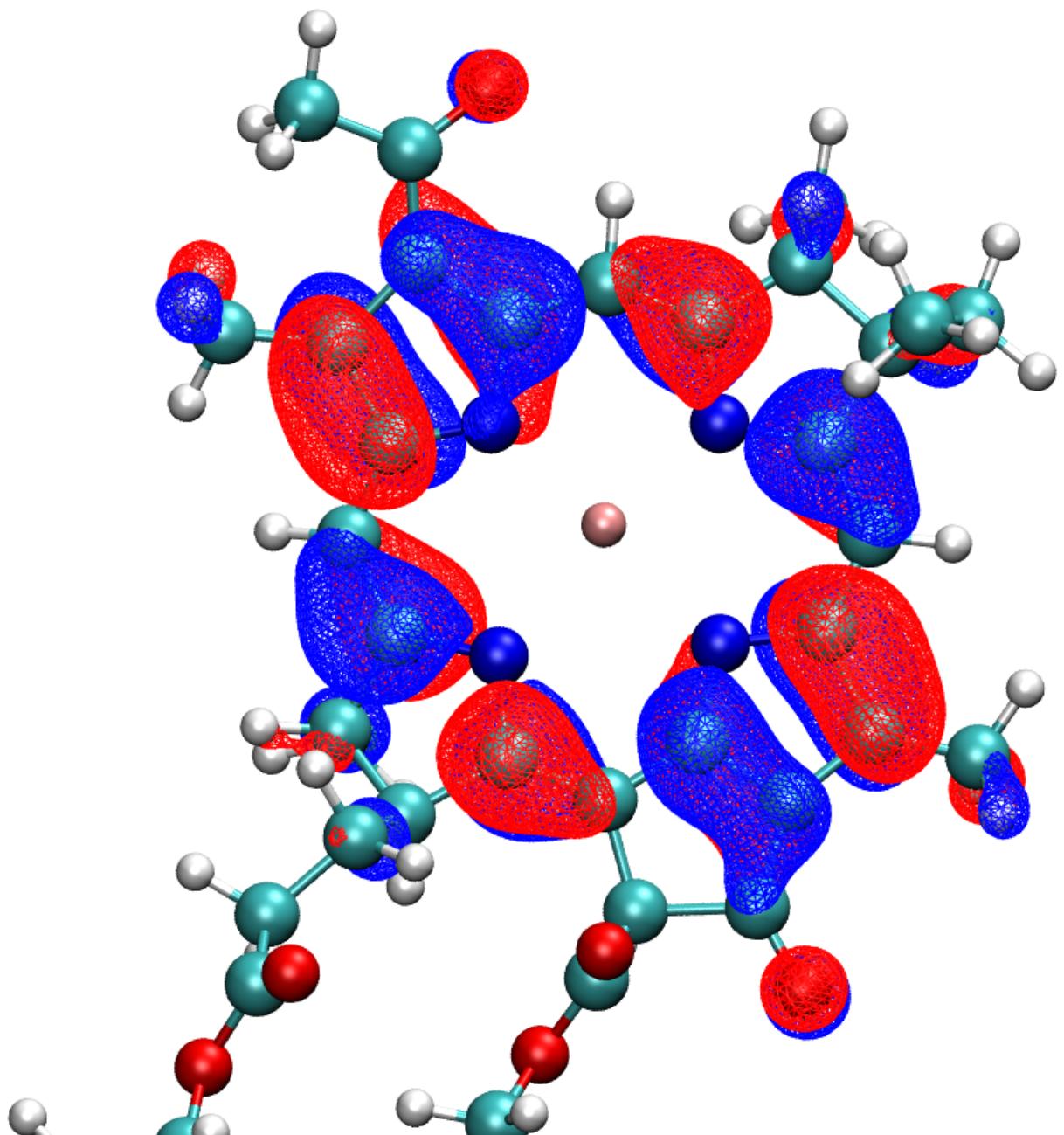


Figure 3.2: The HOMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.

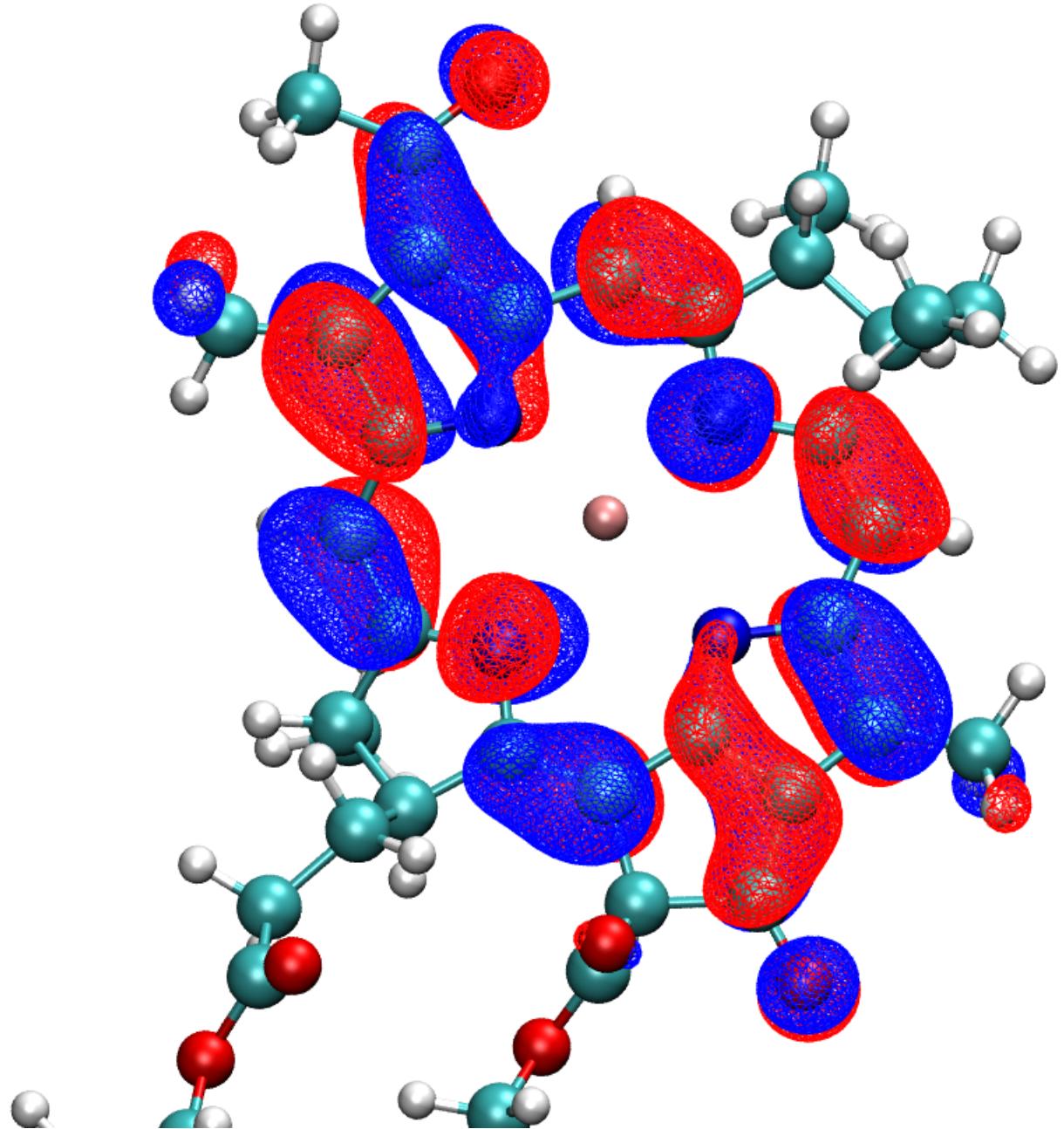


Figure 3.3: The LUMO orbital of bacterial chlorophyll a from PBE0/Def2-SVP DFT.

and transition charges, this would still not be efficient enough to calculate the large enough volume 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. 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.

3.1.6 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:

$$(3.13) \quad \mathbf{D}^{01} = |\Psi^0\rangle\langle\Psi^1|$$

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 p^0 , p^1 for the ground and excited state respectively. The ground state and excited state density can be similarly calculated as:

$$(3.14) \quad \mathbf{D}^{00} = |\Psi^0\rangle\langle\Psi^0|$$

$$(3.15) \quad \mathbf{D}^{11} = |\Psi^1\rangle\langle\Psi^1|$$

. 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 of the form eq. 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 twice as large as 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.

3.2 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 for chl-xTB to be optimised to is also required. All of these elements are important in minimising the amount of error in the final method, as well as determining how well chl-xTB performs for different problems and other properties.

3.2.1 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

$$(3.16) \quad f_{\text{RMSE}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_i^N (\Delta E_i - \Delta E_{i,\text{ref.}})^2}$$

where \mathbf{x} 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:

$$(3.17) \quad R^2 = 1 - \frac{\sum_i^N (\hat{y}_i - y_i)^2}{\sum_i^N (\hat{y}_i - \bar{y})^2}$$

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:

$$(3.18) \quad f_{\text{full}}(\mathbf{x}) = \lambda_1 \text{RMSE}(\Delta E) + \lambda_2 \text{RMSE}(|\mu|) + \lambda_3 (1 - R^2(\Delta E)) + \lambda_4 (1 - R^2(|\mu|))$$

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.

3.2.2 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. 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.

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

$$(3.19) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

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

$$(3.20) \quad f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$$

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:

$$(3.21) \quad \mathbf{x}_r = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1})$$

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:

$$(3.22) \quad \mathbf{x}_e = \mathbf{x}_0 + \gamma(\mathbf{x}_r - \mathbf{x}_0).$$

. 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:

$$(3.23) \quad f(\mathbf{x}) = \begin{cases} \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_r - \mathbf{x}_0) & \text{if } f(\mathbf{x}_r) < f(\mathbf{x}_{n+1}) \\ \mathbf{x}_c = \mathbf{x}_0 + \rho(\mathbf{x}_{n+1} - \mathbf{x}_0) & \text{otherwise } f(\mathbf{x}_r) \geq f(\mathbf{x}_{n+1}) \end{cases}$$

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 possiblity 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:

$$(3.24) \quad \mathbf{x}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$$

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.

3.2.2.2 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:

$$(3.25) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

however with the added constraints:

$$(3.26) \quad c_i(\mathbf{x}) = 0$$

$$(3.27) \quad c_j(\mathbf{x}) \leq 0$$

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:

$$(3.28) \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

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

$$(3.29) \quad \mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_n^m \lambda_n g_n(\mathbf{x})$$

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

$$(3.30) \quad \min_d f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 \mathcal{L}(\mathbf{x}_k, \lambda_k) d$$

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:

$$(3.31) \quad \min_d \|(\mathbf{D}_k)^{\frac{1}{2}} (\mathbf{L}_k)^T d + (\mathbf{D}_k)^{-\frac{1}{2}} (\mathbf{L}_k)^{-1} \nabla f(\mathbf{x}_k)\|$$

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

$$(3.32) \quad \mathbf{L}_k \mathbf{D}_k (\mathbf{L}_k)^T = \mathbf{B}_k$$

. 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 of the Nelder-Mead and SLSQP algorithms are shown in figures 3.4 and 3.5 respectively. Vertices and iterations of solutions for the HimmelBlau function (a standard

benchmark for optimisation algorithms) are plotted in the \mathbf{x} space. It can be seen that the Nelder-Mead algorithm requires both 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 3.2.4.

3.2.3 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 [27]. 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 or sequence.

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 3.1. 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 little importance on the accuracy [27].

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 [27]. 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.

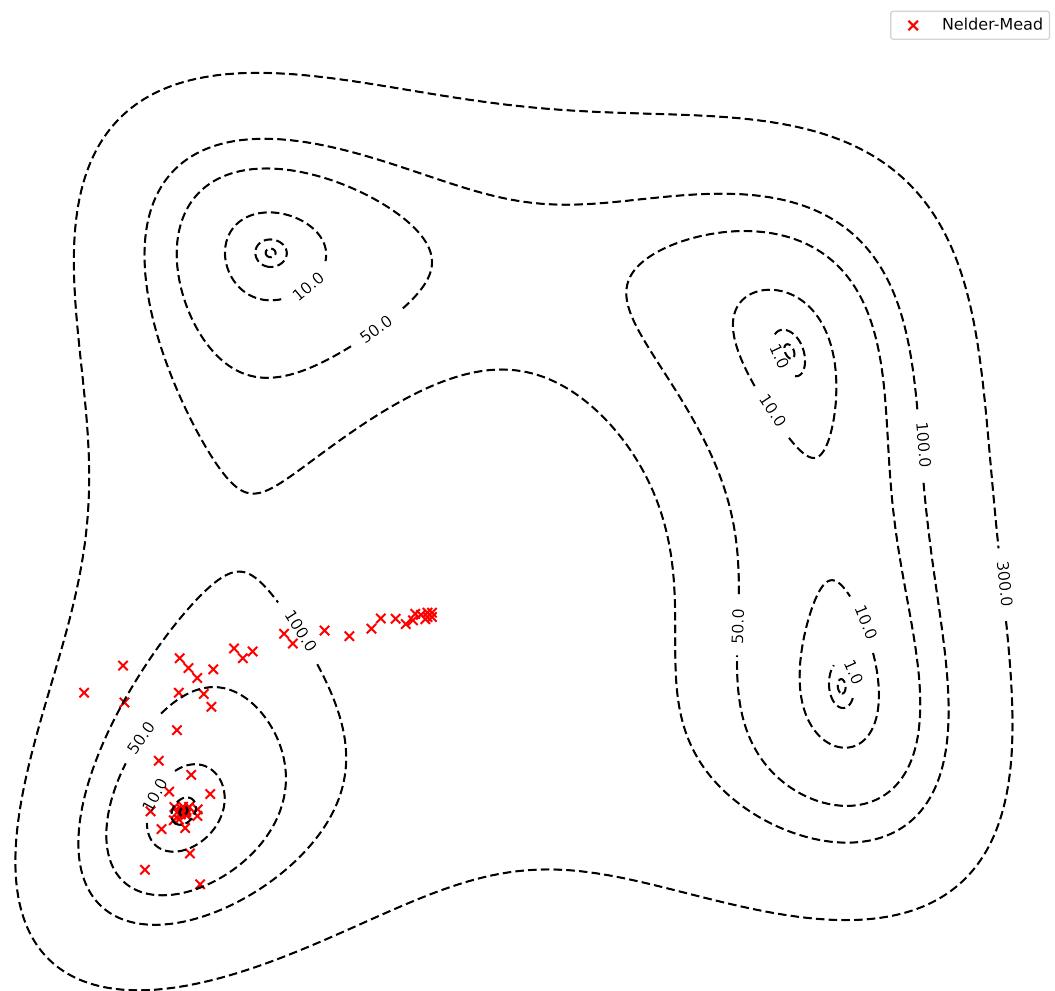


Figure 3.4: Positions of the Nelder-Mead simplex vertices for the HimmelBlau function.

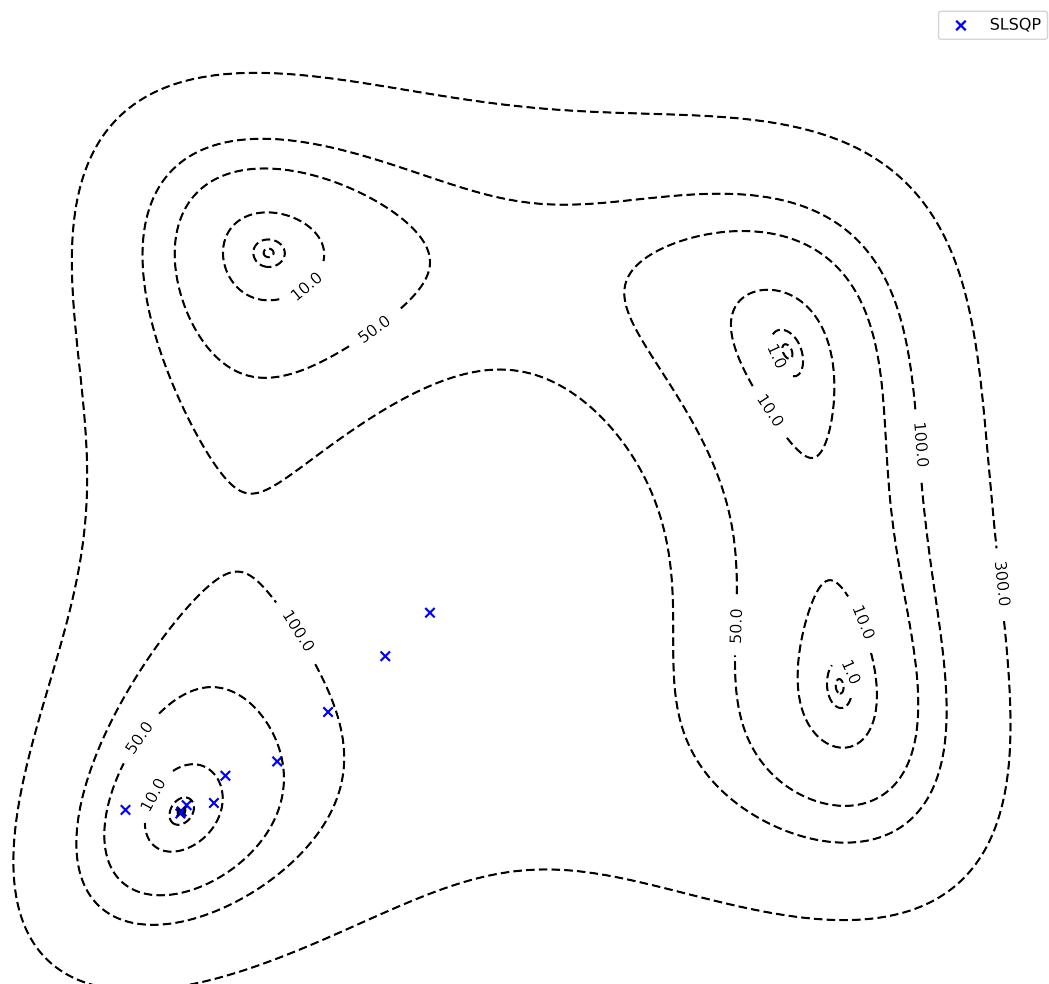


Figure 3.5: Positions of iterative solution vector \mathbf{x} from SLSQP for the HimmelBlau function.

	RMSE(ΔE)	$R^2(\Delta E)$	RMSE(μ)	$R^2(\mu)$
CAM-B3LYP	0.15	0.90	0.26	0.55
BLYP	0.06	0.84	0.45	0.09
Δ -SCF	0.22	0.84	1.54	0.33
$\Delta\epsilon$	0.22	0.83	1.59	0.30

Table 3.1: ref data vals

There is a large variation in correlation and RMSE values between the reference methods. This variation sets a reasonable expectation of how well a new method might perform. The PBE0 and CAM-B3LYP excitation energies have RMSE of 0.15 eV and an R^2 value of 0.9. The geometry variations have a very similar effect on transition properties calculated on these two properties. The lower level BLYP functional has a smaller RMSE of 0.06 eV, but is not as correlated with an R^2 value of 0.84. This highlights how RMSE is not always a full picture of how well two methods agree. However this correlation is still relatively good, compared to the R^2 value of 0.84. Hence we should expect a well trained method to be on this level of accuracy. The Δ -SCF and eigenvalue difference methods, whilst still well correlated, have much higher RMSE values, both around 0.22 eV. Having established in the last chapter the suitability of the Δ -SCF method, this then gives one lower bound on the accuracy, with the CAM-B3LYP and BLYP being an upper bound.

The agreement of transition dipole magnitudes is much lower than excitation energies. The RMSE of Δ -SCF and eigenvalue difference methods is significantly higher (1.54 a.u. and 1.59 a.u. respectively) than CAM-B3LYP (0.26 a.u.). 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. CAM-B3LYP has a more reasonable RMSE of 0.26 a.u., with the average value being similar to PBE0. It is harder to state what a reasonable accuracy for transition dipole magnitudes could be. Additionally, there is also little correlation between methods. CAM-B3LYP and PBE0 has a correlation of 0.55, which is surprisingly low for two comparatively similar and high level methods. The best agreement is between Δ -SCF and eigenvalue difference, which might be due to the similarity between the ground and excited states from these two methods.

Overall, whilst it might be expected for transition energies to agree with an accuracy between 0.15 eV and 0.22 eV, the transition dipoles may have vary more. It is hard to say whether this variation is due to inaccurately describing the transition density of chlorophyll geometries, as none of the reference methods are especially correlated.

Hamiltonian		
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		
y_K	2.147	2.000
y_J	4.012	4.000
a_x	0.067	0.500
D_{scale}	0.636	-

Table 3.2: optimized parameters from SLSQP procedure.

3.2.4 Results

The final parameters for the chl-xTB method are given in table 3.2. 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, as and the R^2 of transition energy is equally good. While the correlation is low, it is similar to the correlation of other methods and so it is expected to be at around this value.

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.

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 3.2), with the exception of the a_x parameter. This parameter is far lower than the sTDA-xTB equivalent, which has a value of 0.500, but is in line with other methods that use similar MNOK approximations for coulomb-type integrals in response methods

[5].

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.

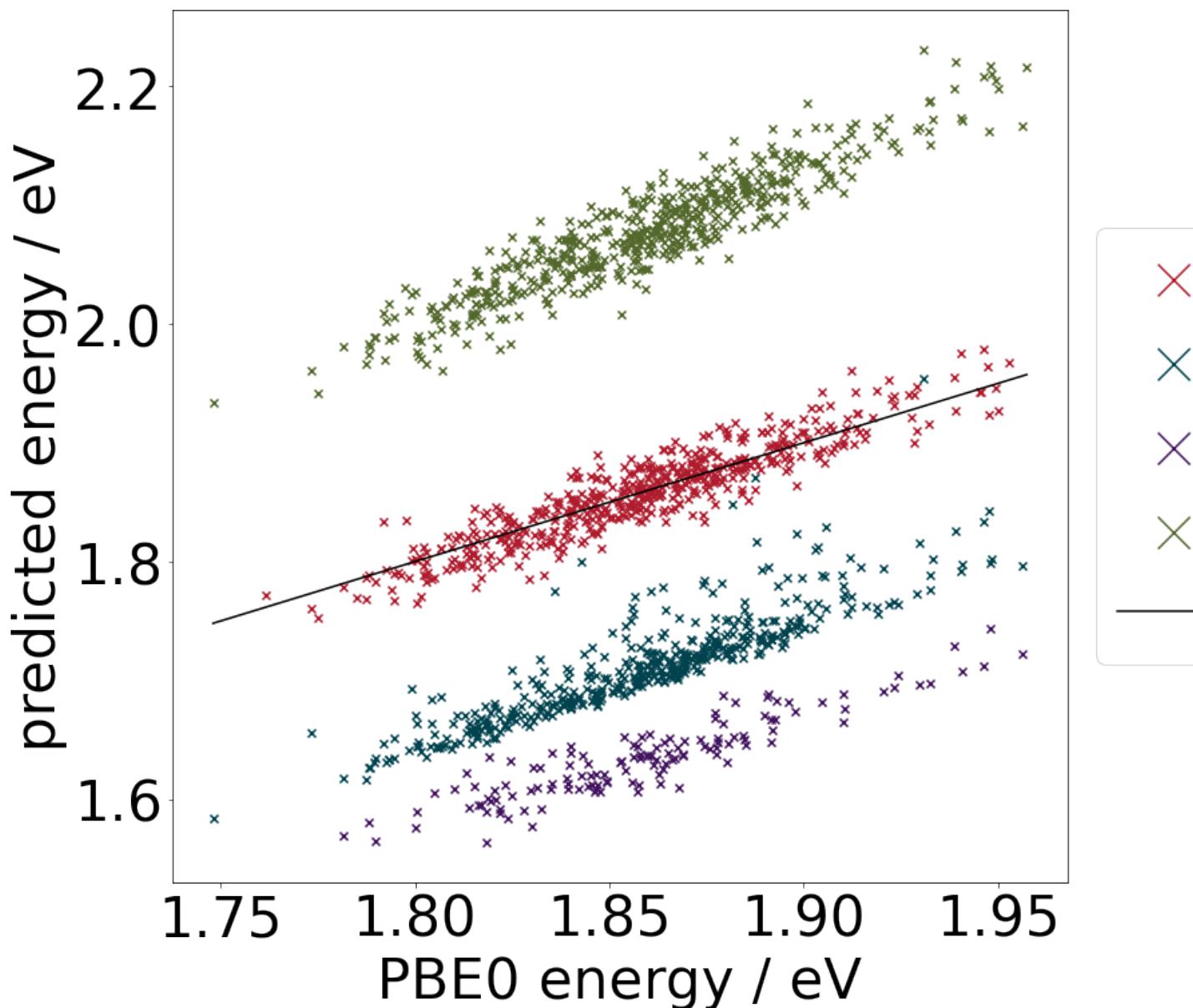


Figure 3.6: fig:training scatter

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 along vibrational modes. 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, giving 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.

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 3.7, and have frequencies calculated by an GFN1-xTB at. This is inline with other calculated frequencies.

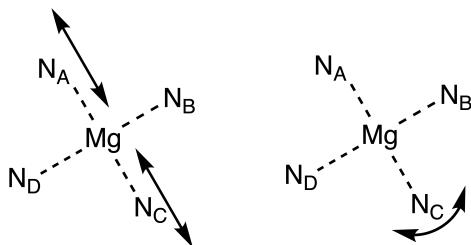


Figure 3.7: Normal modes of the nitrogen-magnesium centre of chlorophyll that have $D_{4h}-D_{2h}$ (left) and $D_{4h}-C_s$ (right) symmetry breaking components.

Other normal modes that would also have this symmetry breaking component were identified by the amount of N_A-N_C displacement for each normal mode. 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 approximate the peaks in the BChla absorption spectrum.

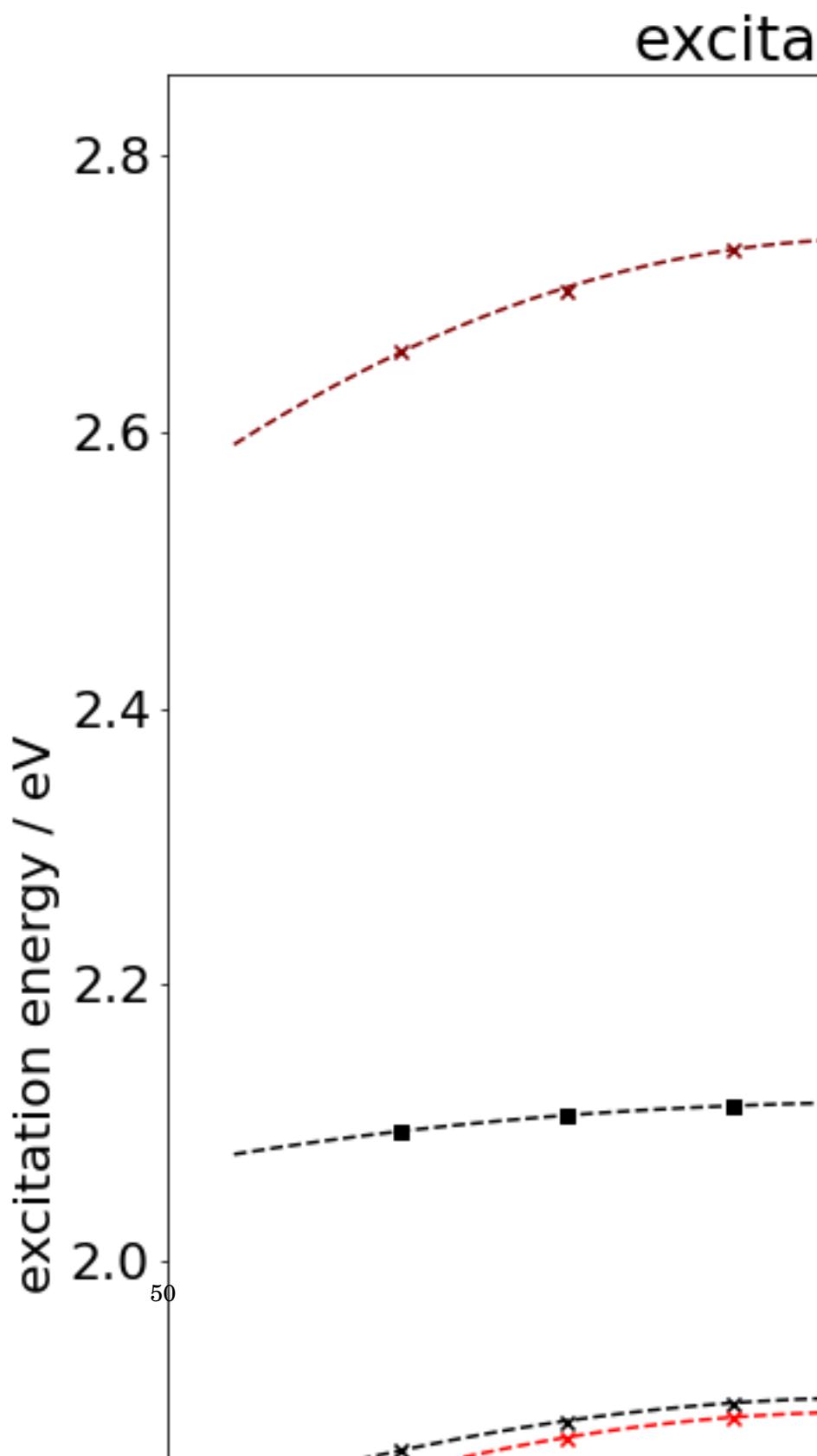
The modes with highest displacement were then chosen for scans, and the results are shown in the set of figures.

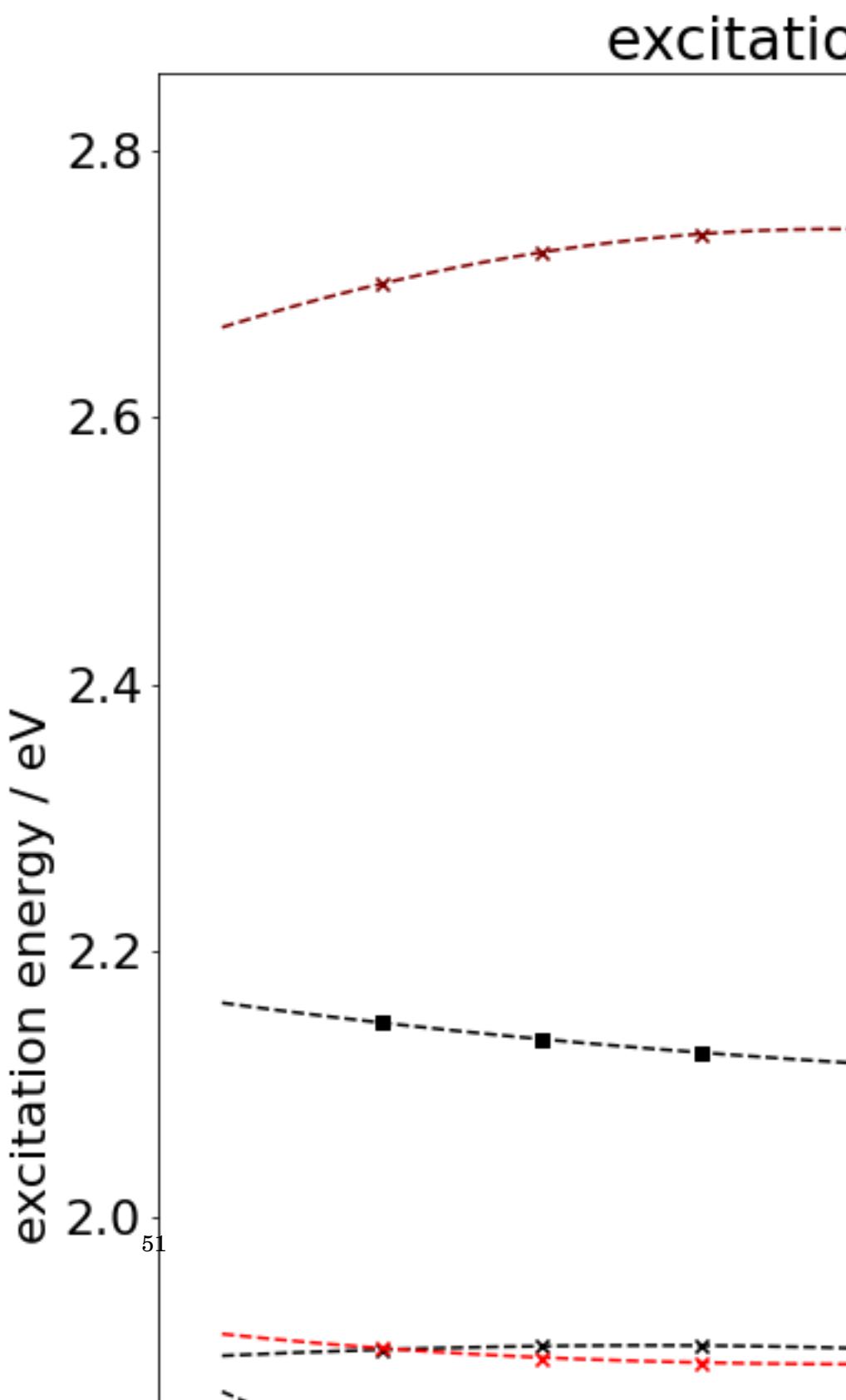
*1320.3, vibration 162
*1361.8, vibration 169

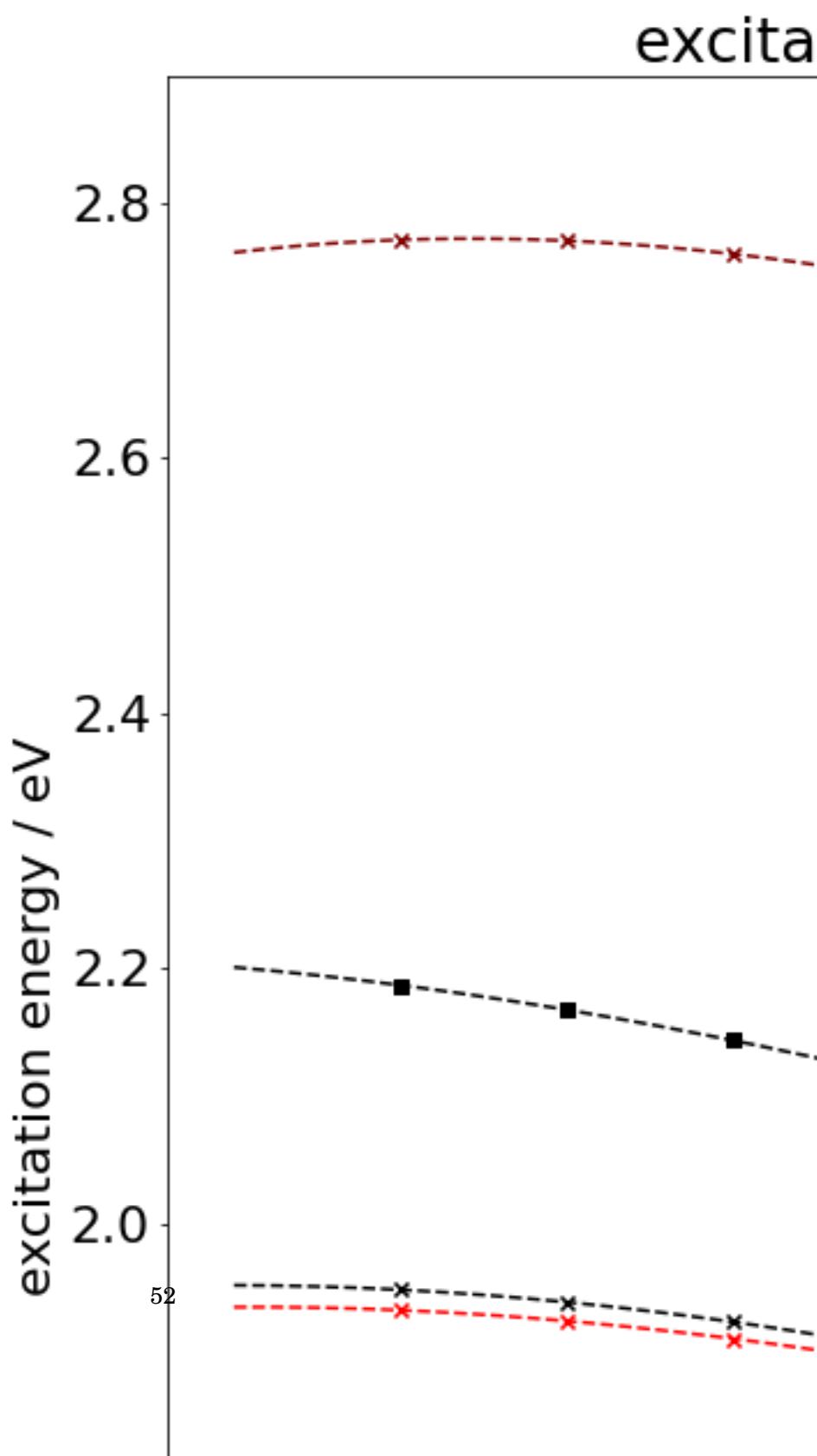
*119142.9, vibration 130
*1130

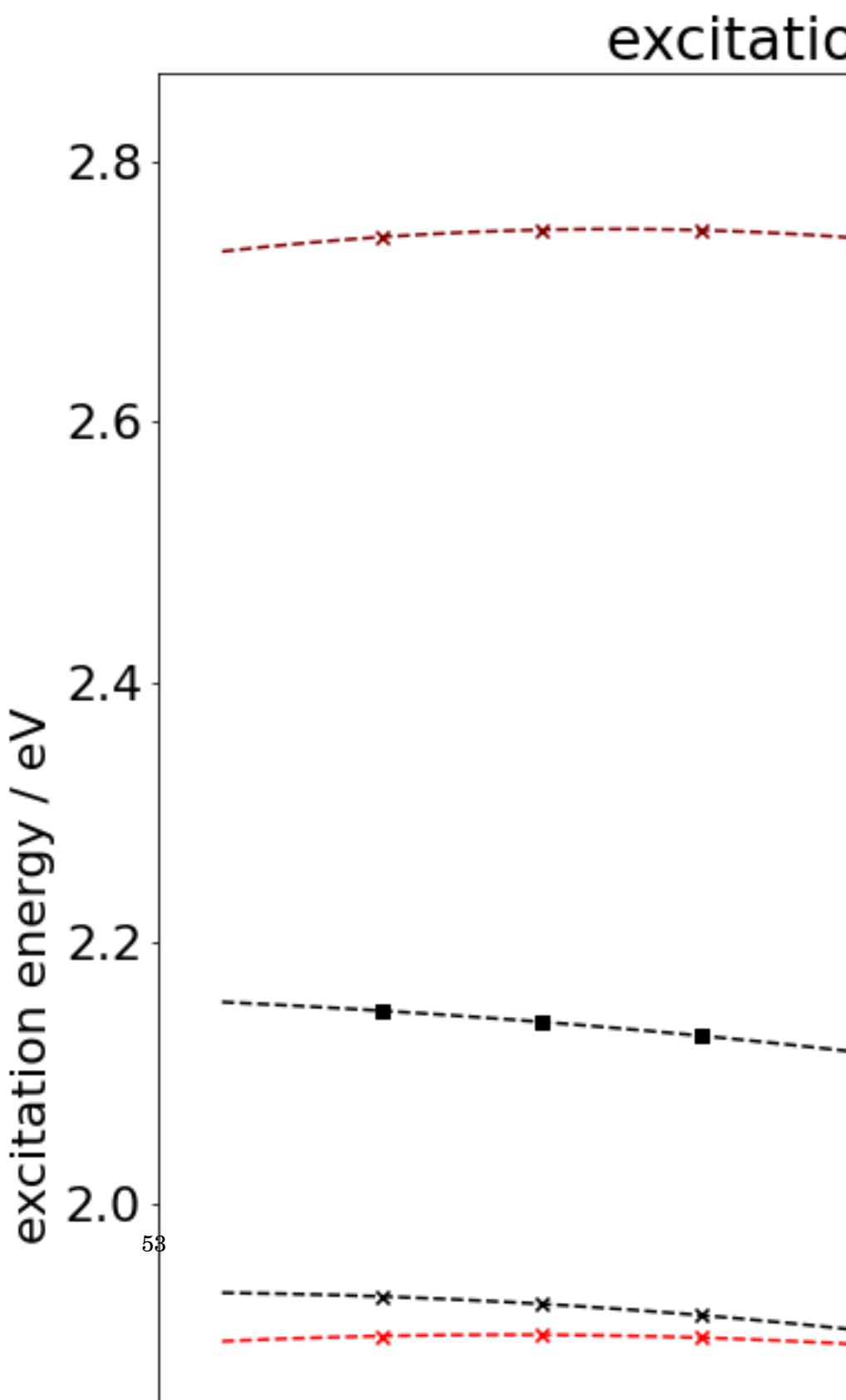
*1122.2, vibration 132
*1105.0, vibration 129

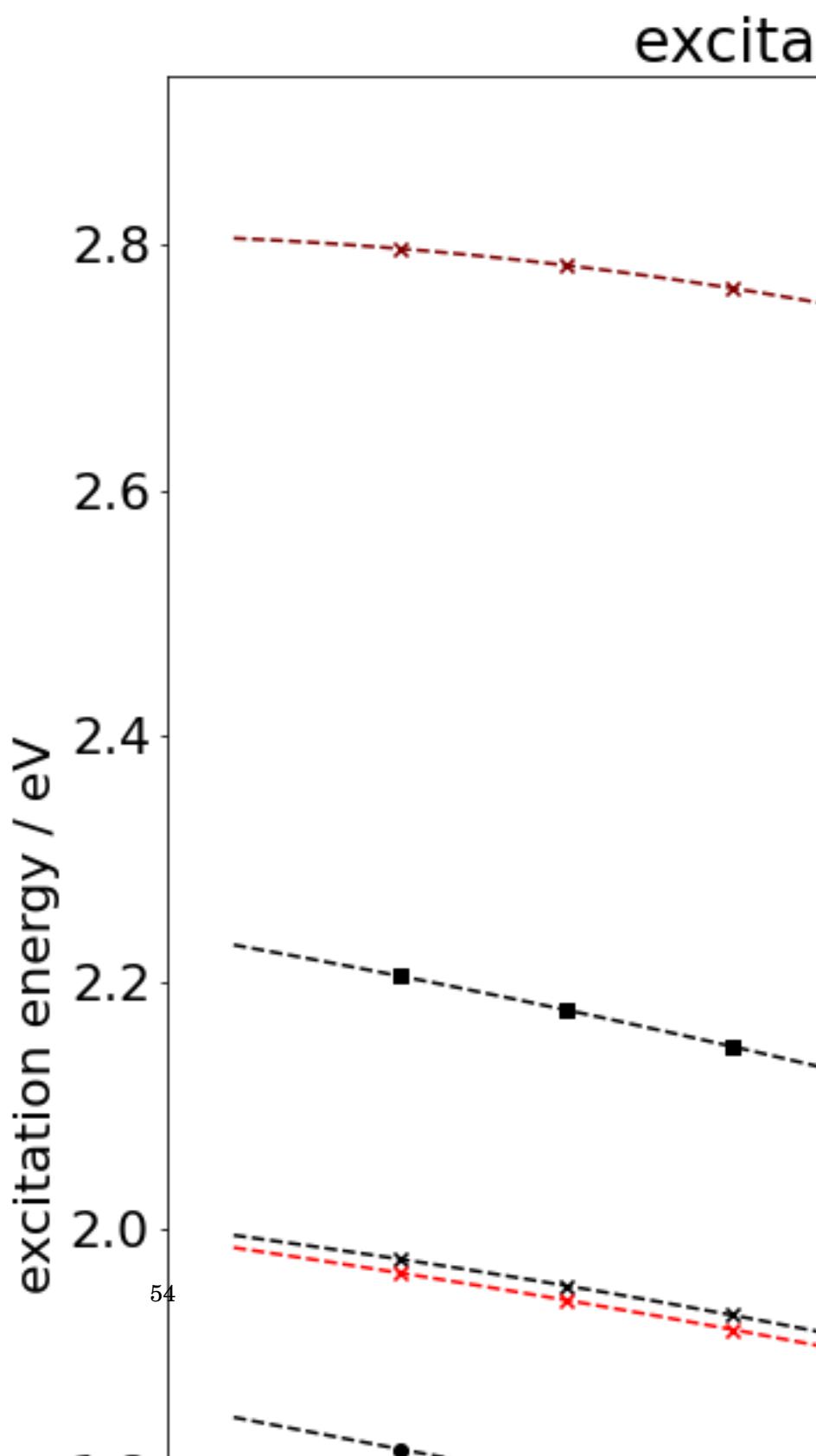


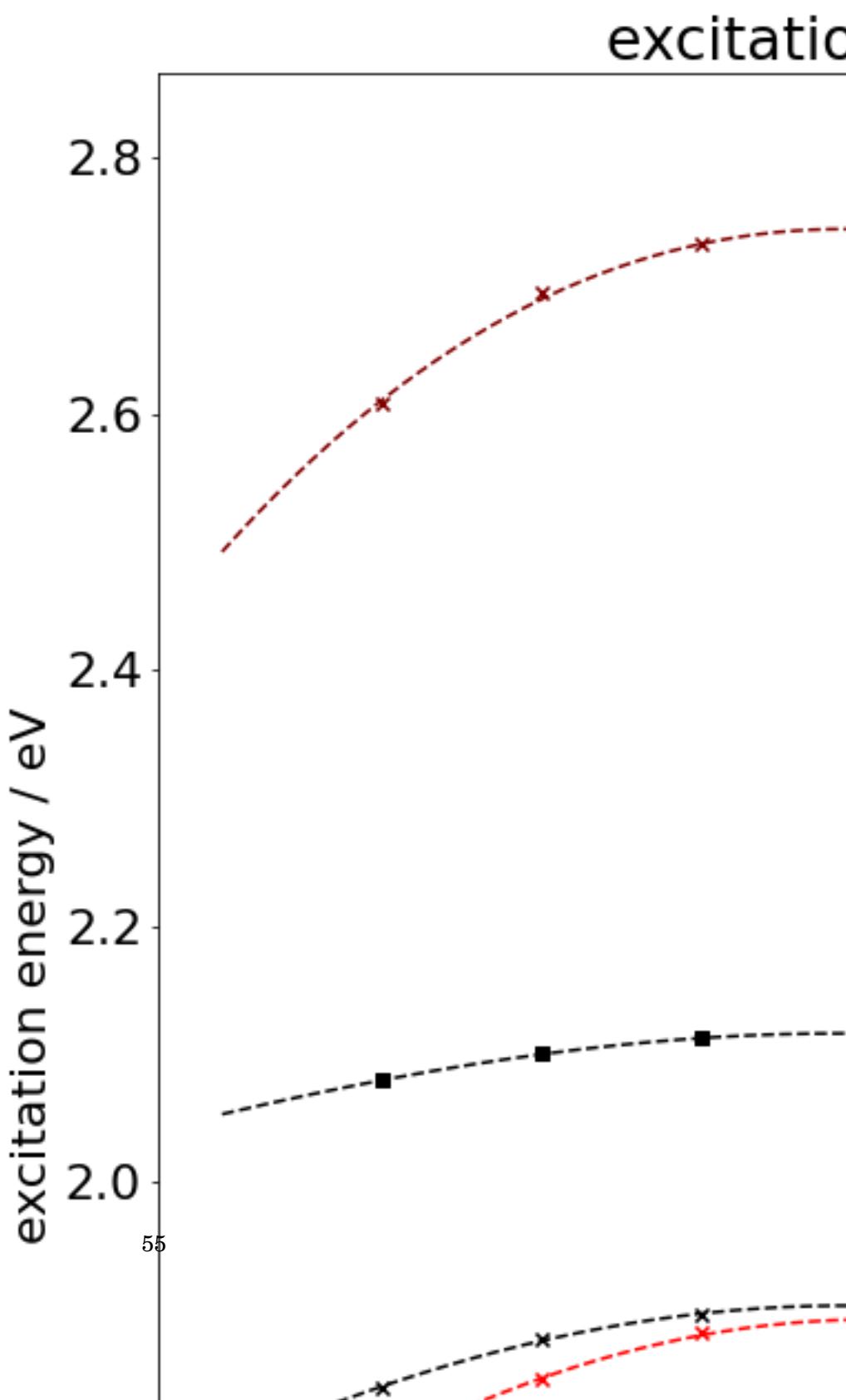


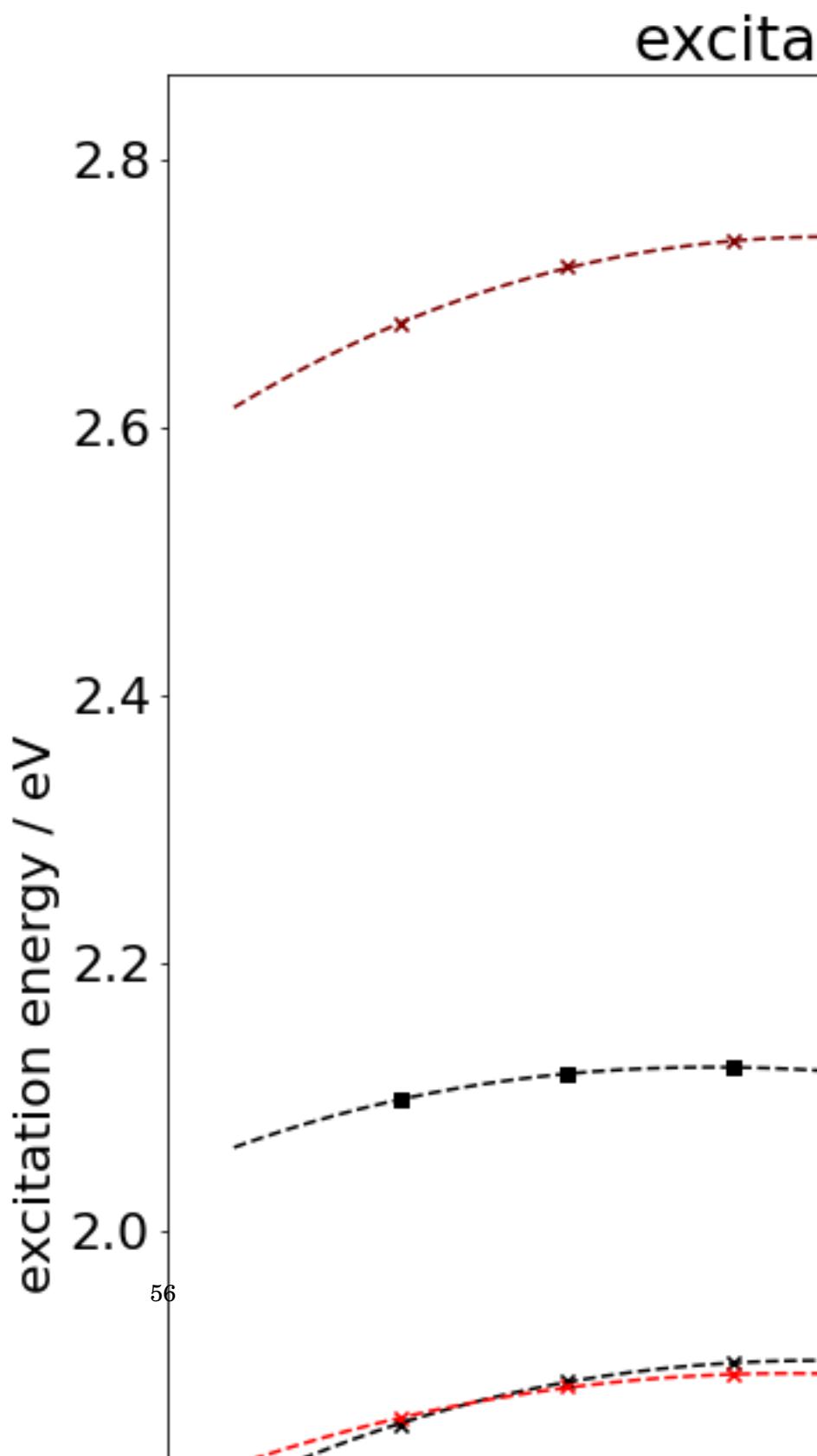


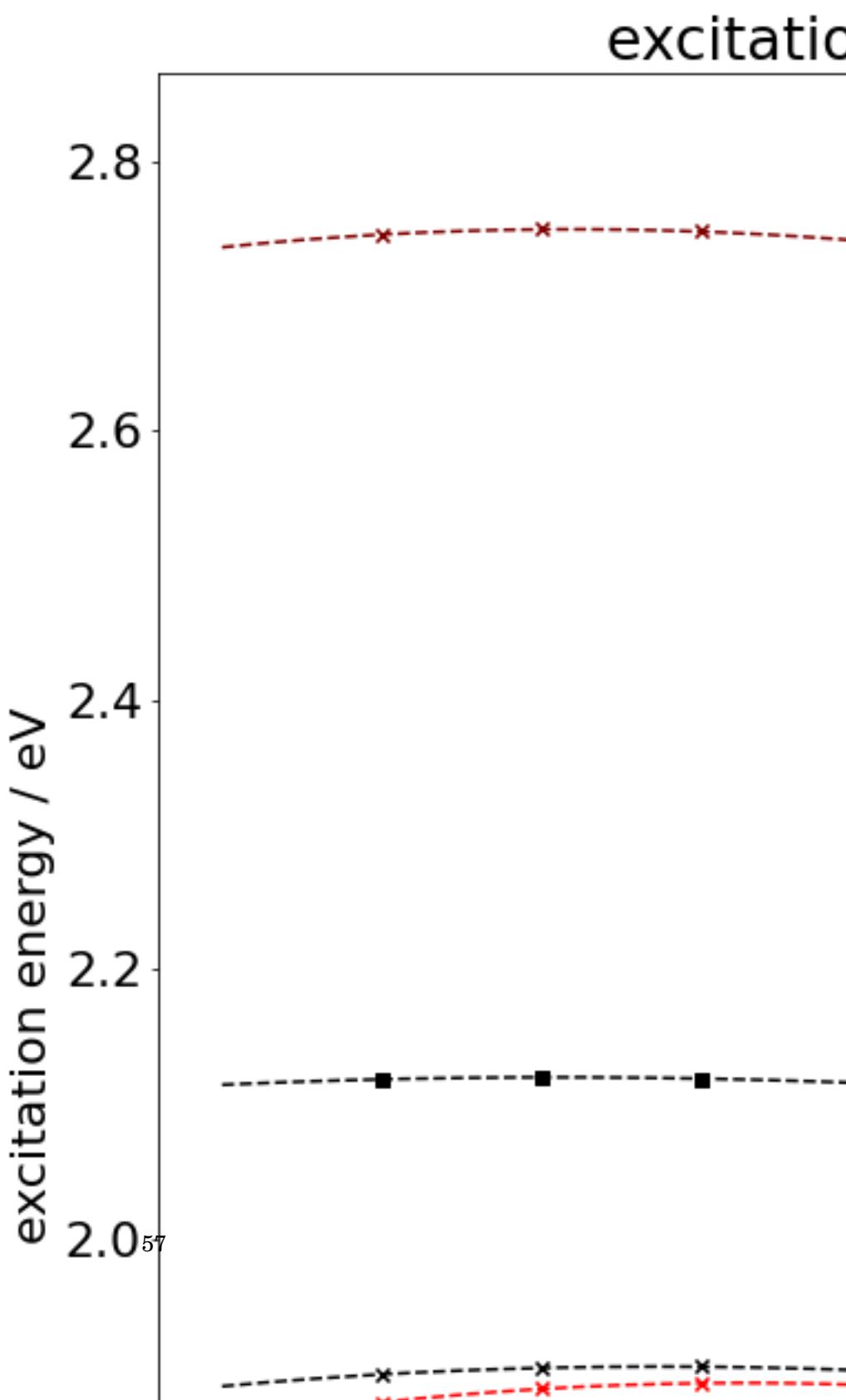


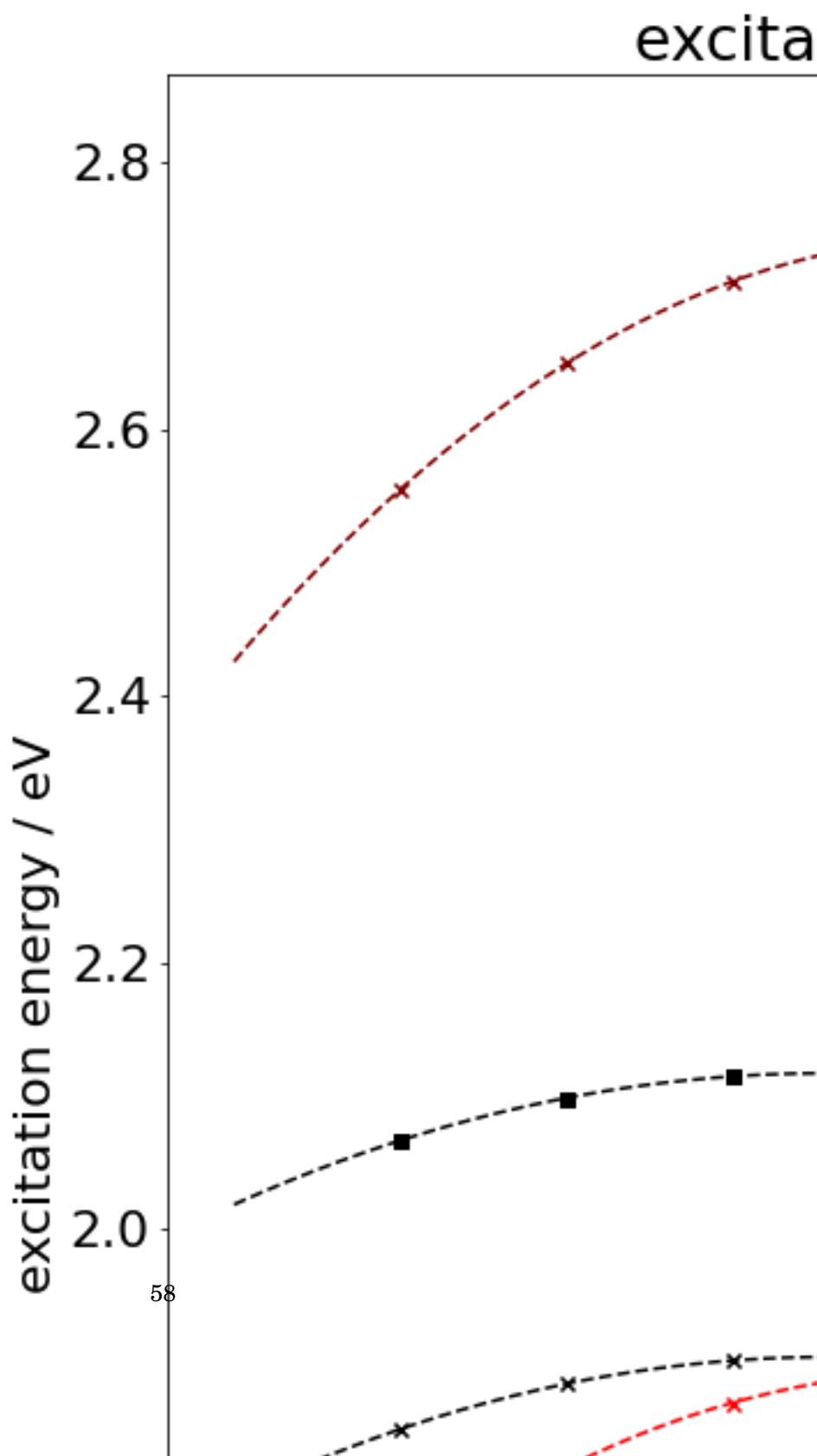


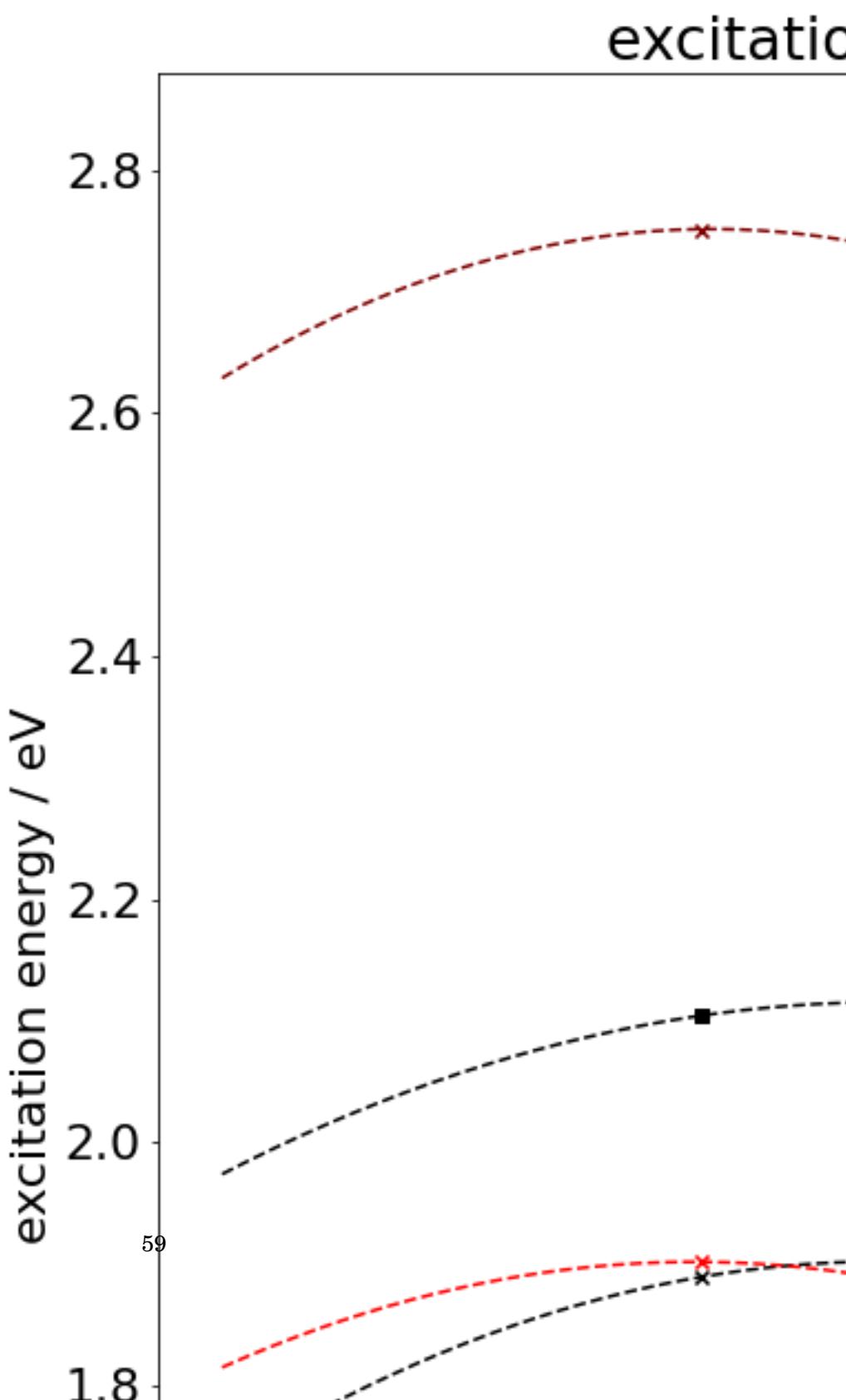


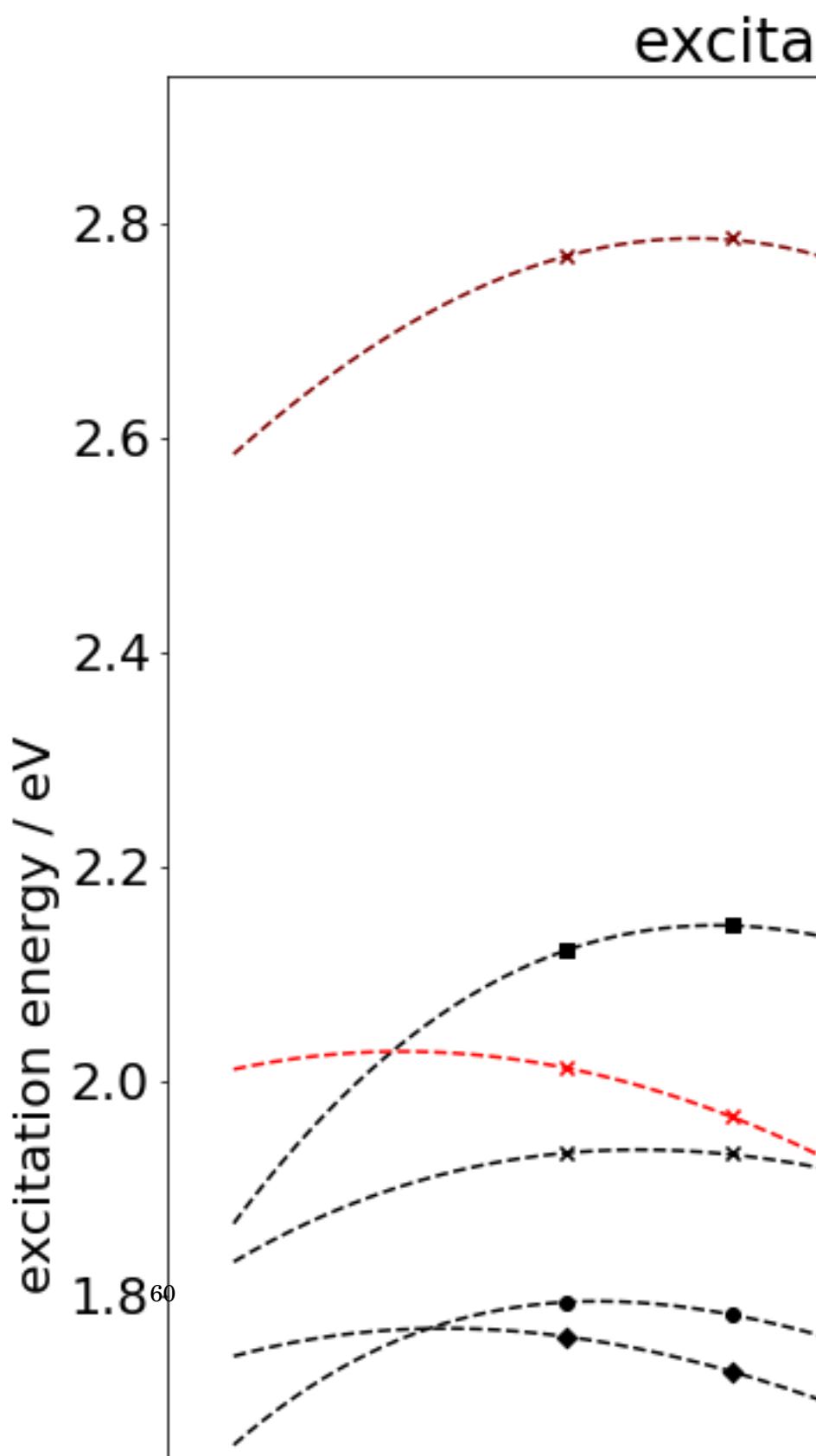












3.3.3 Absorption Spectra

An absorption spectrum for BChla was also calculated with chl-xTB. In order to account for inhomogeneous broadening, the chlorophyll structures were taken from an MD simulation of BChla in an explicit ether solvent.

The MD was performed with the OpenMM toolkit. Forcefield parameters for the chlorophyll were taken from a bespoke parameterisation for photosystem II [29], 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, both using the Def2-SVP basis set. Experimental data for the absorption spectra was also found from.

The effect of including embedding effects was also investigated. As implemented in the QCORE package, the ground state xTB calculation could include a QM-MM polarisation term in the Fock matrix, for both periodic and non-periodic embedding. However the size of the solvated chlorophyll system required a more efficient implementation of the real-space Ewald summation, as well as implementing a particle mesh Ewald method, which was done by using the HelpME library.

Absorption spectra, both with and without a single-parameter shift of excitation energies, can be seen in ref.

chl-xTB performs equally well as TD-DFT methods at simulating absorption spectra, although constrained by the quality of the training data. It can be seen that the chl-xTB lineshape is similar in position and width to the PBE0 method. Although both lineshapes are much wider than the experimental spectrum, the CAM-B3LYP line also fails to capture this broadening effectively, and so the accuracy of the broadening from chl-xTB is within reasonable bounds. The more complex shape of the experimental spectrum, with shoulders at 610 nm, 580 nm and 530 nm are not captured by either chl-xTB. This can be explained by PBE0 also not having these shoulders in the Q_y transition, and so the cause of this behaviour might not be included in the training data. The CAM-B3LYP does show a little of these shoulders with an additional peak, however it does not capture all of the shape. The embedded chl-xTB spectra is both red-shifted and broadened slightly, which is in line with changes in simulated spectra from other methods.

Overall, chl-xTB performs as well as can be expected from the training data. From this it could be expected that better training data would yield better accuracy from chl-xTB. 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

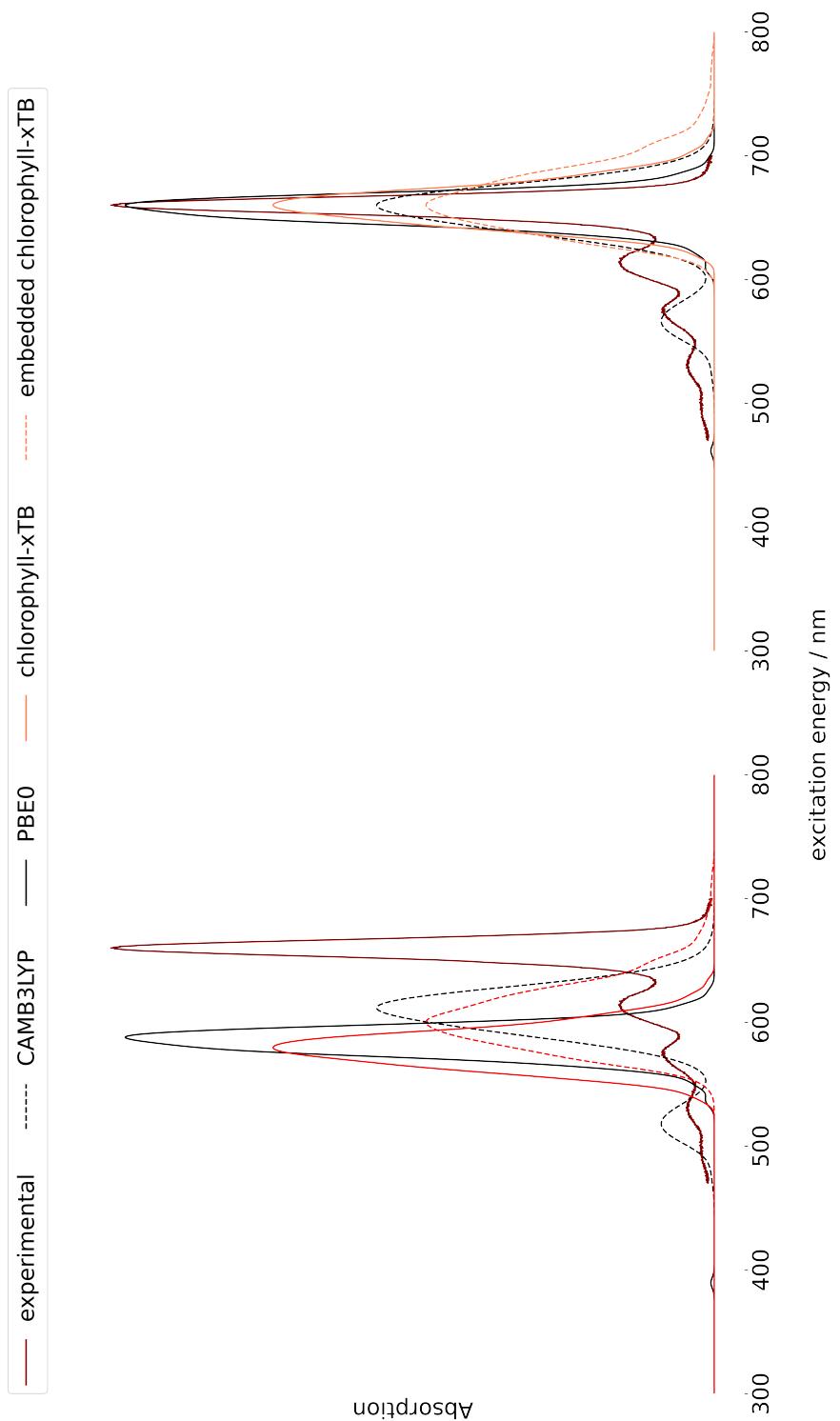


Figure 3.8: Chlorophyll in diethyl ether.

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.

3.4 Conclusions

EXCITON METHOD

P^{reamble}

4.1 Theory

4.1.1 Exciton States

4.1.2 Embedding

4.2 Truncated Chlorophylls

4.2.1 Rotation

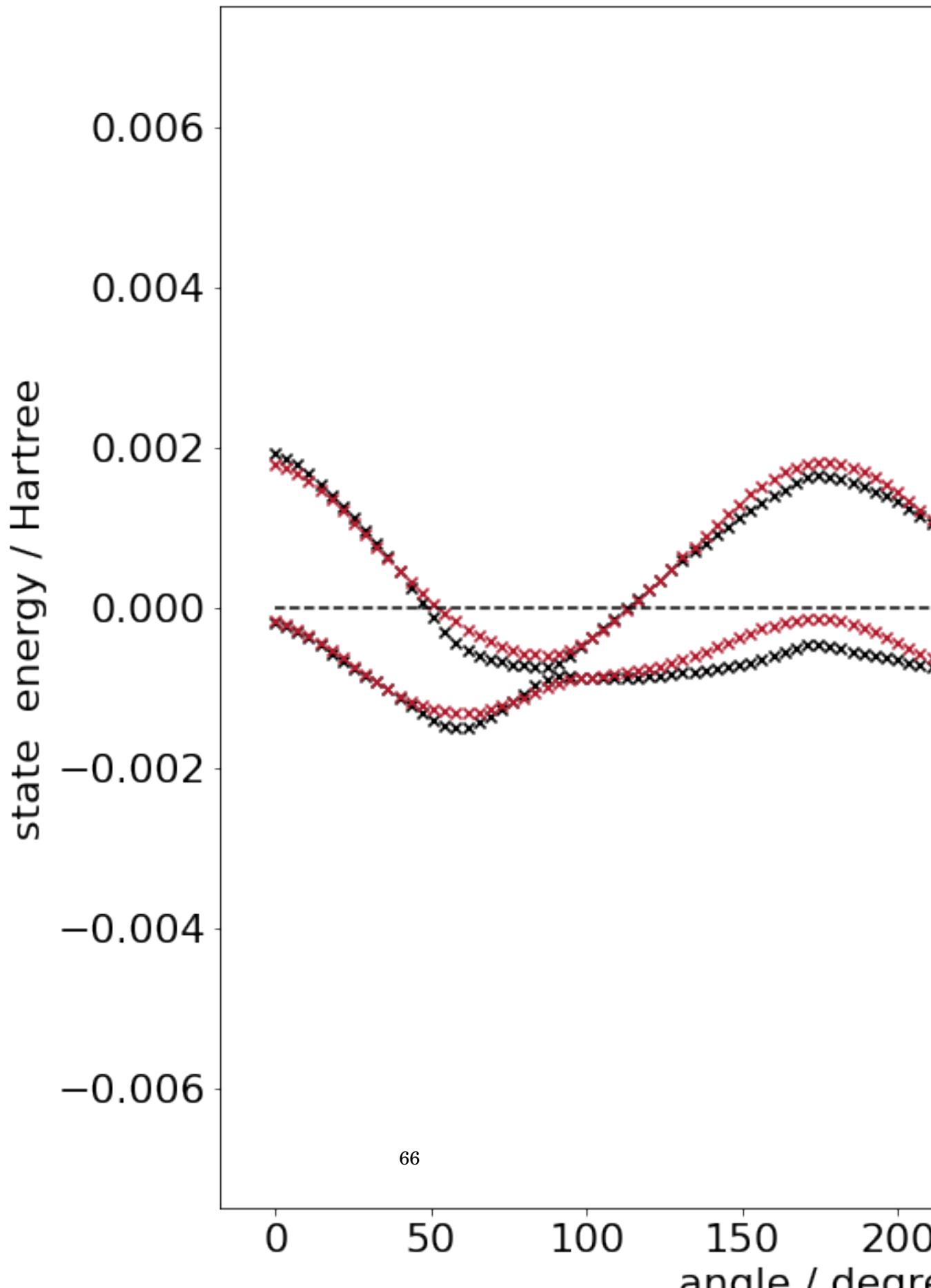
4.2.2 Distance

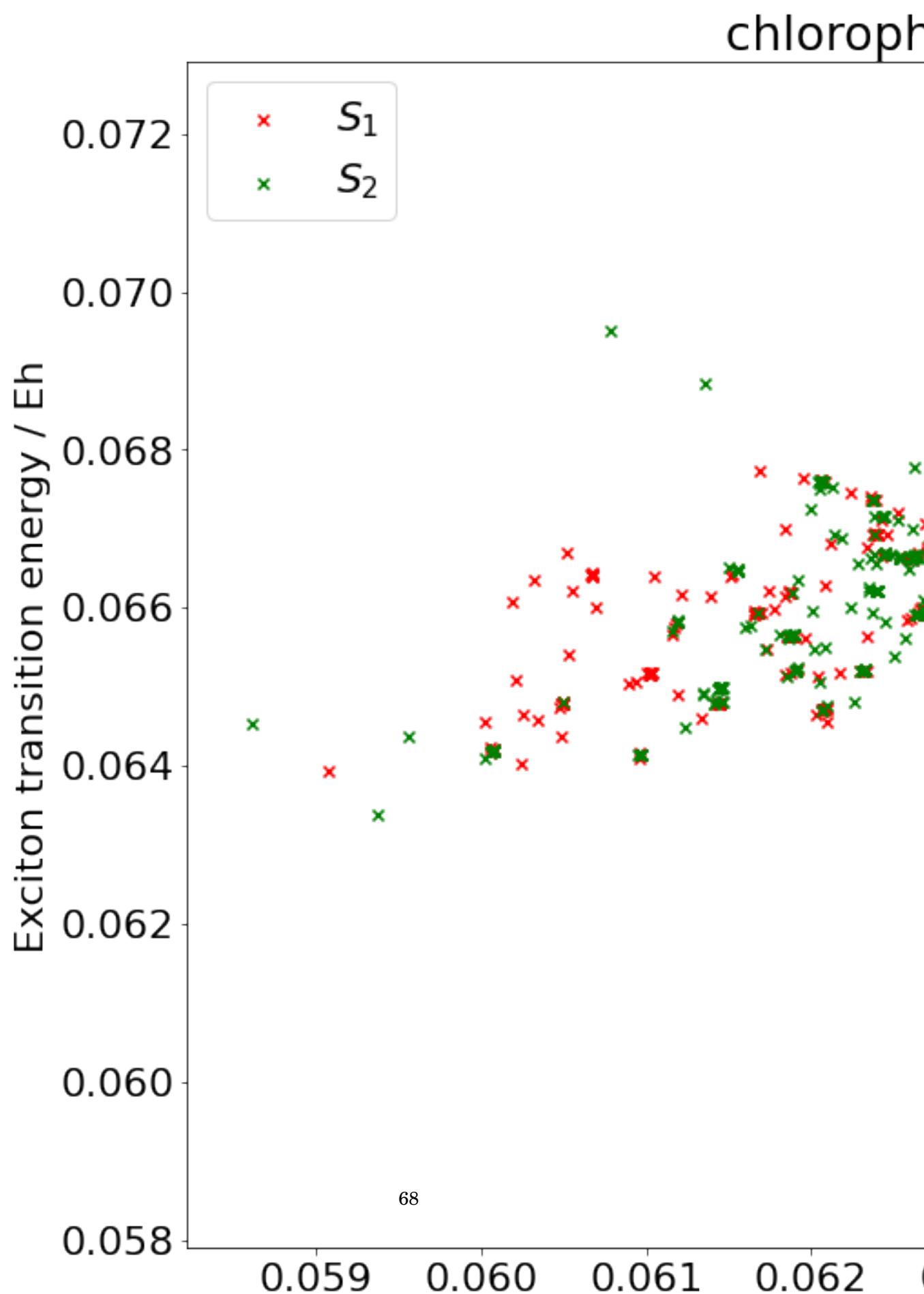
4.3 LHII Pairs

4.3.1 Assignment of States

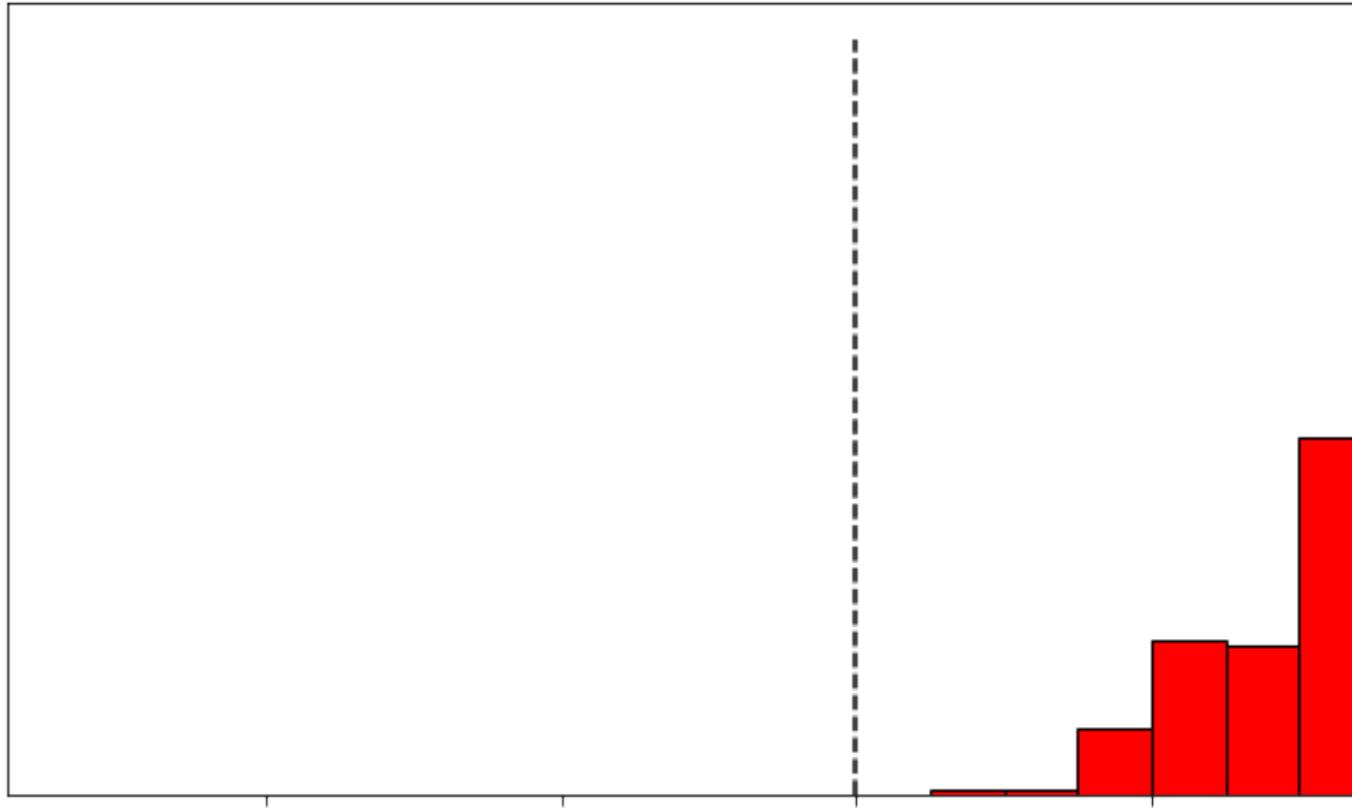
4.3.2 Comparison

Excited states of a ro

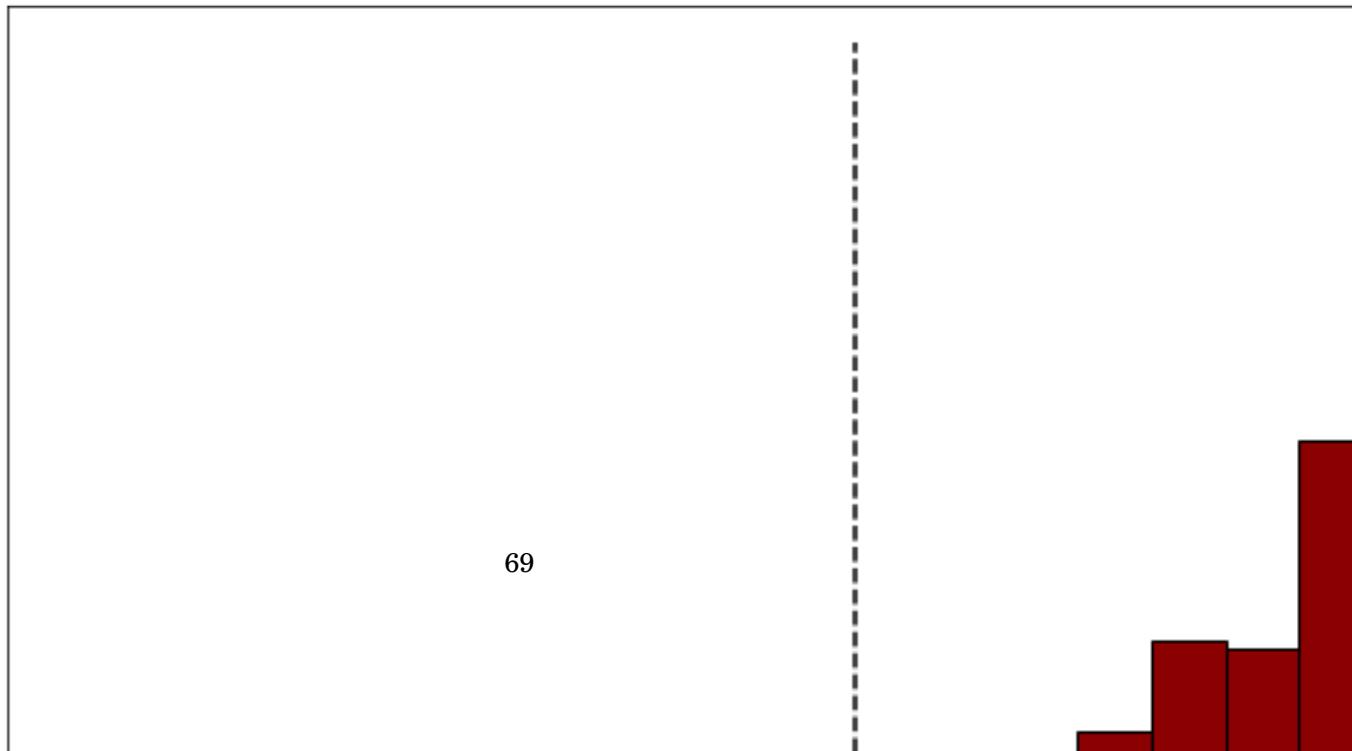




energy ordered



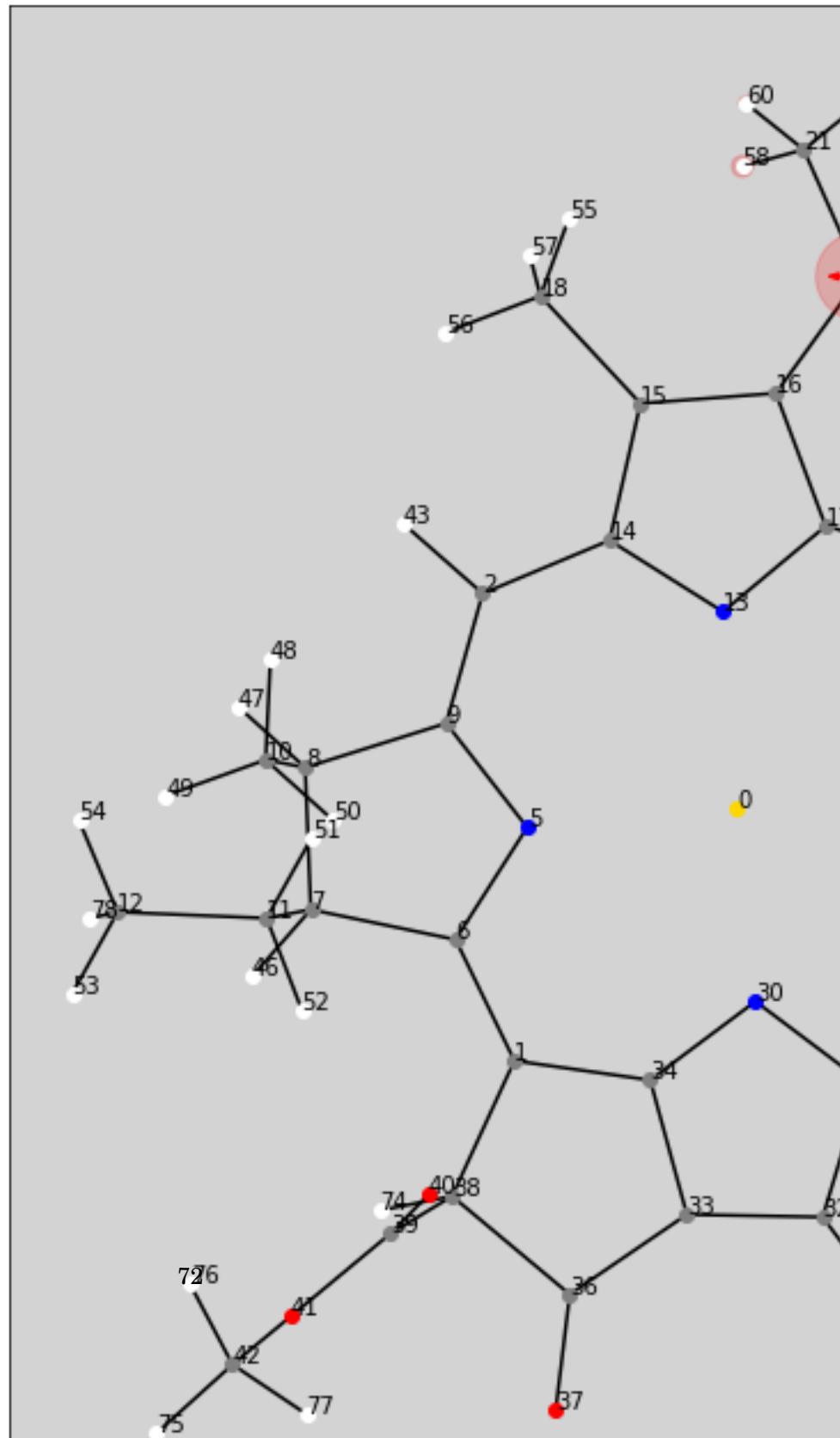
centre ordered

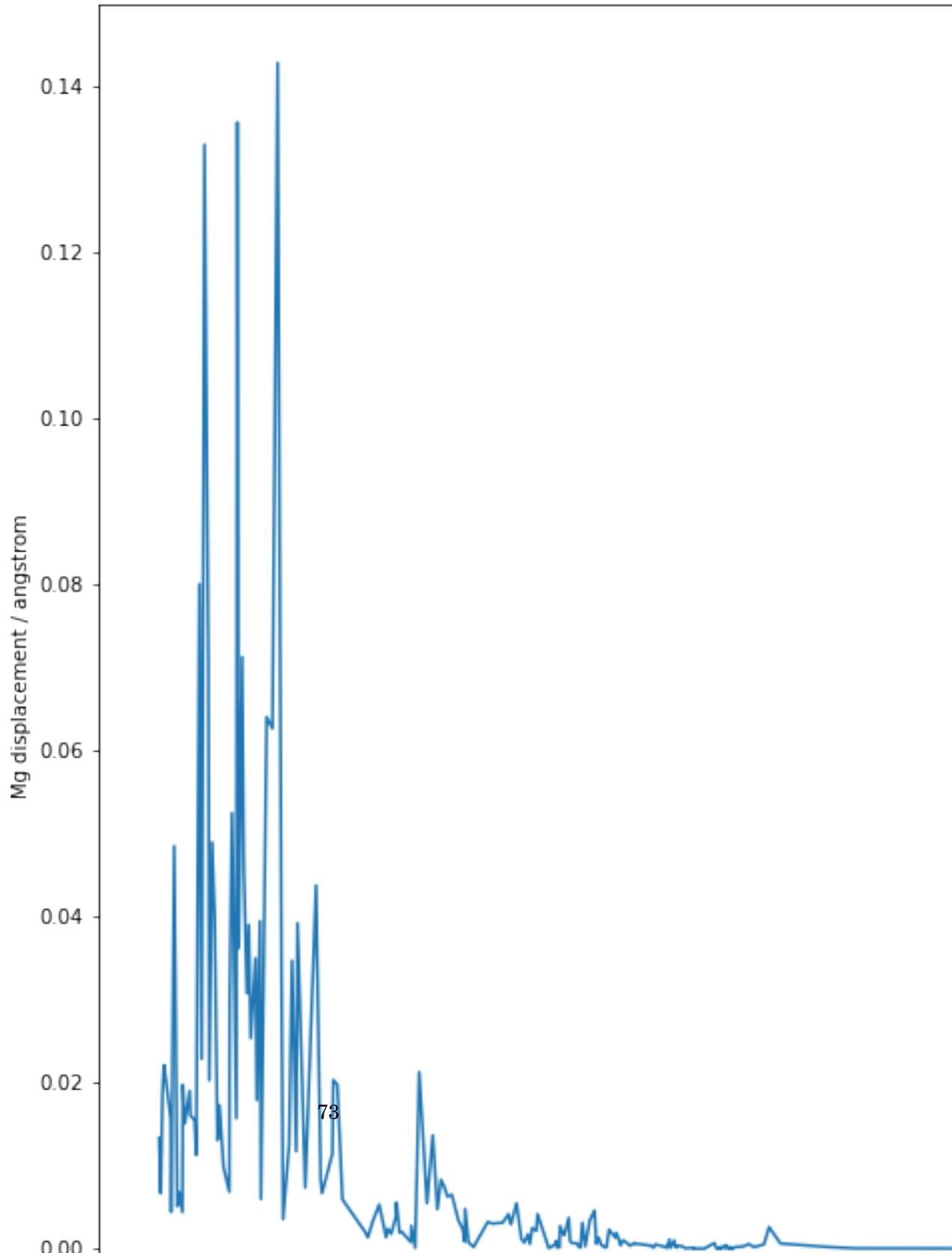


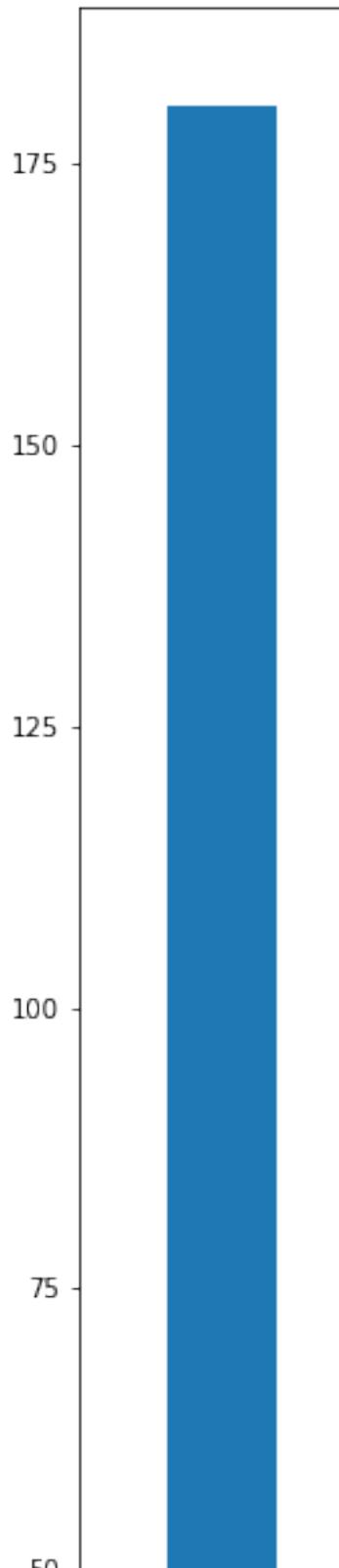
C H A P T E R

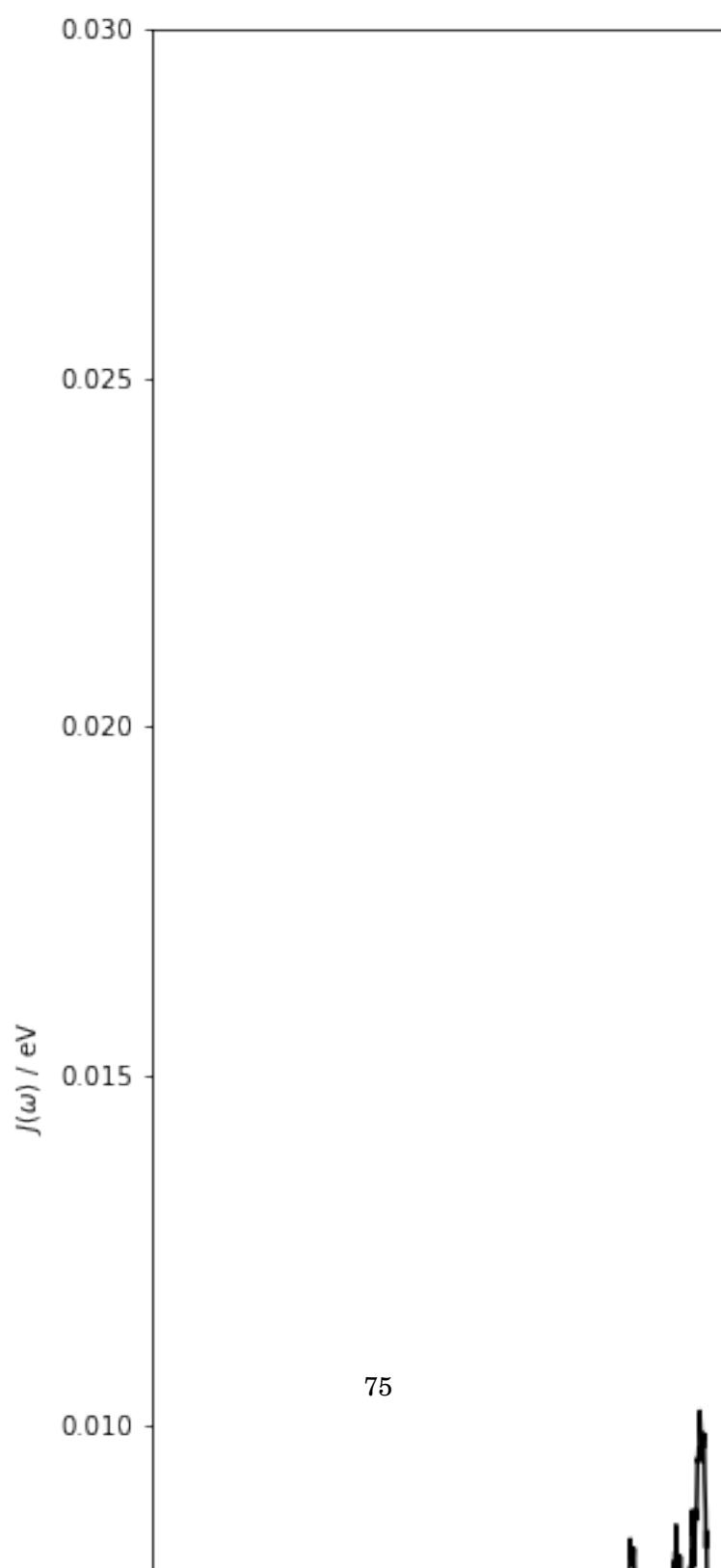


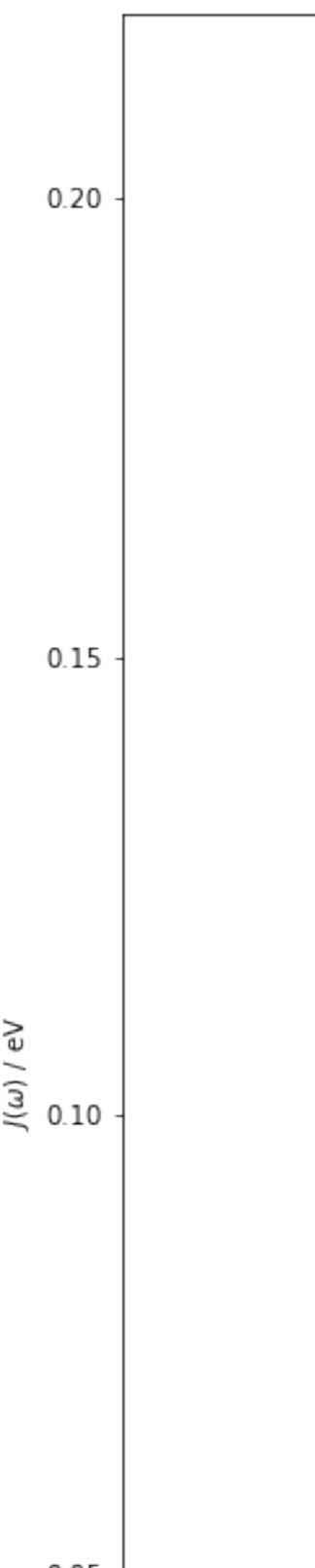
ATOMISTIC MODELLING OF LIGHT HARVESTING COMPLEXES

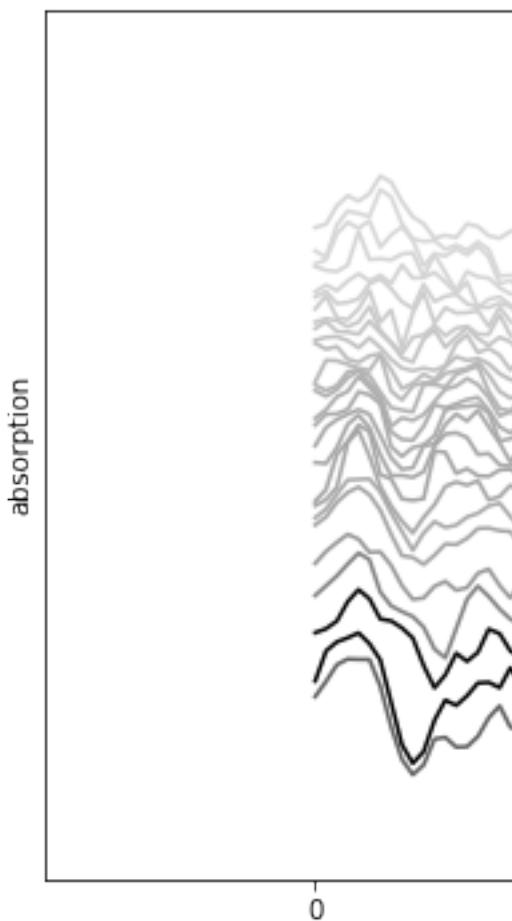


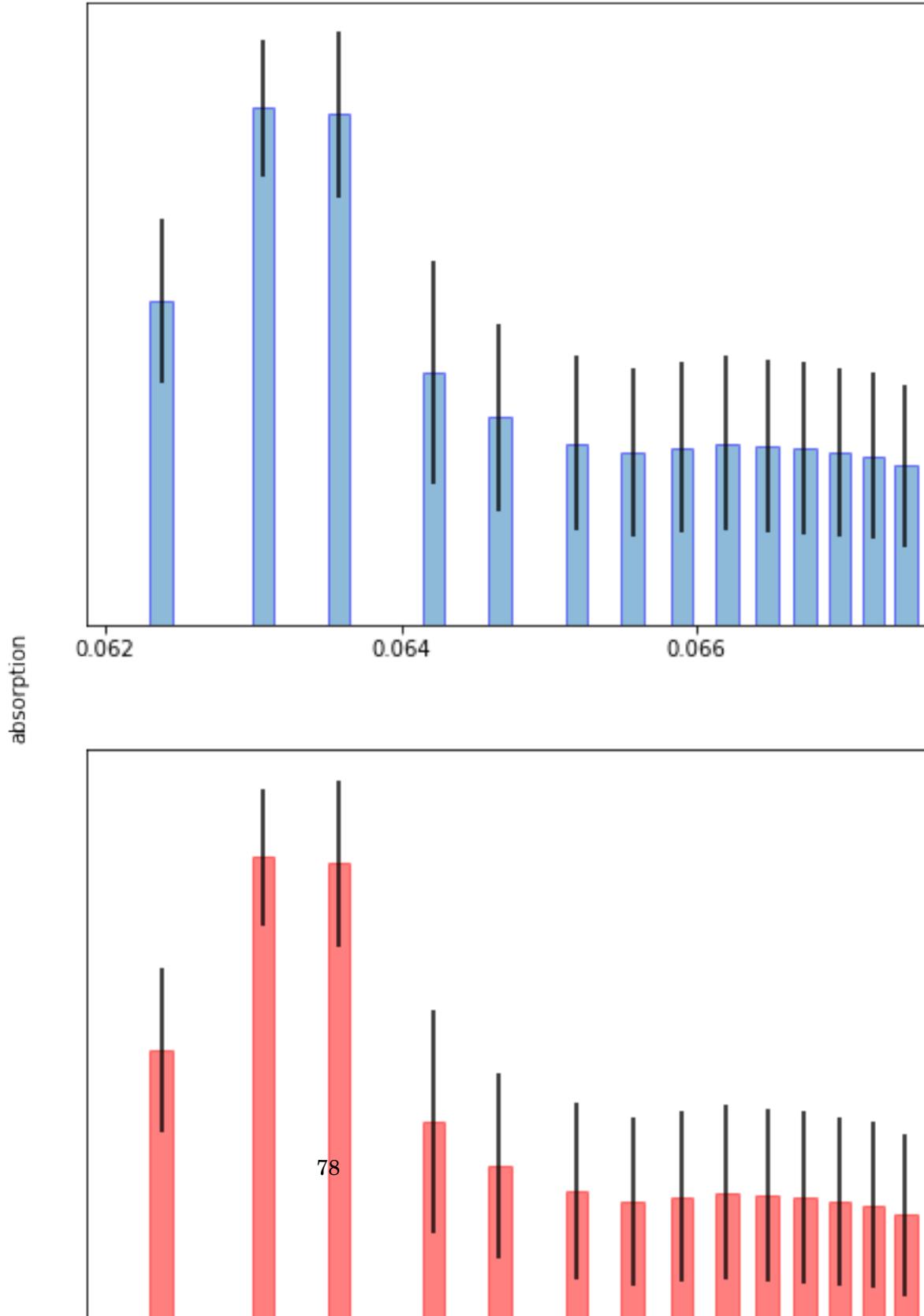












5.1 LHII

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

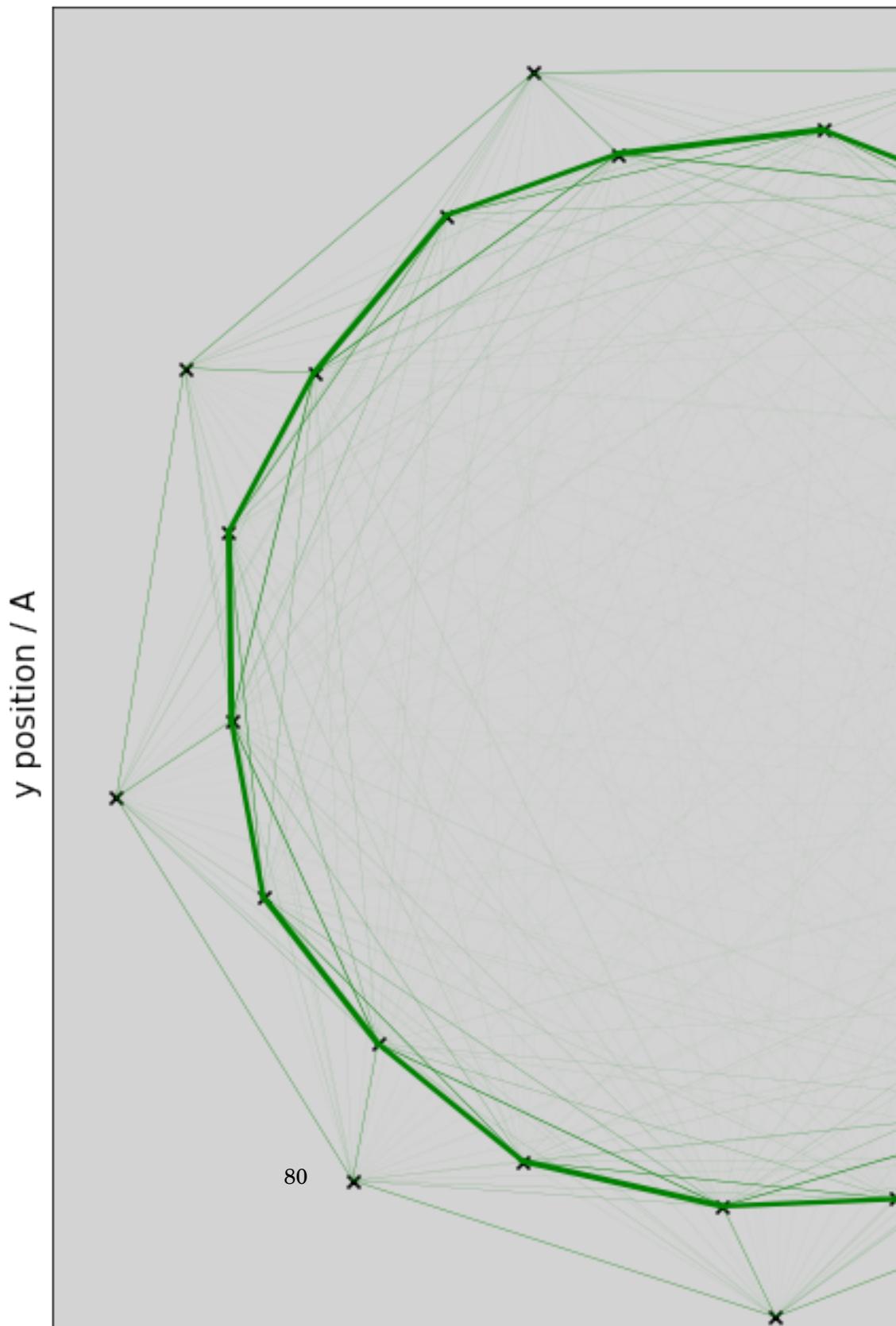
5.4 Sites, states and couplings

5.4.1 Sites

5.4.2 Exciton states

5.4.3 Coupling

5.5 Excitation Energies



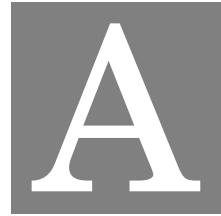
DISCUSSION

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] 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. GRUHN, 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] 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.
- [19] 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.
- [20] 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.
- [21] 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.

BIBLIOGRAPHY

- [22] 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).
- [23] K. NISHIMOTO AND N. MATAGA, *Electronic Structure and Spectra of Some Nitrogen Heterocycles*, Zeitshrift fur Phys. Chemie Neue Folge, 12 (1957), pp. 335–338.
- [24] K. OHNO, *Some Remarks on the Pariser-Parr-Pople Method*, Theor. chim. Acta, 2 (1964), pp. 219–227.
- [25] C. C. J. ROOTHAAN, *New Developments in Molecular Orbital Theory*, Rev. Mod. Phys., 23 (1951), p. 69.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.

