Towards Understanding Underlying Principles for Efficient Energy Transport in Biological Complexes

by

Kirsten Louise Claridge

Thesis submitted towards the award of the degree of Doctor of Philosophy.

University of Liverpool, Department of Chemistry

October 2019

Contents

List of Figures

List of Tables

List of Abbreviations

2DES	Two-dimensional electronic spectroscopy
ACF	Autocorrelation function
BCL	Bacteriochlorophyll-*a*
CLX	Chlorophyll-*x*
CT	Charge transfer
CYC	Phycocyanobilin
DFT	Density functional theory
EET	Electronic energy transfer
ES	Electronic structure
FF	Forcefield
FMO	Fenna-Matthews-Olson complex
FT	Fourier transform
IPR	Inverse participation ratio
LH 2/II	Light harvesting complex 2/II
LHC	Light harvesting complex
LO	Local oscillator
MD	Molecular dynamics
MDFF	Molecular dynamics forcefield
MM	Molecular mechanics
NMA	Normal modes analysis
PDB	Protein data bank
PEB	Phycoerythrobilin
PE545	Phycoerythrin 545
PPC	Pigment protein complex
QC	Quantum chemical
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
QM/MMPol	Quantum mechanics/molecular mechanics with polarised MM part
RMS	Root mean square
TDDFT	Time-dependent density functional theory

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Alessandro Troisi, I am truly grateful for the opportunity to carry out this research and for all his support and guidance over the course of my PhD. His insight and scientific judgment have proved invaluable and our many discussions about this work have provided me with far more than just a deeper understanding of the subject.

I must also acknowledge the European Research Council for financial support.

I'd also like to thank Dr. Daniele Padula for all his efforts in our collaborative work. I would also like to thank him along with all the other members of the Troisi research group (past, present and visiting) for their friendship over the past 3 and a half years.

From the Warwick days of my studies there are a couple people I must thank, these are Max Saller for patiently answering my (many) questions regarding various topics and for being a consistent lunch companion and Jack King for his friendship over the entire course of my studies.

Finally, I must mention my partner Alex Cousins for everything he's done for me throughout this time and my parents (Susan and Gordon) for their unending well of emotional support and belief in me, without which I could not have reached this point. (Also my cat for de-stressing pets.)

Declaration

This thesis is submitted to the University of Liverpool in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by me except in the cases outlined below.

Kirsten Claridge, October 2019

In chapter 1 Figure 1.4 was reproduced from: Panitchayangkoon, G.; Hayes, D.; Fransted, K. A.; Caram, J. R.; Harel, E.; Wen, J.; Blankenship, R. E.; Engel, G. S. Long-Lived Quantum Coherence in Photosynthetic Complexes at Physiological Temperature. *Proc. Natl. Acad. Sci.* **2010**, *107*, 12766–12770.

In chapter 4, Dr. Daniele Padula generated Figures 4.1, 4.2, 4.4 and 4.5 (from data provided by the author).

Parts of this thesis have been published by the author:

1. Claridge, K.; Padula, D.; Troisi, A. How Fine-Tuned for Energy Transfer Is the Environmental Noise Produced by Proteins around Biological Chromophores? *Phys. Chem. Chem. Phys.* **2018**.
2. Claridge, K.; Troisi, A. Developing Consistent Molecular Dynamics Force Fields for Biological Chromophores via Force Matching. *J. Phys. Chem. B* **2019**, *123*, 428–438.
3. Claridge, K.; Padula, D.; Troisi, A. On the Arrangement of Chromophores in Light Harvesting Complexes: Chance versus Design. *Faraday Discuss*. **2019**, DOI: 10.1039/C9FD00045C

Other published work:

- Padula, D.; Lee, M. H.; Claridge, K.; Troisi, A. Chromophore-Dependent Intramolecular Exciton−Vibrational Coupling in the FMO Complex: Quantification and Importance for Exciton Dynamics. *J. Phys. Chem. B* **2017**, *121*, 10026–10035.

Abstract

Light harvesting complexes are a type of biological complex found in photosynthetic bacteria and plants. They perform a critical role in the photosynthetic pathway – transporting energy absorbed from sunlight to the reaction centre – and do so with exceptional efficiency. Understanding the root of this however has proved difficult due to the large size and the complexity of the systems. The wide structural variability of these complexes further complicates matters, raising the question of optimality of the complexes: is the exceptional efficiency by chance or design? This work aims to aid the description of the underlying principles of energy transport in these complexes and move towards answering the question of chance versus design. A force matching method to develop molecular mechanics forcefields for use in studies combining molecular dynamics with higher level methods (a strategy regularly employed in biophysical research) is outlined. This method reduces errors resulting from the methodology and lessens computational demand of developing forcefields, making the study of multiple complexes more feasible. Resulting forcefields lead to notable differences in subsequent spectral density computations. Structure-function relationships and the role of environment is investigated by examining spectral densities of rigid chromophores of several complexes with different functions. Analysis of the results reveals the environment to be unspecific in regard to function. Examining the exciton dynamics of several light harvesting complexes may uncover important characteristics of light harvesting. In order to study this the excitonic Hamiltonian of several complexes is computed and their dynamics propagated using a Lindblad master equation. The results indicate the protein likely plays little role in exciton dynamics with the most influential component being the solvent. The exciton dynamics are found to be determined solely by the static disorder of the system and the results suggest relative arrangement of LHCs is more important for efficiency than internal arrangement of chromophores.

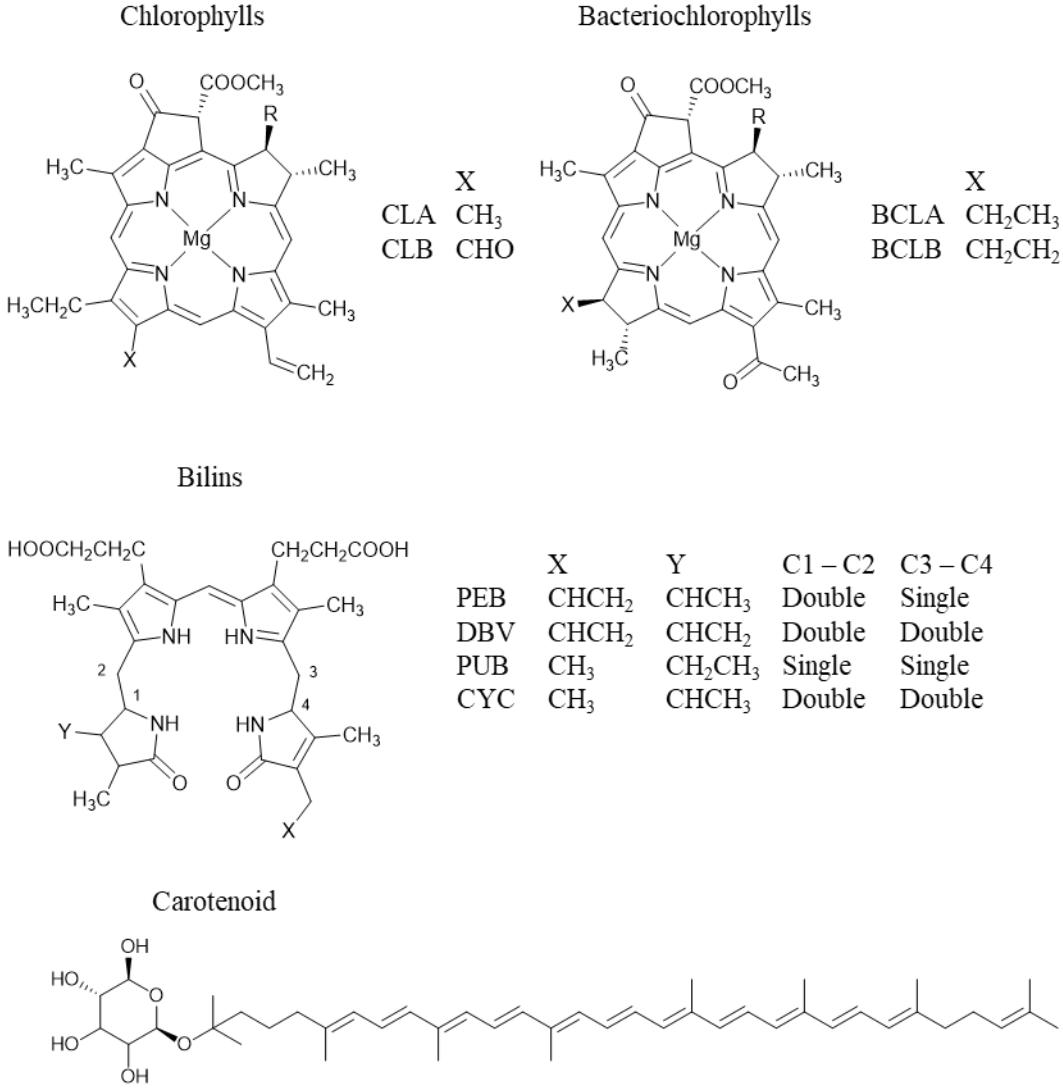
1 Introduction

This chapter gives an overview of the research area of pigment protein complexes (PPCs) including a brief description of the complexes and their structure as well as the background on why these proteins have gained such research interest and a general account of where current research is now focused. Finally, the aims of this work are introduced with a brief outline of the chapters that follow.

1.1 The PPCs and their structure

The PPCs are a group of complexes found in plants and photosynthetic bacteria. The complexes consist of a protein scaffold containing multiple pigment molecules, typically chromophores and carotenoids. The light harvesting complexes (LHCs) are a specific subset of the PPCs; those involved in transporting energy from absorbed light to the reaction centre – a vital process in photosynthetic organisms. The photosynthetic reaction centre is a complex of several proteins in which charge separations occurs and is responsible for the primary energy conversion reactions in photosynthesis. Unlike other ubiquitous vital biological components, the structures of LHCs vary considerably. Firstly, there are the pigments themselves. The chromophores directly involved in light-harvesting are classified into three groups: the chlorophylls (chls) – typically found in plant LHCs; the bacteriochlorophylls (bchls) – typically found in bacteria LHCs and the bilins – typically found in cryptophyte algae. Each group contains several different types of chromophore with slight structural variations. Representative structures detailing the differences between the chromophore groups are shown in Fig. 1.1. A LHC will contain light-harvesting chromophores from only one group but may contain several different types of chromophore from that group. In addition to the light-harvesting chromophores some LHCs also contain another type of chromophore, the carotenoid pigments. An example of a carotenoid molecule is shown in Fig. 1.1. Varying numbers of these pigments are arranged differently in the assorted LHCs. Next, there is the variability of the protein scaffold to consider. Generally, the main structure of the protein is formed of pairs of α helices which are associated with a pigment molecule.t[1] These α helices pairs are arranged to form the overall quaternary structure which can be exceedingly varied.

For example, large numbers of these pairs can group together to form very dense disc-like structures, or they can stand next to one another to form cylindrical structure, or a small number of pairs may form a smaller, less symmetrical shape; such examples can be seen in Fig. 1.2. However, whilst many LHC protein scaffolds fit this description, there are also structures that do not follow it at all. For example, the Fenna-Matthews-Olson complex $(FMO)^2$ – an LHC found in green sulphur bacteria – has a structure that is better described as a trimer of sandwich-like units, where the pigments are held between two layers of β sheets, also shown in Fig. 1.2.



	X
CLA	CH_3
CLB	CHO

	X
BCLA	CH_2CH_3
BCLB	CH_2CH_2

	X	Y	C1 – C2	C3 – C4
PEB	$CHCH_2$	$CHCH_3$	Double	Single
DBV	$CHCH_2$	$CHCH_2$	Double	Double
PUB	CH_3	CH_2CH_3	Single	Single
CYC	CH_3	$CHCH_3$	Double	Double

Figure 1.1 Chromophore types – depicted is the general structure with a table denoting the variations of the chromophores. Chlorophylls: chlorophyll-*a* (CLA) and chlorophyll-*b* (CLB); bacteriochlorophylls: bacteriochlorophyll-*a* (BCLA) and bacteriochlorophyll-*b* (BCLB); bilins: phycoerythrobilin (PEB), dihydrobiliverdin (DBV), phycourobilin (PUB) and phycocyanobilin (CYC). Structure of a carotenoid (Rhodopin glucoside).

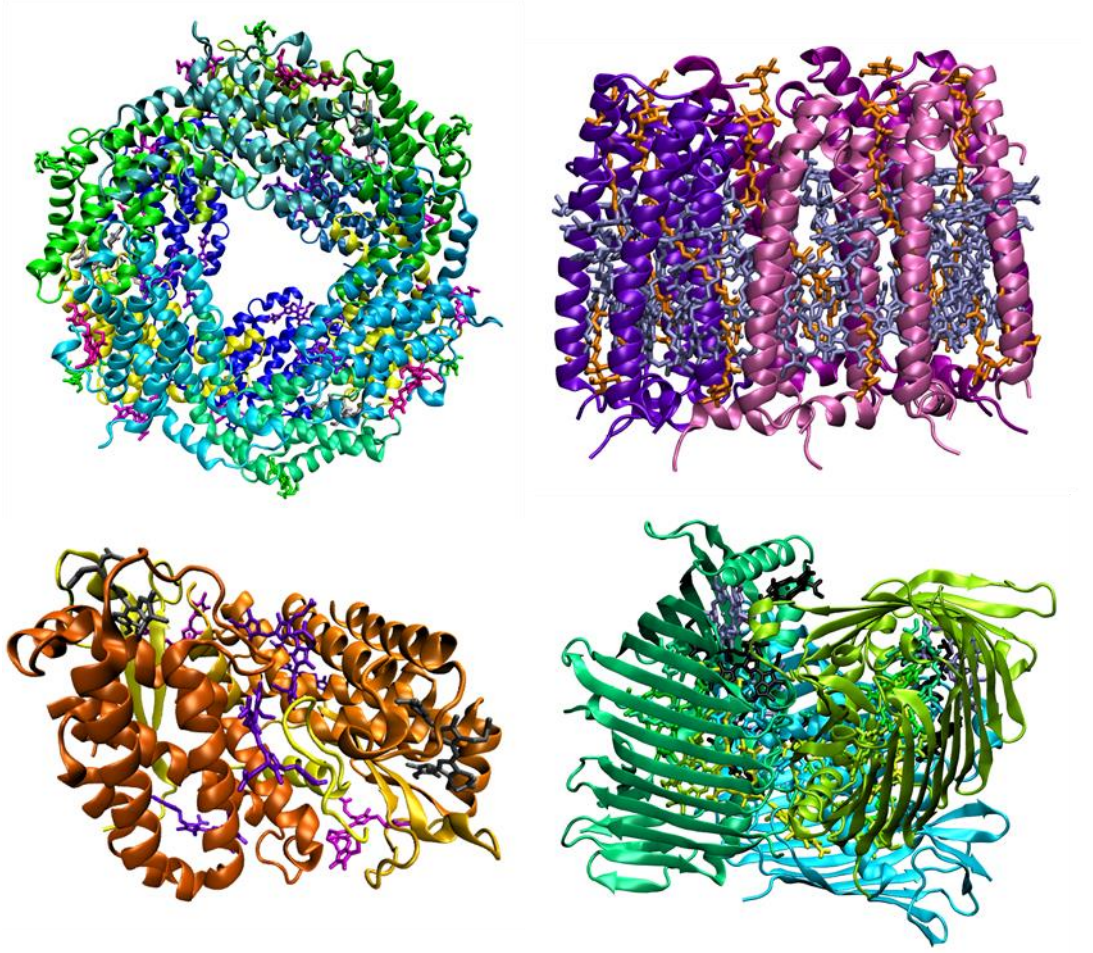


Figure 1.2 Protein structures of different LHCs: phycoerythrin – disc like (top left), LH2 – cylindrical (top right), phycoerythrin 545 (bottom left) and FMO (bottom right)

1.2 Quantum mechanical coherence in LHCs

1.2.1 Initial experimental reports

In 2007 Engel et. al. reported long-lived electronic quantum mechanical coherence at 77 K in FMO[3] observed from two-dimensional electronic spectroscopy (2DES) experiments. Later experiments also reported long-lived coherence in FMO at ambient temperature[4] as well as other observations of long-lived coherence.[5,6] These findings generated an intense research interest as they contradicted the previously held assumption that energy transport in LHCs occurred incoherently due to the magnitude of the fluctuations of the protein environment being sufficient to destroy any coherence. Though more recent evidence contradicts these findings,[7] LHCs remain an interesting topic for research. This is for a couple of reasons:

firstly, they have exceptional quantum efficiency[8,9] – in higher plants it has been reported that up to 90% of absorbed photons reach the reaction centre[10] – and avoid concentration quenching despite the dense arrangement of the pigment molecules;[11,12] secondly, the great structural variability of LHCs across light-harvesting organisms. The combination of these qualities of LHCs are interesting from two standpoints: materials design and scientific understanding. Elucidating the source of this high quantum efficiency potentially provides design principles for artificial light harvesters with practical applications, e.g. solar cells.[11] This structural variability combined with great quantum efficiency also poses interesting questions: are these structures by chance or design? What is the source of this high quantum efficiency? Do we understand the mechanism of energy transport occurring? Is there underlying commonality between the different complexes? Due to this combination LHCs present a very interesting field of research.

1.2.2 Exciton transport models

As it was the experimental reports of long-lived electronic coherence that generated interest into LHCs, it is unsurprising that initial research focused on coherence: how to model it and if, how and why it aids excitation energy transport (EET).[11,13–17] A system evolves coherently if it can be described as a superposition of two or more states, i.e. at a given time there is a finite probability that the system is in any of the states, thus the full density matrix is required to describe the state. Conversely incoherent evolution of a system is when the system can be described by a single quantum mechanical state at any given time and so only the populations of the density matrix are required. Before discussing the descriptions of coherent EET, it is important to outline the incoherent description as detailed by Förster theory[18] as this was initially assumed to be the correct description of EET and also to provide comparison. Qualitatively Förster theory describes EET as the 'hopping' of an excitation through the chromophore network, via the simultaneous de-excitation of a donor chromophore and excitation of an acceptor. The energy transfer rate k is given by:

$$k = \frac{2\pi}{\hbar} |V|^2 \int_0^\infty d\varepsilon \, I(\varepsilon_D, \varepsilon_A, \varepsilon), \tag{1.1}$$

where V is the electronic coupling between the donor chromophore, D, and acceptor chromophore, A, and $I(\varepsilon_D,\varepsilon_A,\varepsilon) = f(\varepsilon_D,\varepsilon)a(\varepsilon_A,\varepsilon)$, the spectral overlap defined as the product of the area-normalised fluorescence spectra of the donor f and absorption of acceptor a. Originally this theory was assumed to appropriately describe EET in LHCs but upon the discovery of long-lived coherences,[3] other descriptions were proposed.

Redfield theory[19] is a reduced density matrix approach where the interaction between the system and the environment (in the case of LHCs, the chromophore network and protein scaffold, respectively) is treated as a perturbation. As coherences are directly included in the density matrix description, Redfield theory was considered as a potential method to model and study coherence in EET. However, comparison of the Redfield and Förster theories when applied to two LHCs revealed Redfield strongly overestimated rates when the electronic coupling was weak, whilst Förster underestimated although gave more realistic results.[20] These results can be understood by considering the relative strengths of the electronic coupling V and the reorganisation energy (the energy dissipated due to the relaxation of an excitation causing nuclear reorganisation of the environment) λ and how these relate to the outcome of each model. When the electronic coupling is much smaller than the reorganisation energy ($V < \lambda$) Förster theory performs well, conversely Redfield theory is best when the electronic coupling is much larger than the reorganisation energy ($V > \lambda$) but, in LHCs the electronic coupling and reorganisation energy cover a range of relative magnitudes and typically have a similar relative magnitude ($V \sim \lambda$) and in this intermediate regime neither theory performs particularly well, as several studies have demonstrated.[21,22]

Modelling the intermediate regime was then the aim to describe EET and so modified Redfield theory[23,24] was considered. In modified Redfield theory the inclusion of the system-bath interaction is split so that the diagonal elements of the system-bath Hamiltonian are included non-perturbatively and only the off-diagonals are treated perturbatively. Consequently, modified Redfield theory was proposed as being capable of describing the intermediate regime, as well as the limiting regimes and this was found to be true. Modified Redfield resulted in an appropriate rate for the intermediate regime and correctly collapsed to either Förster or Redfield under their respective conditions.[24] But there are still some problems with modified Redfield theory. Firstly, in splitting the perturbation, direct examination of the off-diagonal elements (i.e. those relating to coherences) becomes impossible, so whilst the rate can be modelled with coherence directly included, no insight can be gained about it – which is ultimately the aim

The Markov approximation is an important underlying assumption of many reduced density matrix approaches, like Redfield. LHCs are open quantum systems as the chromophore network (quantum system) interacts with the protein scaffold (environment) and so not only can information flow from the system to the environment, it can also flow from the environment into the system. Consequently, the environment may retain information from the system and then pass this information back to the system later in time – the results of

which are known as memory effects. To disregard these effects, the Markov approximation assumes that the timescale of the environment fluctuations and the timescale of the evolution of the quantum system, are clearly separable. In other words, if the timescale of the process is much longer than the timescale of the environment fluctuations, it can be interpreted as the environment 'forgetting' the information before it can be passed back. This is can be a good approximation, but it has been argued that it is not valid for EET in LHCs as exciton dynamics can take place between 0.1 and 10 ps which is of a similar timescale for environment motion. Thus it has been argued that the inclusion of non-Markovian effects is required to fully and correctly account for the generation and survival of coherence.[25,26]

There have been several proposed methods to go beyond the Markov approximation in the study of LHCs. The hierarchical equations of motion (HEOM),[27] provide a formally exact solution which include the non-Markovian dynamics through a memory kernel. However, this is an extremely computationally expensive method and so is not considered routine. The reduced HEOM method was developed[28] for use in studying EET as it was asserted that in order to fully understand the role of coherence and how the environment influences it, the environmental phonon dynamics must be described in more detail. Application to a dimer system under various conditions showed that the model was capable of replicating both wave-like coherent and "hopping" incoherent energy transport mechanisms as well as reproducing Förster and Redfield theories under conditions of the Markov approximation. From the results, it was concluded that the long-lived coherence arose from sustaining of coherent oscillations by slow environmental fluctuations in a region with small reorganisation energy. In Redfield (and modified Redfield) the Markov approximation results in infinitely fast fluctuations which destroy the coherence. Another example of a non-Markovian method is the quantum master equation combined with the polaron transformation.[28] Application of this method to an LHC found in cryptophyte algae highlighted the importance of quantized vibrations for efficient EET and concluded that near-resonance of vibrational frequencies with the energy levels of excitons was the main feature responsible for promotion of EET.[29]

1.2.3 Debate on the functional role of coherence

Alongside study into how to model coherence, the questions of how and to what extent coherence aids transport, have also been a topic of interest. There have been arguments for both opposing limits: that coherence is essential for efficient transport[6,30] and that it is not essential and even reduces efficiency if it occurs at all,[7,31] as well as more balanced views: that it perhaps is not essential for efficiency but is nonetheless important.[4,13,17,32–35] The

arguments against the relevance of coherence predominantly relate to lab vs natural conditions. First, some argue that the ensemble averaging done in spectroscopy lab experiments is problematic as it washes out mechanistic detail of individual processes. Additionally, it has been proposed that vibrational coherences may have been misidentified as electronic in nature.[36] And finally, the fact that natural light is very different to the polarised excitation light sources used in lab experiments thus the experiments are not representative of real world conditions and therefore the observed results do not depict what occurs in natural light harvesting.[30] The argument is that as the intensity of sunlight is constant across all relevant timescales, all points in time are equivalent due to this constant illumination which means the system is essentially in a steady state, therefore excitation by sunlight can never lead to wavelike motion.[17] There has also been several studies that have shown there is no improvement in efficiency between the previously mentioned intermediate regime and Förster theory, indicating that coherence does not aid efficiency, for example HEOM calculation of the full 24-site FMO trimer dynamics found the timescale of transport was the same as that predicted by Förster theory.[37]

There have been several different explanations for how coherence may increase efficiency in EET. Once this intermediate coupling regime was identified, it was proposed that this led to competition between purely dissipative dynamics (as described by Förster theory) and unitary evolution (as described by Redfield) in which the delocalization of the excitation and superposition of molecular sites are robust to dissipation by the bath.[35] An increased rate then occurs as a result of this site coherence being maintained during EET. In this case, coherence is deemed necessary for efficient transport as it is the sustained coherence that is said to be increasing rate. Another popular explanation of how coherence aids transport evolved from the idea that the environment may be fine-tuned in some way to promote coherence rather than destroy it.[21] Dephasing can counteract destructive interference which would reduce trapping (where excitation energy is lost due to the exciton becoming localised on a site and eventually relaxing) which would in turn increase efficiency. From this idea of a fine-tuned environment arose the idea of a "promoting mode".[38] It was found that many LHCs contain high-energy, well-localised vibrations which were typically intramolecular and resonant with some electronic transitions.[29] It was suggested that coupling with these modes promoted coherent, reversible, electronic transitions which result in the regeneration of coherence and so there was actually protection of coherence. Vibrations of this type were identified in the LHC phycoerythrin 545 (PE545) and their existence was found to improve efficiency.[39]

Another avenue of research focused on the source of coherence in order to elucidate how coherence increases efficiency. The spectral density contains information of the system-environment interaction and as it was proposed that this interaction is key for coherence it became the focus of several studies.[39–45] There are two methods for computing the spectral density, these are combining molecular dynamics (MD) with some quantum chemical (QC) calculation, typically combined quantum mechanics/molecular mechanics (QM/MM) or normal modes analysis (NMA). A study of the light harvesting PE545 protein combined both these methods using MD with QM/MMPol to compute the spectral densities and then an NMA to determine the origin of any important nodes.[43] This found both protein and solvent played a role in modulating intramolecular vibrations and thus promoting or supressing certain interactions to enhance transport. However, another study combining MD and QC to compute spectral density compared the spectral densities of FMO and PE545 and found that the low frequency range (thought to be most responsible for influencing dynamics) of the two was very different.[39] In the case of FMO the origin of this low frequency part was predominantly the environment whilst for PE545 it was from the pigment molecules, contrary to the findings of the previous study. Another study focused on the spectral densities of the individual pigments of FMO,[42] in order to identify if there was any specificity to their environments but found very similar spectral densities for all pigments.

1.3 Moving beyond coherence

More recent research has diverged from a sole focus on coherence and is broadly covered by the following four categories: experimental spectroscopic studies – using spectroscopic techniques to gain insight into LHC properties; computational studies and method development – creating or improving methodology to study LHCs; phenomenological and protein specific models – developing models (with different levels of detail) to describe LHCs; feature investigation – investigating the importance of a particular quality of a LHC. These categories can be intersecting, for example one may perform experimental studies to inform a protein specific model or one may use feature investigation to determine what to include in a phenomenological model or one may develop a method to investigate a specific feature, etc.

1.3.1 Recent experimental work

The debate on the importance of electronic coherence [7,13,17,21,26,31,46] was initiated by reports of its existence from 2DES experiments and has evolved through the proposal of different possible interpretations of these experimental results. Therefore, it is useful to provide an

outline of 2DES experiment method. It is a third-order four-wave-mixing spectroscopic technique,[47–49] where the first three pulses generate the signal whilst the fourth pulse is used for detection (and is called the local oscillator, LO). Pulses 1 and 2 excite the system and the delay time between them is the coherence time, τ. The system then evolves over the waiting time, T. Pulse 3 then creates another coherence. Finally, the time between pulse 3 and signal emission is the detection time t. This is represented in Fig 1.3. The 2D spectrum is then obtained from a Fourier transform (FT) of the signal across τ and t for a given waiting time T. The initial coherence (created by pulses 1 and 2) can be thought of as excitation-energy absorption and the final coherence (created by pulse 3) can be thought of as excitation-energy emission. Thus, the resulting spectrum represents a map of excitation and emission energies as a function of the waiting time, T, an example of this result is given in Fig. 1.4. Any off-diagonal elements in this spectrum (known as cross peaks) indicate coupling between electronic states as if there were no coupling, the excited state absorption and emission would completely cancel one another out eliminating the presence of cross peaks. Oscillation of these cross-peaks over time indicates coherence and is called quantum beating. A more detailed explanation and details of the experimental set up can be found in ref. 50.

In the first report of long-lived coherences in FMO, quantum beating of a cross-peak lasting for up to 660 fs at 77 K was observed.[3] This cross-peak was assigned as the 1 – 3 exciton cross-peak, named so due to its assignment to BCLs 1 and 3 in FMO. Further 2DES experiments then reported the presence of quantum beating for this cross peak for over 1500 fs at 77 K and up to 300 fs at 277 K.[4] The conclusion drawn from these results was that electronic coherence existed long enough to be involved in exciton dynamics and therefore could be the potential source of high quantum efficiency in LHCs. However, in 2017, motivated by a lack of corroborating experimental evidence for long-lived coherence, further spectroscopic

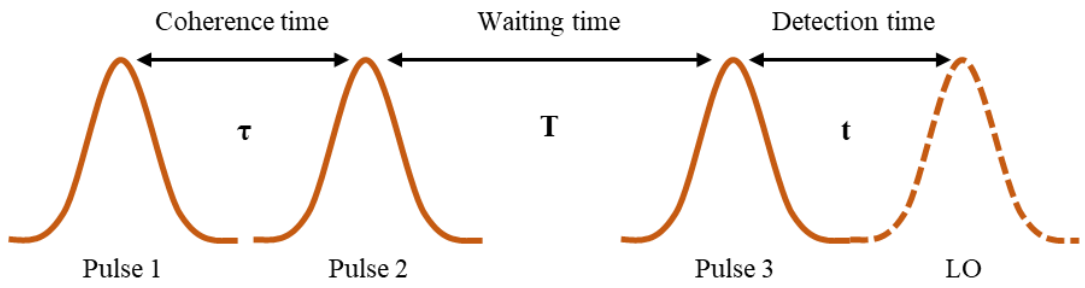


Figure 1.3: Representation of 2DES pulses and delay times.

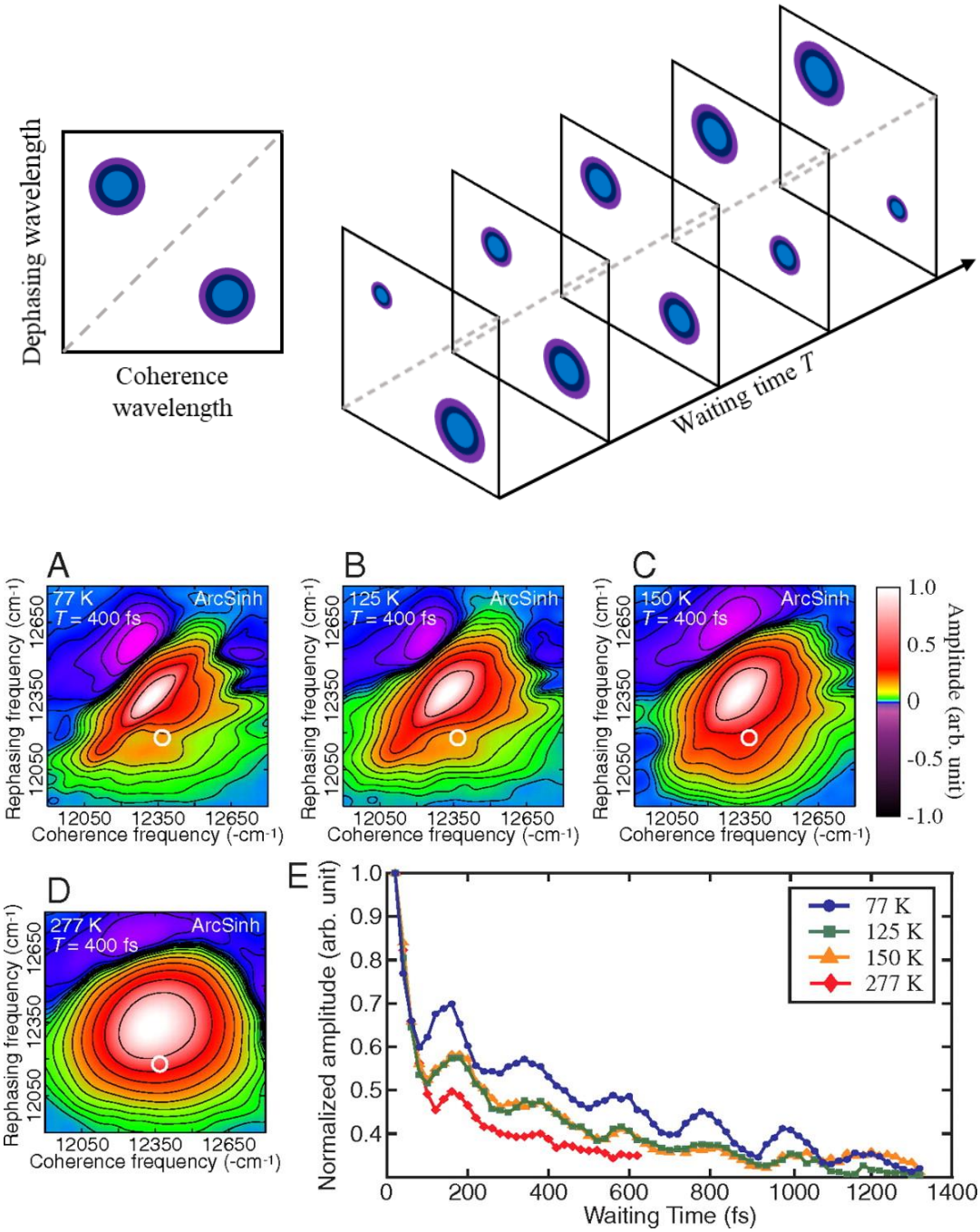


Figure 1.4 Schematic example of cross peaks beating in time (top). Real example of 2DES results taken from ref. 4 (bottom).

study of FMO was carried out under physiologically realistic conditions (296 K), in combination with theoretical modelling. The results of which concluded that electronic decoherence happened on a timescale of 60 fs, in line with a rapid decay of quantum coherence as opposed to long-lived. The authors also proposed that this result could be considered generalisable to other LHCs, due to the similarity in inter-pigment separation of FMO and other LHCs. Since this contradiction to long-lived coherence was reported, other

work has sought to corroborate or further contradict long-lived electronic coherence. 2DES study of the peridinin-chlorophyll complex also found short (< 100 fs) coherence times.[50] However, the research also found substitution of chlorophyll-*a* with chlorophyll-*b* in the complex led to slower energy transport. It was proposed that this was due to even faster decoherence owing to the formyl group of chlorophyll-*b* increasing the coupling to the bath motions due to an increased dipole moment resulting from the electron withdrawing nature of the moiety. Thus, it was concluded that even on these short-lived timescales coherence aids transport efficiency. Additionally, further studies have utilised a combination of experimental spectroscopy and theoretical modelling techniques and concluded that the observed quantum beating is in fact vibrational in origin as opposed to electronic.[36,51] Such findings highlight the role of theoretical work in the understanding of such complexes processes by aiding analysis, assignment and interpretation of complicated experimental results.

1.3.2 Computational chemistry studies

Studying LHCs computationally is difficult due to the large and complex nature of the systems. Even studying the "closed" system can be tricky as it typically consists of multiple chromophores, which are relatively large molecules (> 100 atoms). To this already complex system even more detail is added, as the "open" system is of interest – where the chromophores are interacting with the protein environment. Consequently, it is impossible to simulate or model these systems completely quantum mechanically – for example recently an quantum mechanics/molecular mechanics molecular dynamics simulation (QM/MM-MD) was performed on the PC645 protein.[52] In this QM/MM-MD simulations the dynamics of the system is run by propagating part of the system at a quantum level and part of the system at a classical level simultaneously. In this case the quantum part of the system consisted of 75 – 80 atoms and 8 sets of 50 ps trajectories were run. These simulations took 9 months and cost 2 million CPU hours to run, and whilst an impressive undertaking, the utility of the results is questionable due to the relatively short timeframe, which is too short for the system to be fully equilibrated. This study demonstrates the limits of what is currently possible for dynamics simulations of LHCs.

However open quantum systems may still be examined through combining quantum mechanics (QM) with molecular mechanics (MM) – usually through taking a classical MM trajectory as input for a QC calculation. Despite the conceptual simplicity, this approach contains a large number of approximations as well as many methodological choices which can be difficult to validate. There is the choice of forcefield (FF) for the MM part, including

whether to include polarizability, as well as the level of theory for the QM part, e.g. semiempirical or density functional theory (DFT) based, and the number of states to include. The sampling also requires consideration; a ns timescale is required for protein conformational change whilst fast motions and EET processes occur on ps timescale. Even the fundamental decision of how to split the system requires consideration: the most obvious choice is to have the protein environment as the MM part and the chromophores as the QM part, but perhaps there is some amino acid side chain in close proximity to the chromophore that also warrants QM treatment. It is not always possible to foresee the consequences of these decisions. For example, a now well-known issue when combining QM and MM methods stems from the inconsistency of the methods was discovered only once many studies utilising this method has been performed – the result has been dubbed the "geometry mismatch problem".[41] It is the result of the inconsistency between the methods which leads to different equilibrium structures, an issue that will be addressed in chapter 3.

1.3.3 Recent phenomenological studies and relation to atomistic description

Protein specific and phenomenological studies have essentially the same goal: to understand EET in LHCs but they employ different methods. In phenomenological research the focus is on the underlying physical principles and using simplified model systems to investigate and demonstrate EET. In the study of LHCs, a typical phenomenological model consists of a simplified description of the chromophore network and some simplified description of its interaction with the bath. For example, the chromophore network can be described by it excitonic Hamiltonian matrix, where the diagonal elements are the excitation energies of the chromophores and the off-diagonal elements are the couplings between the chromophores. The interaction with environment can then be brought in in a number of ways, e.g. as Lindblad operators.[53]

Several of the most recent phenomenological models have focused on the role of vibrations,[54–56] considering the more recent experimental evidence suggesting quantum effects arise from vibrations. Comparison of inclusion of under-dampened vibrational modes in either the quantum sub system or the classical bath of a model 3 site system found no significant difference in result suggesting that if quasi-resonant vibrational modes aid EET, their inclusion in models can be achieved through inclusion in the quantum subsystem or as part of the classical bath.[55] Another phenomenological study examined the effect of site specificity of vibrations on a two-site model system connected to a source and sink.[54] It found

that when both sites coupled to a global bath, the coupling led to reduced coherences and lower transport rate whilst each site coupling to individual vibrational environments led to enhanced resonance and increased transport. Suggesting that inclusion of the specific environment in the study of LHCs is important.

In the case of protein specific studies, typically atomistic detail of the LHC is included somehow or model inputs and parameters are tuned to represent the LHC in an attempt to gain more specific insights into that individual complex.[57–60] The results are sometimes intended to be expanded or compared to other complexes or for application to materials design. There is much wider variety in the nature of these sorts of studies, due to the wider variance of systems that are available for investigation. For example, one protein specific study aimed to simulate the 2DES of the light harvesting II complex[57] (LH II) whilst another used a generalised master equation to examine the robustness of transport between light harvesting 2 (LH2, different to LH II) units to identify the robustness of transport to differing orientations and separations of the units.[59] There is also a wider range of tools available to use in protein specific studies experimental methodology can be implemented for protein specific studies, for example hole burning spectroscopy experiments of FMO and a mutated FMO showed that the single point structural mutations resulted in more than single site energy shifts, indicating the effect of the system-environment interaction.[58] Valuable insights can be gained from both phenomenological and protein specific studies to aid the understanding of EET and uncover principles for the fine tuning and improvement of efficiency in artificial materials.

1.3.4 Investigating features influencing EET

A number of studies have focussed on discovering the relationship between a particular feature (structural or electronic) and the efficiency of transport, to help understand the origin and to inform design principle for artificial light harvesters. If some important feature can be identified, it can elucidate the biological function or even provide means to improve artificial light harvesters. Additionally, such studies can clarify what details are necessary to include in LHC models – if some feature is found to be completely unimportant or conversely predominantly deterministic of dynamics, then the model can be simplified by excluding it (in the former case) or excluding many other components other than the feature (in the latter case). The role of vibrations in EET have been the subject of study,[61,62] finding that intramolecular vibrations improve transport efficiency in some way. Other properties have also been investigated, for example to understand aspects of the absorption spectra of LH2, the effect of charge transfer (CT) states were studied by including the mixing of excitons and

CT states in the electronic Hamiltonian used to simulate the absorption spectrum.[63] It was discovered that this led to the broadening of a particular peak at 850 nm in the spectrum, believed to be key for efficient transport. Another study focused on the macrostructure of a couple of LHCs with identical chromophore arrangements but different protein scaffolds, in order to uncover the source of a 30 nm blue shift in one of the spectra.[64] The conclusion was that the curvature of the macrocycle ring of the protein was the most important factor. Such a discovery has the potential to inform design principle for the tuning of absorption spectra in artificial light harvesters. The extensive variability in structure and properties of LHCs creates a vast feature space for potential examination.

1.4 Objectives of this work

The research carried out on LHCs so far has undoubtedly provided insight into EET and properties of LHCs in general. However, a unified underlying description of transport mechanism remains elusive. Due to the size and complexity of these systems, studying them in detail is often computationally demanding, making it difficult to study more than one or two complexes at a time. Cross comparison of studies is not straight forward as there is such a range of methodologies in use when examining these systems. Resulting in challenges in identifying commonalities and adequately answering questions surrounding optimality of these complexes and what aids efficient transport.

In order to move toward a more coherent description of transport in LHCs and to identify key principles in light harvesting, in this work the aim is to study a couple of crucial elements of EET for a range of LHCs in a unified way. The three research chapters of this work address different problems related to this goal: a computational chemistry problem, a biochemistry problem and a biophysics problem. Firstly, in chapter 3, a new solution is proposed for one of the predominant issues in combining QM and MM methods – the geometry mismatch problem. A method for molecular dynamics forcefield parameterisation is introduced which utilises gradient descent to produce a forcefield whose parameters give rise to consistent forces with a corresponding quantum chemical calculation. Next in chapter 4, motivated by the speculation that the protein environment of LHCs is optimised in some way to promote efficient EET, the interaction of the environment with the chromophores of several PPCs with different functions (one light harvesting and two chromophore transporters) are examined. By examining the role of environment for these different complexes with different functionalities, the existence of some specialised interaction or vibrational mode should be clear. To include atomistic detail of the protein environment QM/MM calculations were performed using

snapshots from an MD trajectory as input. Furthermore, to focus solely on the role of the environment the chromophores were kept rigid in the MD. Finally, in chapter 5, a homogeneous method is used to examine the exciton dynamics of five LHCs and use statistical analyses to examine underlying factors influencing dynamics. Through these works the aim is to gain an understanding of the underlying principles of EET in LHCs and identify any important feature, if they exist, through a direct comparison of results obtained.

2 Theoretical background

In this chapter the details of the various computational and theoretical techniques utilised in the rest of this thesis will be outlined. The topics covered are: QM/MM, MD, TD-DFT, exciton dynamics and Lindblad master equation, spectral densities and excitonic couplings. Sections 2.1-3 and 2.5 are relevant for chapters 3 and 4. For chapter 5, sections 2.1 and 2.4 are relevant.

2.1 QM/MM

In a perfect world there would exist unimaginably powerful computers, which could be used to quantum mechanically simulate any system regardless of its size. Sadly, it is not a perfect world and treating protein systems like the LHCs completely quantum mechanically is simply impossible. However, often it is only some smaller part of the total protein system that is of chemical interest and it is not necessary to model the rest at a quantum mechanical level. Simulating a system by treating part of it with an appropriate QM level and the rest using MM is much more feasible and this is the approach of hybrid QM/MM simulation methods. The method was first introduced in 1976 by Warshel and Levitt[65] and has since become a very popular method of simulation of systems too large to be treated entirely at an *ab initio* level but involve some property that cannot be adequately described by simple MM methods. QM/MM is ideal for studying the electronic properties of chromophores embedded in their protein environment.

In QM/MM three types of interactions contribute to the potential energy: interaction between MM region atoms, interaction between QM region atoms and interaction between QM region and MM region atoms. The underlying method is simple: obtain the energy for the MM region utilising MM methods, the QM region with QM methods and calculate the QM/MM interaction somehow. However, the actual implementation of QM/MM can be very different from case to case. For the MM part there is many traditional FFs to choose from, as well as more complex polarisable FFs. Similarly, the QM part may be treated semi-

empirically (e.g. ZINDO) or with DFT. Once the MM and QM methods have been chosen, their treatment is then straightforward, it is the QM/MM interaction that is most difficult to describe. The approaches for treating the cross-region interactions can be divided into two categories: subtractive coupling schemes and additive coupling schemes.

In the subtractive coupling scheme, the MM energy of the total system is calculated first, then the QM energy of the isolated QM system is added. Finally, to correct for the double inclusion of the QM system interactions, the MM energy of the QM system is subtracted (hence subtractive) to give the QM/MM energy of the total system

$$E_{QM/MM}^{total} = E_{MM}^{total} + E_{QM}^{QM} - E_{MM}^{QM}. \tag{2.1}$$

The subscripts denote the level of treatment whilst the superscripts denote the section of the system. The ONIOM model[66,67] is the most commonly used subtractive method. As the subtractive coupling scheme requires no direct communication between the QM and MM subsystems it is relatively simple to implement. However, this absence of direct communication also means that the MM environment cannot polarise the QM electron density, something that can play an important role in biological charge transfer processes. Additionally, to use the subtractive coupling scheme it is necessary to have a FF that sufficiently describes the QM subsystem which is not always trivial to obtain.

In additive schemes the total energy is a sum (hence additive) of MM energy, QM energy and a QM/MM coupling

$$E_{QM/MM}^{total} = E_{QM}^{QM} + E_{MM}^{MM} + E_{QM-MM}^{total}. \tag{2.2}$$

Here the interaction of the subsystems is treated explicitly. There are several methods for describing these interactions: mechanical embedding, electrostatic embedding and polarisable embedding. In this work the only additive approach used is electrostatic embedding (chapters 3 and 4), so only this approach will be outlined here. In electrostatic embedding, the electrostatic interactions between the QM and MM subsystems are treated during the computation of the electronic wavefunction through an electrostatic contribution to the Fock matrix. The MM environment is treated as classical point charges which enter the QM Hamiltonian as on-electron operators

$$h_i^{QM-MM} = h_i^{QM} - \sum_J^M \frac{e^2 Q_J}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_j|}, \tag{2.3}$$

where $\mathbf{r}_i$ is position of electron i, $\mathbf{R}_J$ is the position of MM atom J, h_i^{QM} is the original one-electron operator for electron i and M is the number of MM atoms that have a partial charge Q_J. So, the electrons experience these MM atoms as special nuclei with non-integer charges.

The QM and MM subsystems may be connected by chemical bonds. In such cases simply cutting the bond would impact the QM wavefunction as it would add one or more unpaired electrons to the QM subsystem which do not actually exist. Again, there are several approaches to treat such instances; using link atoms is one such solution. Introducing a monovalent link atom at an appropriate position along the region crossing bond is the simplest solution. Typically, hydrogen is used, but there is no restriction on the atom type and fragments, such as methyl groups, are often used too. The link atoms appear only in the QM calculation and not the MM. A fixed position along the bond is given to the link atom at every step of the simulation in order to remove any additional degrees of freedom added by the insertion of the link atom.

2.2 MD

MD was formulated in the late 1950s as a way to model the dynamics of chemical systems.[68] It is based on the assumption that atoms behave according to classical mechanics and quantum mechanical features can be ignored. Thus, each atom is treated as a point particle with a corresponding mass. The classical Hamiltonian is

$$H(p,q) = T(p) + V(q), \tag{2.4}$$

where p denotes momentum and q denotes position. And so, $V(q)$ is the potential energy of the system and $T(p)$ is then the classical kinetic energy of the system, given by:

$$T(p) = \sum_{i=1}^{N} \frac{|\, p_i \,|^2}{2m_i}, \tag{2.5}$$

where N is the total number of atoms in the system and m_i is the mass of atom i. The time evolution of the positions and momenta of the atoms are then given by the following equations of motion:

$$\frac{dp}{dt} = -\frac{\delta H(p,q)}{\delta q} \quad \Rightarrow \quad \frac{dp_i}{dt} = -\frac{dV(q)}{dq_i}, \tag{2.6}$$

$$\frac{dq}{dt} = -\frac{\delta H(p,q)}{\delta p} \quad \Rightarrow \quad \frac{dq_i}{dt} = \frac{p_i}{m_i}. \tag{2.7}$$

Note that these equations simply correspond to Newton's second law. Equation (2.6) corresponds to $F = ma$ and equation (2.7) is simply the definition of velocity. Thus, in MD the goal is to solve Newton's equations of motion.

There are two main methods for integrating Newton's equations of motion, these are the Verlet algorithm[69] (also referred to as the Leapfrog algorithm) and the Velocity-Verlet algorithm.[70] The Verlet algorithm can be derived using Taylor expansions

$$r(t+\delta t) = r(t) + \frac{dr(t)}{dt}\delta t + \frac{1}{2}\frac{d^2r(t)}{dt^2}\delta t^2 + \frac{1}{3!}\frac{d^3r(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \tag{2.8}$$

$$r(t-\delta t) = r(t) - \frac{dr(t)}{dt}\delta t + \frac{1}{2}\frac{d^2r(t)}{dt^2}\delta t^2 - \frac{1}{3!}\frac{d^3r(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4). \tag{2.9}$$

Adding these together gives

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \frac{d^2r(t)}{dt^2}\delta t^2 + \mathcal{O}(\delta t^4). \tag{2.10}$$

The Verlet algorithm requires the position vector at $t - \delta t$ and so is not self-starting. The Velocity-Verlet algorithm explicitly incorporates velocity in order to avoid the issue of not self-starting. The integration cycle is as follows, first the momenta are updated at a mid-step

$$p\left(t + \frac{1}{2}\delta t\right) = p(t) + \frac{1}{2}\delta t F(t) \tag{2.11}$$

the positions at the next step are then updated

$$q(t+\delta t) = q(t) + p\left(t + \frac{1}{2}\delta t\right)\delta t \tag{2.12}$$

the new forces are then computed at the new position

$$F(t+\delta t) = -\frac{dV(q)}{dq} \tag{2.13}$$

finally, the momenta are updated

$$p(t+\delta t) = p\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t F(t+\delta t). \tag{2.14}$$

Utilising either of these algorithms provides a trajectory of the atoms in the system.

Another important component of MD is the FF. A FF is an equation for the potential energy of the system and its parameters. Typically, it is given as a sum of intra- and inter-molecular terms

$$V = V_{intra} + V_{inter}. \tag{2.15}$$

The intramolecular (or bonded) part of the potential energy is composed of terms relating to bond stretching, angle bending and torsional potentials

$$V_{intra} = \sum V_{bond} + \sum V_{angle} + \sum V_{tors}, \tag{2.16}$$

and the intermolecular (or non-bonded) part is given by the sum of van der Waals interactions and electrostatics

$$V_{inter} = V_{vdw} + V_{el}. \tag{2.17}$$

The exact for of each of these terms depends upon the FF used. Some examples of the most popular FFs in MD are AMBER,[71] OPLS,[72] CHARMM.[73] As the only FF utilised in the following chapters is the all-atom CHARMM FF,[74] only the exact form for that particular FF is given here, which is

$$\begin{aligned} V = & \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\varphi [1 + \cos(n\varphi - \delta)] \\ & + \sum_{impropers} k_\omega (\omega - \omega_0)^2 + \sum_{UB} k_u (u - u_0)^2 \\ & + \sum_{non-bonded} \epsilon \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}} \end{aligned} \tag{2.18}$$

where b denotes bond length, θ denotes bond angle size, and φ and ω denote dihedral and improper dihedral size respectively, u denotes the 1-3 atom distance of the Urey-Bradley component of angle bending from non-bonded interactions. The k_X parameters are force constants and X_0 parameters are equilibrium length/angle size. For the dihedral term, n is the multiplicity of the dihedral function and δ is the phase shift. (Note the torsional terms are split into dihedrals and improper dihedrals). The non-bonded term uses a standard 12-6 Lennard Jones potential for the van der Waals, where R_{min} corresponds to the point at which the potential crosses the x-axis and the electrostatic component is given by a Coulombic potential.

2.3 Electronic structure theory

The Schrödinger and time-dependent Schrödinger equations

$$H\Psi = E\Psi; \quad H\Psi(t) = i\frac{\partial}{\partial t}\Psi(t) \tag{2.19}$$

govern systems composed of electrons and nuclei, where H is the Hamiltonian of the system, E is the energy and Ψ is the many particle wavefunction. Thus, the energy (or time-dependent property) of a system can be found by solving the time-independent or time-dependent

Schrödinger equation. The general Hamiltonian (using atomic units) of such a system comprised of N nuclei and n electrons is given by

$$\begin{aligned} H &= -\sum_{I=1}^{N}\frac{\nabla_I^2}{2m_I} + \sum_{I<J}^{N}\frac{q_I q_J}{\left|\mathbf{r}_i - \mathbf{R}_J\right|} - \sum_{i=1}^{n}\frac{\nabla_i^2}{2} + \sum_{i<j}^{n}\frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} - \sum_{i=1}^{n}\sum_{J=1}^{N}\frac{q_J}{\left|\mathbf{r}_i - \mathbf{R}_J\right|} \\ &= T_N(\mathbf{R}) + V_{NN}(\mathbf{R}) + T_e(\mathbf{r}) + V_{EE}(\mathbf{r}) + V_{Ne}(\mathbf{R};\mathbf{r}) \end{aligned} \tag{2.20}$$

where m is mass, q is charge and $\mathbf{r}$, $\mathbf{R}$ is coordinate. The T terms correspond to kinetic energies and the V terms correspond to interaction energies. In the Born-Oppenheimer approximation[75] the electronic and nuclear motions are decoupled as a result of approximating the nuclei as fixed. The basis of this approximation is the significant weight difference between electrons and nuclei; nuclei are much heavier and therefore move much slower and so the electrons respond instantaneously to any nuclear motion. The nuclei can therefore be treated as fixed and the system described in terms of the electrons only and the Hamiltonian then becomes the electronic Hamiltonian

$$H_e = T_e + V_{Ne} + V_{ee}\left(+V_{NN}\right). \tag{2.21}$$

As the nuclei are fixed V_{NN} is a constant and thus it does not affect the wavefunction. The electronic Schrödinger equation can be solved analytically for a few special cases but in general analytical solutions are precluded by the electron interaction term V_{ee}. Therefore, the solution must be found via some suitable approximate method.

2.3.1 Hartree-Fock theory

Hartree-Fock theory is a wavefunction based method to approximately solve the Schrödinger equation. First, note that the sum over nuclei in the V_{Ne} term is identical for all atoms and so can be treated as an external field and therefore $T_e + V_{Ne}$ is a sum over single-particle Hamiltonians

$$\begin{aligned} H_e &= -\sum_{i=1}^{n}\frac{\nabla_i^2}{2} - \sum_{i=1}^{n}\sum_{J=1}^{N}\frac{q_J}{\left|\mathbf{r}_i - \mathbf{R}_J\right|} + \sum_{i<j}^{n}\frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} \\ &= \sum_{i}^{n} h(\mathbf{r}_i) + \sum_{i<j}^{n}\frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|}. \end{aligned} \tag{2.22}$$

The Hartree product wavefunction is then obtained by defining each electron spin orbital, $\varphi(\mathbf{x})$, as an eigenstate of h

$$\begin{aligned} \varphi(\mathbf{x}) &= \psi(\mathbf{r})\alpha(\sigma) \text{ or } \psi(\mathbf{r})\beta(\sigma) \\ h(\mathbf{r}_i)\varphi_i(\mathbf{x}) &= \epsilon_i\varphi_i(\mathbf{x}) \end{aligned} \tag{2.23}$$

and treating the electrons as non-interacting, essentially ignoring the V^{ee} term

$$H^{HP} = \sum_{i=1}^{n} h(\mathbf{r}_i). \tag{2.24}$$

Thus, the eigenfunction of H^{HP} is simply the product of all the single-electron spin orbitals

$$\begin{aligned} \Psi^{HP}(\{\mathbf{x}_n\}) &= \varphi_i(\mathbf{x}_1)\varphi_j(\mathbf{x}_2)\ldots\varphi_x(\mathbf{x}_n) \\ E^{HP} &= \epsilon_i + \epsilon_j + \ldots + \epsilon_x. \end{aligned} \tag{2.25}$$

However, ignoring electron interactions is a terrible approximation, so the Hartree product is substituted into the full electronic Hamiltonian to include electron interactions through a mean-field potential.

There is another issue with the Hartree product wavefunction, it is not anti-symmetric with respect to interchange of electrons, i.e. it violates the anti-symmetry principle. As the sign of the wavefunction changes when two electrons exchange, this is known as the exchange interaction. In order to rectify this the wavefunction can instead described by a Slater determinant

$$\Psi_{SD} = \frac{1}{\sqrt{n!}} \begin{bmatrix} \varphi_i(\mathbf{x}_1) & \varphi_j(\mathbf{x}_1) & \cdots & \varphi_x(\mathbf{x}_1) \\ \varphi_i(\mathbf{x}_2) & \varphi_j(\mathbf{x}_2) & \cdots & \varphi_x(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_i(\mathbf{x}_n) & \varphi_j(\mathbf{x}_n) & \cdots & \varphi_x(\mathbf{x}_n) \end{bmatrix} \tag{2.26}$$

Then the total Hartree-Fock energy is given by

$$\begin{aligned} \hat{H}_e &= \sum_{i=1}^{n} h(\mathbf{r}_i) + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ E_{HF} &= \langle \Psi_{SD} \,|\, \hat{H} \,|\, \Psi_{SD} \rangle \\ &= \sum_{i=1}^{n} \langle \varphi_i \,|\, h_i \,|\, \varphi_i \rangle + \frac{1}{2} \sum_{i,j}^{n} \Big(\big[\varphi_i \varphi_i \,|\, \varphi_j \varphi_j \big] - \big[\varphi_i \varphi_j \,|\, \varphi_j \varphi_i \big] \Big) \\ \big[\varphi_i \varphi_j \,|\, \varphi_k \varphi_l \big] &= \int \frac{\varphi_i(\mathbf{x}_1)\varphi_j(\mathbf{x}_1)\varphi_k(\mathbf{x}_2)\varphi_l(\mathbf{x}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2. \end{aligned} \tag{2.27}$$

Where the first term of E_{HF} is the expectation value for the one electron operator. The first two-electron integral is the coulomb integral as it corresponds to the coulomb repulsion between electron densities

$$\mathcal{J}_{ij} = \int\int \frac{|\varphi_i(\mathbf{x})|^2 \, |\varphi_j(\mathbf{x}')|^2}{|\mathbf{r} - \mathbf{r}'|} \tag{2.28}$$

and the second is the exchange integral

$$\mathcal{K}_{ij} = \int\int \frac{\varphi_i(\mathbf{x})\varphi_j(\mathbf{x})\varphi_j(\mathbf{x}')\varphi_i(\mathbf{x}')}{|\mathbf{r}-\mathbf{r}'|} \tag{2.29}$$

which arises as a result of the anti-symmetry principle and has no classical analogue.

Then following the variational principle

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0, \tag{2.30}$$

the wavefunction that best describes the system is found by minimising the energy. This is achieved by solving a set of one-electron Fock equations called Fock operators,

$$\begin{aligned} f(\mathbf{x})\varphi_i(\mathbf{x}) &= \epsilon_i \varphi_i(\mathbf{x}) \\ f(\mathbf{x}) &= h(\mathbf{x}) + \sum_{j=1}^{n} [\mathcal{J}_j(\mathbf{x}) - \mathcal{K}_j(\mathbf{x})], \end{aligned} \tag{2.31}$$

where $\mathcal{J}_j$ is the coulomb operator and $\mathcal{K}_j$ is the exchange operator and they are given by

$$\begin{aligned} \mathcal{J}_j(\mathbf{x})\varphi_i(\mathbf{x}) &= \left[\int \frac{|\varphi_j(\mathbf{x}')|^2}{|\mathbf{r}-\mathbf{r}'|} \, \mathrm{d}\mathbf{x}' \right] \varphi_i(\mathbf{x}) \\ \mathcal{K}_j(\mathbf{x})\varphi_i(\mathbf{x}) &= \left[\int \frac{\varphi_j(\mathbf{x}')\varphi_i(\mathbf{x}')}{|\mathbf{r}-\mathbf{r}'|} \, \mathrm{d}\mathbf{x}' \right] \varphi_j(\mathbf{x}) \end{aligned} \tag{2.32}$$

To solve the Fock equations a basis set is introduced. It is assumed the molecular orbital of the system can be written as a linear combination of atomic orbitals

$$\varphi_j = \sum_{k=1}^{m} c_{kj} \chi_k. \tag{2.33}$$

Substituting in to the Fock equation and formulating the expansion coefficients *c* in terms of matrices gives the Hartree-Fock-Roothaan equation

$$\mathbf{Fc} = \mathbf{Sc}\epsilon \tag{2.34}$$

where **F** is the Fock matrix, **c** is the expansion coefficient matrix and **S** is the overlap matrix

$$\begin{aligned} \mathbf{F} = F_{ij} &= \int \chi_i(\mathbf{x}) \, f_i \chi_j(\mathbf{x}) \\ \mathbf{S} = S_{ij} &= \int \chi_i(\mathbf{x}) \chi_j(\mathbf{x}). \end{aligned} \tag{2.35}$$

The Hartree-Fock-Roothaan equation must then be solved using the self-consistent method to find the values of the expansion coefficient matrix **c** which result in the lowest energy and therefore best wavefunction of the system.

However, there is still some remaining error in the Hartree-Fock result arising from electron correlation. As Hartree-Fock is a mean-field approach each electron experiences an averaged interaction with all other electrons and any instantaneous interactions between individual electrons due to electron-electron repulsion are neglected. These instantaneous interactions are often referred to as dynamic correlations. Additionally, approximating the wavefunction as a single Slater determinant can lead to physically unrealistic behaviour especially if several electronic states of the system have very similar energies. For example, a H_2 molecule dissociates into a pair of radicals but describing H_2 with a single Slater determinant predicts some ionic character (i.e. H^- $1s^2$, H^+ $1s^0$) at all interatomic separation distances. This is often referred to as static correlation.

2.3.2 Time-independent density functional theory

In density functional theory, the electron density, $\rho(\mathbf{r})$, is used to solve the electronic Schrödinger equation, rather than the many particle wavefunction. The density appears a preferable object to work with as it depends on only one variable, whilst the wavefunction is a function of $3N$ coordinates. In order to use the density to solve the Schrödinger equation, it must be true that the density uniquely determines the external potential, V_{Ne}, (and thus the properties) of the system. This is the first Hohenberg-Kohn theorem.[76] The proof of the theorem proceeds by considering the opposite to be true: the same ground state density, $\rho(\mathbf{r})$, belongs to two systems with two distinct external potentials, $V_1(\mathbf{r})$ and $V_2(\mathbf{r})$, and thus with distinct Hamiltonians, H_1 and H_2, and wavefunctions, ψ_1 and ψ_2. Then, according to the variation principle,

$$E_1 = \langle \Psi_1 | H_1 | \Psi_1 \rangle < \langle \Psi_2 | H_1 | \Psi_2 \rangle. \tag{2.36}$$

Also

$$\langle \Psi_2 | H_1 | \Psi_2 \rangle = \langle \Psi_2 | H_2 | \Psi_2 \rangle + \langle \Psi_2 | H_1 - H_2 | \Psi_2 \rangle = E_2 + \int \rho(\mathbf{r}) \left[V_1(\mathbf{r}) - V_2(\mathbf{r}) \right] \mathrm{d}\mathbf{r} \tag{2.37}$$

And therefore

$$E_1 < E_2 + \int \rho(\mathbf{r}) \left[V_1(\mathbf{r}) - V_2(\mathbf{r}) \right] \mathrm{d}\mathbf{r}. \tag{2.38}$$

Swapping the indexes in equations (2.36) and (2.37) produces the same relations and so

$$E_2 < E_1 + \int \rho(\mathbf{r}) \left[V_2(\mathbf{r}) - V_1(\mathbf{r}) \right] \mathrm{d}\mathbf{r}. \tag{2.39}$$

Adding these inequalities produces

$$E_1 + E_2 < E_1 + E_2 \tag{2.40}$$

which is nonsense. The assumption that the theorem is false is clearly invalid and therefore the theorem is true. Thus, we can write the energy as a functional of the density. The second Hohenberg-Kohn theorem[76] is a consequence of the variation principle and states that for ground state density $\rho(\mathbf{r})$ and any other density $\tilde{\rho}(\mathbf{r})$, corresponding to a given external potential, the energy of the ground state will be lower

$$E[\tilde{\rho}] \geq E[\rho] \tag{2.41}$$

So then, from the Hohenberg-Kohn theorems it follows that there is some functional of the density which gives all properties of the system. However, the exact form of this functional is not known. Considering the Hamiltonian in equation (2.21), the interaction V_{Ne} is the external potential and is thus

$$V_{Ne} = \int \rho(\mathbf{r}) V(\mathbf{r})\, \mathrm{d}\mathbf{r}, \tag{2.42}$$

and the coulombic repulsion in V_{ee} is

$$J[\rho] = \frac{1}{2} \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'. \tag{2.43}$$

But it is difficult to express the kinetic energy T_e in terms of the density. Additionally, there is nothing to account for the exchange interaction, nor electron correlation which are crucial components of the electron-electron interactions (V_{ee}). Thus, Kohn and Sham proposed[77] that the ground state of a fictitious system of non-interacting particles can be formed which is identical to that of the ground state density of the true interacting system. As in Hartree-Fock theory this non-interacting system is composed of a Slater determinant of molecular orbitals

$$\rho(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{x})|^2 \tag{2.44}$$

and so as before the kinetic energy is the sum of the individual kinetic energies

$$T[\rho] = \sum_i \left\langle \varphi_i \left| \frac{\nabla^2}{2} \right| \varphi_i \right\rangle. \tag{2.45}$$

Note that this is not the exact kinetic energy of the true system as the electrons are non-interacting. The Kohn-Sham energy is then

$$E_{KS}[\rho] = T[\rho] + J[\rho] + \int \rho(\mathbf{r}) V(\mathbf{r}) + E_{xc}[\rho] \tag{2.46}$$

where $E_{xc}[\rho]$ is the exchange-correlation functional, which by definition contains the difference between the kinetic energy of the true (interacting) system and that of the non-interacting system as well as the difference between the full electron-electron interaction energy and the Coulomb interaction energy. Then similarly to Hartree-Fock theory, the Kohn-Sham orbitals (and thus the energy) are found by self-consistently solving the Kohn-Sham equations

$$\left[-\frac{\nabla^2}{2}+\frac{1}{2}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}'+V(\mathbf{r})+V_{xc}(\mathbf{r})\right]\varphi_i(\mathbf{r})=\epsilon_i\varphi_i(\mathbf{r}) \tag{2.47}$$

where V_{xc} is functional derivative of the exchange correlation potential. However, the exact form of exchange-correlation functional $E_{xc}[\rho]$ is not known and thus must be approximated. Many (many) approximate exchange-correlation functionals have been developed, some use reference fitting data and parameters, whilst others rely upon physical reasoning and constraints. In general, most functionals can be categorised by three main groups: local density functionals, generalised gradient functionals and hybrid functionals.

The exchange-correlation energy density, $e_{xc}(\rho(\mathbf{r}))$, has a nonlocal dependence on the density; that is at a particular point $\mathbf{r}$, $e_{xc}(\rho(\mathbf{r}))$ is determined by the density at all other points, $\rho(\mathbf{r}')$. The local density approximation allows the removal of this nonlocal dependence through the approximation that for inhomogeneous systems with slowly varying density, the system is similar to the homogeneous electron gas

$$E_{xc}^{LDA}[\rho]=\int\rho(\mathbf{r})e_{xc}(\rho(\mathbf{r}))\,d\mathbf{r}. \tag{2.48}$$

The exchange-correlation energy density at a given point now depends only on the density at that point. Now the exchange-correlation energy density can be split into the exchange energy density and correlation energy density

$$e_{xc}(\rho(\mathbf{r}))=e_x(\rho(\mathbf{r}))+e_c(\rho(\mathbf{r})) \tag{2.49}$$

where the exchange energy density can be calculated exactly from Hartree-Fock theory

$$e_x(\rho(\mathbf{r}))=\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho(\mathbf{r})^{4/3} \tag{2.50}$$

and thus, the local density approximation exchange potential is

$$V_x^{LDA}(\mathbf{r})=\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}(\mathbf{r}). \tag{2.51}$$

A nice simple expression. Whilst the correlation energy density is not known exactly, quantum Monte Carlo calculations have provided very accurate numerical results which have been used to derive parameterisations[78–80] for the correlation energy in the local density approximation. The local density approximation performs well for solid-state and materials applications, however it suffers from self-interaction (interaction energy contributions resulting from electrons interacting with themselves) which consequently leads to generally underestimated values for the Kohn-Sham energy eigenvalues. As a result, the local density approximation is not suitably for many chemical applications.

The local density approximation assumes a homogeneous density based upon the assumption that the density varies slowly, for atoms and molecules this is not an optimal assumption. To overcome this, the gradients of the electron density are included in E_{xc} – this is the generalised gradient approximation

$$E_{xc}^{GGA} = \int f\left(\rho(\mathbf{r})\nabla\rho(\mathbf{r})\right)\mathrm{d}\mathbf{r}. \tag{2.52}$$

There is a wide variety of generalised gradient approximation functionals available as there is no unique formulation or parameterisation procedure. Popular examples of such functionals include PBE[81] and BLYP[82,83] (composed of the Becke exchange functional and Lee-Yang-Par correlation functional). The generalised gradient approximation has led to great improvements in practical application of DFT, owing to its accuracy and relative computational simplicity.

The search for even greater accuracy continues beyond the generalised gradient approximation. Recall the expression for $J[\rho]$ equation (2.43)

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)}{r_{ij}}\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j = \frac{1}{2}\sum_{i,j}\int\int \frac{\left|\varphi_i(\mathbf{r}_i)\right|^2\left|\varphi_j(\mathbf{r}_j)\right|^2}{r_{ij}}\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \tag{2.53}$$

thus, when $i = j$ there is a contribution to the energy from an electron interacting with itself. This is the source of the self-interaction error. In Hartree-Fock the self-interaction is exactly cancelled by the exchange energy. Additionally, it is known that Hartree-Fock exchange overestimates whilst the self-interaction error underestimates, and so hybrid functionals were conceived. In hybrid functionals a fraction of Hartree-Fock exchange is mixed with the DFT exchange

$$E_{xc}^{hybrid} = aE_x^{HF} + (1-a)E_x^{DFT} + E_c^{DFT}. \tag{2.54}$$

Perhaps the most famous of the DFT functionals, B3LYP is of this variety, consisting of the Lee-Yang-Par correlation functional mentioned previously and Becke's three-parameter hybrid[84] for exchange. This is just one example of a great many hybrid functionals that have been produced. There is some criticism that the level of empiricism introduced by the choice of functionals to be mixed and the weighting of the terms is too much and consequently, it is easy to produce a functional which perfectly reproduces some reference data set but is very poorly generalisable. Nonetheless, B3LYP and other hybrid functionals have proven to be successful in many practical applications of DFT. Another class of hybrid functionals are the long-range corrected (also known as range separated functionals) in which the Coulomb interaction is separated into short-range and long-range parts

$$E_{xc}^{LRC} = E_x^{DFT} + E_x^{HF} + E_c^{DFT}. \tag{2.55}$$

Where the short-range interaction is treated using a "standard" DFT functional (such as generalised gradient) and the long-range is treated using Hartree-Fock. The separation parameter can be determined using physical arguments or chosen empirically. The advantage of such hybrid functionals is that at long-range the local density and generalised gradient approximations are invalid but at short-range they are perfectly fine; thus, it can take advantage of the good short-range behaviour and account for the long-range asymptotic behaviour. And still ever more types of functional exist and continue to be developed, including meta-generalised gradient approximation, self-interaction correction, optimized effective potential and whatever other things can be though up by DFTers.

2.3.3 Time-dependent density functional theory

Since DFT for the stationary ground state is both accurate and computationally efficient, a similar formalism for time-dependent properties and electronic excited states would be ideal – a time-dependent density functional theory (TDDFT). However, there are two key differences between the time-independent ground state and the time-dependent case. Firstly, there is no variation principle in the time-dependent case and secondly, the time-dependent Schrödinger equation is an initial value problem. So then, it must be proven that two *n*-electron systems, starting from the same initial state, will have time-dependent densities that differ by more than a time-dependent constant if they are subject to two different time-dependent potentials. Runge and Gross[85] proved that for potentials with Taylor-expandable time series this was true. The proof first establishes that at times infinitesimally later than the initial time, different potentials result in different currents densities. It then goes on to

demonstrate that different current densities must have different densities. And so, for the time-dependent case the potential may be written as a functional of the density

$$V(\mathbf{r},t) = V[\rho, \Psi_0](\mathbf{r},t). \tag{2.56}$$

Since (from time-independent DFT) the ground state is a functional of the density, the formal dependence of the initial state can be removed by using the system's ground state as the initial state; meaning the time dependent potential, $V(\mathbf{r},t)$, can be written as functional of the density only.

A time-dependent formalism of Kohn-Sham DFT exists with the same premise as its time-independent counterpart; that the density may be calculated from a non-interacting system representative of the true system

$$\rho(\mathbf{r},t) = \sum_i^n \left|\varphi_i(\mathbf{r},t)\right|^2, \tag{2.57}$$

and the time-dependent Kohn-Sham equation is

$$\left[-\frac{\nabla^2}{2} + V(\mathbf{r},t)\right]\varphi_i(\mathbf{r},t) = i\frac{\partial}{\partial t}\varphi_i(\mathbf{r},t). \tag{2.58}$$

The time-dependent effective potential is then

$$V[\rho, \Psi_0, \Phi_0](\mathbf{r},t) = V(\mathbf{r},t) + V_H(\mathbf{r},t) + V_{xc}[\rho, \Psi_0, \Phi_0](\mathbf{r},t). \tag{2.59}$$

Here the time-dependent exchange-correlation potential formally has a dependence on the density and the initial state of the exact interacting system, as well as the initial state of the non-interacting system. Though as previously mentioned, typically the system of interest is initially in the ground-state and so, the exchange-correlation potential becomes a functional of the density only $V_{xc}[\rho](\mathbf{r},t)$. However, the density-dependence of the exchange-correlation potential is not straightforward moreover, as in the time-independent case it is nonlocal. This means that at some specific space-time point (**r**, *t*), the exchange-correlation potential depends on the densities at all other points in space at all previous times $\rho(\mathbf{r}', t')$ where $t' \le t$. Thus, in order to simplify the exchange-correlation potential, the adiabatic approximation is typically employed in TDDFT. The nonlocality is removed by replacing the ground-state density with the instantaneous time-dependent density in the exchange-correlation potential

$$V_{xc}^{adiabatic} = V_{xc}^{gs}[\rho_0](\mathbf{r})|_{\rho_0(\mathbf{r})=\rho(\mathbf{r},t)}. \tag{2.60}$$

Whilst solving the time-dependent Kohn-Sham equations is a viable route for determining time-dependent properties of a system, it is not always necessary to seek a full solution. In many practical applications the system does not strongly deviate from its initial state as it is only subjected to a small perturbation. Consequently, in such situations, it is possible to compute the change of a variable/observable in the perturbation rather than calculating the change of the wavefunction directly. This is the aim of linear response theory. Consider a system initially in the ground state where at t_0 a time-dependent potential, which causes some small time-dependent changes, is switched on, in turn causing the density to become time-dependent. This time-dependent density can be expanded in terms of the changes induced by this perturbation

$$\rho(\mathbf{r},t) = \rho_0(\mathbf{r}) + \rho_1(\mathbf{r},t) + \rho_2(\mathbf{r},t) + \ldots, \tag{2.61}$$

where the first term is the ground state density, the second term is the first-order change (due to the perturbation) in the density (the linear response), the third term is the second-order change in the density and so on. The linear response dominates when the perturbation is small. Formally the linear density response is

$$\rho_1(\mathbf{r},t) = \int_{-\infty}^{\infty}\int \chi(\mathbf{r},t,\mathbf{r}',t')V_1(\mathbf{r}',t')\,\mathrm{d}\mathbf{r}'\,\mathrm{d}t' \tag{2.62}$$

where χ is the density-density response function. However, typically the frequency-dependent response is of more interest than that of the real-time response and so instead

$$\rho_1(\mathbf{r},\omega) = \int \chi(\mathbf{r},\mathbf{r}',\omega)V_1(\mathbf{r}',\omega)\,\mathrm{d}\mathbf{r}' \tag{2.63}$$

and the FT of the density-density response can be written using the Lehmann representation[86]

$$\chi(\mathbf{r},\mathbf{r}',\omega) = \lim_{\eta\to 0^+}\sum_{n=1}^{\infty}\left\{\frac{\left\langle\Psi_{gs}\middle|\hat{\rho}(\mathbf{r})\middle|\Psi_n\right\rangle\left\langle\Psi_n\middle|\hat{\rho}(\mathbf{r}')\middle|\Psi_{gs}\right\rangle}{\omega-\Omega_n+i\eta} - \frac{\left\langle\Psi_{gs}\middle|\hat{\rho}(\mathbf{r}')\middle|\Psi_n\right\rangle\left\langle\Psi_n\middle|\hat{\rho}(\mathbf{r})\middle|\Psi_{gs}\right\rangle}{\omega+\Omega_n+i\eta}\right\} \tag{2.64}$$

where

$$\Omega_n = E_n - E_0 \tag{2.65}$$

is the nth excitation of the many-electron system. From this it is clear the exact excitation energies of the system are poles of the response function. Thus, knowing the function χ would allow the exact computation of the density response and following that, the observables. As before, a non-interacting Kohn-Sham system can be used to compute the linear density response of the real system by measuring the response of the non-interacting system to an

effective perturbation.[87] This effective perturbation is defined as the real external perturbation plus the Hartree potential and the exchange-correlation potential

$$V_1^{eff}(\mathbf{r},\omega) = V_1(\mathbf{r},\omega) + \int \left\{ \frac{1}{|\mathbf{r}-\mathbf{r}'|} + f_{xc}(\mathbf{r},\mathbf{r}',\omega) \right\} \rho_1(\mathbf{r}',\omega)\, \mathrm{d}\mathbf{r}', \tag{2.66}$$

where f_{xc} is the frequency-dependent exchange-correlation kernel. The response function of the non-interacting system to this effective perturbation is

$$\chi(\mathbf{r},\mathbf{r}',\omega) = \sum_{j,k=1}^{\infty} (f_k - f_j) \frac{\varphi_j(\mathbf{r})\varphi_k^*(\mathbf{r})\varphi_j^*(\mathbf{r}')\varphi_k(\mathbf{r}')}{\omega - \omega_{jk} + i\eta}. \tag{2.67}$$

f_j and f_k are occupation numbers of the ground state configuration of the non-interacting system and $\omega_{jk} = \epsilon_j - \epsilon_k$. This means the response function of the non-interacting systems has poles at the excitation energies.

The excitation energies of a system are the energy differences between the ground-state, E_0, and the higher states, E_n,

$$E_{exc}^n = E_n - E_0. \tag{2.68}$$

But it can also be described as the energy corresponding to a characteristic frequency which describes the rearrangement of the probability density when the system transitions from the ground state to a higher excited state. Thus, the excitation can be defined as an eigenmode. Linear response TDDFT can compute these eigenmodes and therefore the excitation energies of a system using the Casida equation[88,89]

$$\begin{pmatrix} \mathbf{A} & \mathbf{K} \\ \mathbf{K} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{2.69}$$

where **X** and **Y** are the excitation and de-excitations and

$$\begin{aligned} \mathbf{A} &= A_{ia\sigma,i'a'\sigma'}(\omega) = \delta_{ii'}\delta_{aa'}\delta_{\sigma\sigma'}\omega_{ai\sigma} + K_{ia\sigma,i'a'\sigma'}(\omega) \\ \mathbf{K} &= K_{ia\sigma,i'a'\sigma'}(\omega) = \int\int \varphi_{i\sigma}^*(\mathbf{r})\varphi_{a\sigma}(\mathbf{r})\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{r} \times f_{Hxc\sigma\sigma'}(\mathbf{r},\mathbf{r}',\omega)\varphi_{i'\sigma'}(\mathbf{r}')\varphi_{a'\sigma'}^*(\mathbf{r}') \end{aligned} \tag{2.70}$$

where f_{Hxc} denotes the Hartree-exchange-correlation kernel and is

$$f_{Hxc}(\mathbf{r},\mathbf{r}',\omega) = \frac{1}{|\mathbf{r}-\mathbf{r}'|} + f_{xc}(\mathbf{r},\mathbf{r}',\omega) \tag{2.71}$$

and *i* and *a* correspond to the occupied and unoccupied Kohn-Sham orbitals, respectively and σ denotes the spin-dependence. If the excitation frequencies are not close to zero (as is the

case for molecules) then Tamm-Dancoff approximation can be made. This decouples the excitations and de-excitations in eqn. (2.69) by setting **K** to zero.

In order to carry out linear-response TDDFT, there must be an approximation for the exchange-correlation kernel f_{xc}. This can be formally achieved by calculating $f_{xc}(\mathbf{r},t,\mathbf{r}',t')$, from approximating the time-dependent exchange-correlation potential, and performing a FT of the time-dependent kernel f_{xc} to obtain the frequency-dependent kernel $f_{xc}(\mathbf{r},\mathbf{r}',\omega)$. Practically however it is rarely necessary to perform this explicitly. Recalling the adiabatic approximation of the time-dependent exchange-correlation potential, eqn. (2.60), the adiabatic approximation for the exchange-correlation kernel is

$$f_{xc}^{adiabatic}(\mathbf{r},\mathbf{r}') = \frac{\delta V_{xc}^{gs}[\rho_0](\mathbf{r})}{\delta\rho_0(\mathbf{r}')} = \frac{\delta^2 E_{xc}[\rho_0]}{\delta\rho_0(\mathbf{r})\delta\rho_0(\mathbf{r}')} \tag{2.72}$$

which is frequency independent. Adiabatic exchange-correlation kernels can be derived from their standard time-independent DFT exchange-correlation potential counterparts (e.g. LDA, GGA, hybrids etc.), for example the adiabatic LDA exchange-correlation kernel

$$f_{xc}^{adiabatic\ LDA}(\mathbf{r},\mathbf{r}') = \frac{d^2 e_{xc}(\rho_0(\mathbf{r}))}{d\rho_0(\mathbf{r})^2}\delta(\mathbf{r}-\mathbf{r}'). \tag{2.73}$$

Linear response TDDFT using the Tamm-Dancoff approximation with a suitable adiabatic exchange-correlation kernel may then be used to compute excitation energies of many-electron systems (i.e. molecules).

2.4 Dynamics of open quantum systems in the Markovian limit

To begin with, the definition of the density matrix, ρ, must be introduced, for a pure state this is given as

$$\rho \equiv |\varphi\rangle\langle\varphi|. \tag{2.74}$$

Typically, φ is expanded as a linear combination of basis states $|n\rangle$

$$|\varphi\rangle = \sum_n c_n |n\rangle. \tag{2.75}$$

Thus, ρ becomes

$$\rho = \sum_{n,m} c_n c_m^* |n\rangle\langle m|, \tag{2.76}$$

therefore, the matrix elements are

$$\rho_{nm} \equiv \langle n | \rho | m \rangle = c_n c_m^*. \tag{2.77}$$

However, generally systems of interest are in a mixed state, i.e. a statistical ensemble, in such cases the density matrix is defined as

$$\rho = \sum_i P_i | \varphi_i \rangle \langle \varphi_i | \tag{2.78}$$

in which a system comprises of a mix of pure states $|\varphi_i\rangle$ each occurring with probability P_i, in which $P_i \geq 0$ and the values of P_i sum to one. The expectation value of some observable A of the system can then be computed according to

$$\langle A \rangle = \sum_i P_i \langle \varphi_i | A | \varphi_i \rangle = \sum_i P_i \text{ tr}(| \varphi_i \rangle \langle \varphi_i | A) = \sum_i \text{tr } (P_i | \varphi_i \rangle \langle \varphi_i | A) = \text{tr} \left(\sum_i P_i | \varphi_i \rangle \langle \varphi_i | A \right)$$
$$\langle A \rangle = \text{tr } (\rho A) \tag{2.79}$$

and ρ evolves in time according to the time dependent Schrödinger equation

$$\frac{d}{dt} \rho = -i[H, \rho] \tag{2.80}$$

where $\hbar$ is set to 1 for convenience. As per equation (2.77) the matrix elements of the diagonal of the density matrix will be given by $c_n c_n^* = | c_i |^2$ and thus are referred to as populations (as $|c_i|^2$ is the probability of being in state i) and the off-diagonal elements are given by $c_n c_m^*$ and are known as coherences. Therefore $\rho(t)$ gives the population and coherences of a system at the corresponding time t.

When discussing open quantum systems, one may consider some system, S, in which all interesting processes are taking place and some bath or environment, B, that the system is interacting with. The combination of these subsystems $S + B$ forms the total system which is considered closed and thus follows Hamiltonian dynamics. For clarity in the following the reduced system of interest, S, will be referred to as "the system", the bath or environment, B, will be referred to as "the environment" and the combination of these two parts, $S + B$, will be referred to as "the total system". If the system and environment are separable the total density matrix may be written as a tensor product of the system and environment density matrices

$$\rho = \rho_S \otimes \rho_B \tag{2.81}$$

As only the observables occurring in the system are of interest, these operators are of the form

$$A = A_S \otimes I_B \tag{2.82}$$

the expectation value of which is determined by

$$\langle A \rangle = \mathrm{tr}_S \{ A \rho_S \}. \tag{2.83}$$

Thus, it is useful to introduce the definition of the system density matrix ρ_S

$$\rho_S = \mathrm{tr}_B \{ \rho \} \tag{2.84}$$

where tr_B denotes a partial trace over the Hilbert space associated with the environment and ρ is the density matrix of the total system. ρ_S is known as the reduced density matrix.

This reduced density matrix approach is valid if the system and environment are separable. This is true when the environmental correlation times of the total system are short i.e. the characteristic timescales of the environment correlations are smaller than the characteristic timescales of the evolution of the system. Physically this means that the bath is so much faster than the system that its effects are averaged out and information of where the bath is or was, at $t < 0$, is not required. In other words, it is assumed that the future of the system does not depend on the environment and its past – i.e. the Markovian approximation. Assuming this, as in equation (2.81), the total system density matrix $\rho(0)$ may be written as a product tensor of the system and environment. And so, a linear, completely positive map from the density matrix of the system ρ_S into itself may be formulated. This is a dynamical map. The transformation from the initial system (at time $t = 0$) to the system at some other time ($t > 0$) can be written as

$$\rho_S(0) \mapsto \rho_S(t) = V(t)\rho_S(0) \equiv \mathrm{tr}_B \{ U(t,0)[\rho_S(0) \otimes \rho_B] U^{\dagger}(t,0) \} \tag{2.85}$$

where $U(t,0)$ is the unitary-time evolution operator. When the reference state ρ_B and final time t are fixed, this relation defines a map from the space of the system density matrices into itself

$$V(t) : \mathcal{S}(\mathcal{H}_S) \mapsto \mathcal{S}(\mathcal{H}_S). \tag{2.86}$$

$V(t)$ is the dynamical map for some fixed time ($t \geq 0$) and represents a linear, completely positive and trace-preserving quantum operation. If t can vary then a one parameter family of dynamical maps $\{V(t) | t \geq 0\}$ can be obtained where $V(0)$ is the identity map. This family of maps, known as a semigroup, describes the complete time evolution of the system. Thus from the quantity defined in equation (2.84) the total system dynamics may be evolved and as per equation (2.83) any observable may be computed from this system density matrix. As previously stated, this holds for Markovian behaviour which can be formalised with the help of the semigroup property

$$V(t_1)V(t_2) = V(t_1 + V_2),\ t_1, t_2 \geq 0. \tag{2.87}$$

For a quantum dynamical semigroup, there exists a generator of the semigroup, a linear map $\mathcal{L}$, which allows the semigroup to be represented in exponential form

$$V(t) = \exp(\mathcal{L}t). \tag{2.88}$$

Which gives a first-order differential equation for the reduced density matrix of the system

$$\frac{d}{dt}\rho_S(t) = \mathcal{L}\rho_S(t). \tag{2.89}$$

The generator $\mathcal{L}$ of the semigroup is a super-operator and may be regarded as a generalisation of the Liouville super-operator of closed system dynamics. The diagonal form of the generator is:

$$\mathcal{L}\rho_S = -i[H, \rho_S] + \sum_{k=1}^{N^2-1} \gamma_k \left(L_k \rho_S L_k^\dagger - \frac{1}{2}\left\{ L_k{}^\dagger L_k, \rho_S \right\} \right). \tag{2.90}$$

(For derivation see ref. 90) This is the most general form for the generator of a quantum dynamical semigroup. The first term of the generator represents the unitary part of the dynamics generated by the Hamiltonian H, i.e. the system dynamics in the absence of any environment. The operators L_k are usually referred to as Lindblad operators and the corresponding density matrix in equation (2.89) is called the Lindblad equation. Note that the non-negative quantities γ_k have the dimension of an inverse time provided the L_k are taken to be dimensionless.

In exciton dynamics, the unitary part of the dynamics is generated by the excitonic Hamiltonian H^{exc} and the Lindblad operators will correspond to processes in which the exciton is lost. For travel through a chromophore network these are: capture of the exciton by the sink, dephasing in which there is local loss of the exciton due to the coupling of the chromophore site to its vibrational environment and dissipation where the exciton is lost due to decaying. So, for a system comprising N chromophores, the excitonic Hamiltonian H^{exc} is given by

$$H^{exc} = \sum_{i=1}^{N} \varepsilon_i |i\rangle\langle i| + \sum_{m \neq n} V_{mn}(|m\rangle\langle n| + |n\rangle\langle m|), \tag{2.91}$$

where the diagonal ε_i represent the energy landscape and the off-diagonal V_{mn} represent coupling between the chromophore sites (e.g. Coulombic coupling). There is then a collapse operator for each of the processes described earlier, to be used in

$$\mathcal{L}(\rho) \equiv L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}. \tag{2.92}$$

And so, overall

$$\dot{\rho} = -i[H^{exc}, \rho] + \mathcal{L}(\rho). \tag{2.93}$$

2.5 Spectral density

Redfield theory is a reduced density matrix approach and the form of the Redfield equation[19] is similar to that of the Lindblad equation discussed in the previous section

$$\dot{\rho}_{\mu\nu} = -i[H, \rho_{\mu\nu}] + \sum_{\mu'\nu'} R_{\mu\nu,\mu'\nu'} \rho_{\mu'\nu'} \tag{2.94}$$

where μ and ν are eigenstates of the system. The derivation of the Redfield equation also uses the Markovian approximation however, there are important differences between the Redfield and Lindblad derivations. In Redfield the exact time evolution is approximated by assuming a small system bath coupling. The important result of the differences in their derivations is that whilst Redfield is more clearly linked with the system Hamiltonian, the time evolution is no longer guaranteed to be unitary meaning the population can go negative. The first term of equation (2.94) describes the unitary evolution of the system, as previously discussed, and the second term is the Redfield tensor, R, which describes the relaxation of the system due to interaction with the bath. This can be expressed in terms of Γ, the dampening matrix

$$R_{\mu\nu,\mu'\nu'} \equiv \Gamma_{\nu'\nu,\mu\mu'} + \Gamma^*_{\mu'\mu,\nu\nu'} - \delta_{\nu\nu'} \sum_{\kappa} \Gamma_{\mu\kappa,\kappa\mu'} - \delta_{\mu\mu'} \sum_{\kappa} \Gamma^*_{\nu\kappa,\kappa\nu'} . \tag{2.95}$$

The dampening matrix is mainly determined by the bath correlation function,[91,92] $C(t)$, and for the system-bath interaction the dampening matrix can be written as

$$\Gamma_{\mu\nu,\mu'\nu/} = \frac{1}{\hbar} \sum_{i,j} \langle e_\mu \mid i \rangle \langle i \mid e_\nu \rangle \langle e_{\mu'} \mid j \rangle \langle j \mid e_{\nu'} \rangle C_{ij}[\omega_{\nu'\mu'}]. \tag{2.96}$$

where the bath correlation function describes the correlation of the electronic Hamiltonian matrix elements due to the fluctuation of the bath,

$$C_{ij}(t) = \langle \delta H_{ij}(0) \delta H_{ij}(t) \rangle. \tag{2.97}$$

The Fourier transform of $C_{ij}(t)$ gives $C_{ij}[\omega]$ and is expressed in terms of the bath spectral density $J_{ij}(\omega)$. So, the bath spectral density contains information about the relaxation of the system due to interaction with the bath. Phenomenological models often adopt an analytical form of the spectral density dependent on one or two parameters, e.g. Drude-Lorentz,[93] however it is also possible to compute (2.97) explicitly.

One method to achieve this is by using normal modes analysis (NMA) of the system. NMA is based on the hypothesis that the vibrational modes of the protein with the lowest frequencies describe the greatest protein movements and are functionally relevant.[94] In NMA the motions with a small amplitude in a potential well which cannot cross energy barriers are studied. When a system is in a well, i.e. in equilibrium, the potential energy yields a quadratic approximation

$$V = \frac{1}{2}\left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 \eta_i \eta_j = \frac{1}{2} \mathrm{V}_{ij} \eta_i \eta_j \tag{2.98}$$

where q_i are the generalised coordinates and η is the deviation from the equilibrium, $q_i = q_{0i} + \eta_i$. The kinetic energy T is similarly approximated as a quadratic function, then from the Lagrangian, $L = T - V,$ comes the n linear differential equations of motion

$$T_i \ddot{\eta}_i + V_{ij} \eta_j = 0 \tag{2.99}$$

Substituting an oscillatory solution $\eta_i = a_{ik} \cos(\omega_k t + \delta_k)$ into the previous equation give the eigenvalue problem of the amplitudes a_{ik} matrix **A** and the matrix of the potential energy second derivatives

$$\mathbf{A}^T \mathbf{V} \mathbf{A} = \lambda. \tag{2.100}$$

The eigenvectors $\mathrm{A_k}$ and their eigenvalues λ_k give the vibrational normal modes which fully describe the pattern of motions of the protein. The harmonic oscillators computed with NMA can then be used to describe the protein environment in a spectral density calculation for example as described in ref. 40.

Another method to compute the spectral density is to use the snapshots of an MD trajectory as input for QC calculations of the excitation energy of the pigments of the system. Performing a Fourier transform of the autocorrelation function of the fluctuation of the excitation energies along the MD trajectory. For example, assuming the effect of the environment predominantly causes fluctuations of the on-site energy, the spectral density of pigment i, $J_i(\omega)$, in a PPC is given by

$$J_i(\omega) = \frac{\beta \omega}{\pi} \int_0^\infty C_i(t) cos(\omega t)\, \mathrm{d}t \tag{2.101}$$

where β is $1 / k_B T$ and the autocorrelation function C_i is expressed in terms of the fluctuation of the excitation energy of pigment i, $\delta\varepsilon_i$

$$C_i(t) = \langle \delta\varepsilon_i(t) \delta\varepsilon_i(0) \rangle. \tag{2.102}$$

However, as briefly discussed in the introduction the combination of classical molecular dynamics and quantum chemistry calculations leads to errors in the resulting spectral density. This is because the forcefields used in MD produce different equilibrium structures to those of the quantum chemistry calculations, thus they will have different normal modes, with different frequencies which leads to an inaccurate description of the electronic-nuclear interaction.

3 Developing MDFFs for spectral density computations

In order to examine LHCs with atomistic detail, classical mechanics methods must be utilised such as MM or MD due to the size of the complexes. In the study of LHCs, these low-level methods are often combined with higher-level methods. One example of this methodology is the computation of pigment spectral density by utilising MD trajectories as input for QCs. However, the combination of such methods results in errors. The aim of this chapter is to introduce a new method for developing molecular dynamics forcefields (MDFFs) which requires less manual input which will reduce these errors and also make it more feasible to compute the spectral densities of many pigments of LHCs.

3.1 Introduction

3.1.1 Spectral density and the geometry mismatch problem

Spectral density is an important tool for examination of PPCs, as outlined in Chapter 2 this quantity contains information on the system-bath interaction which has been proposed as playing some important role in EET.[4,21,26,30,34,54,95] As such, there are many examples in literature endeavouring to compute this quantity.[29,40,52,93,96–100] However, it became apparent that when using methods combining MD and QC calculations, the result was sensitive to the choice of FF used in the MD portion of the computation.[41,101] It was determined this was due to the so-called geometry mismatch problem which describes the inconsistency between the equilibrium geometries produced by the two-levels of theory. This difference leads to an inconsistency between the computed excitation energies and the corresponding ground state surface from the MD and thus produces inaccuracies in the system-bath interaction. Work by Lee and Coker[41] outlines how the differences in the QM excited state, QM ground state and MD ground state potential energy surfaces affect the spectral density and introduce a method

to avoid this problem in which the quantum intramolecular vibrational modes are computed directly and the lower frequency intermolecular modes are incorporated through analytical electrostatic interactions. An alternative approach to compute an accurate spectral density of a PPC, would be to develop a method which produces FFs consistent with the QC part.

3.1.2 Developing new FFs

The need of FFs for new molecules is not a novel problem and much work has focused on the generation of FF parameters. A standard method to build a new FF is by analogy i.e. identifying analogous parameters from existing FFs based on the structure of the molecule. For CHARMM[71] and AMBER[73] automated by analogy procedures have already been developed, these are CGenFF[102] and GAFF[103], respectively. The methods work in a similar fashion, utilising look-up tables and empirical rules to generate the new parameters. The results of both methods typically require further manual parameterisation before they can be implemented.

Further work has endeavoured to automate parameterisation in order to remove the need of manual adjustment and make the process easier and faster. For example, GAAMP[104] utilises QM data as reference to automatically produce parameters compatible with the CHARMM and AMBER FFs. Other work focuses on automatically parameterising force constants for the AMBER forcefield especially for transition metal complexes through the use of *ab initio* frequency calculations.[105] The JOYCE[106] procedure parameterises the intramolecular part of FF, using an iterative approach to fit to QM energies, gradients and the Hessian matrix, similarly the PICKY[107] procedure employs a least square fitting procedure to parameterise the intermolecular part, using QM energies. The QMDFF[108] constructs FFs from solely QM input; utilising an equilibrium structure, the Hessian matrix, atomic partial charges and the bond orders to generate specific, non-transferable FFs of molecules. The fitting of parameters to a flexible combination of theoretical and experimental data is possible with the ForceBalance[109] method, which automatically generates parameters through an iterative approach which minimises the difference between the FF parameters and target data. This method was in fact employed to improve the AMBER protein FF.[110] Other methods seek to simplify the parameterisation process, such as the FF Toolkit[111] which creates CHARMM compatible parameters from target QM data and the QuickFF[112] which employs a 3 step method to build parameters from *ab initio* data.

However, these methods are general tools for developing new FFs and whilst they are useful, they are not specifically developed for use in spectral density calculations. But there

are examples of FFs which have been generated with the specific use for spectral density calculations in mind. Similar to the general methods, these methods predominantly utilise QM calculations to produce target data. For example FFs compatible with the popular AMBER FF for bacteriochlorophyll-*a* (bhcl) (as well as other cofactors: methyl bacteriopheophytin-*a* and a ubiquinone derivative) have been developed.[113] Further work transformed these parameters to create a chlorophyll-*a* (CLA) FF,[114] compatible with another commonly used FF, CHARMM. Both these BCL and CLA FFs have been utilised in several spectral density studies.[40,97,115] However, other work has criticised the development of these parameters for lacking detailed validation against experimental structure data and so more validated parameters for the AMBER FF of the chromophores (and other cofactors) of the photosystem II (PSII) complex were developed.[116] The resulting FFs of this work have also been used in the calculation of pigment spectral density.[100] However these FFs fall foul of the geometry mismatch problem.

In order to specifically combat the geometry mismatch problem, several FFs of carotenoids[117] and an apocarotenoid FF[118] have been created with the aim of better reproducing the QC structural properties. Although these FFs were found to reduce the geometry mismatch error, the parameterisation requires time consuming QC calculations and manual input. If there was a small number of pigments (e.g. < 5) found in PPCs this would be less of limitation, after investing a fair amount of time one would obtain all the FFs needed and be able to accurately compute spectral densities. However as discussed in the first chapter, this is not the case for PPCs; there are many pigments. Furthermore, the protein scaffold is known to distort and influence the pigment structure and there are many protein environments in PPCs which may all distort identical pigments slightly differently, further increasing the number of FFs required. Thus, in order to study many PPCs, which is necessary to identify any underlying patterns or similarities between complexes, the amount of time and effort required to produce FFs in this way becomes unfeasible. A fast, less manually intensive method is necessary.

3.1.3 Force matching

Force matching is an ideal method to efficiently produce the desired consistency with significantly less manual input. The method was devised by Ercolessi and Adams for the development of interatomic potentials[119] and has since been applied to a range of systems, including developing interatomic potentials for Zr-Cu[120] and Mg,[121] flexible water model potentials,[122] creating FFs for semiconducting polymers,[123] generating free energy surfaces

for chemical reactions[124] and in the reparameterization of the single point charge water model.[125] Force matching has also been combined with combined QM/MM methods to create FFs for moieties/molecular fragments.[126] Force matching is particularly advantageous in this case as FFs can be obtained which take the environment into account, additionally it effectively constrains towards chemically sensible parameters as unphysical parameters will lead to a high force. In this case, the difference in the forces resulting from the FF and the forces resulting from electronic structure (ES), of each atom in a collection of equilibrium structures, generated in the protein environment, are minimised by changing the intramolecular parameter values composing the FF. The aim of this work is to enable the generation of many FFs of biological chromophores embedded in a protein environment for use in spectral density calculations of PPCs.

3.2 Systems

3.2.1 Chromophores

FMO is one of the most popular systems studied so its chromophore bacteriochlorophyll-*a* (BCL) is chosen. In addition, FFs for the common chromophores chlorophyll-*a* (CLA) and chlorophyll-*b* (CLB) are developed, taken from the water soluble chlorophyll-binding protein (WSCP). The structures of all the chromophores are shown in Fig. 3.1. The FF is optimised for the bacteriochlorin/chlorin ring and immediate groups only (highlighted in red in Fig. 3.1), i.e. the phytyl chain and ester group are not included. This is because they are not involved in the conjugated system and so are often not included in the quantum calculation, thus it is not crucial that their geometry be consistent. This also reduces the computational demand.

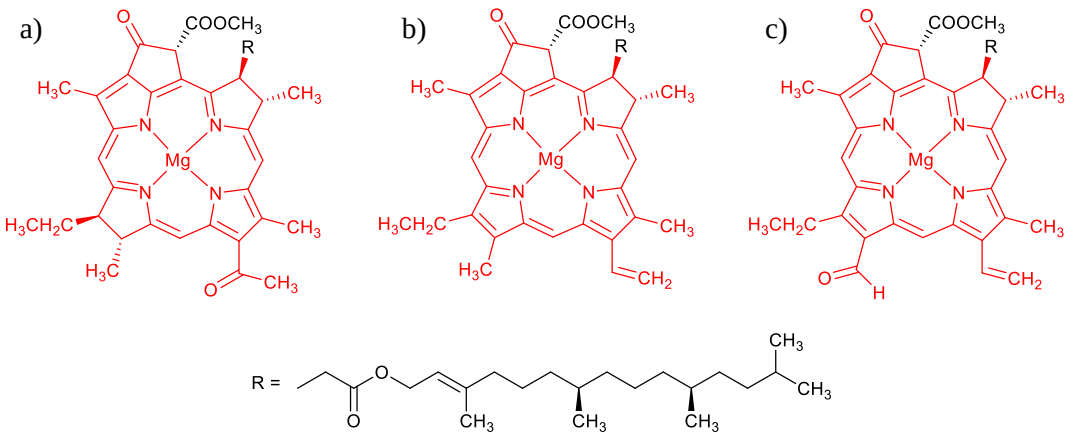


Figure 3.1: Structures of a) BCL b) CLA and c) CLB with portion of molecules for which FF is optimised highlighted in red.

3.2.2 Parameter Sets of Initial FFs

The general functional form of any FF summarised in Chapter 2. Here the CHARMM FF is used, the functional form of which is given in equation (2.18). Only the parameters of the intramolecular terms (bonds, angles, dihedrals and improper dihedrals) are optimised, as these have the greatest effect on the internal structure which is the part desired to be consistent. The multiplicity and phase shift of the dihedral term are pre-set as is equilibrium improper dihedral angle, and so these parameters are not changed in the optimisation. Meaning the overall parameter set to be optimised consists of: the bond; angle; dihedral angle; and improper dihedral angle force constants (k_b, k_θ, k_φ and k_ω, respectively) and the equilibrium bong length and equilibrium bond angle (b_0 and θ_0, respectively), for each atom type.

The initial FF of BCL is taken from the by analogy FF created by CGenFF, whilst the CLA and CLB FFs are taken from literature ref. 147 and 148. The atom types assigned by these FFs are overly degenerate for the desired purpose, i.e. atoms are given the same atom type and so treated as 'structurally similar' when their internal electronic structure is different. This could be rectified by assigning each atom an individual type. However, a greater number of atom types means a greater number of parameters to be optimised which increases the optimisation time. Additionally, due to the near symmetric nature of the molecules many of the parameters would be near degenerate, making the optimisation procedure inefficient. The best solution is then to assign new atom types. To achieve this, first 'structural similarity' must be quantitatively defined. The base framework of the chromophores consists of carbon, as with all organic molecules, and so comparing the lengths of standard C – C single bonds (~ 1.54 Å), double bonds (~ 1.35 Å) and conjugated bonds (~ 1.40 Å) shows significant differences in bonding structure are those of 0.05 Å and greater. Using this as a guide, an estimate of 'structurally similar' bond lengths can be defined as those within 0.01 Å of one another, whilst bond lengths with a difference of 0.03 Å or more are 'structurally different' and those lying between 0.01 Å and 0.03 Å must be considered further. Additionally, the chemistry of the atom can be considered, i.e. what bond type is it involved in e.g. C – O, C – N, C – C, and so on. Finally, any matching bond types within the 0.01 Å – 0.03 Å region can be categorised based on their bond angles. An example of this assignment, for BCL, is shown in Fig. 3.2. After this assignment was done, the total number of parameters to be optimised for each chromophore were: BCL – 399 parameters; CLA – 446 parameters; CLB – 455 parameters. The atom type definition of each chromophore is included in the appendix in Figs. A3.1-3.3.

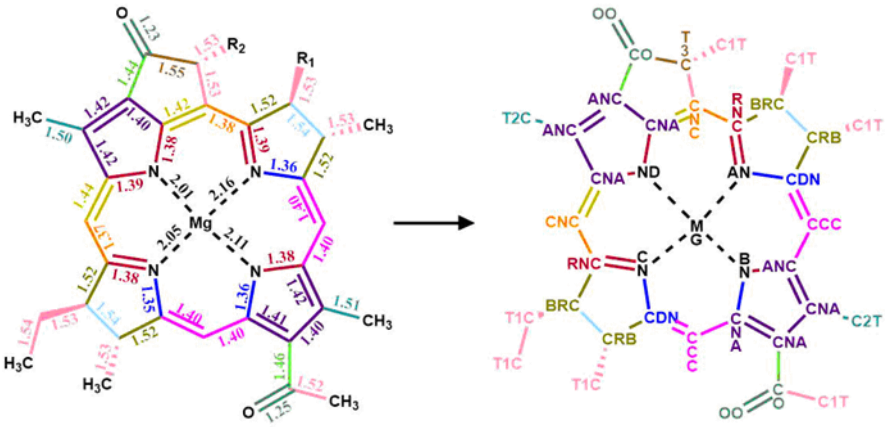


Figure 3.2: Colour coded illustration of assignment of atom type based on structure depicting the bond lengths (taken from optimisation BCL geometry) and the resulting atom type.

Finally, it is important that the electrostatics are consistent with the rest of the FF. The partial charges for the CHARMM FF are determined using minimum HF water interaction energies and distances as target data as described in the supporting information of ref. 72. The charges for the chromophores were taken from literature.[36] It should be noted that whilst in this work the parameters of a CHARMM FF are optimised, the method can be applied to any molecular mechanics FF.

3.3 Force Matching Procedure

The FF is optimised by minimising an objective function. For the FF of a molecule comprised of a set of parameters, [*p*], the objective function is the root mean squared (RMS) displacement of the ES forces and MD forces, given by

$$O([p]) = \sqrt{\frac{1}{3MN}\sum_{i=1}^{N}\sum_{j=1}^{M}|\mathbf{f}_{ij}^{DFT} - \mathbf{f}_{ij}^{MD}|^2}. \qquad (3.1)$$

Where N is the number of atoms in the molecule, M is the number of reference structures and $\mathbf{f}_{ij}$ denotes the forces on atom i of structure j from DFT or MD.

Gradient descent minimisation is based on the observation that from a point, r, of a function, $f(r)$, the function will decrease fastest moving from r in the direction of the negative gradient of the function, $\nabla f(r)$. In other words, to minimise a function, one follows

$$r = r - \alpha \nabla f(r), \qquad (3.2)$$

where α is a scaling factor to determine the step size of the descent. Previous work utilising force matching with a Monte Carlo optimisation showed that the objective function is convex without secondary minima[123] thus a gradient descent method is appropriate.

To calculate the descent direction of the FF parameter set, the gradient, G_i, of the objective function for each parameter, p_i, must be obtained numerically

$$G_i \approx \frac{O(p_1,...,p_i+\Delta p_i,...,p_n)-O(p_1,...,p_i,...,p_n)}{\Delta p_i} \tag{3.3}$$

Then each parameter is changed using the negative of its gradient

$$p_i^{new} = p_i - \alpha G \tag{3.4}$$

to create a new parameter set. The process continues until the objective function is minimised and the FF is considered optimised. In principle, one could use this new FF to generate new structures and iterate the process, however previous work[49] has shown the majority of improvement (93 – 96 %) is achieved in the first iteration and so no iteration is implemented here.

3.4 Computational details

3.4.1 Initial geometries

Structures of the chromophores were obtained from MD simulations of the chromophore in its protein environment, solvated by water. Utilising structures from the chromophores embedded in their environment constrains the geometries to those accessible in the protein environment. The proteins used were FMO, the water-soluble chlorophyll-a binding (WSCP-a) protein and water-soluble chlorophyll-b binding (WSCP-b) protein for BCL, CLA and CLB, respectively. Exploring geometries in the protein environment as opposed to in vacuum is important for FFs to be used in the study of PPCs, as the environment distorts the chromophore geometry. All MD was carried out using GROMACS 5.0.5 software.[127]

The initial structures for FMO, WSCP-a and WSCP-b were taken from the Protein Data Bank, PDB: 3BSD and 5HPZ (used for both WSCPs), respectively. The missing residues of WSCP were built with the CHARMM-GUI website.[128] In FMO there are 4 BCL molecules for which the Mg atom coordinates to a histidine residue (HIS). These were assigned a protonation state to allow this coordination, all other HIS residues were assigned protonation in position *ε*. The FF description of Mg coordination is well known to be difficult to determine. As axial coordination to the Mg atom of chromophores affects their spectroscopic

properties[129,130]. In this case it is assumed their treatment in the initial FF is valid. The rest of the residues in all the proteins were assigned their standard protonation state. For all proteins, the CHARMM36[74] FF and TIP3P[131] water model were used. The systems were solvated with water in a cubic box with edge length of 120 Å and 110 Å for FMO and the WSCPs, respectively. After a steepest-descent energy minimisation, a 200 ps NVT equilibration (300 K; Berendsen thermostat[132]; τ_T = 0.1 ps) was run, followed by 200 ps NPT equilibration (1 bar; Berendsen barostat[132]; τ_P = 0.1 ps), using a 2 fs integration step. A subsequent 10 ns of dynamics was run. The structures for the chromophores were then taken every 500 fs from a final 80 ps run (NVT ensemble). Each protein contains multiple chromophores; so the structures were taken cycling through each chromophore, i.e. the first BCL structure was taken as BCL 1, the second as BCL 2, etc. for each structure.

3.4.2 Reference force calculation

The reference forces were calculated with density function theory (DFT) using the B3LYP functional and 6-311G* basis set with Gaussian 03. The forces (both DFT and MD) were computed in vacuum in order to preserve separation between the intramolecular and intermolecular part of the FF. As noted before for the type of FF used, the reference force calculation can be computed with any functional or basis set desired.

3.4.3 Minimisation Procedure

The procedure is as follows: the initial objective function, $O([p^{initial}])$, is calculated as given in equation (3.1), then for each parameter, p_i, of the set, the corresponding gradient, G_i, is computed as in equation (3.3), where the value of Δp_i is taken as 0.001% of the parameter. A new parameter set, $[p^{new}]$, is then generated, by updating each parameter as described in equation (3.4), and its objective function computed, $O([p^{new}])$. The minimisation is considered complete once the decrease in the objective function, $\Delta O([p])$, is consistently less than 0.01 kcal mol^{-1} Å^{-1}.

As each parameter set consists of 100s of parameters, calculating each gradient numerically is time consuming. To circumvent this bottleneck, in the implementation of the algorithm the gradients are only recalculated if the objective function increases, i.e. if $O([p^{new}]) > O([p])$, otherwise the minimisation continues using the same set of gradients. In addition, the five parameter types: b_0, θ_0, k_φ, k_b k_θ; each affect the forces to a different magnitude, leading to gradients of differing magnitudes. The order of their affect, in the case of chromophores, is:

$$b_0 \gg \theta_0 > k_\varphi \approx k_b \approx k_\theta \,. \tag{3.5}$$

The b_0 parameters have a significantly greater effect on the forces and thus a much greater gradient. For a multi-variable function where one variable has a significantly greater gradient, the gradient descent method will rapidly minimise the function with respect to that variable and then the minimisation will progress exceedingly slowly. Thus, to avoid this slow down, the b_0 parameters are optimised separately first. Once decrease in the objective function is less than 0.0024 kcal mol^{-1} Å^{-1} this is considered complete. For the same reason, once the θ_0 parameters are optimised α is increased; once all their gradients are less than 0.5 degrees^{-1} α is increased by a factor of 250. Furthermore, to reduce the objective function before the minimisation procedure, the b_0 and θ_0 parameters were set to those of the DFT optimised structure performed at the same level as the reference structure calculations. An alternative solution to this issue of differing gradients problem would to be use an algorithm that automatically adjust α for example the momentum optimisation[133] or Nesterov[134].

3.4.4 Spectral Densities

The correlation functions were calculated according to equations (2.102) and (2.101) (implemented as detailed in ref. 135). A subsequent Fourier transforms in the $0 \leq t \leq 16$ ps range was used to obtain the spectral densities. For each protein system, MD trajectories were obtained utilising the original FF and the optimised FF of the chromophore. The initial protein setup, energy minimisation and equilibration steps were performed as outlined in 3.4.1. A further 25 ns of dynamics (NVT) was run. The excitation energies for the spectral densities were computed from snapshots taken every 2 fs from the last 16 ps of this MD trajectory, using a QM/MM scheme within TDDFT linear response theory,[136] performed using the QChem 4.2 software.[137] The MM part comprised all residues within a 35 Å radius of the pigment; included additively as point charges that affect the QM system. B3LYP[84]/3-21G* was used for the QM part, which was reduced through the insertion of a link atom between carbon atoms 1 and 2 of the phytyl chain. The 3-21G* basis set was used to reduce computational time as it has been found (Fig. S2 in ref. 138) that there is good correlation between 6-31G* and 3-21G* for these systems. Four roots were computed but only the lowest was considered. This is a standard approach similar to those used in refs. 39 and 101. As each of the proteins contains multiple pigments one pigment was arbitrarily selected from each, for the computation of the spectral density; these were: BCL 4, CLA 2 and CLB 2.

3.5 Results

3.5.1 The final forces

The FFs were considered optimised to a local minimum once the change in objective function was less than 0.01 kcal mol^{-1} Å^{-1} for 50 steps or more. The initial and resultant values of the objective functions are shown in Table 3.1 along with the overall change and the total number of steps of the algorithm. Interestingly, the parameters of the CLs, taken from literature have greater initial objective functions than the BCL generated by analogy using CGenFF. This is surprising as even for traditional uses of MD, the automatically generated by-analogy FFs require further parameterisation before use and thus the literature values would be expected to be superior. The average of the final objective functions is 21.84 kcal mol^{-1} Å^{-1} indicating a minimum of a roughly similar value for all. In comparison to similar methods the final values are slightly worse (larger by a factor of 3-4) however they are within the same range as the mismatch observed between forces of different electronic structure calculations.[123]

Visualising the average error in force on each atom before and after optimisation, illustrates the improvement of the FF. The average error in force on each atom, Δf_i, is defined as:

$$\Delta f_i = \sqrt{\frac{1}{3M}\sum_{j=1}^{M} |\mathbf{f}_j^{DFT} - \mathbf{f}_j^{MD}|^2} \,. \tag{3.6}$$

This is depicted in Fig. 3.3 for each of the pigments. This visualisation provides further information than the final objective function alone; considering BCL and CLA, it is clear the remaining error measured by the objective function is caused by a few atoms which still have a large error whilst the rest are well optimised. In contrast, for CLB the remaining error is spread more evenly over the total molecule, indicating some further optimisation may be

Table 3.1: Initial and final objective function, total decrease and the number of steps of the optimisation, for each pigment.

Chromophore	Initial O([p]) /kcal mol^{-1} Å^{-1}	Final O([p]) /kcal mol^{-1} Å^{-1}	Number of steps
BCL	33.01	22.31	6919
CLA	40.49	21.49	5289

CLB	41.13	21.73	6694

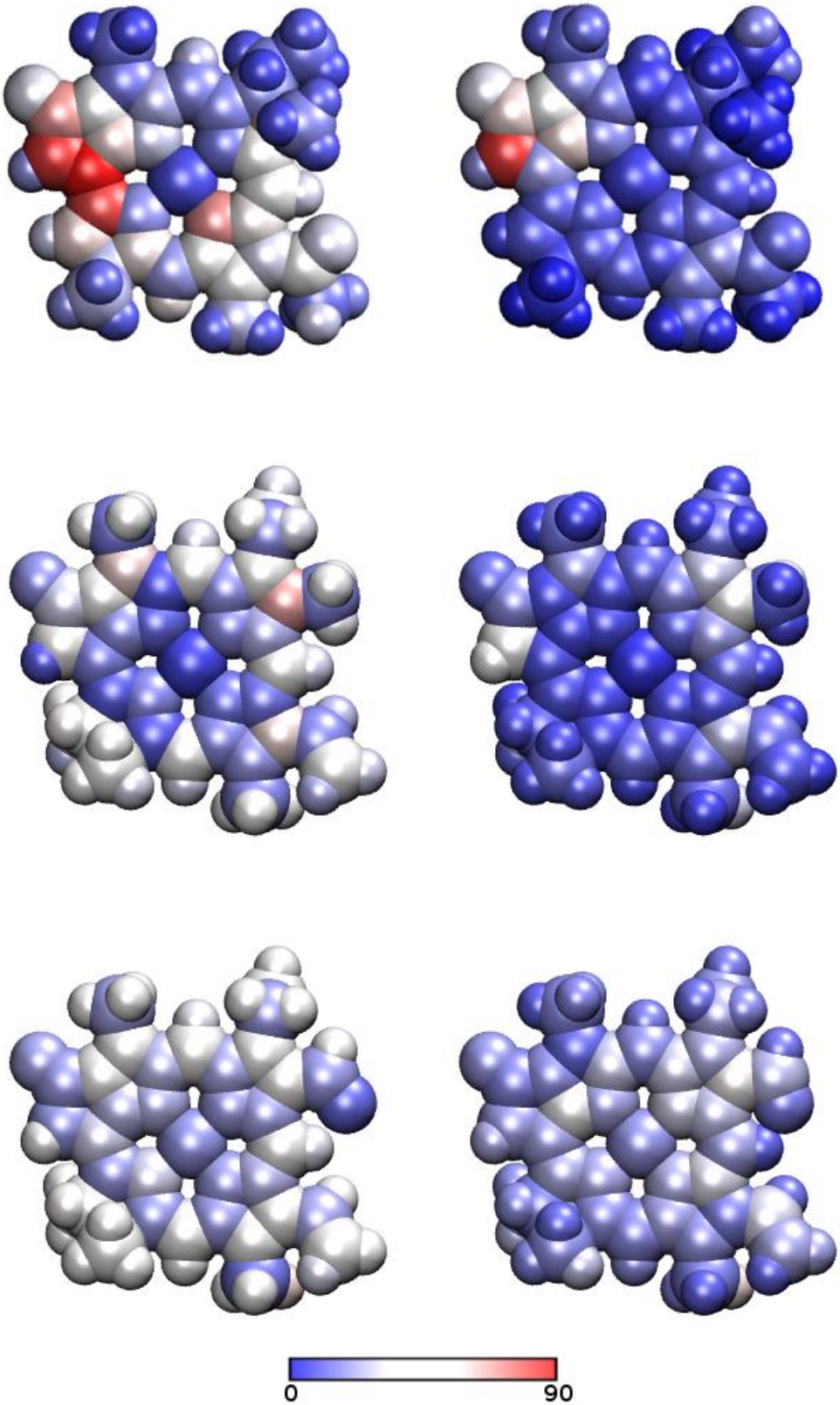


Figure 3.3: The atomic force error (in kcal mol^{-1} $Å^{-1}$) before (left) and after (right) optimisation for (top to bottom) BCL, CLA and CLB.

required. Thus, visualisation of the average force error per atom is recommended to ensure full optimisation. For BCL and CLA, the atoms where the force error remains large is likely the result of a poor atom type assignment. If an atom is assigned unsuitably, i.e. equivalent to another atom when it is in fact different, assigning parameter values that describe both atoms in the structure well is problematic. The result is either parameters describing one atom well and the other poorly or describing both poorly, and thus the error in the force(s) on the atom(s) will be greater. This observation provides potential routes of improvement for the algorithm. A process periodically calculating the average error per atom could be added to identify when the overall error is being caused by only a few atoms and then reassign these atom types. Additionally, a procedure could be developed to automatically assign atom types based on bond and angle types and sizes from the optimised structure to prevent any human error in atom type assignment.

Examining the final average error in force for each atom of BCL, it is evident there is one atom with a notably large error and three to four others also with a smaller but still significant error. In Fig. 3.4 the values of the average error in forces are shown and those with $\Delta f_i > 50$ kcal mol^{-1} Å^{-1} are highlighted (atom numbers 56, 57, 62, 64); it can be seen these are the same atoms with notably high force after optimisation as in Fig. 3.3. Further optimisation after reassigning the atom types of these atoms lead to a decrease of 1.5 kcal mol^{-1} Å^{-1} of the objective function. This example also clarifies that the main origin of the residual mismatch between empirical and *ab initio* force is the functional form of the empirical potential, unable to capture all the details of the electronic structure calculations.

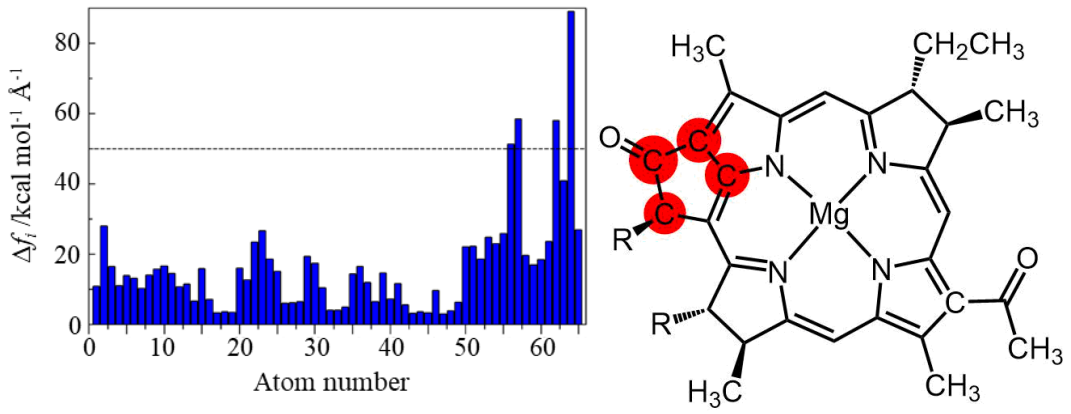


Figure 3.4: Δf_i values for atoms in BCL (left). BCL structure with atoms with $\Delta f_i > 50$ kcal^{-1} Å^{-1} highlighted in red (right).

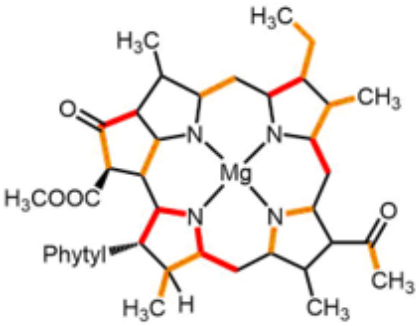
3.5.2 FF structural change

To understand the structural change between the original and optimised FFs the change in the b_0 parameter can be examined. Furthermore, as the electronic properties of a molecule are highly dependent on its internal structure, such changes will give an indication of how deficient the original FF is in describing the molecular structure. Firstly, considering the bond length change defined as 'structurally different' previously in this chapter (> 0.03 Å), the equilibrium bond parameters with a change of greater than 0.03 Å were identified. Next, changes in the bonding structure were determined by considering the typical bond lengths for C – C and C – N bonding structures (noted in Table 3.2). These results are summarised in Fig. 3.5, which contains tables for each pigment of the initial by-analogy value (i.e. before they were set equal to those of the optimised structure), b^i, final value, b^f, and the overall change, b^{change}, of the equilibrium bond length, as well as a depiction of all the bonds in the pigment these atom types correspond to. The bonds with a change greater than 0.03 Å are highlighted in orange and those with a change in bonding structure are highlighted in red. In majority of the b_0 changes in the BCL FF are decreases, conversely the CLB optimised FF predominantly results in increases in bond lengths and CLA has a more balanced combination of increases and decreases. This indicates there is no pattern of over- or under- estimation of bond lengths by the original FFs in comparison to the ES. The greatest number of changes, as well as the changes with greatest magnitude occur in the BCL FF.

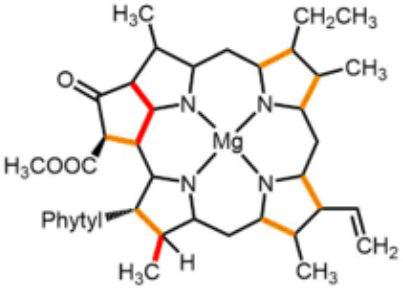
Table 3.2: The typical bond lengths of different bonding structures for C – C and C – N bonds.

Bond	Single /Å	Conjugated /Å	Double /Å
C – C	1.54	1.40	1.35
C – N	1.47	1.34	1.25

BCL			
Parameter	b^i /Å	b^f /Å	b^{change} /Å
C – C bonds			
CNR CRB	1.54	1.46	-0.08
CDN CCC	1.42	1.35	-0.07
CNA CO	1.44	1.50	0.06
CRB C1T	1.55	1.51	-0.04
CNC C3T	1.42	1.38	-0.04
CC C1T	1.54	1.50	-0.04
CNA CNC	1.41	1.37	-0.04
CO C3T	1.54	1.50	-0.04
CRB CDN	1.51	1.47	-0.04
C1T C1T	1.54	1.51	-0.03
C – N bonds			
NA CNR	1.44	1.33	-0.11
NA CDN	1.39	1.31	-0.08
NB CNA	1.30	1.35	0.05



CLA			
Parameter	b^i /Å	b^f /Å	b^{change} /Å
C – C bonds			
CPM2 CPA3	1.40	1.46	0.06
CPBN CPA3	1.40	1.45	0.05
CT2N CTN	1.53	1.48	-0.05
CT2N CT2N	1.54	1.58	0.04
CTN CPM2	1.53	1.49	-0.04
CPAN CPBN	1.40	1.43	0.03



CLB			
Parameter	b^i /Å	b^f /Å	b^{change} /Å
C – C bonds			
CT2N CT2N	1.54	1.61	0.07
CN CPBN	1.49	1.54	0.05
CPBN CPA3	1.40	1.45	0.05
C2 CTN	1.58	1.62	0.04
CPM2 CPA3	1.40	1.44	0.04
CT2N CTN	1.53	1.49	-0.04
CPBN CPAN	1.40	1.44	0.04
CTN CPM2	1.53	1.49	-0.04
CTR CPBN	1.50	1.53	0.03
CPBN CPA2	1.41	1.44	0.03

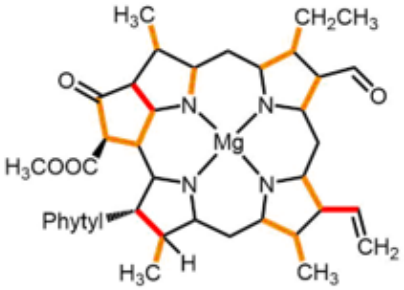


Figure 3.5**:** Tables of b_0 details for each pigment containing: the atom types of the bond; initial by-analogy value, b^i; final value, b^f; and the overall change, b^{change}, for bond types with b^{change} ≥ 0.03 Å. The corresponding bonds are highlighted in the pigment structure, change in bonding structure (red) and structural change (orange). Note as some atoms have degenerate types, multiple bonds may be described by the same parameter hence the number of bonds in the tables and number of bonds highlighted is not necessarily equivalent.

3.5.3 Spectral densities with new FFs

To briefly illustrate the possible effects of this revised forcefield on the exciton physics, the spectral densities of each pigment were computed with the original and optimised FFs as outlined in section 3.4.4. The results are depicted in Fig. 3.6. It has previously been suggested that the existence of an intramolecular vibrational mode that is quasi-resonant with transition energy increases transport efficiency in PPCs, through vibronic coupling.[139] As these optimised FFs generate more consistent structures, the differences in the high frequency part of the spectral density become especially important as they may correspond to intramolecular vibrations potentially capable of improving transport efficiency through vibronic coupling. Whilst other work has shown that no such vibrations are present in FMO,[140] a quasi-resonant vibration has been found in the phycoerythrin PPC: PE545.[29] Thus, for PPCs for which it is unknown if an intramolecular vibration capable of vibronic enhancement exists, the differences due to the optimised FF, particularly in the high frequency (> 500 cm^{-1}) region, are especially important. From the spectral density the reorganisation energy, λ, can be calculated via

$$\lambda = \frac{1}{\pi}\int_0^\infty \frac{J(\omega)}{\omega}\,\mathrm{d}\omega\,. \tag{3.7}$$

For each pigment these are reported in Table 3.3, for the original FF, $\lambda_{original}$, and optimised FF, $\lambda_{optimised}$. The reorganisation energy is often used as a measure of the 'total strength' of the pigment-protein interaction; the reorganisation energies of all 3 pigments are lower with the optimised FF, suggesting the original FFs may systematically overestimate the value of λ and thus the strength of the pigment-protein interaction, further supporting the need to develop consistent FFs.

Table 3.3: Reorganisation energy computed from the spectral density with the original FF, $\lambda_{original}$, and optimised FF, $\lambda_{optimised}$.

Chromophore	$\lambda_{original}$ /cm^{-1}	$\lambda_{optimised}$ /cm^{-1}
BCL 4	9	5
CLA 2	59	44
CLB 2	262	86

As the variance of the excitation energy is related to the integral of the spectral density, the difference between the original and optimised FFs is illustrated by computing the average of the excitation energy, E_{avg}, of a chromophore from each system (the same chromophore as used in the spectral density computation) and the fluctuation of the excitation energy, σ_{exc}, along a long MD trajectory (100 ns). The initial set up of the MD was as outlined in 3.4.1. The excitation energy was then computed along the trajectory every 500 ps, using the same method and functional/basis set combination as outlined for the excited state calculations of the spectral density. The results are summarised in Table 3.4. BCL and CLB both have notable changes in the average energy, with BCL 4 having a smaller average excitation energy by 0.07 eV and CLB having a greater average excitation energy by 0.08 eV. The difference between the averages resulting from the 2 FFs of CLA is much less, only 0.02 eV but the variance is notably reduced. The energy distributions of the excitation energies of each chromophore over the 100 ns are depicted in Fig. 3.7. As one would expect the energy distributions to fit a normal distribution, Gaussian distributions have been fit to each (using E_{avg} and σ_{exc} from Table 3.4 as μ and σ respectively). This data demonstrates the magnitude of the affect that the FF has on the resultant spectral density.

Table 3.4: Average excitation energy, E_{avg}, and its fluctuation, σ_{exc}, for each BCL 4, CLA 2 and CLB 2, computed over 100 ns using the original FF and optimised FF.

Chromophore	FF	E_{avg} /eV	σ_{exc}
BCL 4	Original	1.94	0.02
	Optimised	1.87	0.05
CLA 2	Original	2.08	0.07
	Optimised	2.06	0.04
CLB 2	Original	1.96	0.08
	Optimised	2.04	0.06

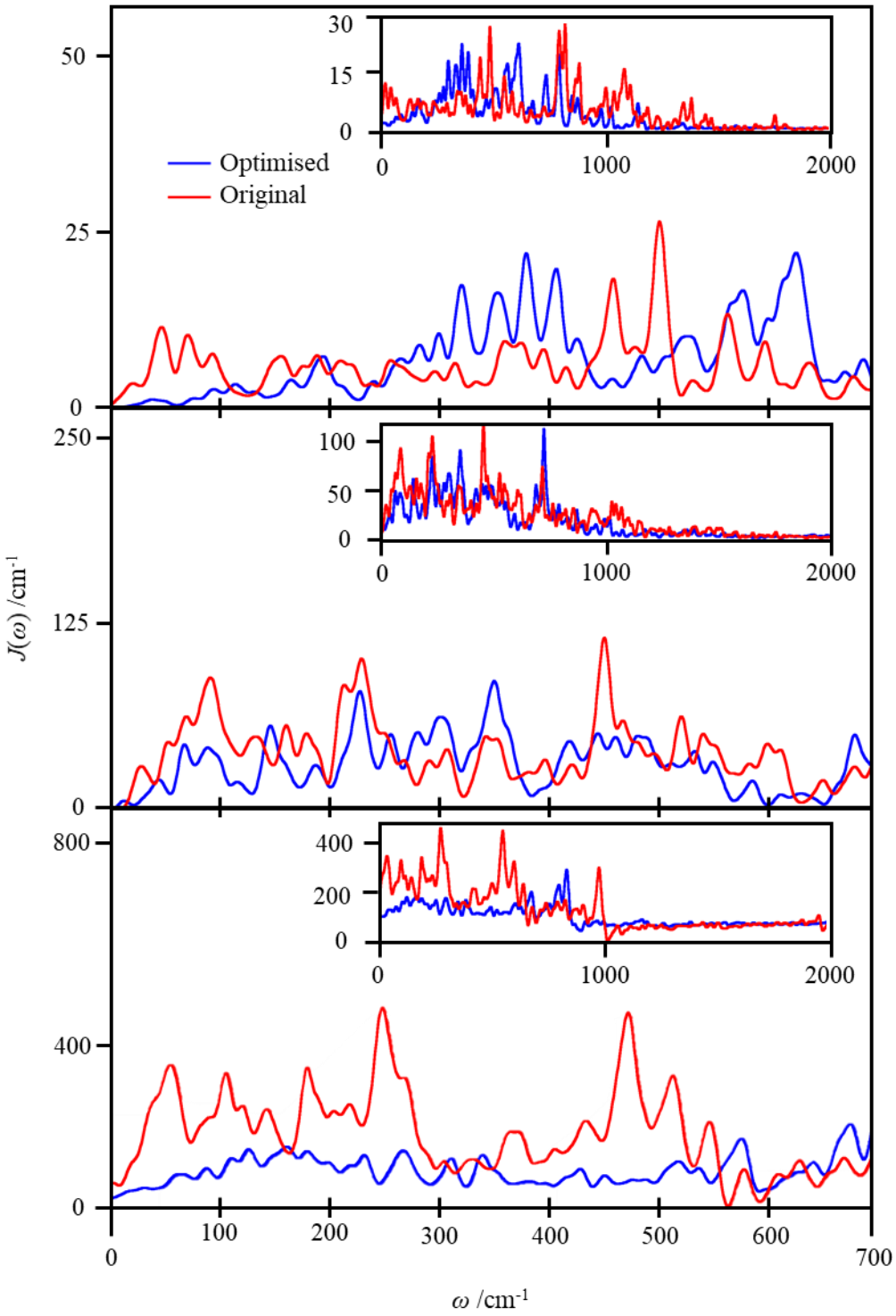


Figure 3.6: Spectral densities computed using the original FF (red line) and optimised FF (blue line) for the three chromophores: BCL 4 (top panel), CLA 2 (middle panel), CLB 2 (bottom panel).

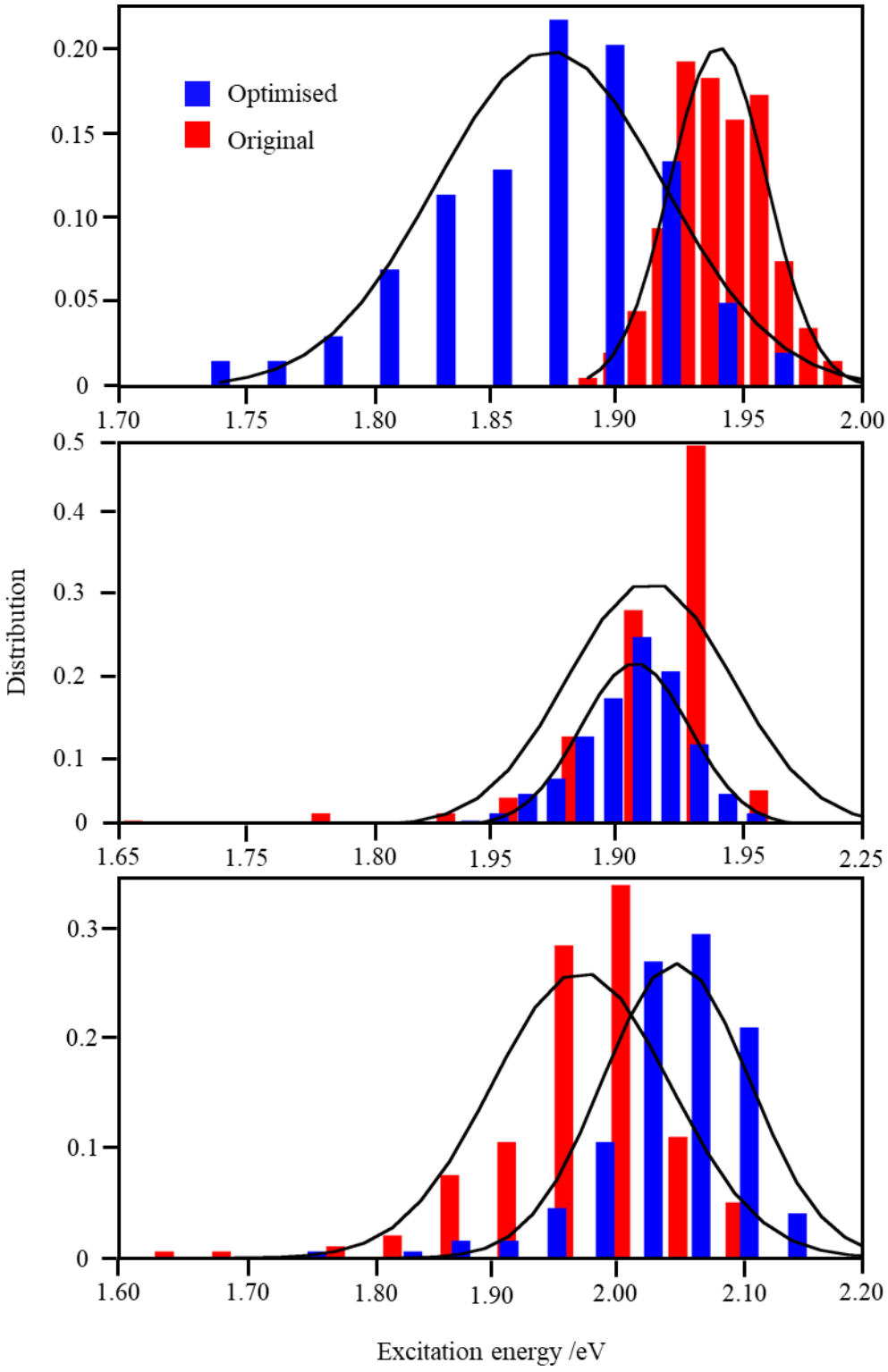


Figure 3.7: Excitation energy distributions of the original (red bars) and optimised (blue bars) FF over 100 ns run with their fitted normal distribution (black lines) for BCL 4 of FMO (top), CLA 2 of WSCP-a (middle) and CLB 2 of WSCP-b (bottom).

3.6 Conclusions

A method ensuring the consistency of the empirical FF with the quantum chemical method of choice for spectral density calculations of LHCs has been introduced and applied to the calculation of the spectral densities of three pigments: BCL, CLA and CLB. This force matching method benefits from its automatic nature: once the initial FF is generated and both sets of forces are computed it requires no further manual input unlike other parameterisation methods, allowing the efficient generation of multiple FFs and therefore the study of multiple PPCs. However, the implementation of the algorithm could be improved by utilising a central difference in the gradient calculation to reduce errors and using a quadratically convergent second order optimisation method such as a Newton-CG method to reduce the required number of iterations. The method is tuneable, applicable to both different molecular mechanics FFs and QC functionals, enabling the study of a range of PPCs with any desired MD and QC combination which is required to compare and contrast different systems in order to understand underlying mechanisms. It is shown that it is possible, within the methodology, to verify whether the parametrization of the forcefield is sufficiently accurate or new atom types ought to be defined.

The spectral densities obtained from the optimised FF have been compared with those obtained from the original FF demonstrating the important changes introduced by improving the FF aligning it as much as possible with the quantum chemical component of the calculation. A step forward could be considered the calculation of the spectral density entirely from an *ab initio* QM/MM trajectory, an attempt that has been made very recently, however was exceptionally resource intensive and time consuming[52]. Thus, such a method is still too expensive for efficient computation of multiple spectral densities, with additional concerns on the lack of proper equilibration in such short simulation time. The most statistically meaningful comparisons are performed on very long trajectories, showing that the reorganisation energies computed from the optimised FF are consistently lower than the original FFs. The observed differences in the high frequency region of the spectral density are potentially important to correctly identify the presence (or lack of) intramolecular vibrational modes capable of vibronic enhancement.

4 Specificity of local protein environment in LHCs

The role of the LHC environment in EET is a topic of much interest, particularly whether the environment is somehow optimised to promote efficient transfer. If this were the case, it would be expected that the system-environment interaction would be markedly different for LHCs than for PPCs not involved in light-harvesting. In this chapter, the aim is to uncover the existence or non-existence of such specialised interactions by examining the spectral densities of a few PPCs with different functions (light-harvesting and non-light-harvesting). To separate out the environmental effects the spectral density of the rigid chromophore is examined, to eliminate contributions from the internal modes.

4.1 Introduction

4.1.1 Spectral density studies of LHCs

As previously outlined, it has been proposed that the environment of LHCs is specialised to promote or protect coherence in some way,[13–15,45] and the spectral density is the ideal quantity for studying system-bath interactions. As noted in the last chapter there is much work on computing this quantity. Previous examination of the spectral densities of individual pigments in FMO found they are relatively similar.[97] Other PPCs have also been the focus of such studies. Examination of the spectral density of the PE545 complex identified the origin of relevant peaks and differences between pigments, concluding both protein and solvent play an important role in influencing coherence.[43] However much of this work utilises the combined MD and QC method which suffers from the geometry mismatch problem (previously discussed in 3.1.1) as it was carried out prior to this issue being highlighted. Additionally, whilst there are some studies that compare spectral densities of different complexes,[39] it is still unclear whether any environment modulation is optimised for

promoting coherence and how fine-tuned this may be. A first step toward addressing such a question would be to compare the environments of light-harvesting and non-light harvesting PPCs. The presence of regions in the light-harvesting PPC which amplify or reduce the exciton-nuclear coupling, would indicate the environment is adjusting the decoherence to some beneficial level, furthermore the absence of similar regions in non-light-harvesting PPCs would suggest some optimisation. Conversely identifying such interactions do not exist would signify there is no specific adjustment and thus the environment behaves as a glassy low dielectric medium as commonly assumed by biophysicists.[141,142]

4.1.2 Rigid chromophores

Here the environmental effects on a rigid chromophore are examined in order to focus solely on the role of environment. Freezing the intramolecular modes is the ideal method to do this and is advantageous for several reasons: firstly, it removes the contribution of the intramolecular vibrations of the chromophore which obscure those contributed by the environment, so modes in the spectral density arising from environment become clear. Next, higher frequency motions are predominantly due to the intramolecular motions and research has shown that for FMO this high frequency part (> 500 cm^{-1}) of the spectral density plays a small role in the exciton dynamics,[140,143–145] as these modes have much larger vibrational energy than the excitonic transition energy. Furthermore, the intramolecular modes are a main source of inaccuracy in computed spectral densities due to the problematic treatment of quantum vibrations in a classical framework[41,140] and the geometry mismatch.[16,146] The RMS of the heavy atoms oscillations at room temperature around their equilibrium positions are lower than 0.20 Å from QM/MM computation of normal modes and the intramolecular modes have larger characteristic frequencies[25] making this frozen chromophore system an acceptable method. Additionally, in the study of electron transfer isolating the environmental effect through freezing the chromophore is common, it is considered equivalent to assuming additivity of internal and external reorganization energy.

In contrast to other detailed studies of the effect of the environment on the absolute (average) position of energy levels, here the aim is to study the effect of the environment exclusively on the fluctuations of the excitation energies with the aim to identify specific interactions that affect the low-frequency part of the spectral density. Furthermore, through the examination of several PPCs with different functionality in order to identify any specialisation of the environment. In total three PPCs are examined, these are FMO, which is involved in light harvesting, and two variants of the water-soluble chlorophyll binding

protein, which are involved in transporting chromophores in the organism: WSCP-a and WSCP-b, in other words, the protein scaffold is identical, but the chromophore contained is different. Additionally, the analysis of a range of systems is advantageous because it helps ruling out accidental correlations that might be identified by analysing only a more modest amount of data, or to highlight patterns that were not arising frequently in other cases.

4.2 Systems

The three systems examined in this work are summarised in Table 4.1. FMO seemed a good candidate to represent the light-harvesting proteins as it is one of the most studied proteins in this field, due to its crystal structure being available for many years[2] and (as previously mentioned) it being the system for which long-lived coherence was initially observed.[3] Whilst the WSCPs are chromophore transporters and are not actively involved in light-harvesting.

FMO is a trimer containing a total of 24 chromophores; 8 per monomer unit. Each unit has 7 chromophores inside the monomer protein scaffold, with the eighth chromophore sandwiched between two adjacent monomer units. The WSCPs each contain 4 chromophores, organised as two adjacent parallel displaced pairs.

Table 4.1: Proteins and chromophores analysed in this work

Protein	Chromophores	Abbreviation	# of Chromophores
FMO	Bacteriochlorophyll-*a*	BCL	8
WSCP-a	Chlorophyll-*a*	CLA	4
WSCP-b	Chlorophyll-*b*	CLB	4

4.3 Computational Details

4.3.1 Protein MD

The crystal structures of FMO and WSCP are from the Protein Data Bank (PDB: 3BSD and 5HPZ, respectively). Any missing residues at the edges of the protein chain have been ignored, while any missing atoms were added using the CHARMM-GUI website.[147] Preparation and MD simulations were performed with the GROMACS 5.0.5 software.[148] The proteins were embedded in cubic boxes with side 120 Å and 110 Å for FMO and WSCP, respectively. Histidine residues have been assigned the appropriate protonation state to allow coordinating Mg atoms of chromophores, otherwise they have been assigned the proton in

position ε. Water molecules were added, and the ionic strength of the starting box was set to 150 mM by adding potassium and chloride ions.[149,150] The system has been described using the TIP3P model for water molecules, the CHARMM36 forcefield for the protein,[74] and literature parameters for the chromophores. Initial minimisation of the system was performed keeping all chromophores frozen, with 2000 steps of Steepest Descent. For all the following steps, a 2 fs integration step was used and constrained all the bonds with the LINCS algorithm implemented in GROMACS. Equilibration of the system was done in two steps of 500 ps each, first NVT (heating up to 300K) and then NPT conditions (Berendsen barostat), respectively. Finally, a separate equilibration of several nanoseconds was carried out for each frozen chromophore.

4.3.2 Spectral Density Computations

The last 16 ps of the several nanosecond trajectories were used to compute the excitation energies every 20 fs, using a QM/MM scheme within TDDFT linear response theory.[136] An MM radius of 30 Å around the chromophore was included in the calculation as this is needed to neglect the uncertainty arising from the fluctuation of excitation energy as a function of MM boundary, which is consistent with similar considerations in the literature.[44] Additionally, there is a high correlation between 6-31G* and 3-21G* data, with a slope close to 1 ($R^2 \approx 0.98$, see Fig. A4.1), thus the standard deviations at the two levels of theory are comparable and so the 3-21G* basis set is used as it is a good compromise between accuracy and computational cost. Comparison of 4 different functionals (B3LYP, CAM-B3LYP, M06-2X, ωB97X-D), showed consistent behaviour ($R^2 > 0.90$) among CAM-B3LYP, M06-2X, ωB97X-D (see Fig. A4.2). So the ωB97X-D/3-21G* combination, including all the residues within a 35 Å radius around the QM chromophore, was used for the QM/MM calculations an approach similar to the ones used, for example, in refs. 39 and 101. Also, the number of atoms of the chromophores in the QM region was reduced by inserting a link atom between the first and second carbon atoms of the phytyl chain. The QChem 4.2 software was used for the QM/MM calculations.[151] Fourier transforms (FTs) have been computed over 800 points, obtained on a 16 ps time window at regular intervals of 0.02 ps.

4.4 Results

4.4.1 Spectral Density Results

The spectral density is obtained from an FT of the autocorrelation function (ACF) of the fluctuation of the excitation energy of the first excited state of the chromophore along the MD

trajectory. Table 4.2 reports the statistics of the QM/MM calculations and Figs. 4.1-2 show the data processing in terms of the ACF calculation and their FTs.

Identifying patterns in the FTs and ACFs leads to two classification groups: a group showing tails in the ACFs with only one peak at low frequency ($\omega < 10$ cm^{-1}) in the FT (Group 1) and a group with clearly identifiable periodicities in the ACF leading to one or more strong peaks in the FT with frequencies on a significant time scale ($\omega > 10$ cm^{-1}) in exciton lifetime terms (Group 2). Group 1 contains half the FMO chromophores: BCLs 367, 371, 372 and 373

Table 4.2: QM/MM average excitation energies and their fluctuations along the MD trajectory for the chromophores belonging to proteins reported in Table 4.1.

Chromophore	Avg. (eV)	σ (eV)	Group
BCL367	1.7581	0.0119	1
BCL368	1.7908	0.0123	2
BCL369	1.7839	0.0088	2
BCL370	1.7613	0.0072	2
BCL371	1.7735	0.0073	1
BCL372	1.7972	0.0054	1
BCL373	1.7920	0.0089	1
BCL400	1.7710	0.0105	2
CLA1	2.0938	0.0067	2
CLA2	2.1316	0.0072	1
CLA3	2.0270	0.0071	2
CLA4	2.1145	0.0151	2
CLB1	2.1398	0.0070	2
CLB2	2.1130	0.0082	2
CLB3	2.0631	0.0126	1
CLB4	2.0901	0.0150	1

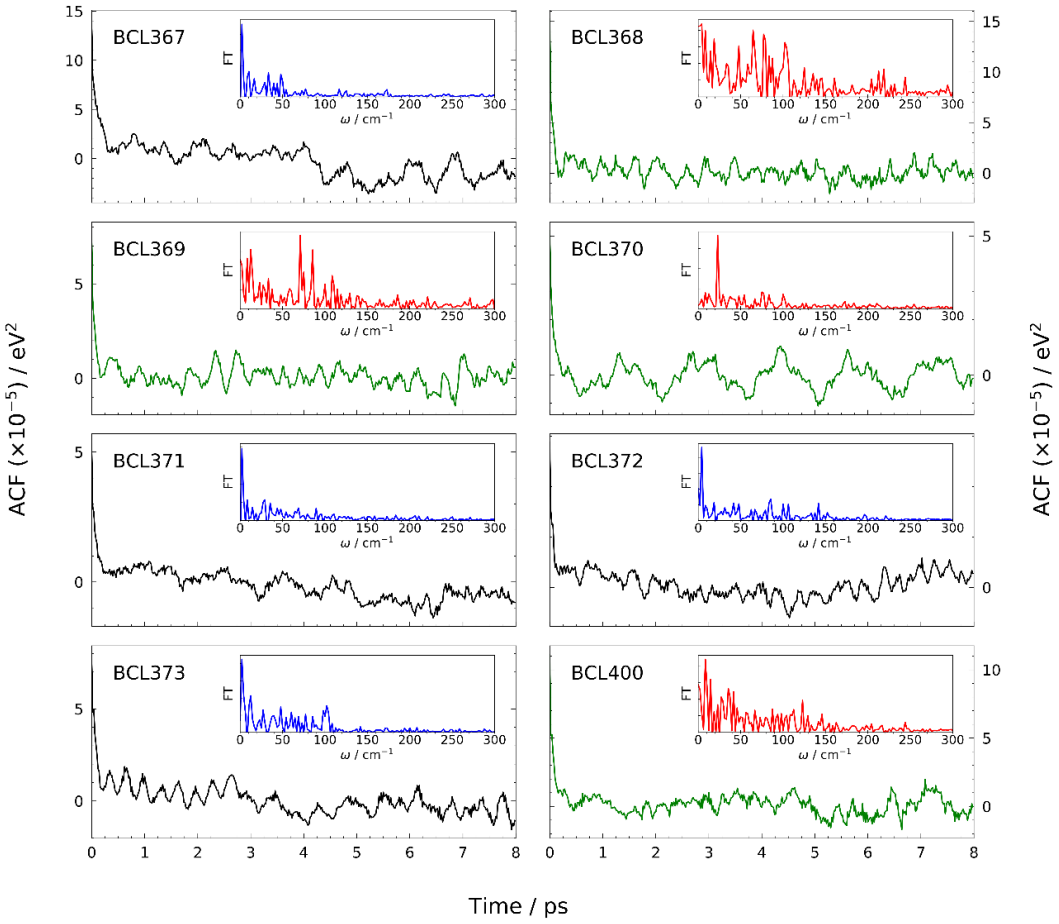


Figure 4.1: Autocorrelation functions (black lines for Group 1, green lines for Group 2) and their FTs (insets, blue lines for Group 1, red lines for Group 2) for FMO chromophores.

and three chromophores of the WSCP variants: CLA 2, CLBs 3, 4. Tails in ACFs can indicate non-convergence of the ACF due to long-time protein dynamics. In order to determine if this was the case here, the ACF was recomputed over a longer timescale however this resulted in no significant improvement. Thus, it was decided that repeating the calculation on the other chromophores on a longer timescale would not be of benefit. The motions characterising Group 1 are significantly slower than the exciton dynamics and act as inhomogeneous broadening: in a sample there will be different conformations of the protein, and to a good approximation the experiment can be modelled as static along these degrees of freedom. Considering the symmetry of the WSCP systems, it is surprising to note the discrepancy of CLA 2 in regard to the other CLA chromophores. This is assigned to the symmetry breaking effect of the solvent which will be discussed in more detail later, as will the importance of the solvent. Group 2 contains the remaining chromophores: BCLs 368, 369, 370, 400, and CLAs

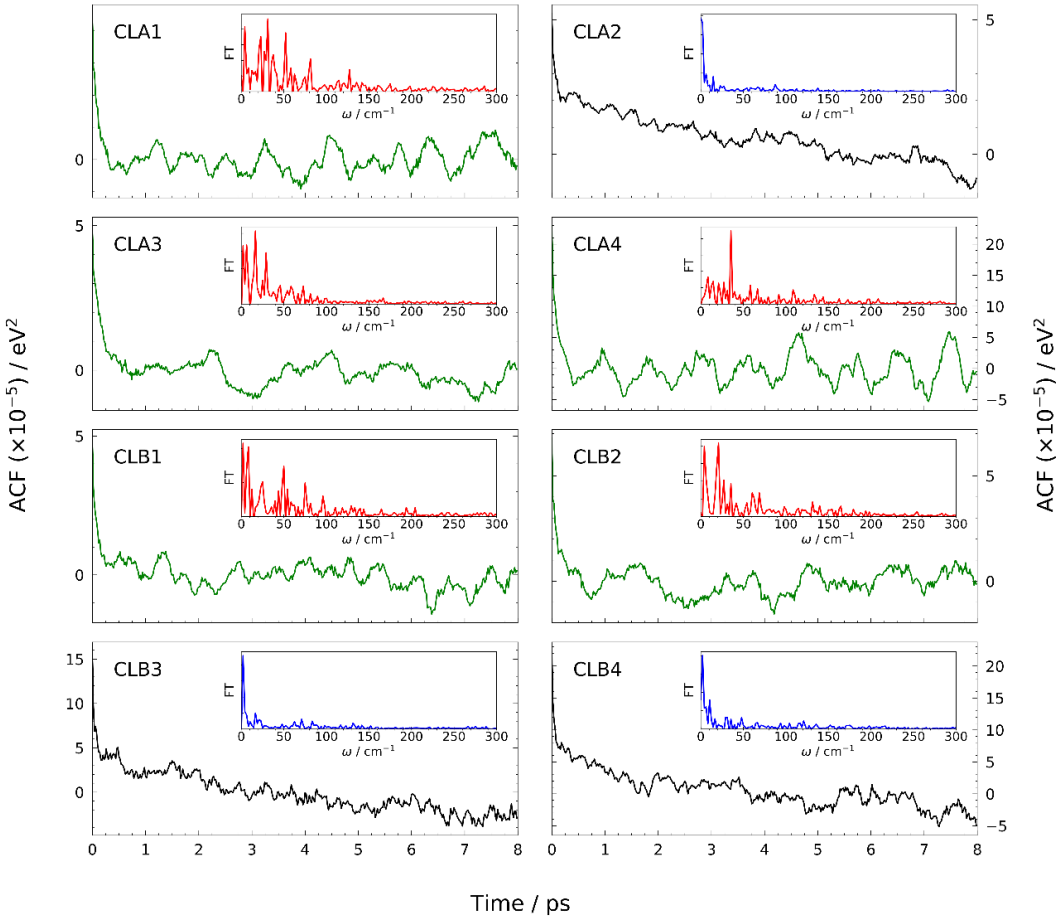


Figure 4.2: Autocorrelation functions (black lines for Group 1, green lines for Group 2) and their FTs (insets, blue lines for Group 1, red lines for Group 2) for WSCP chromophores.

1, 3, 4, and CLBs 1, 2. The spectral densities for these chromophores show the motions are coupled in a complex way with the exciton dynamics.

To determine if belonging to Group 1 or 2 was correlated with fluctuation of the excitation energy (i.e. the overall exciton-vibration coupling strength) a Kruskal-Wallis statistical test was performed.[152] A Kruskal-Wallis test determines whether samples originate from the same distribution. The result (Kruskal-Wallis-p = 0.60) indicates that the σ values of the two groups are from the same distribution and therefore there is no correlation between magnitude of excitation energy fluctuation and belonging to a certain group. The same test was used for fluctuations arising from protein type to determine if the magnitude of fluctuations of excitation energy are protein specific, and therefore related to biological function or if similar fluctuations occur due to a generic protein environment. First it was confirmed that the fluctuations of the two WSCP variants belong to the same distribution (Kruskal-Wallis-p =

0.56) as would be expected due to the identical nature of the protein scaffold. Then in order to determine if there was significant difference between the WSCPs and FMO, the Kruskal-Wallis test of the fluctuations of all three proteins was performed. Resulting in the conclusion that the fluctuations are not specific to FMO (Kruskal-Wallis-p = 0.81) and so are not related to the biological function of the proteins.

4.4.2 Correlations between environmental motion and excitation energy

Next, to gain chemical insight possible of aiding materials design and to detect any special function of the environment, identification of motions correlated with fluctuations of excited energy was attempted. Pearson's correlation coefficient (r) was computed between the TDDFT excitation energies and the coulombic interaction energies with each of the residues within 10 Å of the chromophore along the MD trajectory. The coulombic interaction energy was computed by representing the chromophore as a set of point charges from the difference between ground and excited state atomic charges within a Merz-Kollman scheme.[153,154] The results of BCL 368 and CLA 3 are reported in Table 4.3 as an example. Reported in the table for each residue with a correlation $r > 0.30$ is: the fluctuations of the coulombic interaction energy (σ_E), the distance (d) of interacting atoms and its fluctuation (σ_d), the atom names of the interacting atoms, and assignment of the interaction to a particular component of the environment (solvent, S, protein, P, or chromophore, C). The complete series of tables (Tables A4.1-3) containing the results of this analysis for all chromophores are reported in the appendix for this chapter. Note solvent residues are formally identical and can exchange. However, here the order number of the simulation is used to refer to them.

Table 4.3: Correlation analysis for residues surrounding chromophores.

Residue	r	σ_E (eV)	d (Å)	σ_d (Å)	Interaction	Class
BCL 368						
SOL20272	0.373	2.89E-04	1.95	0.05	OW-MG	S
SOL16680	0.335	3.21E-04	2.19	0.28	HW1-ND	S
SER73	0.318	2.17E-04	1.77	0.19	HG-OBB	P
CLA 3						
ALA34	0.612	2.06E-04	1.93	0.17	O-HE1	P
ALA33	0.516	1.15E-04	1.91	0.15	O-HN	P

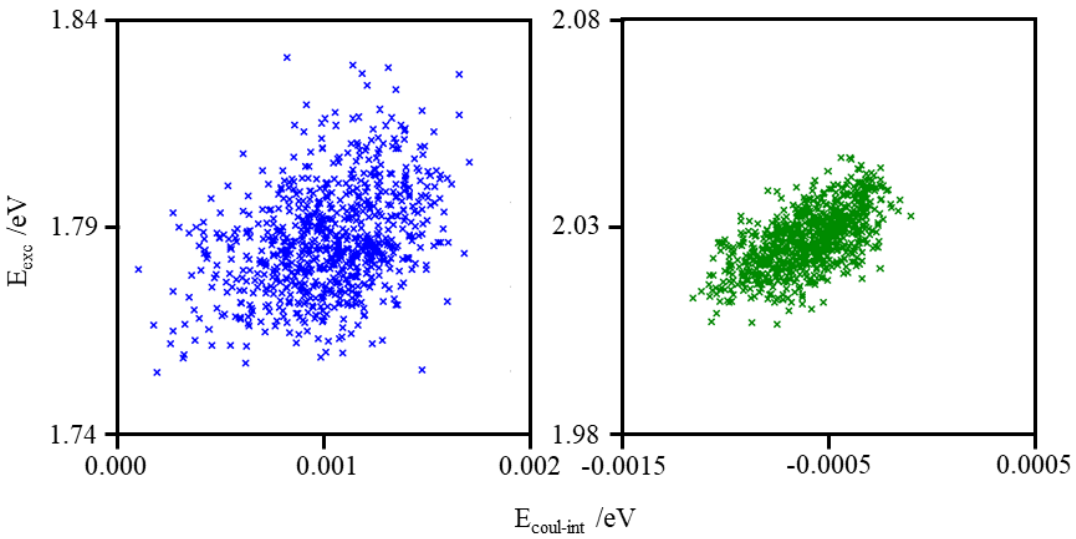


Figure 4.3 Coulombic interaction energy – excitation energy correlation for SOL20272 and BCL 368 (r = 0.373, blue) and ALA34 and CLA 3 (r = 0.612, green).

In Fig. 4.3 is shown the scatter plots of two correlations representative of moderate ($0.3 < r < 0.5$) and good ($r > 0.5$) correlations, to highlight the significant difference between them. The summary of interaction classification data reported in Table 4.3 (and Tables A4.1-3 in the appendix) is shown in Fig. 4.4 (see also Fig. A4.3 in the appendix). It demonstrates that the solvent predominantly influences the excitation energies. Also, the residue with the largest fluctuation of coulombic interaction energy is typically the most correlated, confirmed by Tables A4.1-3. Visual analysis of the MD trajectory confirms proximity and interaction (e.g.

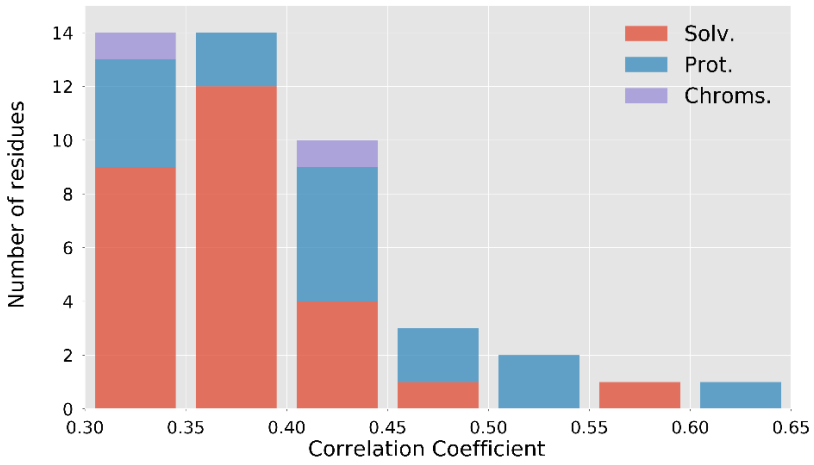


Figure 4.4: Distribution of the correlation coefficients for chromophore/environment interactions.

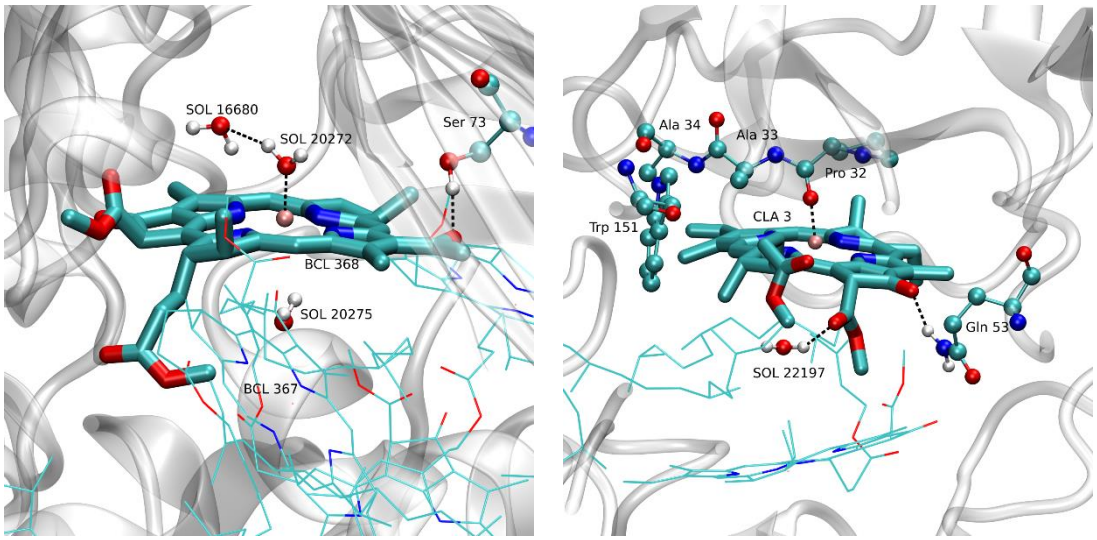


Figure 4.5: location of the most correlated residues reported in Table 4.3 BCL 368 (left) and CLA 3 (right).

hydrogen bonds or coordination) of the residues with chromophores, exemplified in Figs. 4.5 and 4.6.

To identify specific motions of the residues identified as most responsible for excitation energy fluctuations, the FT of the coulombic interaction energies (used to calculate *r*) were computed and compared to those of the TDDFT data. This analysis was performed on CLA 4 and BCL 370 due to their well-defined peak in the frequency domain in their spectral densities (Figs. 4.1-2). In theory this well-defined peak may be assigned to a specific motion. However, this comparison scarcely useful for analysing specific motions, as the frequency resolved signal also contains the components of many protein motions when a small residue or solvent molecule is considered. The results are shown in the appendix in Fig. A4.4.

The most correlated residue for BCL 370 is a solvent molecule (r = 0.48, SOL 26496) which is not coordinated to the chromophore directly but is involved in a H-bonded solvent network, holding it in place in proximity of the chromophore but still allowing it to rotate. This motion modulates the excitation energy. Similarly, for CLA 4 a solvent molecule is one of the most correlated residues (r = 0.36, SOL 23254), which again is free to rotate and modulate the chromophore excitation energy whilst being held nearby by a solvent network. This is demonstrated for both cases in Fig. 4.6. This freedom to rotate suggests a dipole-dipole interaction mechanism as the dipole is free to change orientations.

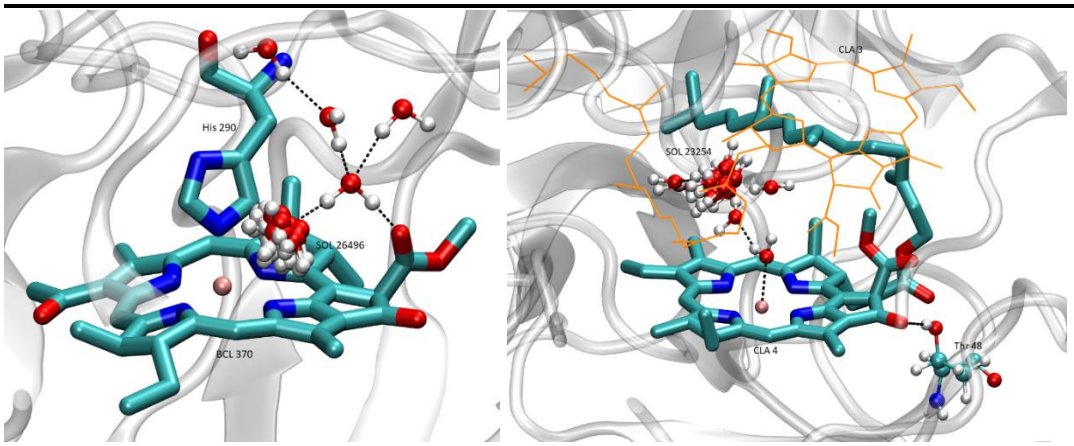


Figure 4.6: Solvent chain and SOL 26496 motions modulating the excitation energy of BCL 370 (left). Solvent chain, containing highly correlated solvent SOL 23254, mediating the interaction of the phytyl chains of CLA 3 (orange) and CLA 4 (right).

4.4.3 Average environmental composition of chromophores

Finally, the radial distribution function between the centre of mass of the chromophore and centres of mass of surrounding residues of certain types was computed and integrated along the MD trajectory to obtain the average number of that residue. This was to analyse the average composition of the surrounding environment of the chromophores, to understand if there are any hydrophobic/hydrophilic pockets and if these are important. The results are reported in Table 4.4 and are uniform amongst the chromophores, with the only exception being BCL 400. However, this is expected as the BCL 400 chromophore is located at the edge of the protein scaffold and is much more exposed to solvent. The environment component that is most variable is the solvent but there is not strong correlation between the average fraction of solvent in the 10 Å sphere around the chromophore and the fluctuation of the excitation energy along the trajectory (Fig. A4.5 in the appendix). This is backed up by additional statistical tests (Pearson's $r = 0.22$, Spearman's $\rho = 0.30$).

Table 4.4: Average number and type of residues in a 10 Å shell from the chromophore.

	σ (eV)	Chrom	Prot	Solv	Tot
BCL367	0.0119	1.00	16.56	5.82	23.38
BCL368	0.0123	0.00	12.88	11.31	24.18
BCL369	0.0088	1.00	15.14	8.19	24.33
BCL370	0.0072	0.00	15.30	13.97	29.27
BCL371	0.0073	0.00	13.06	13.65	26.71
BCL372	0.0054	0.00	15.22	3.26	18.49
BCL373	0.0089	1.00	14.78	9.11	24.89
BCL400	0.0105	0.00	18.15	32.73	50.88
CLA1	0.0067	1.00	14.96	9.91	25.86
CLA2	0.0072	1.00	15.96	11.71	28.68
CLA3	0.0071	1.00	15.43	9.62	26.06
CLA4	0.0151	1.00	16.11	12.41	29.53
CLB1	0.0070	1.00	17.20	7.19	25.39
CLB2	0.0082	1.00	17.72	12.45	31.17
CLB3	0.0126	1.00	16.65	8.28	25.94
CLB4	0.0150	1.00	16.98	14.59	32.57

4.5 Conclusions

In this work the fluctuations of the excitation energy of a set of similar chromophores (BCL, CLA and CLB) embedded in different protein environments (FMO, WSCP) were studied using an approach combining MD with subsequent QM/MM calculation on a number of the snapshots from the trajectories. To isolate the role of environment the chromophore whose spectral density was to be computed was kept frozen during the MD simulation. This also circumvents inaccuracies arising from the geometry mismatch problem. ACFs of the QM/MM excitation energies and their FTs were computed, analysis of which led to the identification of two groups of chromophores. A group showing only very low frequency (ω < 10 cm^{-1}) contributions (Group 1) and a group showing frequencies that can couple with the exciton dynamics in a complex way (Group 2). It has previously been shown that evolution of the spectral density occurs slowly over a nanosecond timescale as the protein samples different environments[140] and this timeframe is well beyond what is computable as time series of excitation energy. These motions do not influence the exciton dynamics as they are essentially

stationary in comparison. The exciton dynamics are dictated by a specific spectral density from within a limited time interval as opposed to one averaged over long times.[155] It would be expected that any interactions and/or environmental fluctuations specifically optimised to enhance exciton dynamics would be present in any sampled trajectory interval on a timescale of exciton dynamics. Hence the spectral densities examined here are not averages over several trajectories. Additionally, it is noted that due to the time window used, the reliability of the $\omega < 5$ cm^{-1} vibrations is questionable as their oscillation occurs over a time window greater than that of those involved in the FT computation. However, as previously discussed, these very low frequency vibrations act as inhomogeneous broadening and have a negligible role in the reorganisation energy

Some authors have proposed that the specific frequency of environmental motions can be important if they become resonant with specific energy level differences, although there is still no consensus on this issue.[25] The fluctuations of excitation energies due to environment reported here are around half that of the fluctuations due to the intra-chromophore modes of FMO which have been reported as around 0.02 eV.[140] The magnitude of fluctuations reported are comparable to the energy differences among localised excitons, highlighting how fluctuations may determine energy level order, thus improving (or impairing) the level alignment. However, statistical tests revealed no significant difference between the magnitude of fluctuations of the excitation energy of the two groups, nor a significant difference for the different proteins. This lack of distinct difference between the environment fluctuations of proteins with different functions indicates a lack of specific tuning of environmental noise for the promotion of energy transfer. The results observed here suggesting a general value could be obtained from any protein and do can help to rule out hypotheses stating that local environment has naturally evolved to be fine-tuned to promote energy transport.

Analysis of residues whose coulombic interaction energy with the chromophore is most correlated with the fluctuation of the excitation energy has highlighted the importance of solvent in these systems, with proximal rotating solvent molecules being able to modulate excitation energy through rotation of their dipole. These solvent molecules are held in place by strong interaction networks involving hydrogen bonding but are free to rotate, allowing them to continuously modify the extent of the interaction. A final analysis of the environmental composition of each chromophore ruled out any correlation between the number of a certain type of residue in the proximity of the chromophore and its excitation energy fluctuation.

The analysis of three PPCs appears to indicate an "unspecific" nature of protein environment that is independent of biological function. Further extension of these findings to more complexes could be useful in ruling out hypotheses involving the target of natural evolution. Overall the only obvious tunability would be to decrease exciton-environment coupling through reducing pockets accessible by solvent in the vicinity of the chromophore. The findings of complimentary studies focusing on the role of relative arrangement of chromophores indicate that there is some optimisation of this property in naturally occurring LHCs.[53,156] Additionally, the exciton is more strongly coupled with the motions of the chromophore itself[140] than that of environment motions, it is reasonable to argue that the exciton dynamics is likely not substantially affected by the chemical detail of surrounding environment. In summary, these results indicate that the most effective way to control the energy transfer process (either in artificial systems or by natural evolution) is by controlling the chromophore-chromophore interaction rather than the chromophore-environment interaction.

5 Exciton dynamics of multiple LHCs

In order to gain an understanding of the underlying principles of EET in LHCs it is necessary to compare multiple complexes in a consistent way as this will reveal factors important for efficiency. That is the aim of this chapter; to study the exciton dynamics of multiple LHCs using the same method.

5.1 Introduction

5.1.1 Atomistic models vs phenomenological models

When examining the underlying details of biophysical processes for many cases there are two contrasting methods of modelling: highly detailed models which include as much accurate system information as possible and phenomenological modelling which describe fundamental processes in a more simplified manner. The study of the underlying biophysical processes of LHCs is the same. As discussed in the first chapter, due to the system size of LHCs the only feasible way to include atomistic detail is the use of classical MM methods (typically then combined with some higher-level method), whilst phenomenological model examples are things like Förster, Redfield and phonon-mediated models.

Whilst such studies have provided insight into LHCs, both methods suffer their own drawbacks. When using highly detailed system specific models, the conclusions are not necessarily applicable to other systems therefore making it difficult to derive general principles and also increasing the risk of overestimating the importance of some feature/observation. In the case of phenomenological models, they can be difficult to validate if the parameters can simply be tuned to match experiment or highlight some feature. Moreover, these errors may compound one another as the direction of research is influenced by previous findings. A detailed atomistic study may uncover some interesting feature and thus (possibly incorrectly) deem it important and this feature may then be used as the focal point of phenomenological studies. For example, vibronic coupling was identified in an

LHC[157] this feature has then been included in several phenomenological studies.[52,139,157,158] However, as investigated in the previous chapter, vibronic coupling to environment is unlikely to be important. In order to avoid incorrectly assigning importance to some feature or property, there has been research which compares real LHCs to artificial analogues.[53,159–161] The results of such studies have indicated evidence of optimisation toward pigment density[160] and orientation.[53] However, results also suggested that symmetry is important for efficient EET as symmetry breaking of LH2 structures leads to decreased efficiency,[64] but FMO is low symmetry and still highly efficient.

The variability of the structure of LHCs makes comparison of atomistic studies difficult when the aim is to identify important characteristics. There is already a broad range of characteristics to consider within a single complex, e.g. specific interactions, chromophore orientations, inter- or intramolecular vibrations and more. Comparing all these things across several structurally different complexes is too complicated. In phenomenological models the underlying description of the different models is the same, making identification of fundamental components easier. These underlying principles can then be used as a basis to develop a more complex atomistic understanding of the process.

5.1.2 Examining exciton dynamics

Defining key components of exciton dynamics in LHCs is desirable for two main reasons. Firstly, the result may provide insight for materials design to improve efficiency in artificial light harvesters. Also, understanding the fundamentals of exciton dynamics could determine the source of efficiency. In other words, it may be possible to differentiate specialised functions of the complexes and intrinsic properties of molecular aggregates.

To understand the underlying biophysical principles of LHCs important for efficient EET, a uniform examination of the exciton dynamics of several complexes must be undertaken. Utilising the same phenomenological model allows direct comparison of the complexes, clarifying any fundamental commonalities. In this chapter, five PPCs will be examined by computing their excitonic Hamiltonian and their exciton dynamics approximated using a Lindblad master equation. Analysis of the Hamiltonians and resulting dynamics will identify any relationships between simple properties. Understanding the relationship between these simple properties, for example static disorder and relative coupling strengths of the chromophores and the exciton dynamics, is key in determining underlying mechanisms.

5.2 Systems

There are two criteria that must be met for PPCs to be included in this study: they must be involved in light and their structure must be available. In addition to these criteria, as the role of carotenoids is not fully clear – they may be actively involved in light-harvesting[162–164] or may be for photoprotection[165,166] – it was decided to exclude carotenoid containing LHCs. Also, irreversible energy transfer from high to low energy excited states may occur in complexes with more than one chromophore type and whilst this is biologically relevant, it obscures the ultrafast energy transfer that is the focus here. Thus, the final criterion for LHC selection was that the complex contained only one type of chromophore. Applying these criteria to proteins available from the RCSB PDB and removing low resolution, degenerate and non-natural light-harvesters gives roughly ten LHCs to be considered, from which the following 5 were selected: APC, B-APC, CPC, FMO and PE. Table 5.1 summarises their names, chromophore types and number and PDB code. The structures of the complexes are depicted in Fig. 5.1. The set is fairly representative of LHCs with only one chromophore type. It includes two (bacteriochlorophylls and bilins) of the three chromophore types (for the third type complexes are either too large or too low resolution for the purpose here) as well as complexes with a variety of symmetries.

Table 5.1 Summary of systems investigated, and abbreviations used.

LHC name	LHC abbr.	Chrom.	Chrom. abbr.	n chrom.	PDB code
Allophycocyanin	APC	Phycocyanobilin	CYC	12	1KN1
B-Allophycocyanin	B-APC	Phycocyanobilin	CYC	6	4PO5
C-Phycocyanin	CPC	Phycocyanobilin	CYC	18	3L0F
Fenna-Matthews-Olson	FMO	Bacteriochlorophyll-a	BCL	8	3ENI
Phycoerythrin	PE	Phycoerythrobilin	PEB	30	3V57

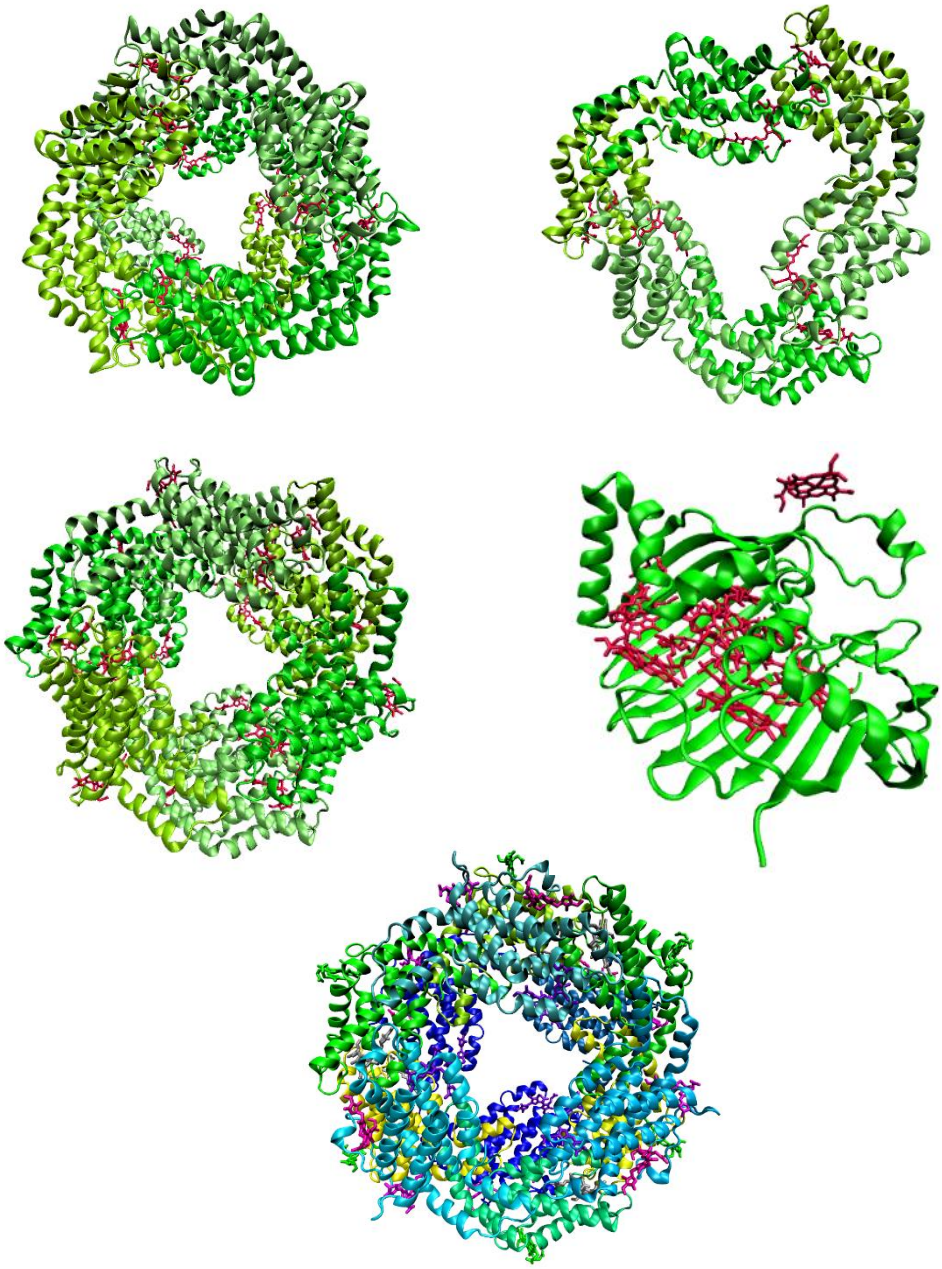


Figure 5.1 Structures of proteins: APC (top left), B-APC (top right), CPC (middle left), FMO (middle right), PE (bottom).

5.3 Computational Details

5.3.1 Structure optimisation

The structure of each LHC was obtained from the RCSB Protein Data Bank and then set up by placing the protein in the centre of a cubic box with 10 Å to the edge, using GROMACS 5.0.5 software.[148] The system was solvated with water molecules and the ionic strength set to 150 mM by adding potassium and chloride ions. The CHARMM36[74] forcefield was used for

the protein, TIP3P for the water, literature parameters for bcla[167] and CGenFF[168,169] parameters for PEB and CYC. The energy of the system was minimised using steepest decent followed by NVT (300 K) then NPT (1 Bar) equilibrations for 200 ps. The result was then used for input into a two-layer ONIOM model (as outlined is section 2.1) implemented with Q-Chem 4.2[151] for geometry optimisation of each chromophore of the protein. The total system was taken as the chromophore surrounded by 5 Å of its environment, with the chromophore as the QM part and the environment (all protein, solvent and ions) as the MM part. The QM part was treated with the B3LYP functional and 6-31G* basis set and the MM part using the same forcefield parameters as the MD setup.

5.3.2 Excited states and coulombic couplings

The excitonic coupling is conventionally separated into a long-range contribution described by the Coulombic coupling and short-range contributions. The Coulombic coupling[170–173] is given by

$$J^C = \iint \frac{\rho^a(\mathbf{r}_i)\cdot\rho^b(\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|}\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \cong \sum_{i,j}^{N_{grid}} \frac{\rho_i^a\cdot\rho_j^b}{|\mathbf{r}_i-\mathbf{r}_j|} \tag{5.1}$$

where $\rho^{a(b)}(\mathbf{r})$ is the transition density of localised exciton *a* (*b*). The integral in equation (5.1) is numerically evaluated as a sum over a grid of points $\{\mathbf{r}_i\}$, known as the transition density cube (TDC) where $\rho_i^{a(b)} = \rho^{a(b)}(\mathbf{r}_i)$. The results of the geometry optimisations were used as input into TD-DFT to compute the first four excited states at the B3LYP/6-31G* level and the transition density cube (TDC) of the first excited state (also implemented using Q-Chem 4.2). The resulting TDCs were the used to compute the coulombic couplings between each pair of chromophores in the complex as detailed in refs. 174 and 175.

5.3.3 Exciton dynamics

The dynamics of a single exciton through the chromophore network of each protein were modelled. This was performed using a Lindblad master equation:

$$\mathcal{L}\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L,\rho\} \tag{5.2}$$

which gives (as outlined in section 2.3):

$$\dot{\rho} = -i[H,\rho] + \mathcal{L}\rho. \tag{5.3}$$

The first term is given by the electronic Hamiltonian:

$$H = \sum_{i=1}^{N} \varepsilon_i |i\rangle\langle i| + \sum_{i \neq j} V_{ij} (|i\rangle\langle j| + |j\rangle\langle i|) \quad (5.4)$$

where N is the number of chromophores, ε_i is the first excited state energy of chromophore i and V_{ij} is the coupling between chromophore i and chromophore j and overall H describes the closed system dynamics. The second term contains Γ, the dephasing rate

$$L = \sqrt{2\Gamma}\, |i\rangle\langle i| \quad (5.5)$$

where Γ is set to be 2.1 ps^{-1} to be in the correct range to be consistent with experimental observations,[53] though its value it is not critical in this study.

This approach assumes the coupling between chromophores and their local environment is roughly similar across chromophores and LHCs, as determined in the previous chapter. Additionally the coupling with high frequency modes is neglected due to its effect on simulated exciton dynamics of FMO being found to be negligible.[140] It can be expected that such modes are important in complexes containing multiple chromophores by being acceptors of electronic energy during irreversible energy transfer toward lower energy excitations, however this situation is not present in the systems considered here.

5.4 Results

5.4.1 Structure of excitonic Hamiltonian

Statistical descriptors of the Hamiltonian matrix are summarised in Table 5.2 and the distribution of the absolute values of the Hamiltonian are illustrated in Fig. 5.2. The disorder of the excitation energy is comparable with the excitonic coupling for all the complexes; thus, the important descriptors are the range of excitation energies and the standard deviation of excitation energies. Comparing the value of the standard deviations in Table 5.2 it is evident that the standard deviation is reasonably similar for all complexes, with those of APC, B-APC and CPC being slightly smaller. Additionally, their value is very close to the thermal energy at room temperature and in-site energy fluctuation due to intramolecular modes[140] (~20 – 25 meV for both). In the previous chapter it was shown the intermolecular motions modulate the on-site energy to a lesser degree (standard deviation of ~ 10 meV) and that this modulation was similar for chromophores not involved in light harvesting. The combination of these observations suggests the static and dynamic disorder experienced by chromophores in LHCs is fairly similar and unspecific.

Table 5.2 Statistical descriptors of the excitonic Hamiltonian of each LHC: minimum excitation energy, $\min\left(\{H_{ii}\}\right)$, maximum excitation energy, $\max\left(\{H_{ii}\}\right)$, excitation energy standard deviation, $\sigma\left(\{H_{ii}\}\right)$, the largest absolute excitonic coupling, $\max\left(\{|H_{ij}|\}_{i\neq j}\right)$ and the smallest of the largest absolute coupling for each chromophore $\min_i\left(\max_j\left(\{|H_{ij}|\}_{i\neq j}\right)\right)$.

LHC	$\min\left(\{H_{ii}\}\right)$	$\max\left(\{H_{ii}\}\right)$	$\sigma\left(\{H_{ii}\}\right)$	$\max\left(\{\lvert H_{ij}\rvert\}_{i\neq j}\right)$	$\min_i\left(\{H_{ij}\}_{i\neq j}^{\max(i)}\right)$
APC	2.2620	2.3934	0.0316	0.0269	0.0233
B-APC	2.2923	2.3782	0.0277	0.0228	0.0174
CPC	2.4314	2.5338	0.0265	0.0201	0.0009
FMO	1.7177	1.8736	0.0558	0.0996	0.0066
PE	2.2197	2.4891	0.0623	0.0191	0.0060

To compare the strength of the excitonic couplings (off-diagonal elements of the matrix), the largest absolute excitonic coupling may be examined. From these couplings it is evident that they are similar for all complexes except FMO, which is 4 – 5 times larger. This is explained by the interchromophore separation i.e. the compactness of the LHCs as it is well known that coupling scales inversely with distance as illustrated in Fig. 5.3. Examining the smallest absolute coupling, however, is not worthwhile as there are many distant chromophores whose coupling will be close to zero. Instead the largest coupling for each chromophore *i* is identified,

$$\{H_{ij}\}_{i\neq j}^{\max(i)} = \max_j\left(\{|H_{ij}|\}_{i\neq j}\right) \tag{5.6}$$

and the smallest of the largest couplings is reported

$$\min_i\left(\{H_{ij}\}_{i\neq j}^{\max(i)}\right) = \min_i\left(\max_j\left(\{|H_{ij}|\}_{i\neq j}\right)\right). \tag{5.7}$$

This gives an indication of the chromophore most poorly coupled with the others.

Examining the most poorly coupled chromophore, it is clear there is a notable difference between the pair APC and B-APC and the other complexes (CPC, FMO and PE). In the case of APC and B-APC, there are no chromophores isolated from the others whilst for the others there is one or more chromophore completely decoupled from the rest. These chromophores

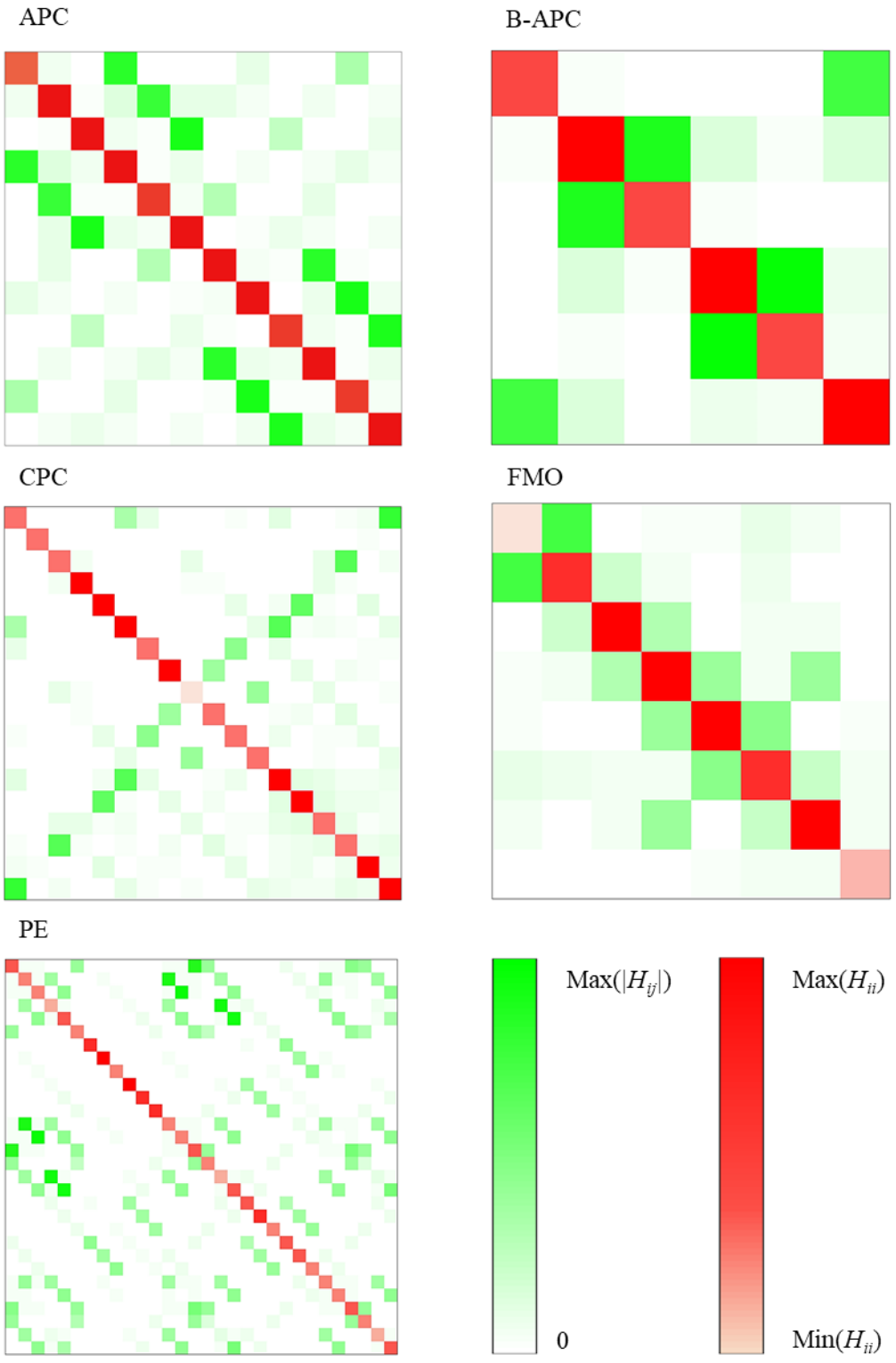


Figure 5.2 Colour scale (different for diagonal and off-diagonal) representation of magnitude of matrix elements of the five complexes. The range of values for each complex can be inferred from the parameters in Table 5.2.

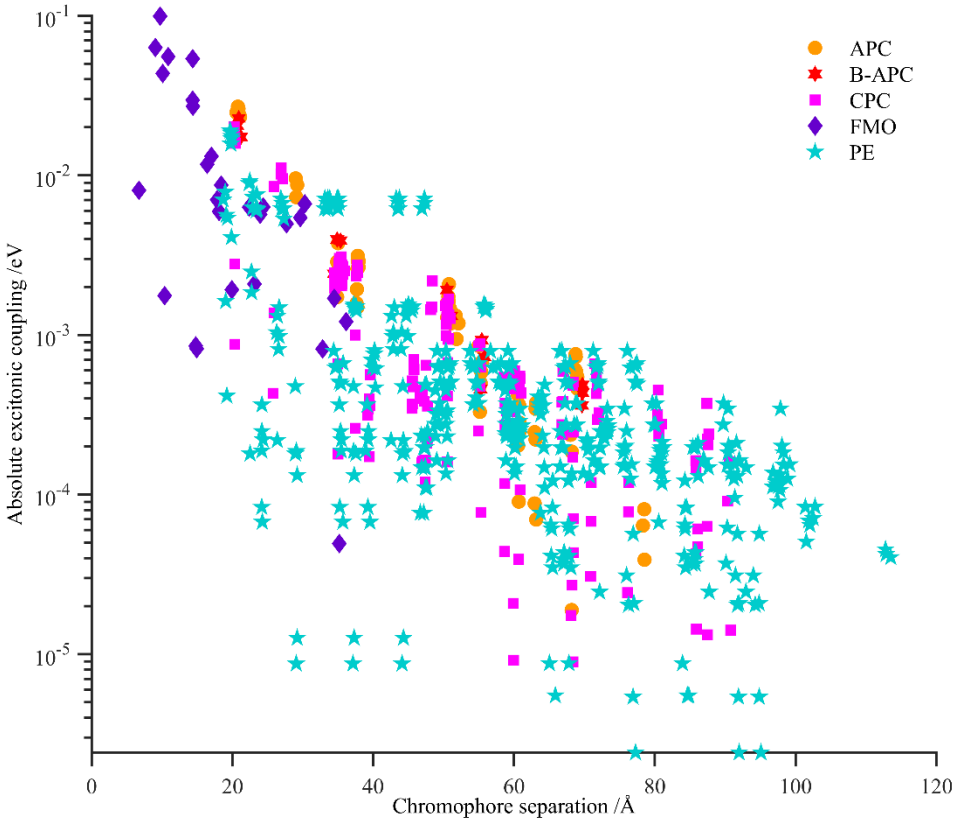


Figure 5.3 Excitonic coupling versus chromophore separation.

can be identified from the complex structures. For example, in FMO there is a chromophore at the edge of the protein, outside of the scaffolding. However, it is not necessarily astute to ascribe substantial significance to these findings as LHCs form larger aggregates and these peripheral chromophores likely couple to chromophores of neighbouring proteins in the aggregate. It is sufficient to bear in mind that such chromophores are a minority and thus should not influence the description of the global dynamics.

5.4.2 A global view of exciton dynamics in LHCs

Even considering the major simplifications of the treatment of the system-bath interaction, the exciton dynamics of the five LHCs is highly complex. The generation of the exciton may come from radiation from the ground state or by transfer from other chromophores or relaxation from higher excited states. There is also the matter of determining when the exciton has reached its "destination" – it is either transferred to the reaction centre complex or onto another LHC towards the reaction centre. In some cases (e.g. FMO) it is possible to determine a site for the "source" and "sink" of the energy transfer however in general this is not known.

For example, Fig. 5.4(a) depicts the difference in the evolution of the diagonal elements of the density matrix for B-APC when the population is initially on chromophore 1 compared to chromophore 5. So, to describe the exciton dynamics in a global fashion, a measure of exciton diffusion independent of final destination is introduced, which is able to quantify the ability to explore more sites from the initial generation of exciton. The time dependent inverse participation ratio[176] (IPR) is defined from the literature of disordered systems by

$$\mathrm{IPR}(t) = \frac{1}{\sum_i \rho_{ii}(t)^2}. \tag{5.8}$$

When the density matrix is expressed in the basis of the localised excitation, the IPR is a number between 1 (where the state is fully localised on a single chromophore) and the number of chromophores, *N*, (where the state is fully delocalised over all chromophores). In Fig. 5.4(b) it can be seen how examining IPR with a log scale for time can be used to monitor the exciton spread with better consideration of the short- and long-term dynamics. To examine the exciton diffusion in a given LHC regardless of the initial excitation site, an average IPR(*t*) is defined as the average over all dynamics where the initial population is one on a single chromophore and all other matrix elements are null. It is then possible to compare all LHCs consistently by examining their average IPR(*t*) which is depicted in Fig. 5.5. Note the dynamics does not include thermal relaxation effects and so is representative of the exciton diffusion before thermal relaxation occurs and assumes broadband excitation, i.e. initial exciton generation is non-specific. $\mathrm{IPR}(0) = 1$ following the definition of initial state given; at long times the average IPR(*t*) oscillates around the average of the IPR of the eigenstates, which depends on the (diagonal and off-diagonal) disorder of the system. The average IPR of the eigenstates is reported in Table 5.3.

Table 5.3 IPR values of the LHCs.

LHC	IPR
APC	3.21
B-APC	2.04
CPC	1.78
FMO	1.85
PE	2.33

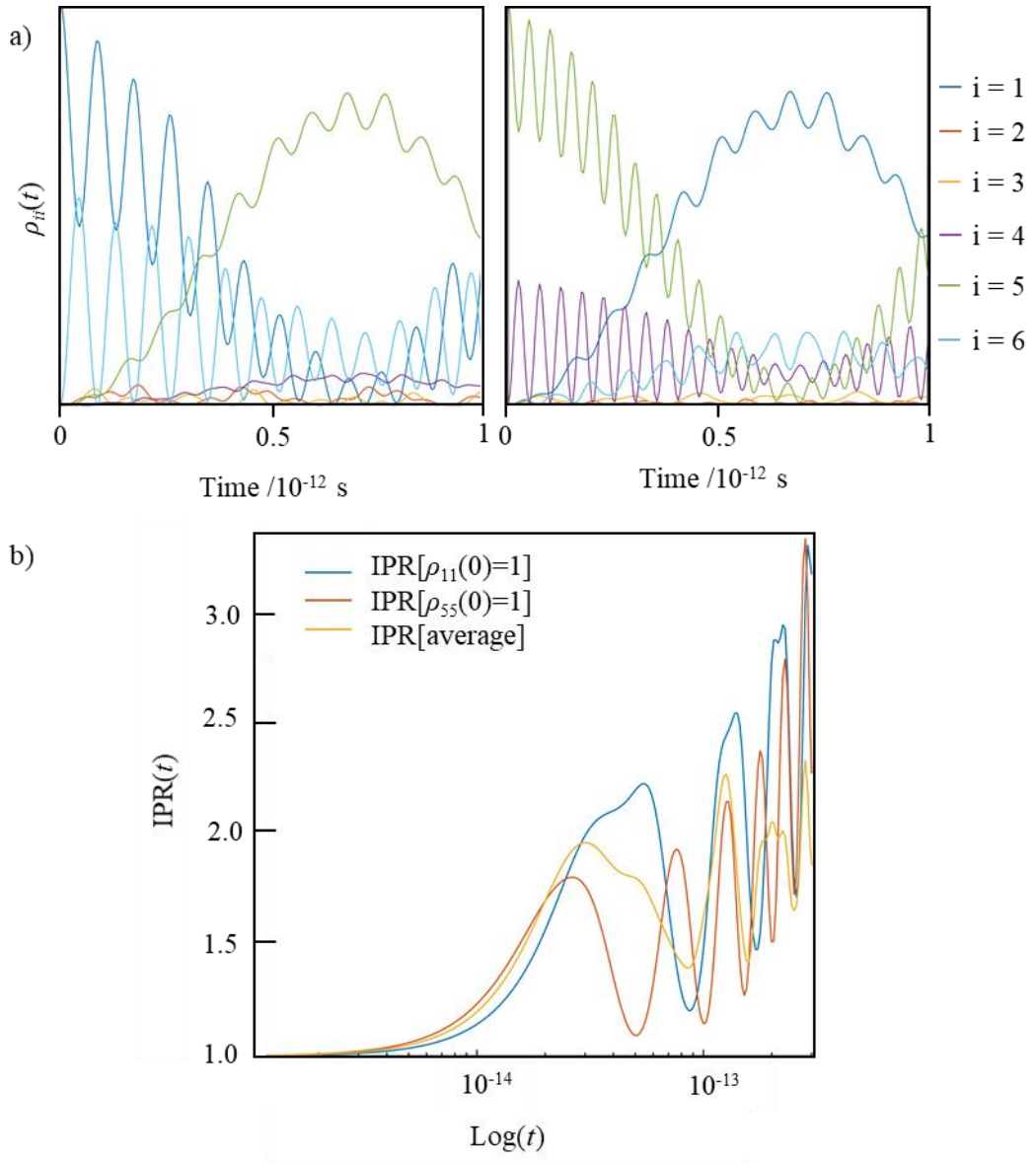


Figure 5.4 a) Time evolution of populations for B-APC with different initial conditions: population on chromophore 1 (left) and population on chromophore 5 (right). b) IPR(t) for initial conditions of the corresponding top panels and average IPR(t) over the six initial conditions.

Their appears to be a slim difference in the dynamics of APC, B-APC, CPC and PE; on a similar timescale the exciton spreads over a similar number of chromophores. FMO deviates from this trend marginally – the larger coupling (shown in Fig. 5.3 and Table 5.2) results in an initial increase in IPR that is about five times faster and the IPR is generally higher at long times due to more delocalised eigenstates.

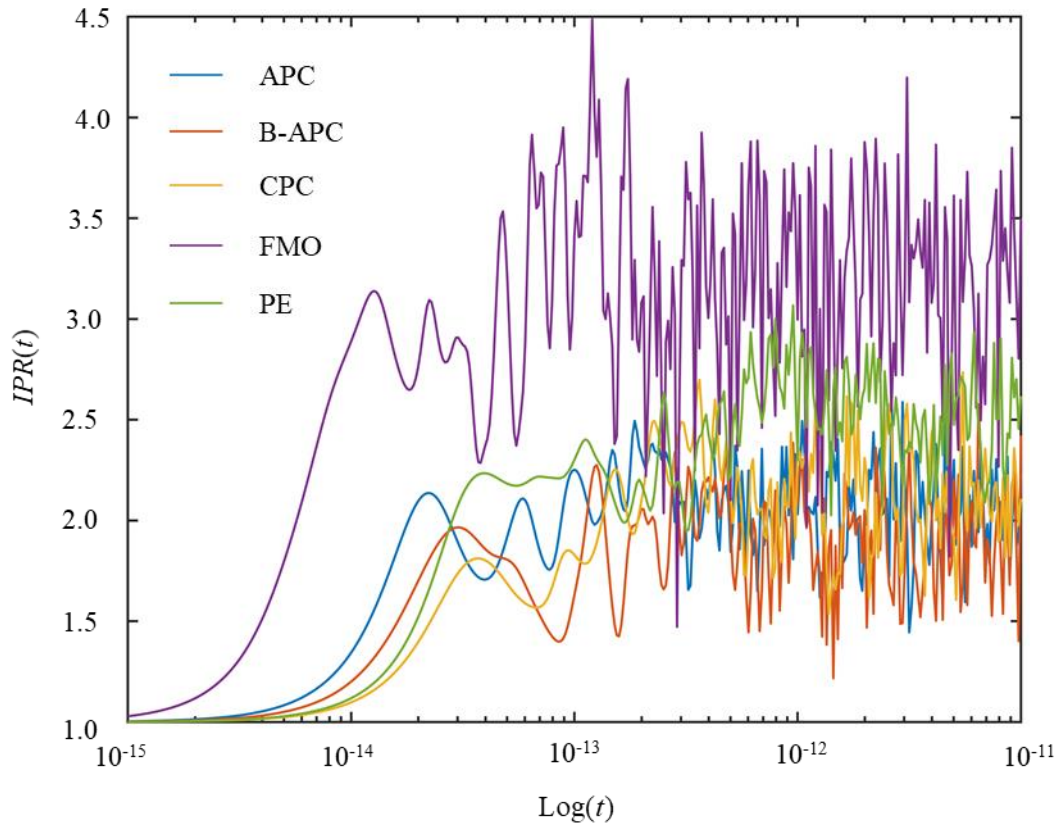


Figure 5.5 Average IPR(t) versus log(t) for the five LHCs.

5.4.3 Attempting to define a "generic" LHC

The modest difference in the resulting dynamics of this group of LHCs suggest it may be possible to define a generic model which describes these chromophore aggregates. This generic model can be used to capture the essence of LHCs and to differentiate between intrinsic features of aggregates which are geometrically similar and specialised functions. Such a model can take the form of the excitonic Hamiltonian: equation (5.4). For different LHCs the on-site disorder range is less varied than that of the observed excitonic coupling. Thus, the on-site energy disorder is set to a fixed value and the excitonic coupling strength is allowed to vary. The diagonal matrix elements ε_i are considered a random variable normally distributed with a variance of 50 meV.

As there appears to be no significant differences resulting from the number of chromophores in the aggregate, a "generic" LHC can be defined with a fixed intermediate size. The excitonic coupling must be generated with a plausible distribution, i.e. coupling decreases with distance increase, and a physically motivated relative sign. If each

chromophore(/site) in the aggregate is assigned a randomly oriented unit vector $\mathbf{u}_i$, the excitonic coupling between site *i* and site *j* is

$$V_{ij} = w\frac{\mathbf{u}_i\mathbf{u}_j}{R_{ij}} \tag{5.9}$$

where R_{ij} is the distance between sites and *w* is a parameter to be varied to match the range of observed dynamics. The dipole – dipole approximation is qualitatively valid for chromophores separated by sufficient distance,[174] and so is the basis of the distance – orientation dependence of the excitonic coupling.

A 13-site aggregate can be arranged in two different clusters: a 3D cluster based on the face-centred cubic lattice and a 2D cluster based on a hexagonal lattice, where distance between the unit vectors of the lattices are set to unity. Comparison of the resulting IPR(*t*) of the clusters as *w* is varied[177] showed IPR reaches larger values in the 3D cluster than in the 2D cluster with similar values of *w*. This result is expected as it mirrors 2 and 3 dimensional Anderson localisation theory[178] in solid state. The results also found the IPR(*t*) for different *w* values were non-crossing meaning at long times the average dynamics are determined solely by the average IPR, which is dependent on the disorder of the system. A larger and faster spread of the wavefunction is caused by decreased disorder. At large values of *w* disorder and IPR saturate. When *w* is large the only relevant source of disorder is that of the coupling V_{ij} (i.e. the off-diagonal matrix elements) resulting from the orientation of the unit vectors.

The results of this generic model indicate a possible explanation for the similarity in observed dynamics of APC, B-APC, CPC and PE: the structures are approximately planar with comparable levels of diagonal and off-diagonal disorder. The *w* parameter range of the 2D model with similar the dynamics to the realistic systems results in very limited change in dynamics as result of 50% increase or decrease of excitonic coupling.[177] In other words, the overall dynamics are insensitive to the detail of the coupling, as seen in these four LHCs. These results suggest the specific arrangement of chromophores in individual LHCs is not so important whilst indirectly highlighting the importance of the 3D arrangement. In nature, disk-shaped LHCs like PE surround the reaction centre in rods formed of stacked disks,[179] thus it is more likely that the special function is due to topology of interaction between LHCs as opposed to the relative coupling of chromophores within each LHC.

5.5 Conclusions

A consistent QM/MM methodology has been employed to compute the excitonic Hamiltonian for five LHCs and their dynamics in an attempt to characterise common or distinctive features. From examination of the Hamiltonian and its statistics, no unique features or properties were apparent, but the results may be useful in provision of a parameter range for the development of reduced models. Analysis of the exciton dynamics of the five complexes, using the *IPR*, demonstrated very similar characteristics for four of the five complexes (APC, B-APC, CPC and PE) with slightly faster dynamics for the fifth (FMO). This observation may be explained by relative compactness of FMO and thus larger exciton couplings between its chromophores.

The results of this study suggest relatively similar dynamics are followed by LHCs containing only one chromophore type, despite the structural differences. The comparatively faster dynamics of FMO perhaps indicates some benefit of more compact systems. Building a "generic" model based on these observations indicated the dynamics are ultimately determined by a single parameter, the degree of localisation of the system eigenstates which itself is governed by the static disorder of the system. The realistic parameters for the model span a relatively small range thus it follows that the resulting dynamics of the real systems are similar. This proposed model may be used as the basis for future study of larger scale (multi-aggregate) systems to examine more general features in biological light-harvesting. Finally, it should be noted again that the selection of systems examined here limit the results to complexes containing a single chromophore type and containing no carotenoids.

6 Conclusions and Outlook

This final chapter summarises the work covered in this thesis and the conclusions that can be drawn, as well as considering the outlook and potential further works.

6.1 Summary

The work presented in this thesis has examined three areas related to the endeavour of understanding the underlying principles of EET in LHCs. In chapter 3, a method for developing MDFFs was introduced which ensures consistency between MD computed forces and QC computed forces in order to address the computational chemistry problem of combining these two methods. The method was used to develop FFs for three chromophores, which were then used in spectral density calculations. Comparison of the spectral densities resulting from using the original and optimised FFs showed marked differences.

The biochemical question of the environmental role in EET was the subject of chapter 4. The spectral densities of rigid chromophores of PPCs with different functions were examined and the results identified two groups. One whose only contribution was from a very low frequency ($\omega < 10\ cm^{-1}$) and one with contributing frequencies capable of coupling with exciton dynamics in a complex way. There was no difference found between the PPCs with different functions and the environment component most affecting excitation energy fluctuations appeared to be solvent.

Finally, the focus of chapter 5 was the biophysical examination of the exciton dynamics of five LHCs, propagated using a Lindblad master equation. Utilising the IPR to analyse the dynamics of the systems revealed extremely similar dynamics for four of the five systems (all but FMO) and slightly faster dynamics in the fifth system. The results were used to define a "generic" model which indicated the dynamics are determined by the static disorder of the system which falls within a relatively small range for realistic systems. Comparison of 2D and 3D aggregates described utilising the model also indirectly highlighted the potential importance of the arrangement of LHC complexes in relation to one another whilst demonstrating the unimportance of specific chromophore arrangement.

6.2 Conclusions

The LHCs are an interesting area of study but their relatively large size and complexity result in computational challenges in their examination. The computational demands and sometimes the human time required to investigate these complexes often limits the number of complexes in a study to one or two. This is an issue if the goal is to understand underlying commonalities between complexes as it limits the number of comparable results and in turn their generalisability. Additionally, for the same reasons (system size and complexity) it is necessary to make approximations which then require validation. The force-matching method introduced in chapter 3 addressed the error resulting from using classical MM generated inputs for QC calculations (approximation problem) and the time usually required to produce a FF which prevents this error (computational demand problem).

After initial generation of the FF and the computation of the two levels of forces are complete, the method is automatic and so the parameterisation of the FF requires no human input. Additionally, the method is not limited to a specific FF or QC method, consequently it is possible to generate FFs for many chromophores for use in any sensible MD/QC combination. Spectral densities of three chromophores (BCL, CLA and CLB) in their protein environments computed with of non-optimised FFs and with FFs optimised with the force-matching method were compared. There were notable differences in the high-frequency regions of the spectral densities which relate to the intramolecular modes.

Part of the complexity of the LHCs is due to the fact that they are open quantum systems. The chromophores transporting the excitation energy are not a closed system but in fact are surrounded by a protein scaffold which may influence the dynamics. The question of what role the environment plays was the subject of chapter 4. In order to examine only the contribution of the environment to the spectral density the chromophore was kept rigid and PPCs with two different functions were examined to uncover any specialisation of the environment. The spectral densities were not averaged over multiple trajectories as the exciton dynamics of the system will be dictated by the specific spectral density within a limited time interval as opposed to an averaged one. Furthermore, it would be expected that any environmental motions/interactions specifically optimised to influence EET would occur in any sampled trajectory.

The two groups discovered from the analysis of the spectral densities of the chromophores contained chromophores from proteins of both function types. Additionally, no statistically significant difference was found between the excitation energy fluctuation of the two proteins,

nor for the two groups. Indicating there is no specific tuning of environment for the promotion of energy transfer, despite the potential influence of fluctuations on energy level order (for benefit or detriment) due to the comparability of magnitude of fluctuations and energy difference of excitons.

From analysing the coulombic interaction energies of the residues with the chromophore, the most strongly correlated residues were found to be predominantly solvent molecules. Thus, the main influence on the excitation energy was determined to be due to the dipole-dipole interactions resulting from the rotation of these solvent molecules. The formation of solvent networks with strong interactions involving hydrogen bonds, keep the solvent molecule in the chromophore proximity whilst allowing it to rotate, allowing continuous modification of the extent of interaction.

Considering the results altogether – chromophores of all proteins belonging to both groups, the lack of distinct difference between excitation energy fluctuations of different proteins, solvent being the residue most influencing fluctuations – the protein environment can be deemed unspecific in nature and it may be possible to determine a general value for any protein. Overall controlling the chromophore-environment is likely to have little effect on the exciton dynamics.

In chapter 5 the excitonic Hamiltonian and resulting exciton dynamics were computed for five LHCs. The results revealed no specific features in the Hamiltonian and its statistics. However, analysing the dynamics using the IPR revealed four of the five complexes had exceptionally similarities and the final complex had only marginally faster dynamics. The complex with faster dynamics has a more compact structure, indicating this may be beneficial somehow.

Constructing a “generic” LHC based on the results showed the dynamics are determined by the degree of localisation of the eigenstates which is controlled by the sources of static disorder of the system. The four complexes with very similar dynamics have very similar levels of static disorder hence such similar dynamics despite relatively different chromophore arrangement. Therefore, the specific arrangement of chromophores within an LHC is likely not as critical as previously thought, whilst the arrangement of LHCs with respect to each other and the reaction centre is more likely to be important.

Ultimately, to understand the underlying principles of efficient EET in LHCs it is clear that firstly, it is important to reduce computational and time demands to more easily study

multiple systems, whilst being cognizant of the potential limitations of any approximations. Next, the local protein environment likely has little effect on exciton dynamics and therefore has no role in efficient EET. Most importantly, it is the static disorder that determine dynamics however as this property appears to be relatively consistent across LHCs regardless of internal chromophore arrangement, it is likely that EET efficiency in biological light-harvesting is influenced by the positions of the LHCs in relation to one another and the reaction centre.

6.3 Outlook and future work

There are a number of pursuable ideas for future work for the topics presented here. The implementation of the force matching procedure could be improved to be more efficient by utilising a central difference in the gradient calculation to reduce errors and using a quadratically convergent second order optimisation method such as a Newton-CG method to reduce the required number of iterations. The procedure could be further improved by introducing more automation. The algorithm for assigning atom-type based on structural similarity could be implemented automatically rather than being done by hand. Additionally, the procedure could also monitor the error in forces per atom and suggest reassignment or automatically reassign atom types in cases where there remains a large error.

In order to solidify the conclusion that the protein environment is not specialised in some way to promote efficient EET, it would be beneficial to examine the spectral densities of rigid chromophores of more PPCs. It could also be interesting to repeat the work employing a QM/MM scheme which includes polarizability of the protein environment, as this can have significant effects in processes like EET. It would also be interesting to expand the sample size of the work undertaken in chapter 5; utilising a similar approach to investigate multi-chromophore systems, i.e. those containing more than one type of chromophore or also containing carotenoids, to determine if they have a similar dependence on static disorder. The most important next step would be to examine dynamics of the LHC aggregates, possibly also including the reaction centre.

7 Bibliography

(1) Engelken, J.; Funk, C.; Constance, D. Functional Genomics and Evolution of Photosynthetic Systems; 2012; Vol. 33, pp 265–284.

(2) Fenna, R. E.; Matthews, B. W.; Olson, J. M.; Shaw, E. K. Structure of a Bacteriochlorophyll-Protein from the Green Photosynthetic Bacterium Chlorobium Limicola: Crystallographic Evidence for a Trimer. *J. Mol. Biol.* **1974**, *84*.

(3) Engel, G. S.; Calhoun, T. R.; Read, E. L.; Ahn, T. K.; Mančal, T.; Cheng, Y. C.; Blankenship, R. E.; Fleming, G. R. Evidence for Wavelike Energy Transfer through Quantum Coherence in Photosynthetic Systems. *Nature* **2007**, *446*, 782–786.

(4) Panitchayangkoon, G.; Hayes, D.; Fransted, K. A.; Caram, J. R.; Harel, E.; Wen, J.; Blankenship, R. E.; Engel, G. S. Long-Lived Quantum Coherence in Photosynthetic Complexes at Physiological Temperature. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12766–12770.

(5) Calhoun, T. R.; Ginsberg, N. S.; Schlau-Cohen, G. S.; Cheng, Y. C.; Ballottari, M.; Bassi, R.; Fleming, G. R. Quantum Coherence Enabled Determination of the Energy Landscape in Light-Harvesting Complex II. *J. Phys. Chem. B* **2009**, *113*, 16291–16295.

(6) Collini, E.; Wong, C. Y.; Wilk, K. E.; Curmi, P. M. G. G.; Brumer, P.; Scholes, G. D. Coherently Wired Light-Harvesting in Photosynthetic Marine Algae at Ambient Temperature. *Nature* **2010**, *463*, 644–647.

(7) Duan, H.-G.; Prokhorenko, V. I.; Cogdell, R. J.; Ashraf, K.; Stevens, A. L.; Thorwart, M.; Miller, R. J. D. Nature Does Not Rely on Long-Lived Electronic Quantum Coherence for Photosynthetic Energy Transfer. *Proc. Natl. Acad. Sci.* **2017**, *114*, 8493–8498.

(8) Blankenship, R. E. *Molecular Mechanisms of Photosynthesis*; Blackwell Science Ltd: Oxford, UK, 2002.

(9) Jursinic, P. Temperature Dependence of Delayed Light Emission in the 6 to 340 Microsecond Range After a Single Flash in Chloroplasts. *Photochem. Photobiol.* **1977**, *26*, 617–628.

(10) Wientjes, E.; Van Amerongen, H.; Croce, R. Quantum Yield of Charge Separation in Photosystem II: Functional Effect of Changes in the Antenna Size upon Light Acclimation. *J. Phys. Chem. B* **2013**, *117*, 11200–11208.

(11) Scholes, G. D.; Fleming, G. R.; Olaya-Castro, A.; van Grondelle, R. Lessons from Nature about Solar Light Harvesting. *Nat. Chem.* **2011**, *3*, 763–774.

(12) Beddard, G. S.; Porter, G. Concentration Quenching in Chlorophyll. *Nature* **1976**, *260*, 366–367.

(13) Chenu, A.; Scholes, G. D. Coherence in Energy Transfer and Photosynthesis. *Annu. Rev. Phys. Chem.* **2015**, *66*, 69–96.

(14) Jang, S.; Cheng, Y. C. Resonance Energy Flow Dynamics of Coherently Delocalized Excitons in Biological and Macromolecular Systems: Recent Theoretical Advances and Open Issues. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. John Wiley & Sons, Inc. January 1, 2013, pp 84–104.

(15) Ishizaki, A.; Fleming, G. R. Quantum Coherence in Photosynthetic Light Harvesting.

Annu. Rev. Condens. Matter Phys. **2012**, *3*, 333–361.

(16) Curutchet, C.; Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem. Rev.* **2017**, *117*, 294–343.

(17) Fassioli, F.; Dinshaw, R.; Arpin, P. C.; Scholes, G. D. Photosynthetic Light Harvesting: Excitons and Coherence. *J. R. Soc. Interface* **2014**, *11*, 20130901.

(18) Förster, T. Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Ann. Phys.* **1948**, *437*, 55–75.

(19) Redfield, A. G. The Theory of Relaxation Processes. In *Advances in Magnetic and Optical Resonance*; Academic Press, 1965; Vol. 1, pp 1–32.

(20) Novoderezhkin, V. I.; Van Grondelle, R. Physical Origins and Models of Energy Transfer in Photosynthetic Light-Harvesting. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7352–7365.

(21) Ishizaki, A.; Fleming, G. R. Theoretical Examination of Quantum Coherence in a Photosynthetic System at Physiological Temperature. *Proc. Natl. Acad. Sci.* **2009**, *106*, 17255–17260.

(22) Ishizaki, A.; Fleming, G. R. On the Adequacy of the Redfield Equation and Related Approaches to the Study of Quantum Dynamics in Electronic Energy Transfer. *J. Chem. Phys.* **2009**, *130*, 234110.

(23) Zhang, W. M.; Meier, T.; Chernyak, V.; Mukamel, S. Exciton-Migration and Three-Pulse Femtosecond Optical Spectroscopies of Photosynthetic Antenna Complexes. *J. Chem. Phys.* **1998**, *108*, 7763–7774.

(24) Yang, M.; Fleming, G. R. Influence of Phonons on Exciton Transfer Dynamics: Comparison of the Redfield, Förster, and Modified Redfield Equations. *Chem. Phys.* **2002**, *275*, 355–372.

(25) Dijkstra, A. G.; Tanimura, Y. The Role of the Environment Time Scale in Light-Harvesting Efficiency and Coherent Oscillations. *New J. Phys.* **2012**, *14*, 073027.

(26) Rebentrost, P.; Mohseni, M.; Aspuru-Guzik, A. Role of Quantum Coherence and Environmental Fluctuations in Chromophoric Energy Transport. *J. Phys. Chem. B* **2009**, *113*, 9942–9947.

(27) Tanimura, Y. Stochastic Liouville, Langevin, Fokker–Planck, and Master Equation Approaches to Quantum Dissipative Systems. *J. Phys. Soc. Japan* **2006**, *75*, 082001.

(28) Ishizaki, A.; Fleming, G. R. Unified Treatment of Quantum Coherent and Incoherent Hopping Dynamics in Electronic Energy Transfer: Reduced Hierarchy Equation Approach. *J. Chem. Phys.* **2009**, *130*, 234111.

(29) Kolli, A.; O'Reilly, E. J.; Scholes, G. D.; Olaya-Castro, A. The Fundamental Role of Quantized Vibrations in Coherent Light Harvesting by Cryptophyte Algae. *J. Chem. Phys.* **2012**, *137*, 174109–235102.

(30) Rolczynski, B. S.; Zheng, H.; Singh, V. P.; Navotnaya, P.; Ginzburg, A. R.; Caram, J. R.; Ashraf, K.; Gardiner, A. T.; Yeh, S.-H. H.; Kais, S.; et al. Correlated Protein Environments Drive Quantum Coherence Lifetimes in Photosynthetic Pigment-Protein Complexes. *Chem* **2018**, *4*, 138–149.

(31) Levi, F.; Mostarda, S.; Rao, F.; Mintert, F. Quantum Mechanics of Excitation Transport in Photosynthetic Complexes: A Key Issues Review. *Reports Prog. Phys.* **2015**, *78*, 082001.

(32) Jang, S. J.; Mennucci, B. Delocalized Excitons in Natural Light-Harvesting Complexes. *Rev. Mod. Phys.* **2018**, *90*.

(33) Kassal, I.; Yuen-Zhou, J.; Rahimi-Keshari, S. Does Coherence Enhance Transport in Photosynthesis? *J. Phys. Chem. Lett.* **2013**, *4*, 362–367.

(34) Editor, G.; Harriman, A.; Ishizaki, A.; Calhoun, T. R.; Schlau-Cohen, G. S.; Fleming, G. R.; Chem Chem, P.; Albinsson, B.; Mårtensson, J.; Chem, P.; et al. Physical Origins and Models of Energy Transfer in Photosynthetic Light-Harvesting. *Phys. Chem. Chem. Phys* **2010**.

(35) Scholes, G. D. Quantum-Coherent Electronic Energy Transfer: Did Nature Think of It First? *J. Phys. Chem. Lett.* **2010**, *1*, 2–8.

(36) Thyrhaug, E.; Tempelaar, R.; Alcocer, M. J. P.; Žídek, K.; Bína, D.; Knoester, J.; Jansen, T. L. C.; Zigmantas, D. Identification and Characterization of Diverse Coherences in the Fenna–Matthews–Olson Complex. *Nat. Chem.* **2018**, *10*, 780–786.

(37) Liu, Y. Y.; Yan, Y. M.; Xu, M.; Song, K.; Shi, Q. Exact Generator and Its High Order Expansions in Time-Convolutionless Generalized Master Equation: Applications to Spin-Boson Model and Excitation Energy Transfer. *Chinese J. Chem. Phys.* **2018**, *31*, 575–583.

(38) Womick, J. M.; Moran, A. M. Exciton Coherence and Energy Transport in the Light-Harvesting Dimers of Allophycocyanin. *J. Phys. Chem. B* **2009**, *113*, 15747–15759.

(39) Aghtar, M.; Strümpfer, J.; Olbrich, C.; Schulten, K.; Kleinekathöfer, U. Different Types of Vibrations Interacting with Electronic Excitations in Phycoerythrin 545 and Fenna-Matthews-Olson Antenna Systems. *J. Phys. Chem. Lett.* **2014**, *5*, 3131–3137.

(40) Renger, T.; Klinger, A.; Steinecker, F.; Schmidt am Busch, M.; Numata, J.; Müh, F. Normal Mode Analysis of the Spectral Density of the Fenna-Matthews-Olson Light-Harvesting Protein: How the Protein Dissipates the Excess Energy of Excitons. *J. Phys. Chem. B* **2012**, *116*, 14565–14580.

(41) Lee, M. K.; Coker, D. F. Modeling Electronic-Nuclear Interactions for Excitation Energy Transfer Processes in Light-Harvesting Complexes. *J. Phys. Chem. Lett.* **2016**, *7*, 3171–3178.

(42) Olbrich, C.; Strümpfer, J.; Schulten, K.; Kleinekathöfer, U. Theory and Simulation of the Environmental Effects on FMO Electronic Transitions. *J. Phys. Chem. Lett.* **2011**, *2*, 1771–1776.

(43) Viani, L.; Corbella, M.; Curutchet, C.; O'Reilly, E. J.; Olaya-Castro, A.; Mennucci, B. Molecular Basis of the Exciton-Phonon Interactions in the PE545 Light-Harvesting Complex. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16302–16311.

(44) Olbrich, C.; Kleinekathöfer, U. Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II. *J. Phys. Chem. B* **2010**, *114*, 12427–12437.

(45) Ishizaki, A.; Calhoun, T. R.; Schlau-Cohen, G. S.; Fleming, G. R. Quantum Coherence and Its Interplay with Protein Environments in Photosynthetic Electronic Energy Transfer. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7319.

(46) Kreisbeck, C.; Kramer, T. Long-Lived Electronic Coherence in Dissipative Exciton Dynamics of Light-Harvesting Complexes. *J. Phys. Chem. Lett.* **2012**, *3*, 2828–2833.

(47) Fuller, F. D.; Ogilvie, J. P. Experimental Implementations of Two-Dimensional Fourier Transform Electronic Spectroscopy. *Annu. Rev. Phys. Chem.* **2015**, *66*, 667–690.

(48) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. The Two-Dimensional IR Nonlinear Spectroscopy of a Cyclic Penta-Peptide in Relation to Its Three-Dimensional Structure. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 2036–2041.

(49) Ginsberg, N. S.; Cheng, Y.-C.; Fleming, G. R. Two-Dimensional Electronic Spectroscopy of Molecular Aggregates. *Acc. Chem. Res.* **2009**, *42*, 1352–1363.

(50) Roscioli, J. D.; Ghosh, S.; LaFountain, A. M.; Frank, H. A.; Beck, W. F. Structural Tuning of Quantum Decoherence and Coherent Energy Transfer in Photosynthetic Light Harvesting. *J. Phys. Chem. Lett.* **2018**, *9*, 5071–5077.

(51) Policht, V. R.; Niedringhaus, A.; Ogilvie, J. P. Characterization of Vibrational Coherence in Monomeric Bacteriochlorophyll a by Two-Dimensional Electronic Spectroscopy. *J. Phys. Chem. Lett.* **2018**, *9*, 6631–6637.

(52) Blau, S. M.; Bennett, D. I. G.; Kreisbeck, C.; Scholes, G. D.; Aspuru-Guzik, A. Local Protein Solvation Drives Direct Down-Conversion in Phycobiliprotein PC645 via Incoherent Vibronic Transport. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *115*, E3342–E3350.

(53) Knee, G. C.; Rowe, P.; Smith, L. D.; Troisi, A.; Datta, A. Structure-Dynamics Relation in Physically-Plausible Multi-Chromophore Systems. *J. Phys. Chem. Lett.* **2017**, *8*, 2328–2333.

(54) Goldberg, O.; Meir, Y.; Dubi, Y. Vibration-Assisted and Vibration-Hampered Excitonic Quantum Transport. *J. Phys. Chem. Lett.* **2018**, *9*, 3143–3148.

(55) Kim, C. W.; Lee, W. G.; Kim, I.; Rhee, Y. M. Effect of Underdamped Vibration on Excitation Energy Transfer: Direct Comparison between Two Different Partitioning Schemes. *J. Phys. Chem. A* **2019**, *123*, 1186–1197.

(56) Reppert, M.; Brumer, P. Quantumness in Light Harvesting Is Determined by Vibrational Dynamics. *J. Chem. Phys.* **2018**, *149*, 234102.

(57) Leng, X.; Yan, Y.-M. M.; Zhu, R.-D. D.; Song, K.; Weng, Y.-X. X.; Shi, Q. Simulation of the Two-Dimensional Electronic Spectroscopy and Energy Transfer Dynamics of Light-Harvesting Complex II at Ambient Temperature. *J. Phys. Chem. B* **2018**, *122*, 4642–4652.

(58) Khmelnitskiy, A.; Reinot, T.; Jankowiak, R. Impact of Single-Point Mutations on the Excitonic Structure and Dynamics in a Fenna-Matthews-Olson Complex. *J. Phys. Chem. Lett.* **2018**, *9*, 3378–3386.

(59) Jang, S. J. Robust and Fragile Quantum Effects in the Transfer Kinetics of Delocalized Excitons between B850 Units of LH2 Complexes. *J. Phys. Chem. Lett.* **2018**, *9*, 6576–6583.

(60) Hsieh, S.-T.; Zhang, L.; Ye, D.-W.; Huang, X.; Cheng, Y.-C. A Theoretical Study on the Dynamics of Light Harvesting in the Dimeric Photosystem II Core Complex: Regulation and Robustness of Energy Transfer Pathways. *Faraday Discuss.* **2019**.

(61) Juhász, I. B.; Csurgay, Á. I. Impact of Undamped and Damped Intramolecular Vibrations on the Efficiency of Photosynthetic Exciton Energy Transfer. *AIP Adv.* **2018**, *8*, 045318.

(62) Chen, H.; Wang, X.; Han, C.-M.; Li, H.-R. Phonon-Mediated Excitation Energy Transfer in a Detuned Multi-Sites System. *J. Phys. B At. Mol. Opt. Phys.* **2019**, *52*, 075501.

(63) Cupellini, L.; Caprasecca, S.; Guido, C. A.; Müh, F.; Renger, T.; Mennucci, B.

Coupling to Charge Transfer States Is the Key to Modulate the Optical Bands for Efficient Light Harvesting in Purple Bacteria. *J. Phys. Chem. Lett.* **2018**, *9*, 6892–6899.

(64) De Vico, L.; Anda, A.; Osipov, V. A.; Madsen, A. Ø.; Hansen, T. Macrocycle Ring Deformation as the Secondary Design Principle for Light-Harvesting Complexes. *Proc. Natl. Acad. Sci.* **2018**, *115*, E9051–E9057.

(65) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(66) Maseras, F.; Morokuma, K. IMOMM: A New Integratedab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

(67) Mats Svensson; Stéphane Humbel; Robert D. J. Froese; Toshiaki Matsubara; Stefan Sieber, and; Morokuma*, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. **1996**.

(68) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31*, 459–466.

(69) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(70) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.

(71) Mackerell, A. D.; Brooks, B.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *Encyclopedia of Computational Chemistry*; 1998.

(72) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(73) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(74) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

(75) Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.

(76) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

(77) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(78) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(79) Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional

Approximations for Many-Electron Systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(80) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.

(81) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(82) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(83) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(84) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(85) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(86) Ullrich, C. A. *Time-Dependent Density-Functional Theory : Concepts and Applications*; Oxford University Press, 2012.

(87) Gross, E. K. U.; Kohn, W. Local Density-Functional Theory of Frequency-Dependent Linear Response. *Phys. Rev. Lett.* **1985**, *55*, 2850–2852.

(88) CASIDA, M. E. Time-Dependent Density Functional Response Theory for Molecules; 1995; pp 155–192.

(89) Casida, M. E.; Gutierrez, F.; Guan, J.; Gadea, F.-X.; Salahub, D.; Daudey, J.-P. Charge-Transfer Correction for Improved Time-Dependent Local Density Approximation Excited-State Potential Energy Curves: Analysis within the Two-Level Model with Illustration for H2 and LiH. *J. Chem. Phys.* **2000**, *113*, 7062–7071.

(90) Schaller, G. Dynamics of Open Quantum Systems; Springer, Cham, 2014; pp 1–26.

(91) Nitzan, A. *Chemical Dynamics in Condensed Phases : Relaxation, Transfer and Reactions in Condensed Molecular Systems*; Oxford University Press, 2006.

(92) May, V.; Kühn, O. *Charge and Energy Transfer Dynamics in Molecular Systems*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, Germany, 2011.

(93) Kell, A.; Feng, X.; Reppert, M.; Jankowiak, R. On the Shape of the Phonon Spectral Density in Photosynthetic Complexes. *J. Phys. Chem. B* **2013**, *117*, 7317–7323.

(94) Skjaerven, L.; Hollup, S. M.; Reuter, N. Normal Mode Analysis for Proteins. *J. Mol. Struct. THEOCHEM* **2009**, *898*, 42–48.

(95) Dawlaty, J. M.; Ishizaki, A.; De, A. K.; Fleming, G. R. Microscopic Quantum Coherence in a Photosynthetic-Light-Harvesting Antenna. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2012**, *370*, 3672–3691.

(96) Jia, X.; Mei, Y.; Zhang, J. Z. H.; Mo, Y.; Haak, J. R. Hybrid QM/MM Study of FMO Complex with Polarized Protein-Specific Charge. *Sci. Rep.* **2015**, *5*, 17096.

(97) Shim, S.; Rebentrost, P.; Valleau, S.; Aspuru-Guzik, A. Atomistic Study of the Long-Lived Quantum Coherences in the Fenna-Matthews-Olson Complex. *Biophys. J.* **2012**, *102*, 649–660.

(98) Anda, A.; De Vico, L.; Hansen, T. Intermolecular Modes between LH2 Bacteriochlorophylls and Protein Residues: The Effect on the Excitation Energies. *J. Phys. Chem. B* **2017**, *121*, 5499–5508.

(99) Kim, C. W.; Choi, B.; Rhee, Y. M. Excited State Energy Fluctuations in the Fenna-Matthews-Olson Complex from Molecular Dynamics Simulations with Interpolated Chromophore Potentials. *Phys. Chem. Chem. Phys.* **2018**, *20*, 3310–3319.

(100) Rosnik, A. M.; Curutchet, C. Theoretical Characterization of the Spectral Density of the Water-Soluble Chlorophyll-Binding Protein from Combined Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 5826–5837.

(101) Chandrasekaran, S.; Aghtar, M.; Valleau, S.; Aspuru-Guzik, A.; Kleinekathöfer, U. Influence of Force Fields and Quantum Chemistry Approach on Spectral Densities of BChl a in Solution and in FMO Proteins. *J. Phys. Chem. B* **2015**, *119*, 9995–10004.

(102) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*.

(103) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(104) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556.

(105) Burger, S. K.; Lacasse, M.; Verstraelen, T.; Drewry, J.; Gunning, P.; Ayers, P. W. Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds. *J. Chem. Theory Comput.* **2012**, *8*, 554–562.

(106) Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.

(107) Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. An Automated Approach for the Parameterization of Accurate Intermolecular Force-Fields: Pyridine as a Case Study. *J. Comput. Chem.* **2012**, *33*, 1055–1067.

(108) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

(109) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.

(110) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039.

(111) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770.

(112) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from *Ab Initio* Input. *J. Comput. Chem.* **2015**, *36*, 1015–1027.

(113) Ceccarelli, M.; Procacci, P.; Marchi, M. An Ab Initio Force Field for the Cofactors of Bacterial Photosynthesis. *J. Comput. Chem.* **2003**, *24*, 129–142.

(114) Palencar, P.; Vacha, F.; Kuty, M. Force Field Development on Pigments of Photosystem 2 Reaction Centre. *Photosynthetica* **2005**, *43*, 417–420.

(115) Markovich, T.; Blau, S. M.; Parkhill, J.; Kreisbeck, C.; Sanders, J. N.; Andrade, X.; Aspuru-Guzik, A. Accelerating the Computation of Bath Spectral Densities with Super-Resolution. *Theor. Chem. Acc.* **2016**, *135*, 1–8.

(116) Zhang, L.; Silva, D.-A.; Yan, Y.; Huang, X. Force Field Development for Cofactors in the Photosystem II. *J. Comput. Chem.* **2012**, *33*, 1969–1980.

(117) Prandi, I. G.; Viani, L.; Andreussi, O.; Mennucci, B. Combining Classical Molecular Dynamics and Quantum Mechanical Methods for the Description of Electronic Excitations: The Case of Carotenoids. *J. Comput. Chem.* **2016**, *37*, 981–991.

(118) Andreussi, O.; Prandi, I. G.; Campetella, M.; Prampolini, G.; Mennucci, B. Classical Force Fields Tailored for QM Applications: Is It Really a Feasible Strategy? *J. Chem. Theory Comput.* **2017**, *13*, 4636–4648.

(119) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Eur. Lett* **1994**, *26*, 583–588.

(120) Cheng, Y. Q.; Sheng, H. W.; Ma, E. Relationship between Structure, Dynamics, and Mechanical Properties in Metallic Glass-Forming Alloys. *Phys. Rev. B* **2008**, *78*, 014207.

(121) Liu, X.-Y.; Adams, J. B.; Ercolessi, F.; Moriarty, J. A. EAM Potential for Magnesium from Quantum Mechanical Forces. *Model. Simul. Mater. Sci. Eng.* **1996**, *4*, 293–303.

(122) Sala, J.; Guàrdia, E.; Masia, M. Improving the Force Matching Algorithm: Application to a Simple Point Charge Flexible Model of Water. *Comput. Phys. Commun.* **2011**, *182*, 1954–1957.

(123) Do, H.; Troisi, A. Developing Accurate Molecular Mechanics Force Fields for Conjugated Molecular Systems. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25123–25132.

(124) Kroonblawd, M. P.; Pietrucci, F.; Saitta, A. M.; Goldman, N. Generating Converged Accurate Free Energy Surfaces for Chemical Reactions with a Force-Matched Semiempirical Model. *J. Chem. Theory Comput.* **2018**, *14*, 2207–2218.

(125) Huang, I. S.; Tsai, M. K. Interplay between Polarizability and Hydrogen Bond Network of Water: Reparametrizing the Flexible Single-Point-Charge Water Model by the Nonlinear Adaptive Force Matching Approach. *J. Phys. Chem. A* **2018**, *122*, 4654–4662.

(126) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. Automated Parametrization of Biomolecular Force Fields from Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations through Force Matching. *J. Chem. Theory Comput.* **2007**, *3*, 628–639.

(127) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(128) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.

(129) Evans, T. A.; Katz, J. J. Evidence for 5- and 6-Coordinated Magnesium in

Bacteriochlorophyll a from Visible Absorption Spectroscopy. *BBA - Bioenerg.* **1975**, *396*, 414–426.

(130) Katz, J. J. Coordination Properties of Magnesium In Chlorophyll from IR and NMR Spectra *. In *Developments in applied spectroscopy*; Springer US: Boston, MA, 1968; pp 201–218.

(131) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(132) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(133) Qian, N. On the Momentum Term in Gradient Descent Learning Algorithms. *Neural Networks* **1999**, *12*, 145–151.

(134) NESTEROV; E., Y. A Method for Solving the Convex Programming Problem with Convergence Rate O(1/K^2). *Dokl. Akad. Nauk SSSR* **1983**, *269*, 543–547.

(135) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 1991.

(136) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. Molecular Excitation Energies to High-Lying Bound States from Time-Dependent Density-Functional Response Theory: Characterization and Correction of the Time-Dependent Local Density Approximation Ionization Threshold. *J. Chem. Phys.* **1998**, *108*, 4439.

(137) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184–215.

(138) Claridge, K.; Padula, D.; Troisi, A. How Fine-Tuned for Energy Transfer Is the Environmental Noise Produced by Proteins around Biological Chromophores? *Phys. Chem. Chem. Phys.* **2018**.

(139) Lee, M. H.; Troisi, A. Vibronic Enhancement of Excitation Energy Transport: Interplay between Local and Non-Local Exciton-Phonon Interactions. *J. Chem. Phys.* **2017**, *146*, 174703–174704.

(140) Padula, D.; Lee, M. H.; Claridge, K.; Troisi, A. Chromophore-Dependent Intramolecular Exciton−Vibrational Coupling in the FMO Complex: Quantification and Importance for Exciton Dynamics. *J. Phys. Chem. B* **2017**, *121*, 10026–10035.

(141) Höfinger, S.; Simonson, T. Dielectric Relaxation in Proteins: A Continuum Electrostatics Model Incorporating Dielectric Heterogeneity of the Protein and Time-Dependent Charges. *J. Comput. Chem.* **2001**, *22*, 290–305.

(142) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598–1603.

(143) Vaughan, F.; Linden, N.; Manby, F. R. How Markovian Is Exciton Dynamics in Purple Bacteria? *J. Chem. Phys.* **2017**, *146*, 124113.

(144) Chenu, A.; Christensson, N.; Kauffmann, H. F.; Mančal, T. Enhancement of Vibronic and Ground-State Vibrational Coherences in 2D Spectra of Photosynthetic Complexes. *Sci. Rep.* **2013**, *3*, 2029.

(145) Malý, P.; Somsen, O. J. G.; Novoderezhkin, V. I.; Mančal, T.; van Grondelle, R. The Role of Resonant Vibrations in Electronic Energy Transfer. *ChemPhysChem* **2016**, *17*, 1356–1368.

(146) Kim, C. W.; Park, J. W.; Rhee, Y. M. Effect of Chromophore Potential Model on the Description of Exciton-Phonon Interactions. *J. Phys. Chem. Lett.* **2015**, *6*, 2875–2880.

(147) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(148) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(149) Damjanovic, A.; Kosztin, I.; Schulten, K.; Kleinekathöfer, U. Excitons in a Photosynthetic Light-Harvesting System: A Combined Molecular Dynamics, Quantum Chemistry, and Polaron Model Study. *Phys. Rev. E* **2002**, *65*, 31919.

(150) Guerra, F.; Adam, S.; Bondar, A.-N. N. Revised Force-Field Parameters for Chlorophyll-a, Pheophytin-a and Plastoquinone-9. *J. Mol. Graph. Model.* **2015**, *58*, 30–39.

(151) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184–215.

(152) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621.

(153) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.

(154) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431–439.

(155) Zwanzig, R. Rate Processes with Dynamical Disorder. *Acc. Chem. Res.* **1990**, *23*, 148–152.

(156) Baker, L. A.; Habershon, S. Robustness, Efficiency, and Optimality in the Fenna-Matthews-Olson Photosynthetic Pigment-Protein Complex. *J. Chem. Phys.* **2015**, *143*, 105101.

(157) Perlík, V.; Seibt, J.; Cranston, L. J.; Cogdell, R. J.; Lincoln, C. N.; Savolainen, J.; Šanda, F.; Mančal, T.; Hauer, J. Vibronic Coupling Explains the Ultrafast Carotenoid-to-Bacteriochlorophyll Energy Transfer in Natural and Artificial Light Harvesters. *J. Chem. Phys.* **2015**, *142*, 34311.

(158) Van Der Vegte, C. P.; Prajapati, J. D.; Kleinekathö, U.; Knoester, J.; Jansen, T. L. C. Atomistic Modeling of Two-Dimensional Electronic Spectra and Excited-State Dynamics for a Light Harvesting 2 Complex. *J. Phys. Chem. B* **2015**, *119*, 12.

(159) Mostarda, S.; Levi, F.; Prada-Gracia, D.; Mintert, F.; Rao, F. Structure–Dynamics Relationship in Coherent Transport through Disordered Systems. *Nat. Commun.* **2013**, *4*, 2296.

(160) Mohseni, M.; Shabani, A.; Lloyd, S.; Omar, Y.; Rabitz, H. Geometrical Effects on Energy Transfer in Disordered Open Quantum Systems. *J. Chem. Phys.* **2013**, *138*, 204309.

(161) Baghbanzadeh, S.; Kassal, I. Geometry, Supertransfer, and Optimality in the Light

Harvesting of Purple Bacteria. *J. Phys. Chem. Lett.* **2016**, *7*, 3804–3811.

(162) Magdaong, N. M.; LaFountain, A. M.; Greco, J. A.; Gardiner, A. T.; Carey, A.-M.; Cogdell, R. J.; Gibson, G. N.; Birge, R. R.; Frank, H. A. High Efficiency Light Harvesting by Carotenoids in the LH2 Complex from Photosynthetic Bacteria: Unique Adaptation to Growth under Low-Light Conditions. *J. Phys. Chem. B* **2014**, *118*, 11172–11189.

(163) Liguori, N.; Xu, P.; van Stokkum, I. H. M.; van Oort, B.; Lu, Y.; Karcher, D.; Bock, R.; Croce, R. Different Carotenoid Conformations Have Distinct Functions in Light-Harvesting Regulation in Plants. *Nat. Commun.* **2017**, *8*, 1994.

(164) Durchan, M.; Tichý, J.; Litvín, R.; Šlouf, V.; Gardian, Z.; Hříbek, P.; Vácha, F.; Polívka, T. Role of Carotenoids in Light-Harvesting Processes in an Antenna Protein from the Chromophyte Xanthonema Debile. *J. Phys. Chem. B* **2012**, *116*, 8880–8889.

(165) Kloz, M.; Pillai, S.; Kodis, G.; Gust, D.; Moore, T. A.; Moore, A. L.; van Grondelle, R.; Kennis, J. T. M. Carotenoid Photoprotection in Artificial Photosynthetic Antennas. *J. Am. Chem. Soc.* **2011**, *133*, 7007–7015.

(166) Kodis, G.; Herrero, C.; Palacios, R.; Mariño-Ochoa, E.; Gould, S.; de la Garza, L.; van Grondelle, R.; Gust, D.; Moore, T. A.; Moore, A. L.; et al. Light Harvesting and Photoprotective Functions of Carotenoids in Compact Artificial Photosynthetic Antenna Designs. *J. Phys. Chem. B* **2004**, *108*, 414–425.

(167) Claridge, K.; Troisi, A. Developing Consistent Molecular Dynamics Force Fields for Biological Chromophores via Force Matching. *J. Phys. Chem. B* **2019**, *123*, 428–438.

(168) Vanommeslaeghe, K.; MacKerell, A. D.; Vanommeslaegh, K.; MacKerell Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(169) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(170) Hsu, C.-P. The Electronic Couplings in Electron Transfer and Excitation Energy Transfer. *Acc. Chem. Res.* **2009**, *42*, 509–518.

(171) Chang, J. C. Monopole Effects on Electronic Excitation Interactions between Large Molecules. I. Application to Energy Transfer in Chlorophylls. *J. Chem. Phys.* **1999**, *67*, 3901.

(172) Brent P. Krueger; Gregory D. Scholes, and; Fleming*, G. R. Calculation of Couplings and Energy-Transfer Pathways between the Pigments of LH2 by the Ab Initio Transition Density Cube Method. **1998**.

(173) Magnasco, V. *Methods of Molecular Quantum Mechanics*; John Wiley & Sons, Ltd: Chichester, UK, 2009.

(174) Fornari, R. P.; Rowe, P.; Padula, D.; Troisi, A. Importance and Nature of Short-Range Excitonic Interactions in Light Harvesting Complexes and Organic Semiconductors. *J. Chem. Theory Comput.* **2017**, *13*, 3754–3763.

(175) Padula, D.; Santoro, F.; Pescitelli, G. A Simple Dimeric Model Accounts for the Vibronic ECD Spectra of Chiral Polythiophenes in Their Aggregated States. *RSC Adv.* **2016**, *6*, 37938–37943.

(176) Evers, F.; Mirlin, A. D. Anderson Transitions. *Rev. Mod. Phys.* **2008**, *80*, 1355–1417.

(177) Claridge, K.; Padula, D.; Troisi, A. On the Arrangement of Chromophores in Light

Harvesting Complexes: Chance versus Design. *Faraday Discuss.* **2019**.

(178) Lee, P. A.; Ramakrishnan, T. V. Disordered Electronic Systems. *Rev. Mod. Phys.* **1985**, *57*, 287–337.

(179) Six, C.; Thomas, J.-C.; Garczarek, L.; Ostrowski, M.; Dufresne, A.; Blot, N.; Scanlan, D. J.; Partensky, F. Diversity and Evolution of Phycobilisomes in Marine Synechococcus Spp.: A Comparative Genomics Study. *Genome Biol.* **2007**, *8*, R259.

8 Appendices

Chapter 3 Appendix

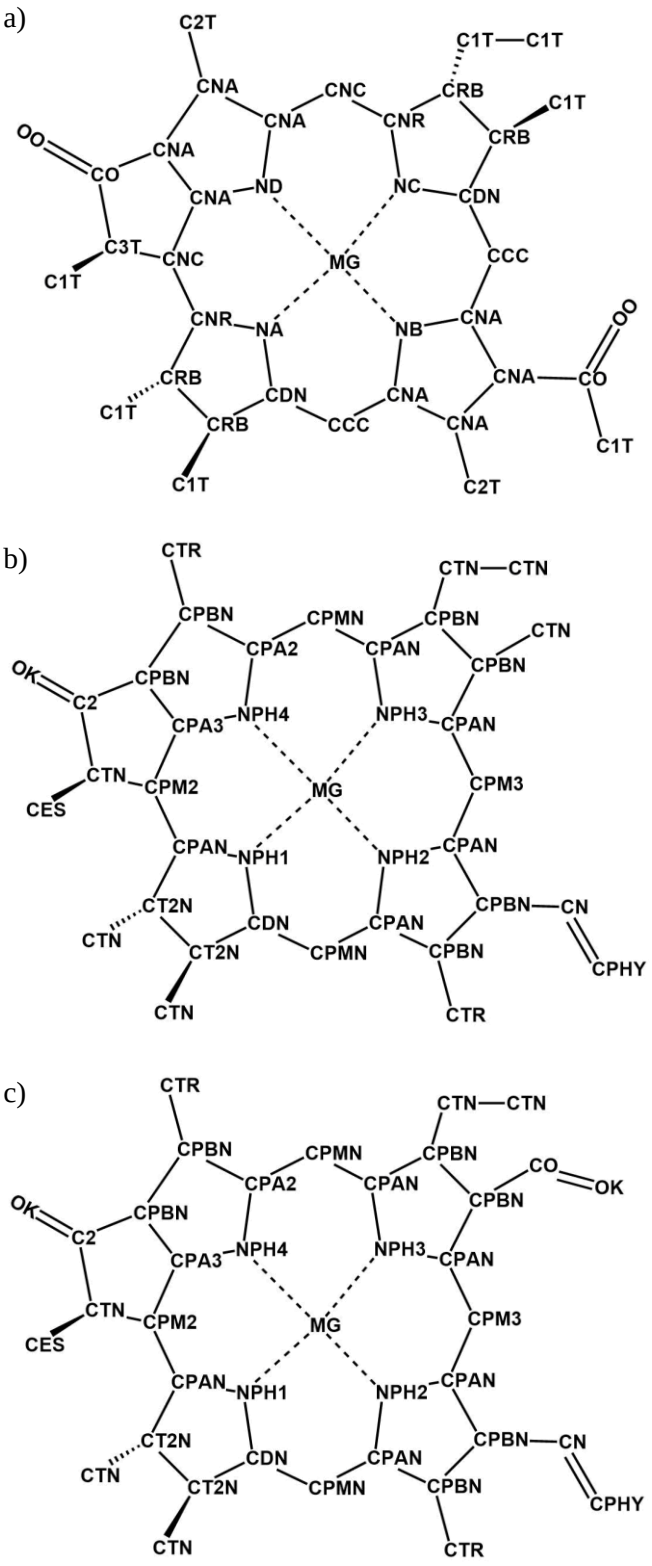


Figure A3 Atom types for chromophores: a) BCL b) CLA c) CLB

Fig. A3 depicts the structures for the three chromophores examined in chapter 3 and their corresponding atom type assignments for definition of FF parameters.

Chapter 4 Appendix

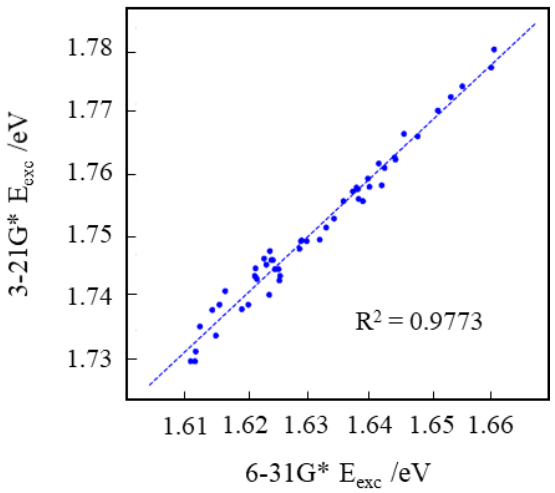


Figure A4.1 Correlation between 3-21G* and 6-31G* excitation energies

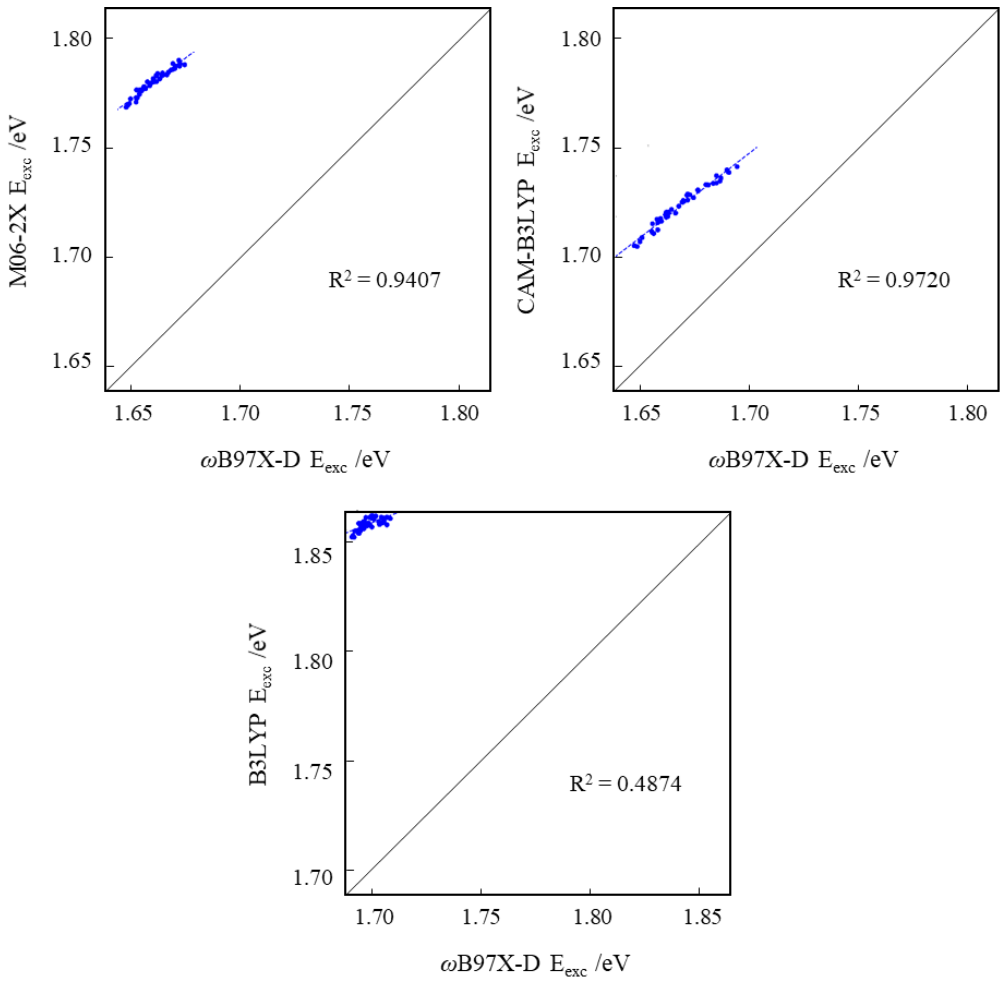


Figure A4.2 Correlation between excitation energies of different functionals

Fig. A4.1 depicts the correlation (0.9773) between excitation energies of BCL computed using the 6-31G* basis set and 3-21G* basis set. Similarly, Fig. A4.2 depicts correlations for excited states computed with different functionals (all with the 6-31G* basis set). The following tables (A4.1-3) contain the results of the interaction analysis detailed in chapter 4, section 4.4.2. For each protein (FMO, WSCP-a, WSCP-b) the table contains the residues top 10 *r* values for each chromophore. The σ_E is also reported, as well as the labels of the interacting atoms, the classification of the interacting residue (solvent, S, protein, P or chromophore, C) and where available the distance between the interacting atoms, *d*, and its fluctuation σ_d.

Table A4.1 BCL interactions

Residue	*r*	σ_E (eV)	Interaction	Class	*d* (Å)	σ_d (Å)
			BCL 367			
SOL19499	0.395	3.29E-04	HW2-O1A	S	1.68	0.15
THR162	0.328	1.10E-04	HG1-CMA	P	3.66	0.41
SOL10024	0.314	9.65E-05	HW-OG	S	2.47	0.58
SER221	0.291	1.37E-04	HG-OBB	P	1.83	0.24
HIS111	0.279	2.24E-04	NE2-MG	P	2.09	0.08
SOL49632	0.249	1.26E-05	HWs-OWs	S	-	-
SOL32676	0.243	4.39E-05	HW-OD1/O/OW	S	-	-
SOL25062	0.237	3.44E-05	HWs-OWs	S	-	-
ASN206	0.237	9.16E-06	OD1-HW+ND2H-OW	P	-	-
SOL38635	0.236	1.85E-05	OW-HN/HWs	S	-	-
			BCL 368			
SOL20272	0.373	2.89E-04	OW-MG	S	1.95	0.05
SOL16680	0.335	3.21E-04	HW1-ND	S	2.19	0.28
SER73	0.318	2.17E-04	HG-OBB	P	1.77	0.19
BCL367	0.234	1.47E-04	-	C	-	-
SOL20275	0.221	2.36E-04	HWs-CHB	S	1.62	0.16
BCL372	0.218	1.46E-05	-	C	-	-
SOL37375	0.218	3.15E-05	HWs-OWs	S	-	-
MET103	0.212	2.00E-05	O-HN	P	2.02	0.16
SOL39517	0.206	1.45E-04	HWs-O1D/O1A	S	-	-
LYS81	0.202	8.97E-05	O-HWs	P	-	-

BCL 369						
SOL19513	0.369	1.82E-04	HW-ND	S	2.76	0.52
SOL19542	0.353	4.84E-04	OW-MG	S	2.19	0.14
BCL370	0.248	9.24E-05	-	C	-	-
PHE307	0.236	4.55E-05	Phenyl-CMC	P	-	-
GLU101	0.227	2.18E-06	O-HWs	P	-	-
SOL19530	0.222	1.13E-04	HWs-OWs	S	-	-
SOL19558	0.220	2.38E-04	HWs-Os	S	-	-
SOL51900	0.177	2.88E-05	HWs-ND	S	2.63	0.55
SOL14934	0.172	2.39E-05	OWs-HWs	S	-	-
HIS298	0.171	2.42E-04	NE2-MG	P	2.12	0.09
BCL 370						
SOL26496	0.485	4.18E-04	HW-NE2	S	2.75	0.51
BCL371	0.410	1.84E-04	-	C	-	-
HIS290	0.243	2.32E-04	NE2-MG	P	2.01	0.06
PRO294	0.220	5.47E-05	HN-O	P	2.09	0.18
TYR364	0.181	5.70E-06	HN-O + HH-OG	P	-	-
ASP48	0.179	6.89E-06	ODs-HWs + HN-O	P	-	-
TYR16	0.172	1.37E-04	HH-OBB	P	1.71	0.19
VAL352	0.150	7.58E-05	O-HN	P	2.09	0.16
ASN289	0.146	2.44E-05	OWs-HWs	P	-	-
BCL373	0.141	3.27E-05	-	C	-	-
BCL 371						
SOL26411	0.306	1.87E-04	HW-OG	S	2.30	0.50
GLU255	0.288	2.46E-05	O-HN/HWs	P	-	-
PRO244	0.274	1.46E-04	O-HBB	P	2.67	0.22
LEU242	0.269	1.48E-04	O-MG	P	1.99	0.09
SER245	0.255	7.19E-05	O-HW	P	1.95	0.18
ARG96	0.250	1.23E-04	HHs-OWs	P	-	-
GLU85	0.237	3.85E-05	OEs -HWs	P	-	-
SOL49632	0.212	7.64E-05	HWs-OWs	S	-	-
LYS247	0.197	4.71E-05	O-HWs	P	-	-
BCL372	0.197	1.33E-04	-	C	-	-
BCL 372						

SOL26466	0.385	1.94E-04	HWs-OWs	S	-	-
HIS146	0.375	2.83E-04	NE2-MG	P	2.06	0.08
TRP184	0.278	1.27E-04	HE1-OBB	P	1.97	0.21
SOL24901	0.228	9.02E-06	HWs-OWs	S	-	-
LYS56	0.203	6.53E-06	HWs-OWs	S	-	-
SER201	0.197	2.97E-05	HN-Os + HG1-OWs	P	-	-
ARG199	0.186	2.18E-05	HHs-OWs/ODs	P	-	-
PHE225	0.185	5.60E-05	O-HN	P	1.92	0.13
LYS247	0.181	2.04E-05	HZ-OD1	P	1.82	0.25
SER98	0.176	7.93E-05	HG1/HN-O	P	-	-
			BCL 373			
SOL27138	0.398	2.66E-04	HW-OE1	S	1.82	0.14
SOL27010	0.324	7.49E-05	HWs-OWs	S	-	-
BCL369	0.318	1.92E-04	-	C	-	-
SOL20275	0.311	5.82E-05	HWs-OWs	S	-	-
SOL19499	0.280	1.29E-04	HWs-OWs	S	-	-
ALA189	0.270	6.18E-05	O-NH	P	-	-
SOL19526	0.249	1.83E-05	HWs-OWs	S	-	-
HIS297	0.238	1.36E-04	ND1-MG	P	2.15	0.07
SOL20286	0.236	4.12E-05	HWs-OWs	S	-	-
SOL26450	0.225	1.22E-04	HW-OBB	S	1.84	0.24
			BCL 400			
SOL32299	0.424	3.28E-04	HW-CHB	S	2.11	0.45
SOL34010	0.301	1.61E-04	HW-HMB	S	3.56	0.61
SOL20922	0.286	3.84E-04	HWs-OWs	S	-	-
TYR124	0.217	1.33E-04	O-MG	P	1.98	0.08
SOL5540	0.212	1.13E-04	HWs-OWs	S	-	-
SOL39842	0.210	3.37E-05	HWs-OWs	S	-	-
SOL20307	0.194	3.34E-05	HWs-OD1/OWs	S	-	-
THR166	0.188	1.06E-04	HG1/HN-O	P	-	-
SOL39022	0.187	4.16E-05	HWs-OWs	S	-	-
SOL8492	0.180	3.36E-05	HWs-OWs	S	-	-

Table A4.2 CLA interactions

Residue	r	σ_E (eV)	Interaction	Class	d (Å)	σ_d (Å)
			CLA 1			
ALA34	0.539	1.95E-04	O-HE1	P	1.84	0.13
ALA33	0.459	1.17E-04	O-HN	P	1.89	0.14
SOL10184	0.340	1.46E-04	HWs-OWs	S	-	-
SOL17152	0.237	4.46E-05	HWs-OWs	S	-	-
TRP151	0.229	7.52E-05	O-HWs + HE1-O	P	-	-
GLY153	0.228	1.12E-06	O-HWs + HN-OWs	P	-	-
GLU85	0.184	3.47E-06	OEs-HWs + HN-OWs	P	-	-
SOL23264	0.178	2.97E-05	HWs-OWs	S	-	-
ASN91	0.177	1.21E-05	O/OD1-HWs + Hs-OWs	P	-	-
PHE40	0.168	4.03E-05	pi-stacking	P	-	-
			CLA 2			
ALA33	0.452	9.99E-05	HN-O	P	2.04	0.19
SOL21499	0.434	6.29E-05	HWs-O	S	-	-
ASN36	0.425	2.75E-05	O -HWs + H-OWs	P	-	-
PRO39	0.424	1.68E-05	O-HW	P	2.11	0.53
SOL33283	0.420	1.63E-05	HWs-O	S	-	-
SOL28386	0.419	1.55E-05	HWs-OWs	S	-	-
PHE40	0.394	5.08E-05	O-HWs	P	-	-
SOL36758	0.390	1.25E-05	HWs-OWs	S	-	-
SOL23262	0.364	8.33E-05	HWs-OE1	S	-	-
SOL35595	0.352	1.32E-05	HWs-OWs	S	-	-
			CLA 3			
ALA34	0.612	2.06E-04	O-HE1	P	1.93	0.17
ALA33	0.516	1.15E-04	O-HN	P	1.91	0.15
SOL20477	0.256	4.28E-06	HWs-OWs	S	-	-
TRP151	0.242	1.31E-06	HE1-O + HN-OWs	P	-	-
GLN53	0.242	1.43E-04	OBD-HE21	P	1.80	0.34
SOL6708	0.219	1.29E-04	HWs-OWs	S	-	-
SOL21943	0.217	4.15E-05	HWs-OWs	S	-	-
SOL12421	0.208	9.50E-06	HWs-OWs	S	-	-
SOL22197	0.200	1.07E-04	HW-O1D	S	1.79	0.19

PRO88	0.193	2.95E-05	O-HW	P	-	-
			CLA 4			
THR48	0.447	1.28E-04	HG1-OBD	P	2.25	0.42
SOL23254	0.361	4.54E-04	HWs-OWs	S	-	-
GLN47	0.268	5.62E-05	O-HBA2	P	3.25	0.31
SOL32487	0.211	1.35E-04	HWs-OWs	S	-	-
ALA33	0.208	1.07E-04	HN-OWs	P	-	-
SOL22243	0.199	5.34E-05	HWs-OE1/O	S	-	-
GLY28	0.192	1.32E-05	HN-OWs + O-HE21	P	-	-
SOL16127	0.174	7.60E-06	HWs-OWs	S	-	-
SOL19783	0.158	1.71E-05	HWs-OWs	S	-	-
ALA34	0.156	1.79E-04	O-HE1	P	1.86	0.12

Table A4.2 CLB interactions

Residue	r	σ_E (eV)	Interaction	Class	d (Å)	σ_d (Å)
			CLB 1			
ALA33	0.430	1.31E-04	HN-O	P	2.04	0.16
ALA34	0.403	2.06E-04	O-HE1	P	1.86	0.12
SOL16413	0.359	2.29E-04	HW-CAB	S	2.44	0.60
GLY153	0.293	1.70E-05	O-HWs	P	-	-
LEU152	0.283	1.60E-05	O-HWs	P	-	-
PRO32	0.264	1.36E-04	O-Mg	P	1.96	0.07
SOL17261	0.236	8.35E-05	HWs-OBD	S	2.12	0.50
SOL23877	0.185	5.70E-06	HWs-OWs	S	-	-
THR48	0.172	9.37E-05	HN/HG1-O	P	-	-
SOL8351	0.170	5.62E-06	HWs-OWs	S	-	-
			CLB 2			
SOL13389	0.369	2.04E-04	HW-C2D	S	1.98	0.21
SOL2821	0.304	4.56E-05	HWs-OWs	S	-	-
ALA33	0.296	1.21E-04	HN-O	P	1.99	0.16
ALA34	0.286	1.92E-04	O-HE1	P	1.96	0.16
GLN53	0.286	1.12E-04	O-HG1 + OE1-HWs	P	-	-
THR48	0.264	1.11E-04	HG1/HN-O + O-HWs	P	-	-
SOL381	0.240	1.99E-05	HWs-OWs	S	-	-
CYS90	0.189	2.14E-05	O-HWs	P	-	-
LEU49	0.188	1.76E-05	HN-O1A	P	1.78	0.16
PRO32	0.183	1.27E-04	O-Mg	P	1.91	0.07
			CLB 3			
GLU85	0.282	1.64E-05	O-HW	P	5.89	0.97
SOL6715	0.271	2.55E-05	HWs-O/OWs	S	-	-
SOL38322	0.270	4.97E-05	HW-OAC	S	1.75	0.23
LEU152	0.265	4.26E-05	O-HWs	P	-	-
GLU85	0.254	1.95E-05	OEs-HWs + HN-OWs	P	-	-
ALA154	0.251	4.88E-05	O-HWs + HN-O	P	-	-
SOL31151	0.236	1.79E-05	HWs-OWs	P	-	-
CHL4	0.234	9.36E-05	-	C	-	-
CYS90	0.227	1.49E-06	HN-OWs + O-HWs	P	-	-

SOL4743	0.225	5.52E-05	OWs-HWs	S	-	-
			CLB 4			
SOL1780	0.582	6.21E-04	HWs-OWs	S	-	-
SOL4246	0.347	3.46E-05	HWs-OWs	S	-	-
LEU152	0.336	1.57E-05	O-HWs	P	-	-
LEU152	0.336	1.57E-05	O-HWs + HN-O	P	-	-
ASN36	0.292	1.19E-05	Hs-OWs + OD1-HWs	P	-	-
LEU93	0.288	1.14E-05	O-HWs	P	-	-
SOL6448	0.283	8.03E-06	HWs-OWs	S	-	-
SOL36798	0.242	9.27E-06	HWs-OWs	S	-	-
SOL2002	0.242	1.34E-05	HWs-OWs	S	-	-
PRO155	0.241	2.67E-05	O-HG1 + O-HN	P	-	-

Fig. A4.3 depicts the distribution of the interaction types reported in Tables A4.1-3 and Fig. A4.4 demonstrates the results of the coulomb interaction energy FT analysis outlined in chapter 4, section 4.4.2. Finally, Fig. A4.5 summarises the environmental analysis described in section 4.4.3.

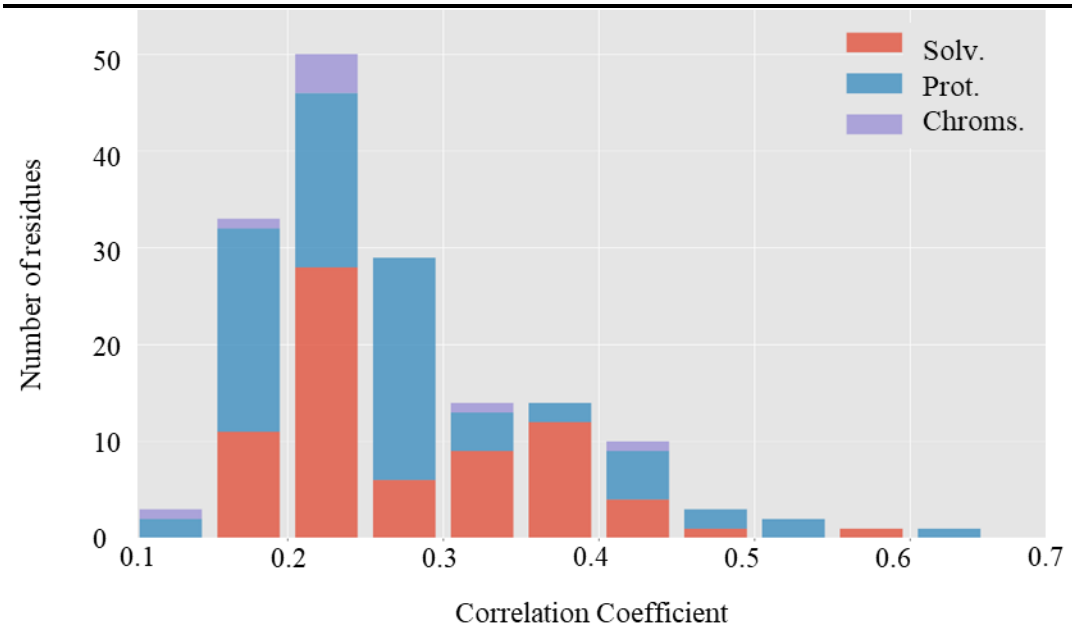


Figure A4.3 Distribution of all the correlation coefficients (reported in Tables A4.1-3) for chromophore interactions with surroundings: chromophore-protein (blue), chromophore-solvent (red) and chromophore-chromophore (purple).

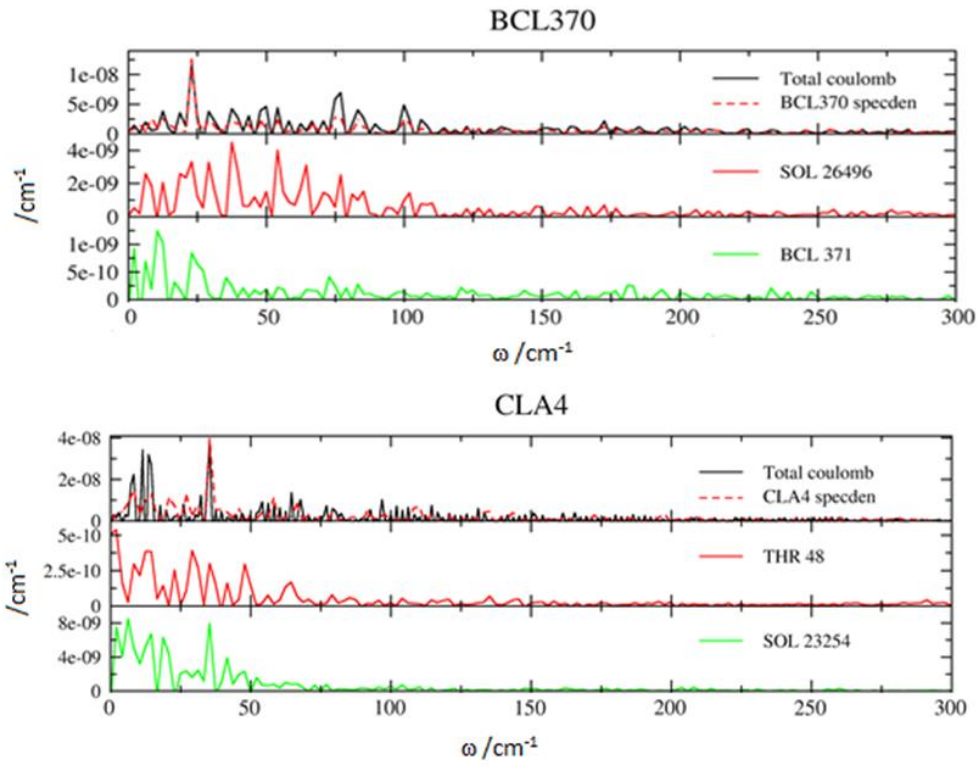


Figure A4.4 FT of coulombic interaction energies of most correlated residues for BCL 370 and CLA 4.

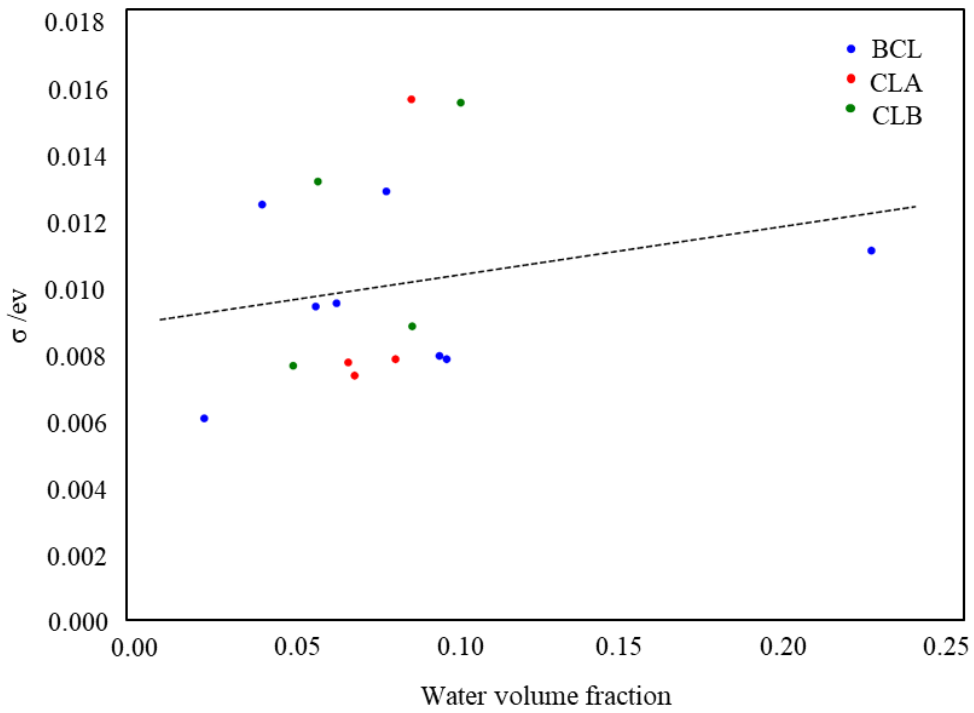


Figure A4.5 Correlation between the fraction of water volume in a 10 Å radius from the chromophore and the fluctuation of the excitation energy along the MD trajectory.