DEVELOPMENT AND APPLICATION OF EFFECTIVE FRAGMENT POTENTIALS FOR (BIO)MOLECULAR SYSTEMS

by

Yongbin Kim

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Chemistry

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Lyudmila V. Slipchenko
Department of Chemistry

Dr. Dor Ben-Amotz
Department of Chemistry

Dr. Sergei Savikhin
Department of Physics and Astronomy

Dr. Mike Reppert
Department of Chemistry

Approved by:
Dr. Christine A. Hrycyna

수고했어

ACKNOWLEDGMENTS

In my first semester at Purdue University, I was supposed to pass the oral English proficiency test; I failed. As a scary consequence, I was no longer eligible for financial support from the Chemistry Department as I was not able to serve as a Teaching Assistant for the 2016 Spring semester. I remember going back home, packing my belongings and e-mailing Lyudmila to say, “I didn’t pass the test, so I should pay my own tuition to continue the Ph.D. program at Purdue”. Right away she replied, “I can cover you with RA in the 2016 Spring semester.” What is more amazing is that this happened even before she actually accepted me as a member of her research Group. The only reason that I was able to embark on the long journey in the Chemistry Department was due to the support from Dr. Slipchenko. She has supported me enormously ever since. Whether it be questions about research or personal matters, she always greeted me with a smile and provided valuable advice. She allowed me to attend countless conferences with enabled me to deepen and broaden my perspective. Lyudmila is my academic parent, she raised me as an independent scientist and for that I am forever in her debt.

I am very grateful to my Theory Group friends. Terri (formerly known as Yen), Pradeep, and Claudia. Terri is a very patient person. When I was dumb in English (like really dumb), she forced me to keep talking and talking and talking. Whenever other international students ask me “how did you pass the oral English proficiency test?”, I would answer “I have a very good friend.” Pradeep was a senior graduate student in the Slipchenko Research Group. When I first joined the group, I didn’t know much about electronic structure theory, so I needed someone to ask questions. Pradeep was my go-to person. I am pretty sure I knocked on his office door more than a million times. Claudia, my friend. She and I joined the Slipchenko Group at the same time. Always a step ahead of me, she successfully completed all the Ph.D. requirements before me, including OP and her defense. I was very lucky to have Claudia to ask questions and she always had right answers. Thank you all for bearing with me.

I must acknowledge my Korean best friend, Dennis Kim (Hyunjin Kim). He served as my translator all the time. Whenever I had to write something for conferences, e-mails, letters, and OP proposals, he helped me correct everything. He sometimes complained, but it was totally understandable because I bothered him a lot. Also, we shared many memories here in West Lafayette. I appreciate that.

More than 80 % of my Ph.D. work was devoted to understanding what happens in the biological systems using the EFP method. The results demonstrated in Chapters 3 through 7 were produced with enormous help from collaborators. Professor Sergei Savikhin, he and Lyudmila make a fantastic duo. He generated all experimental observables in the FMO protein complex. Dr. Dmitri Morozov, when I repeatedly failed to reproduce experimental results by computations, he taught me how important the optimization processes are. Without him, I would not have been able to produce publishable data for sure.

Finally, my mother, Zion Kim. I owe all my achievement to her endless love and support. It is more than I can ever repay her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AA Amino Acid

BChl a Bacteriochlorophyll a

BioEFP Biomolecular Effective Fragment Potential method

CC Coupled Cluster method

CD Circular Dichroism

DFT Density Functional Theory

DMA Distributed Multipole Analysis

EFMO Effective Fragment Molecular Orbital

EFP Effective Fragment Potential

EFPDB Effective Fragment Potential Database

FMO method Fragment Molecular Orbital method

FMO complex Fenna-Matthews-Olson complex

FF Force Field

HF Hartree-Fock

LMO Localized Molecular Orbital

MD Molecular Dynamics

MM Molecular Mechanics

MP2 Second Order Møller-Plesset Perturbation Theory

RMSD Root Mean Square Deviation

SAPT Symmetry Adapted Perturbation Theory

SCF Self-Consistent Field

TBA tert-Butyl Alcohol

TDDFT Time Dependent Density Functional Theory

TDM Transition Dipole Moment

TRESP Transition Electrostatic Potential

QM Quantum Mechanics

VEA Vertical Electron Affinity

VIE Vertical Ionization Energy

ABSTRACT

The Effective Fragment Potential (EFP) is a quantum-mechanical based model potential for accurate calculations of non-covalent interactions between molecules. It can be coupled with *ab initio* methods in so-called QM/EFP models to explore the electronic properties of extended molecular systems by providing rigorous description of surrounding environments. The current EFP formulation is, however, not well suited for large-scale simulations due to its inherent limitation of representing effective fragments as rigid structures. The process of utilizing EFP method for the molecular systems with flexible degrees of freedom entails multiple sets of parameter calculations requiring intensive computational resources. This work presents development of the EFP method for describing flexible molecular systems, so-called Flexible EFP. To validate the applicability of the Flexible EFP method, extensive benchmark studies on the amino acid interactions, binding energies, and electronic properties of flavin chromophore of the cryptochrome protein have been demonstrated. In addition to methodological developments, excitonic properties of the Fenna-Matthews-Olson (FMO) photosynthetic pigment-protein complex are explored. In biological systems where intermolecular interactions span a broad range from non-polar to polar and ionic forces, EFP is superior to the classical force fields. In the present study, we demonstrate excellent performance of the QM/EFP model for predicting excitonic interactions and spectral characteristics of the FMO wildtype complex. We characterize the key factors for accurate modeling of electronic properties of bacteriochlrophyll a (BChl a) photosynthetic pigments and suggest a robust computational protocol that can be applied for modeling other photosynthetic systems. Developed computational procedures were also successfully utilized to elucidate photostability and triplet dynamics in the FMO complex and spectroscopic effects of single-point mutagenesis in FMO. A combination of polarizable EFP molecular dynamics and QM/EFP vibrational frequency calculations were also applied to understanding and interpreting structures and Raman spectroscopy of tert-butyl alcohol solutions.

CHAPTER 1. THE EFFECTIVE FRAGMENT POTENTIAL METHOD

Molecular modeling became an essential tool for investigating and predicting properties of molecular systems. In biological and materials applications, where model molecular systems can contain thousands of atoms, a compromise between computational cost and accuracy of a chosen computational technique becomes a key for obtaining reliable answers. Chemistry of large molecular systems is often driven by noncovalent interactions. Even though correlated quantum mechanical (QM) methods such as coupled cluster (CC) methods and perturbation theory [1,2], as well as modern versions of density functional theory (DFT) [3,4] might describe noncovalent interactions with sufficient accuracy, their computational cost and scaling prevents them from being applied to large-scale simulations. On the other hand, molecular mechanics (MM) force fields (FF) such as AMBER [5] and CHARMM [6] can be applied to macromolecular systems routinely, but they do not explicitly capture important components of noncovalent interactions such as polarization and charge-transfer, making their reliability questionable. To address this challenge, a variety of computational techniques was introduced in recent years, including QM/MM models such as ONIOM [7-9], advanced force fields [10-15], fragmentation schemes [16-22], and semi-empirical methods [23-28]. The Effective Fragment Potential (EFP) [29-34] belongs to a class of first principles-based force fields that provides a physically meaningful description of noncovalent interactions by a fraction of computational cost of analogous QM methods such as the Symmetry Adapted Perturbation Theory (SAPT) [35-38].

The EFP method has been widely applied for modeling structure patterns and binding energies in molecular clusters, such as clusters of water [39], complex liquids [40,41] including water-alcohol [42,43] and water-benzene mixtures [44], and DNA base pairs [45,46]. Previous benchmarks on S22 and S66 data sets demonstrate that EFP reliably describes noncovalent interactions of different types, with the accuracy approaching that of MP2 and exceeding that of classical force fields [47]. Additionally, EFP can be combined with QM models into polarizable QM/EFP schemes, with applications ranging from excitation and ionization properties of solvated chromophores to electron and energy transfer processes in biology [29,30,33,48-51].

The Effective Fragment Potential (EFP) method is a quantum-mechanical (QM) based model potential for describing intermolecular interactions in condensed phase systems. The fundamental formalism of the EFP method is to represent a molecular system as a collection of

rigid fragments. Hence, fragment-fragment interactions determine chemical properties of the whole system. The EFP method describes the total interaction energy ($E_{EFP-EFP}$) of chemical systems by the sum of electrostatic or Coulomb (E_{Ele}), polarization (E_{pol}), dispersion (E_{disp}), and exchange-repulsion (E_{ex-rep}) terms:

$$E_{EFP-EFP} = E_{Ele} + E_{pol} + E_{disp} + E_{ex-rep} \tag{1.1}$$

Charge-transfer term has also been developed by Gordon and co-workers and was shown to be non-negligible in ionic and strongly polar systems [52-54]. However, charge-transfer term is not used in the work described in this thesis. Equation (1.1) particularly refers to EFP-EFP interactions. In the presence of an *ab initio* subsystem, the interactions of EFP fragments with an *ab initio* region are represented by the polarizable embedding model; namely, EFP fragments interact with the QM Hamiltonian via one-electron Coulomb and polarization terms [29,30,33,48-50,55]:

$$\hat{H}_{QM/EFP} = \hat{H}_{QM} + \hat{H}_{EFP} + \hat{H}_{QM-EFP} \tag{1.2}$$

where $\hat{H}_{QM}$ and $\hat{H}_{EFP}$ are the Hamiltonians of the *ab initio* and EFP subsystems, respectively. $\hat{H}_{QM-EFP}$ is the coupling term between QM and EFP subsystems which can be represented as:

$$\hat{H}_{QM-EFP} = \langle p \left| \hat{V}^{Coul} + \hat{V}^{pol} \right| q \rangle p^{\dagger} q \tag{1.3}$$

where p and q represent the atomic orbitals in the QM region. $\hat{V}^{Coul}$ and $\hat{V}^{pol}$ represent electrostatic and polarization perturbations to the *ab initio* Hamiltonian $\hat{H}_{QM}$, respectively.

The electrostatic term, E_{Ele} is computed based on a multipole expansion obtained from the distributed multipole analysis (DMA) [56]. The electrostatic multipoles are centered at fragment atomic centers and bond midpoints of a given molecule. The fragment-fragment electrostatic interaction energy is decomposed into charge-charge, charge-dipole, charge-quadrupole, charge-octopole, dipole-dipole, dipole-quadrupole, and quadrupole-quadrupole terms. The electrostatic energy between multipole expansion points A and B is given by:

$$
\begin{aligned}
E_{Ele}^{AB} = q^B \Bigg[T q^A - \sum_{a}^{x,y,z} T_a \hat{\mu}_a^A + \frac{1}{3} \sum_{a,b}^{x,y,z} T_{a,b} \widehat{\Theta}_{a,b}^A - \frac{1}{15} \sum_{a,b,c}^{x,y,z} T_{a,b,c} \Omega_{a,b,c}^A \Bigg] \\
+ \sum_{a}^{x,y,z} \hat{\mu}_a^B \Bigg[T_a q^A - \sum_{b}^{x,y,z} T_{a,b} \hat{\mu}_b^A + \frac{1}{3} \sum_{b,c}^{x,y,z} T_{a,b,c} \widehat{\Theta}_{b,c}^A \Bigg] \\
+ \sum_{a,b}^{x,y,z} \widehat{\Theta}_{a,b}^B \Bigg[T_{a,b} q^A - \sum_{c}^{x,y,z} T_{a,b,c} \hat{\mu}_c^A + \frac{1}{3} \sum_{c,d}^{x,y,z} T_{a,b,c,d} \widehat{\Theta}_{c,d}^A \Bigg] \\
+ \sum_{a,b,c}^{x,y,z} \widehat{\Omega}_{a,b,c}^B \Bigg[-\frac{1}{15} T_{a,b,c} q^A \Bigg]
\end{aligned}
\tag{1.4}
$$

where q^A is the charge on point A and q^B is the charge on point B. $T, T_a, T_{a,b}$, $T_{a,b,c}$ and $T_{a,b,c,d}$ represent the electrostatic tensors of ranks zero to four. The q, μ, Θ, and Ω refer to the point charges, dipoles, quadrupoles, and octopoles, respectively. These discrete Taylor-series type of expansion approximates the functional electrostatic interactions. Thus, electrostatic interaction energies of multi-fragments system can be achieved using classical interactions between distributed multipolar interactions. The electrostatic contribution to the *ab initio* Hamiltonian due to a multipole expansion point k of an EFP fragment can be written as:

$$
\begin{aligned}
V_k^{Coul}(x) = q_k T(r_{kx}) - \sum_{a}^{x,y,z} \mu_a^k T_a(r_{kx}) + \frac{1}{3} \sum_{a,b}^{x,y,z} \Theta_{a,b}^k T_{a,b}(r_{kx}) \\
- \frac{1}{15} \sum_{a,b,c}^{x,y,z} \Omega_{a,b,c}^k T_{a,b,c}(r_{kx})
\end{aligned}
\tag{1.5}
$$

where r_{kx} is the distance between the expansion point k and the coordinate of an electron x in the QM region.

The multipole expansion accurately describes electrostatic interactions at long separation between fragments (or between a fragment and the QM region). However, charge penetration effects become significant at close separation. To account for the charge-penetration arising from the electron density overlap, short-range damping functions can be employed [56,57]. Gaussian-type damping function is applied to the first (charge) term of the QM-EFP electrostatic potential (Eq. 1.5):

$$
f_k^{gaussian}(x) = 1 - \exp(-\alpha r_{kx}^2) \tag{1.6}
$$

The exponential damping function screens the charge-charge term of the EFP-EFP electrostatic energy (Eq. 1.4):

$$f_{kl}^{exponential} = 1 - \frac{b^2}{b^2-a^2}\exp(-ar_{kl}) - \frac{a^2}{a^2-b^2}\exp(-br_{kl}) \quad (1.7)$$

where r_{kl} is the distance between multipole points k and l, and damping parameters a and b are for the multipole points k and l on fragments A and B, respectively. Damping parameters, α, a, and b in Eqns. (1.6) and (1.7) are computed by minimizing the differences between the electrostatic potentials from the damped multipole expansion and the electronic wave function in the parameter-generating (MAKEFP) calculation for each EFP fragment.

The fragment-fragment polarization energy, E_{pol} is computed as an interaction of induced dipoles of a fragment with a combined field of induced dipoles and static multipoles produced by the other fragments. Hence, polarization energy must be computed in a self-consistent manner because the induced dipole on a fragment is determined by the static multipoles as well as induced dipoles on the other fragments. In EFP, static anisotropic polarizability tensors (α^k, second order tensors) of a fragment are distributed to the localized molecular orbital (LMO) centroids, where the induced dipoles are computed. Polarization energy (in the presence of the *ab initio* subsystem) is then written as:

$$E_{pol} = \frac{1}{2}\sum_a^{x,y,z}\sum_k\left[-\mu_a^k\left(F_a^{mult,k} + F_a^{nuc,k}\right) + \bar{\mu}_a^k F_a^{ai,k}\right] \quad (1.8)$$

where μ^k and $\bar{\mu}^k$ are the induced and conjugated induced dipole on the LMO centroid k. $F^{mult,k}$ is the electric field due to multipole moments on point k; F^{nuc} and F^{ai} are electric fields due to nuclei and electrons of the QM subsystem, respectively. Induced dipoles at polarizable points k are given by

$$\mu_a^k = \sum_k\sum_b^{x,y,z}\alpha_b^k\left(F_{ab}^{mult,k} + F_{ab}^{ind,k} + F_{ab}^{nuc,k} + F_{ab}^{ai,k}\right) \quad (1.9)$$

where $F^{ind,k}$ refers to the field due to induced dipoles of other fragments. Polarization perturbation to the *ab initio* Hamiltonian due to EFP fragments is given as:

$$V_k^{pol}(x) = \frac{1}{2}\sum_a^{x,y,z}(\mu_a^k + \bar{\mu}_a^k)T_a(r_{kx}) \quad (1.10)$$

As aforementioned, the QM wave function (Ψ) and EFP subsystems respond to changes of electron densities between them such that the full consistency can be achieved as depicted in Figure 1.1. The induced dipoles of the EFP region are converged using iterative procedure.

The polarization damping functions can also be applied to the electric fields produced by multipoles and induced dipoles in Eqns (1.8) and (1.9) to avoid polarization catastrophe at close separation of fragments [57].

$$f_{kl}^{pol} = 1 - \exp\left(-\sqrt{c_k d_l} r_{kl}^2\right)\left(1 + \sqrt{c_k d_l} r_{kl}^2\right) \tag{1.11}$$

where c_k and d_l represent the damping parameters on polarizability points *k* and *l*, respectively.

Dispersion energy, E_{disp}, accounts for the long-range electron correlation between effective fragments. In EFP, the dispersion term is modeled as interaction of instantaneous dipoles computed using dynamic polarizability tensors. Dispersion contribution is shown to be dominant in non-polar molecules and aromatic compounds. Similar to static polarizabilities, the dynamic polarizability tensors are computed and distributed on LMO centroids of each fragment. The C_6 dispersion coefficients between a pair of distributed points *k* and *j* on different fragments are computed as integrals over imaginary frequencies $i\omega$ of a product of distributed time-dependent polarizability tensors $\alpha_{\alpha\gamma}(i\omega)$ in the isotropic approximation [58,59]:

$$C_6^{kj} = \int_0^\infty d\omega \alpha_{\alpha\gamma}^k(i\omega)\alpha_{\beta\delta}^j(i\omega) \cong \int_0^\infty d\omega \bar{\alpha}^k(i\omega)\bar{\alpha}^j(i\omega) \tag{1.12}$$

where $\bar{\alpha} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$. The integral in the Eqn (1.12) is computed using a quadrature, and dynamic polarizability tensors on each fragment are pre-computed for a set of 12 imaginary frequencies using the time-dependent HF (TDHF) method. Using C_6 coefficients, the dispersion energy between EFP fragments can be expressed as:

$$E_{disp} = -\frac{4}{3}\sum_{k\in A}\sum_{j\in B}\frac{f_{damp}C_6^{kj}}{R_{kj}^6} \tag{1.13}$$

where f_{damp} corresponds to a damping function for a better description between fragments at close separation. The expression for dispersion energy between QM and EFP subsystems has been also derived and implemented [60].

The exchange-repulsion energy, E_{ex-rep} for each pair of fragments is computed using fragments' localized wave functions, as a sum of exchange-repulsion interactions E_{ex-rep}^{ij} between LMOs of each fragment:

$$E_{ex-rep} = \sum_{i\in A}\sum_{j\in B} E_{ex-rep}^{ij} \tag{1.14}$$

$$E_{ex-rep}^{ij} = -4\sqrt{\frac{-2ln|S_{ij}|}{\pi}}\frac{S_{ij}^2}{R_{ij}} - 2S_{ij}\left(\sum_{k\in A} F_{ik}^{A}S_{kj} + \sum_{l\in B} F_{jl}^{B}S_{li} - 2T_{ij}\right) + 2S_{ij}^2\left(-\sum_{J\in B}\frac{Z_J}{R_{iJ}} + 2\sum_{l\in B}\frac{1}{R_{il}} - \sum_{l\in A}\frac{Z_l}{R_{lj}} + 2\sum_{k\in A}\frac{1}{R_{kj}} - \frac{1}{R_{ij}}\right) \tag{1.15}$$

where i, j, k, l are the LMOs, L, J are the nuclei, S_{ij} and T_{ij} are the overlap and kinetic energy integrals between i^{th} and j^{th} LMOs belonging to different fragments, and F is the Fock matrix elements of each fragment and Z_I is a nuclear charge on atom I. The exchange-repulsion term is the most computationally expensive term in the EFP model (in the absence of the *ab initio* region), as it explicitly evaluates the overlap and kinetic energy integrals between the molecular orbital basis functions of fragments. Two distinct implementations of the QM-EFP exchange-repulsion term have been reported [32,61,62]. Implementation reported in REF [32,61] utilizes rigorous brute-force formulation based on the Hartree-Fock exchange-repulsion energy, but is computationally expensive. In the model from REF [62], the QM-EFP exchange-repulsion term is computed using overlap of *ab initio* wave function with repulsive Gaussian-like potentials positioned at the LMO centroids of each fragment, which provides computationally efficient but less rigorous account for repulsive interactions between the *ab initio* and EFP regions.

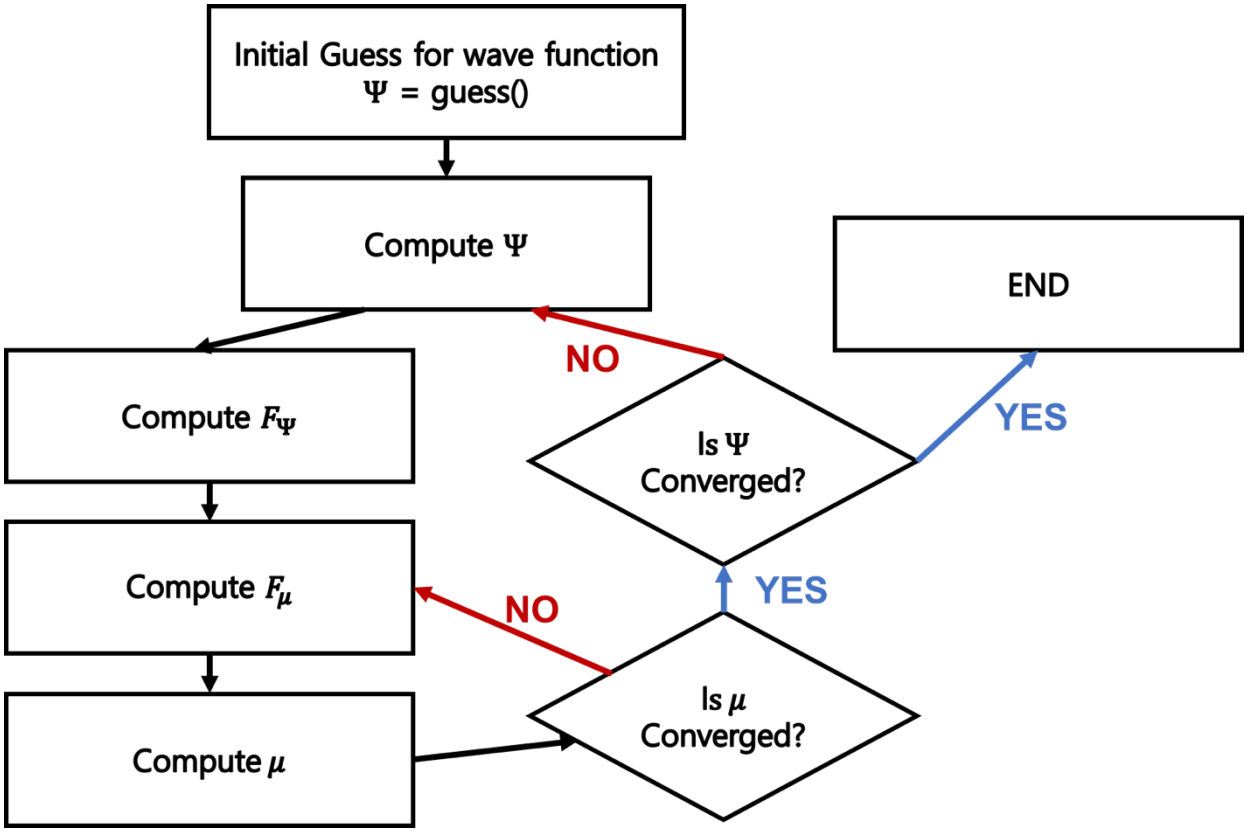

Figure 1.1. The computational procedure describing self-consistent field (SCF) routine in QM/EFP.

CHAPTER 2. EFFECTIVE FRAGMENT POTENTIALS FOR FLEXIBLE MOLECULES: TRANSFERABILITY OF PARAMETERS AND AMINO ACID DATABASE

2.1 Abstract

Accurate but efficient description of noncovalent interactions is a key to predictive modeling of biological and materials systems. The effective fragment potential (EFP) is an *ab initio*-based force field that provides means to obtain physically meaningful decomposition of noncovalent interactions of a molecular system into Coulomb, polarization, dispersion, and exchange-repulsion components. An EFP simulation protocol consists of two steps, preparing parameters for molecular fragments by a series of *ab initio* calculations on each individual fragment, and calculation of interaction energy and properties of a total molecular system based on prepared parameters. As fragment parameters (distributed multipoles, polarizabilities, localized wave function etc.) depend on fragment geometry, straightforward application of the EFP method requires recomputing parameters of each fragment if its geometry changes, for example during thermal fluctuations of a molecular system. Recomputing fragment parameters becomes a computational and human bottleneck and leads to a loss of efficiency of a simulation protocol. An alternative approach, in which fragment parameters are adjusted to different fragment geometries, referred to as "Flexible EFP", is explored here. Parameter adjustment is based on translations and rotations of local coordinate systems associated with fragment atoms. The protocol is validated on extensive benchmarks of amino acid dimers extracted from molecular dynamics snapshots of a cryptochrome protein. Parameter database for standard amino acids is developed to automate Flexible EFP simulations in proteins. To demonstrate the applicability of Flexible EFP in large-scale protein simulations, binding energies as well as vertical electron ionization and electron attachment energies of a lumiflavin chromophore of the Cry1At cryptochrome protein were computed. The results obtained with Flexible EFP are in a close agreement with the standard EFP procedure but provide significant reduction in computational cost.

2.2 Introduction

EFP parameters can be computed in an automated fashion from first-principles calculations for any chemical species. This makes EFP suitable for modeling a wide range of molecular systems, which might be challenging with standard force fields that require cumbersome parameterization of non-standard molecules. On the other hand, the current EFP formalism assumes that an EFP fragment is a rigid molecule, such that one either uses the same parameters for the fragment in all successive computations (if structural changes of the molecule can be ignored) or needs to recompute parameters for each unique molecular geometry. Specifically, the latter scenario becomes unavoidable in biological simulations where both protein backbone and side chains of amino acids possess high structural flexibility. However, recomputing parameters for different fragment geometries requires intensive computational resources, significantly increasing computational and human cost of EFP simulations.

One approach that mitigates a problem of flexible fragments is updating fragment geometries and recomputing fragment parameters on-the-fly, as it is implemented in the Effective Fragment Molecular Orbital (EFMO) method [20,63-65]. In EFMO, which is a hybrid of the EFP and FMO (Fragment Molecular Orbital) methods, interactions between neighboring fragments are computed quantum mechanically with the FMO formalism, while the long-range interactions are obtained from the EFP formalism. The EFMO method has been shown to provide accurate energetics in water clusters, griffithsin-carbohydrate complex and enzymatic catalysis [64,66-69]. Another possibility, which has been partly explored in REF [10], is to adjust parameters computed at one fragment geometry to other geometries without explicitly recomputing them. The goal of the present work is to refine and generalize this approach, and provide essential benchmarks. Specifically, we outline a procedure for adjusting parameters to an arbitrary fragment geometry by introducing local coordinate systems associated with fragment atoms and rotating and translating parameters accordingly. We validate this approach by considering test systems designed from molecular interactions occurring in cryptochrome Cry1At protein. Based on these benchmarks, we design a library of amino acid fragments that can generally be used for modeling noncovalent interactions in any biological protein systems.

2.3 Methods

2.3.1 Flexible EFP scheme

The EFP parameters consist of the following components: (i) atomic and bond midpoint coordinates, (ii) electrostatic multipoles (charges, dipoles, quadrupoles, and octopoles) and the electrostatic damping coefficients at atoms and bond midpoints, (iii) coordinates of the localized molecular orbital (LMO) centroids, (iv) static and dynamic polarizability tensors at the centroids of LMOs, and (v) localized wave function described via atomic basis and molecular orbital coefficients; Fock matrix in the basis of localized molecular orbitals. The location of these parameters for water molecule is shown in Figure 2.1. Multipoles (up to octopole), static and dynamic polarizability tensors and localized molecular orbital wave functions are described as tensors of ranks zero to three.

In this work, we explore whether it is feasible to 'transfer' parameters computed at one fragment geometry to another geometry by translating and rotating the corresponding tensors according to the changes in fragment geometry. We refer to a "flexible fragment" as a molecule that changes its internal geometry but not a chemical connectivity during a considered chemical process. In order to adjust pre-computed potentials to a new geometry, we introduce local translational and rotational frames based on each triplet of neighboring atoms within a molecule. Here we define the triplet of atoms as a group of any three atoms forming a valence angle, or, in other words, the three atoms out of which one is covalently bound to two others. Each triplet of atoms determines a coordinate plane. A rotation/translation matrix is computed for each pair of corresponding atom triplets in geometries A (original structure) and B (new structure). Then, a two-step rotation ($\boldsymbol{R1} = \boldsymbol{R}^{*} \cdot \boldsymbol{R}^{**} = \sum_{k=1}^{n} r_{ik}^{*} \cdot r_{kj}^{**}$ and $\boldsymbol{R2} = \boldsymbol{R}^{*} \cdot \boldsymbol{R}^{***} = \sum_{k=1}^{n} r_{ik}^{*} \cdot r_{kj}^{***}$, where R*, R** and R*** are rotation matrices; i=1,…,m and j=1,…m, m=3, n=3) brings all parameters positioned at the three atoms of structure A to the corresponding parameters in structure B, as shown in Figure 2.2. This three-atom local transformation procedure is repeated until parameters at all atoms of a fragment are shifted to a new geometry on an atom by atom basis. Atoms and corresponding potentials might undergo more than one translation/rotation. In this case, the resulting positions and angles are averaged to produce the final potential. Parameters at bond midpoints are shifted proportionally to changes in the corresponding bond length and rotated using the same rotation matrix as the atoms forming the bond. To adjust parameters positioned at LMO centroids, the following procedure is used. The LMOs are assigned to atom triplets based on the

closest distance either to one of the bond midpoints of the triplet (this is typically the case for LMOs describing valence bonds) or to one of the atoms, in the case of lone-pair LMOs. Then the LMOs are rotated and translated together with the atoms or bond midpoints of the corresponding atom triplet.

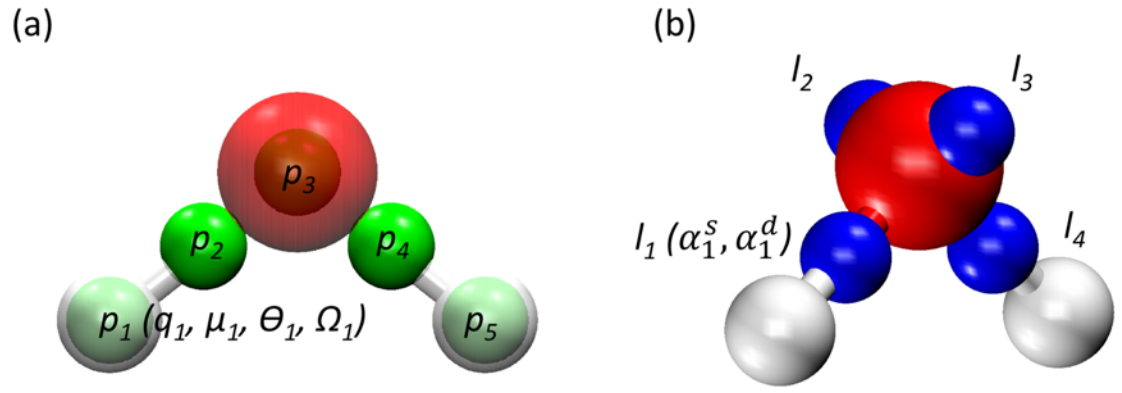

Figure 2.1. Distributed EFP parameters are shown on the example of a water molecule. (a) Multipole moments (charges (q), dipoles (μ), quadrupoles (θ), and octopoles (Ω)) are distributed at atomic centers and bond midpoints p shown with green spheres. (b) static (α^s) and dynamic (α^d) polarizability tensors are placed at localized molecular orbital (LMO) centroids l shown with blue spheres.

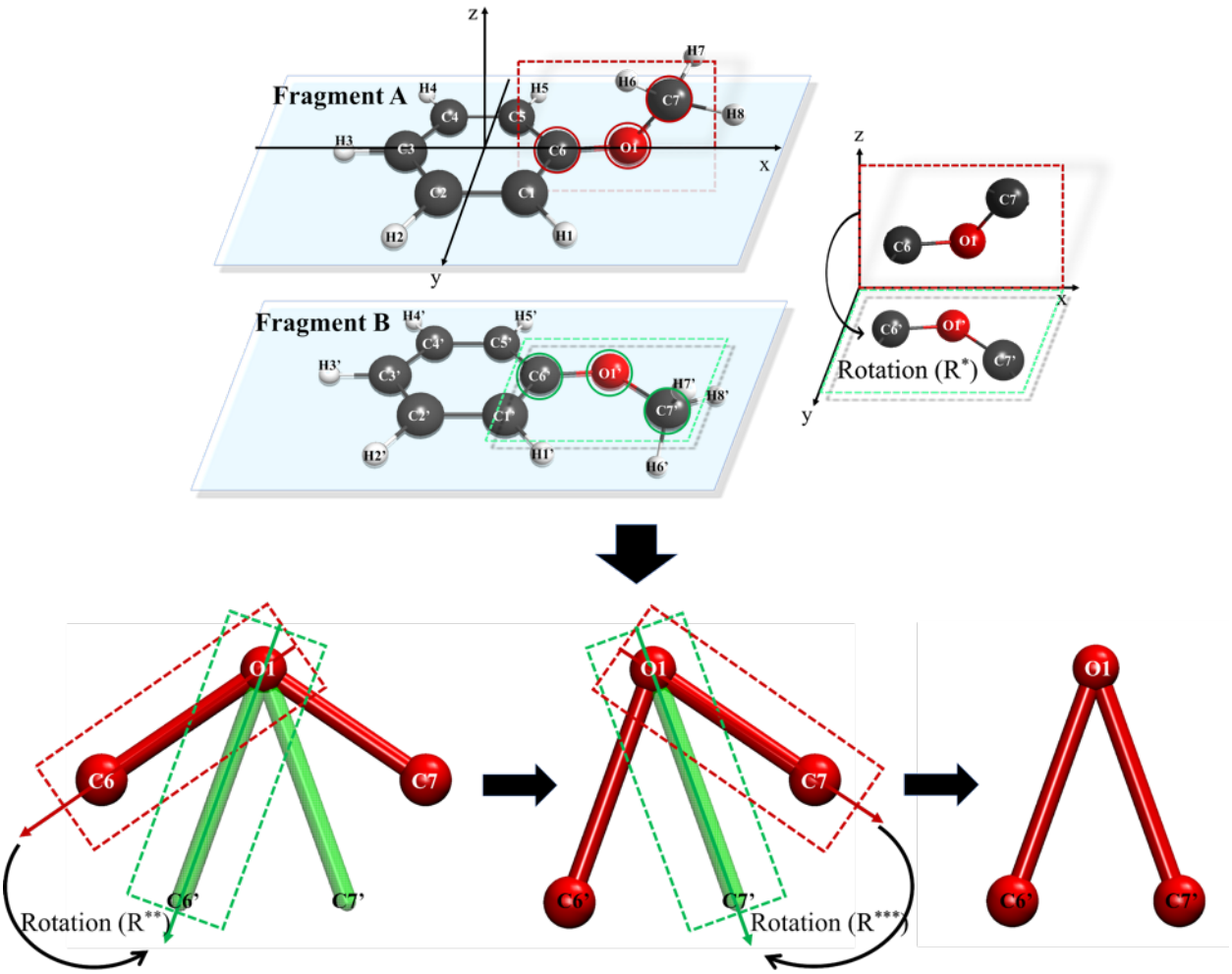

Figure 2.2. Translation of EFP parameters according to changes in fragment's geometry. Fragment A is a pre-computed EFP fragment and B represents a fragment at a different geometry to which the parameters should be adjusted. (Upper panel) Rotation **R*** of a local plane defined by the triplet of atoms C6-O1-C7 between structure A and B. (Lower panel) Rotations **R**** and **R***** bring positions and corresponding parameters of atoms C6 and C7 of fragment with structure A to the positions of atoms in structure B.

2.3.2 Pairwise EFP

Decomposition of the EFP energy of a many-fragment system into pairwise fragment-fragment interactions has been implemented. Electrostatics, dispersion, exchange-repulsion interactions are two-body, so the corresponding system energies are sums of all pairwise fragment-fragment interactions. Many-body polarization energy was decomposed into fragment-fragment interactions in the following way: (i) induced dipoles are self-consistently converged for the whole

system; (ii) using converged dipoles, pair polarization energies between fragments i and j are computed as:

$$E_{ij} = -\frac{1}{2}\sum_{a}^{x,y,z}\left[\sum_{pp\in i}\mu_{pp,a}F_{pp,a}^{i} + \sum_{pp\in j}\mu_{pp}F_{pp}^{j}\right] \quad (2.1)$$

where μ_{pp} is the induced dipole moment at the polarizability point pp, F_{pp}^{j} is the electric field at point pp due to all nuclei and static multipoles of fragment j. In this formulation, the total polarization energy is a sum of all pairwise energies, but each dimer energy implicitly incorporates many-body effects through the induced dipoles self-consistently converged for the whole system. Thus, the pairwise scheme can conveniently decompose the total noncovalent interaction energy into contributions of individual fragment pairs, or extract interactions of a particular fragment with the other fragments in the system, with potential applications in analysis of ligand-binding interactions.

2.4 Computational Details

2.4.1 Construction of EFP parameter database EFPDB

25 protein structures were chosen from molecular dynamics (MD) trajectories of cryptochrome 1 from Arabidopsis thaliana (Cry1At, PDB: 1U3D). A simulation protocol is described in detail in REF [70]. Following the original implementation of the BioEFP fragmentation algorithm [33], each of 25 protein configurations was fragmented into individual amino acid (AA) fragments along $C_\alpha - C$ bonds, resulting in 12125 BioEFP fragments.

To test the Flexible EFP scheme, the following procedure was employed. We noticed in a preliminary work (data not shown) that the additional fragmentation of AAs into the backbone and residue groups improves the accuracy of the Flexible EFP. Thus, in addition to the original BioEFP scheme, each AA fragment was further fragmented along $C_\alpha - C_\beta$ bond into a backbone group ($N - C_\alpha - C = O$) and a side chain group. In this fragmentation, shown in Figure 2.3, a glycine is represented by a backbone fragment. The only exception from the above scheme is proline which is described as a single fragment. Additionally, disulfide bridges are fragmented along $S - S$ bond.

To recombine the backbone and side chain fragments in a complete AA moiety, parameters at a bond midpoint and LMO centroid between C_α and C_β were excluded to ensure the stability of polarization self-consistent procedure (see Figure 2.4). Hence, compared to the original BioEFP

scheme, the Flexible EFP scheme has one less parameter point for each AA and is expected to show consistently smaller magnitudes of all energy terms than the original BioEFP. However, as will be demonstrated in dimer benchmarks, the resulting discrepancies between two EFP schemes are minor in practice. In principle, parameters at the boundary points between fragments can be included if neighboring fragments do not polarize (or do not fully polarize) each other. Such models will be considered in the future work.

EFP parameters for all unique AA structures (i.e., 281 total fragments) were computed with a hybrid 6-31G(d)/6-31(,+)G(3df,2p) (for aryl residues) and 6-31G(d)/6-311++G(3df,2p) (for other residues) basis using MAKEFP module of GAMESS. It was previously noticed that the accuracy of EFP interaction energies improves significantly with the use of a hybrid basis set scheme for the parameter preparation step [42,43,47]. Our earlier studies indicate that electrostatic multipoles are more accurate (and non-divergent) when computed with a small basis set (6-31G(d)), while the rest of the EFP terms (polarization, dispersion, and exchange-repulsion) are more accurate when the parameters are computed in a larger basis set (6-311++G(3df,2p)) [47].

As expected, AA residues possess high flexibility, with lysine producing the largest number of local minima (see Figure 2.5). A selection of a particular fragment from the database is based on the root mean square deviation (RMSD) between geometry of the target molecule and available database fragments. All fragment atoms including hydrogens were used in computing RMSD. Our preliminary investigations show that the accuracy of Flexible EFP noncovalent interaction energies correlates with RMSD between initial and target structures (data not shown). Therefore, fragments with the smallest RMSD to the target structure are selected and subjected to rotation and shifting to match the target geometry. Figure 2.6 shows a spread in geometrical differences between fragments from the MD trajectory and the closest (in the RMSD sense) database fragments for each AA residue. Only for two residues, GLU and GLN, geometries of original fragments might deviate from geometries of data base fragments by more than 0.5 Å. For other residues, typical RMSD differences are between 0.1 and 0.3 Å.

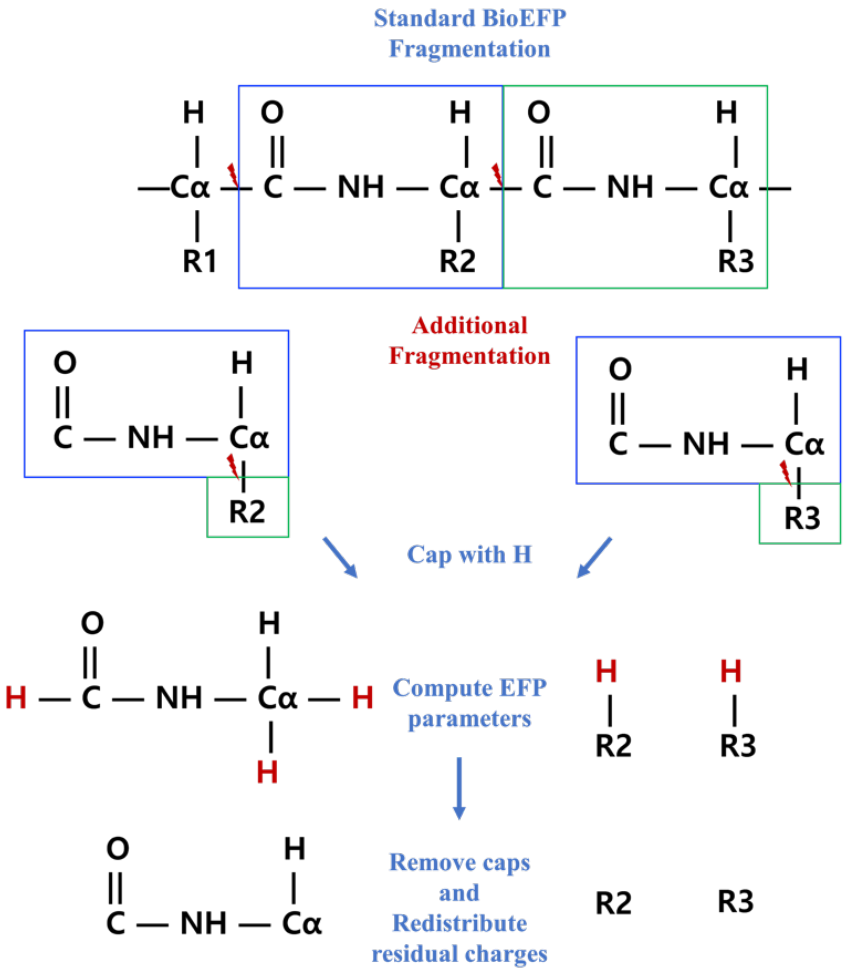

Figure 2.3. Fragmentation of a polypeptide chain into EFP fragments.

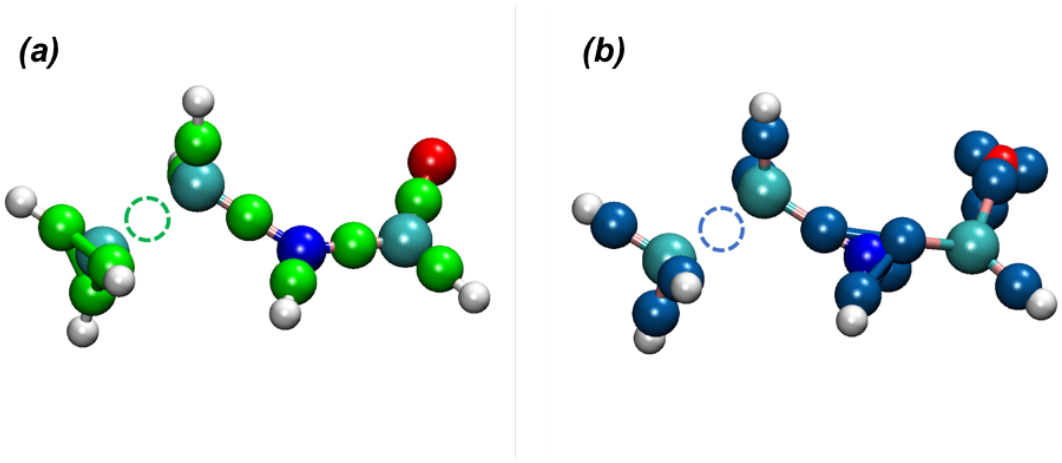

Figure 2.4. Fragmentation into the backbone and side chain groups on an example of the alanine residue. (a) Positions of bond midpoints (shown in green) and (b) positions of LMO centroids (shown in sapphire). Empty circles denote excluded parameters.

	AA	# geom		AA	# geom
	GLY	7		PRO	2
	CYS	3	POLAR	THR	9
	MET	22	POLAR	ASN	18
ALIPHATIC	ILE	27	POLAR	GLN	9
ALIPHATIC	LEU	27	POLAR	SER	3
ALIPHATIC	VAL	9	CATIONIC	ARG	26
ALIPHATIC	ALA	1	CATIONIC	LYS	54
ARYL	TRP	3	CATIONIC	HIP	1
ARYL	HID	4	ANIONIC	ASP	17
ARYL	HIE	1	ANIONIC	GLU	23
ARYL	PHE	3	ANIONIC		
ARYL	TYR	8		**TOTAL**	**281**

Figure 2.5. Unique structures of standard amino acid fragments. Peptide backbone group is represented by glycine (GLY). HID, HIE, and HIP are δ-protonated, ε-protonated, and both δ-, ε-protonated (positively charged) isomers of a histidine residue. EFP fragments are characterized as aryl (TRP, HID, HIE, PHE, and TYR), polar (THR, ASN, GLN, and SER), aliphatic (ILE, LEU, VAL, and ALA), cationic (ARG, LYS, and HIP), and anionic (ASP and GLU) residues.

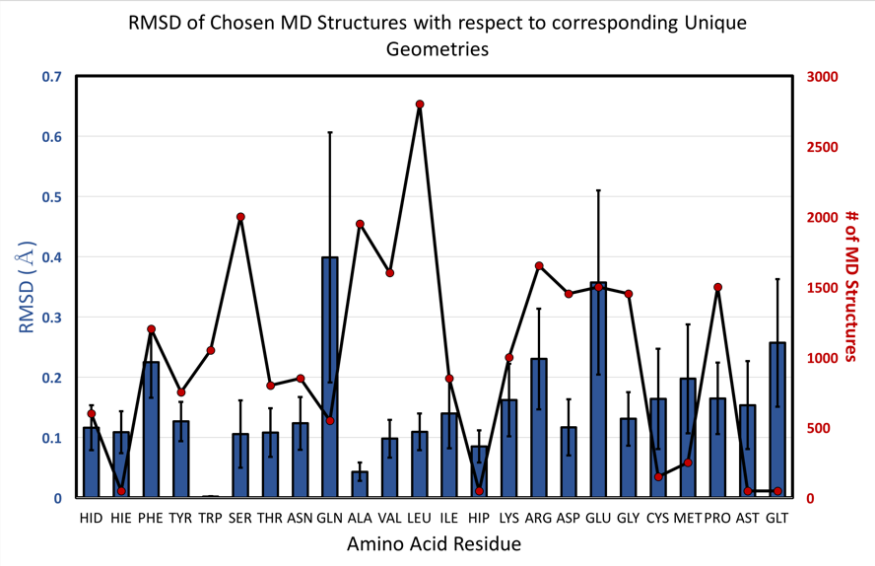

Figure 2.6. Average RMSD values (blue columns) and standard deviations of RMSD for each AA residue. The number of structures for each AA residue is shown with red dots.

2.4.2 Dimer calculations

To test the quality of the Flexible EFP scheme, we considered a set of dimers, in which each monomer is composed of a peptide group (backbone) and an AA residue (side chain), extracted from 25 Cry1At protein snapshots based on the closest atom distance between the AA residues. A threshold of 5 Å between the AAs resulted in 240 aryl-aryl, 533 aryl-aliphatic, 105 aryl-cationic, 49 aryl-anionic, 108 anionic-anionic, 305 anionic-cationic, and 388 polar-polar types of AA pairs (total 1728 pairs). Examples of considered dimers are shown in Figure 2.7. While the considered set does not include all possible interactions, we found it to be sufficiently large and diverse to be representative of protein noncovalent interactions. The total interaction energies and individual energy components of these dimers are compared between Flexible EFP, standard BioEFP, and the symmetry adapted perturbation theory (SAPT).

SAPT calculations: Sherrill and coworkers suggested the gold, silver, and bronze standards of SAPT methods based on the accuracy and computational cost of intermolecular interaction energy computations in several databases (S22, HBC6, NBC10, and HSG) [38]. Taking into consideration a low computational cost and similarity of the description of noncovalent interactions to EFP (Hartree-Fock level for fragments), we utilized sSAPT0/jun-cc-pVDZ (bronze standard) as a reference method [38]. SAPT interaction energies were computed with the PSI4 quantum chemistry package [71].

Standard BioEFP: For performing standard EFP calculations, individual EFP parameters were generated for all monomers in the dimer complexes, totaling in $1728 \times 2 = 3456$ fragment parameters. Each monomer was represented by a single fragment containing a peptide backbone and AA residue groups. In other words, EFP parameters were explicitly computed for each fragment geometry following the original BioEFP procedure as reported in REF [33]. Parameters were prepared in hybrid basis sets: 6-31G(d)/6-31(, +)G(3df,2p) for aryl compounds and 6-31G(d)/6-311++G(3df,2p) for the other monomers.

Flexible EFP: For Flexible EFP calculations, the parameters of each monomer were generated from the EFP parameter database as discussed in 2.4.1. Each monomer (except glycine and proline) is constructed from the peptide backbone and side chain fragments found in the database, using the smallest RMSD between the geometries of the target molecule and available database fragments as a selection criterion. Parameters of the selected database fragments are subjected to rotation and shifting to match the target geometry.

A comparison of workflows of standard BioEFP and Flexible EFP calculations is presented in Figure 2.8.

All EFP interaction energies were computed with the EFPMD module of the LibEFP software library [72,73].

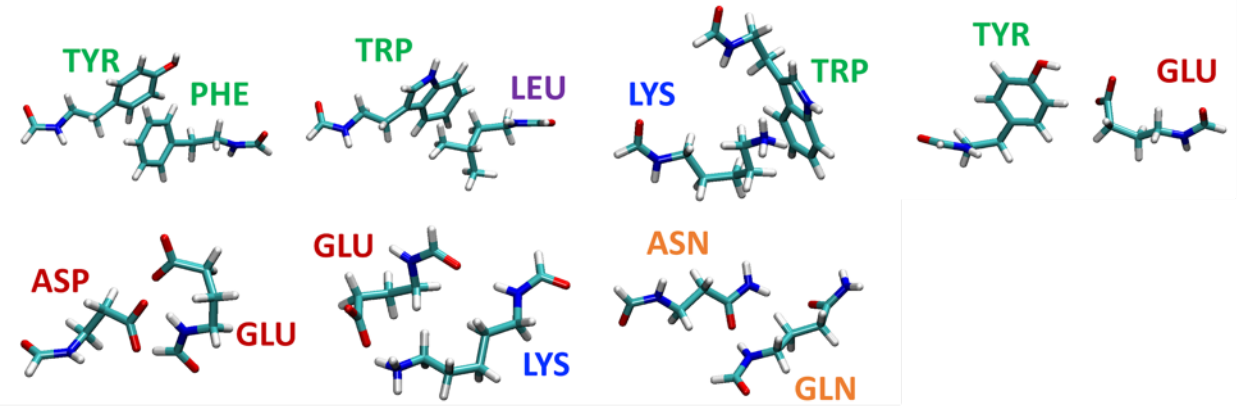

Figure 2.7. Representative structures of amino acid dimers.

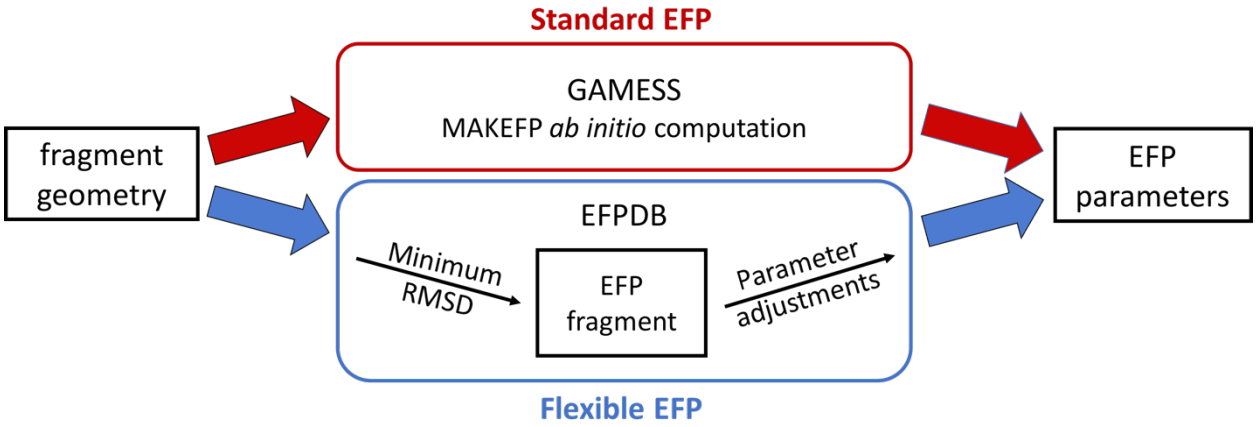

Figure 2.8. Workflow of Standard (Bio)EFP and Flexible EFP calculations.

2.4.3 EFP and QM/EFP calculations of cryptochrome Cry1At protein

To demonstrate the applicability of EFP to large-scale simulations, we performed a series of calculations on properties of cryptochrome Cry1At protein. First, using the pairwise energy decomposition, we computed the interaction energies of lumiflavin (a chromophore responsible for redox activity of Cry1At protein) with 483 AAs of the protein, excluding the interactions with terminal AAs, solvent molecules (water, Na^+ and Mg^{2+}) and cofactors (FAD^+ and ATP). In the

pairwise energy calculations, lumiflavin is represented as an EFP fragment, with the parameters prepared with hybrid 6-31G(d)/6-311++G(3df,2p) basis using the Standard EFP protocol. Standard EFP and Flexible EFP calculations were performed at 25 MD snapshots. AA parameters were prepared in hybrid basis sets: 6-31G(d)/6-31(,+)G(3df,2p) for aryl compounds and 6-31G(d)/6-311++G(3df,2p) for the other monomers as utilized in AA dimer calculations.

Then, ensemble-averaged values of vertical electron affinities <*VEA*> and vertical ionization potentials <*VIE*> were computed based on the same 25 MD snapshots that were utilized for EFP-only benchmarks. For the exact comparison of <*VEA*> and <*VIE*> with the values from REF [74], the 6-31G(d) basis set was employed to compute both electrostatic and polarization parameters of all EFPDB fragments. Additionally, the electrostatic expansion was truncated at the quadrupole level, eliminating octopoles from the expansion. The fragment-fragment electrostatic interactions were damped using exponential screening function, while the QM-fragment interactions were damped with Gaussian-type screening [33]. Gaussian-type polarization damping was employed with POLAB values of 0.3 for AAs and 0.1 for ions to avoid over-polarization of neighboring AAs [57]. Water molecules were described with EFP potentials prepared with hybrid 6-31+G(d)/6-311G++(3df,2p) basis sets. To speed-up QM/EFP calculations, the multipole parameters of water fragment were truncated at the quadrupole moments and distributed to atom centers only, while polarizability was represented with a single polarizability tensor located at the water center of mass as described in REF [74]. Following the Flexible EFP protocol, the EFP parameters of the protein AAs at 25 MD structures were adjusted based on the fragments from the EFPDB. Terminal AAs were prepared with the Standard EFP protocol. EFP parameters of solvent (water, Na^+, and Mg^{2+}) and cofactors (FAD^+ and ATP) were taken from the REF [74].

The ensemble-averaged vertical electron affinity <*VEA*> is computed in the ensemble of the oxidized form of lumiflavin; the ensemble-averaged vertical ionization energy <*VIE*> is computed in the ensemble of the reduced form as:

$$< VEA >=< E(OX) - E(RED) >_{OX} \quad (2.2)$$

$$< VIE >=< E(OX) - E(RED) >_{RED} \quad (2.3)$$

where $E(OX)$ and $E(RED)$ are total QM/EFP energies for oxidized and semireduced states of lumiflavin. The vertical energy gaps were computed at ωB97X-D/6-31G(d) level of theory using Q-CHEM electronic structure package [75].

2.5 Results

2.5.1 Interaction energies in amino acid dimers

Figure 2.9 and 2.10 provide a comparison of total interaction energies of original (Standard) and Flexible EFP with sSAPT0/jun-cc-pVDZ (referred to as "SAPT" in the following discussion) for aryl-aryl, aryl-aliphatic, aryl-cationic, aryl-anionic, anionic-anionic, anionic-cationic, and polar-polar dimers. Mean absolute errors (MAE), maximum errors (MAX), and minimum errors (MIN) of both EFP schemes against SAPT are also reported in Figure 2.9 and 2.10. Plots showing energy decomposition of these dimers are provided in Figure 2.11 – 2.17. The total interaction energies computed with the Flexible and Standard EFP methods exhibit similar values and trends (e.g., complexes with larger errors in Standard EFP typically have similarly larger errors in Flexible EFP). It is noteworthy that the two EFP schemes also exhibit similar values of energy components, meaning that shifting and rotating of the corresponding parameters does not introduce significant errors.

As follows from Figure 2.9 and 2.10, both Standard and Flexible EFP schemes overestimate attractive interactions in aryl-containing complexes, while interactions in other analyzed complexes (polar-polar, cationic-anionic, and anionic-anionic) are overly repulsive. The reason can be deduced from analysis of corresponding energy components. As can be seen in Figure 2.11 – 2.17, the polarization component is underestimated by magnitude (i.e., not sufficiently attractive) in most complexes, due to missing charge-transfer component and inaccuracies in charge-penetration term, resulting in systematic underestimation of the EFP binding energies. On the other hand, as follows from Figure 2.11 – 2.14, both Standard EFP and Flexible EFP significantly underestimate the exchange-repulsion component in complexes containing aryl-type residues, while exchange-repulsion is accurately represented in other complexes (Figure 2.15 – 2.17). This is due to the fact that the parameters for the exchange-repulsion term for aryls were prepared using 6-31(,+)G(3df,2p) basis while 6-311++G(3df,2p) basis was employed for the other compounds. As Figure 2.18 shows, employing 6-311++G(3df,2p) basis for the exchange-repulsion term in aryl compounds results in accurate exchange-repulsion energies. However, we noticed that rotation of diffuse *s* and *p* functions that compose exchange-repulsion parameters resulted in large-non-systematic errors in the exchange-repulsion term of Flexible EFP. Thus, we opted for using a smaller basis for aryl compounds when creating the parameter database, by a price of systematically underestimating exchange-repulsion energies (and

somewhat overestimating total binding energies) in aryl containing complexes. As a side note, the exchange-repulsion term is not included in a typical QM/EFP polarizable embedding model, such that inaccuracies in this component would not affect the performance of the Flexible EFP scheme combined with the QM region, as discussed below.

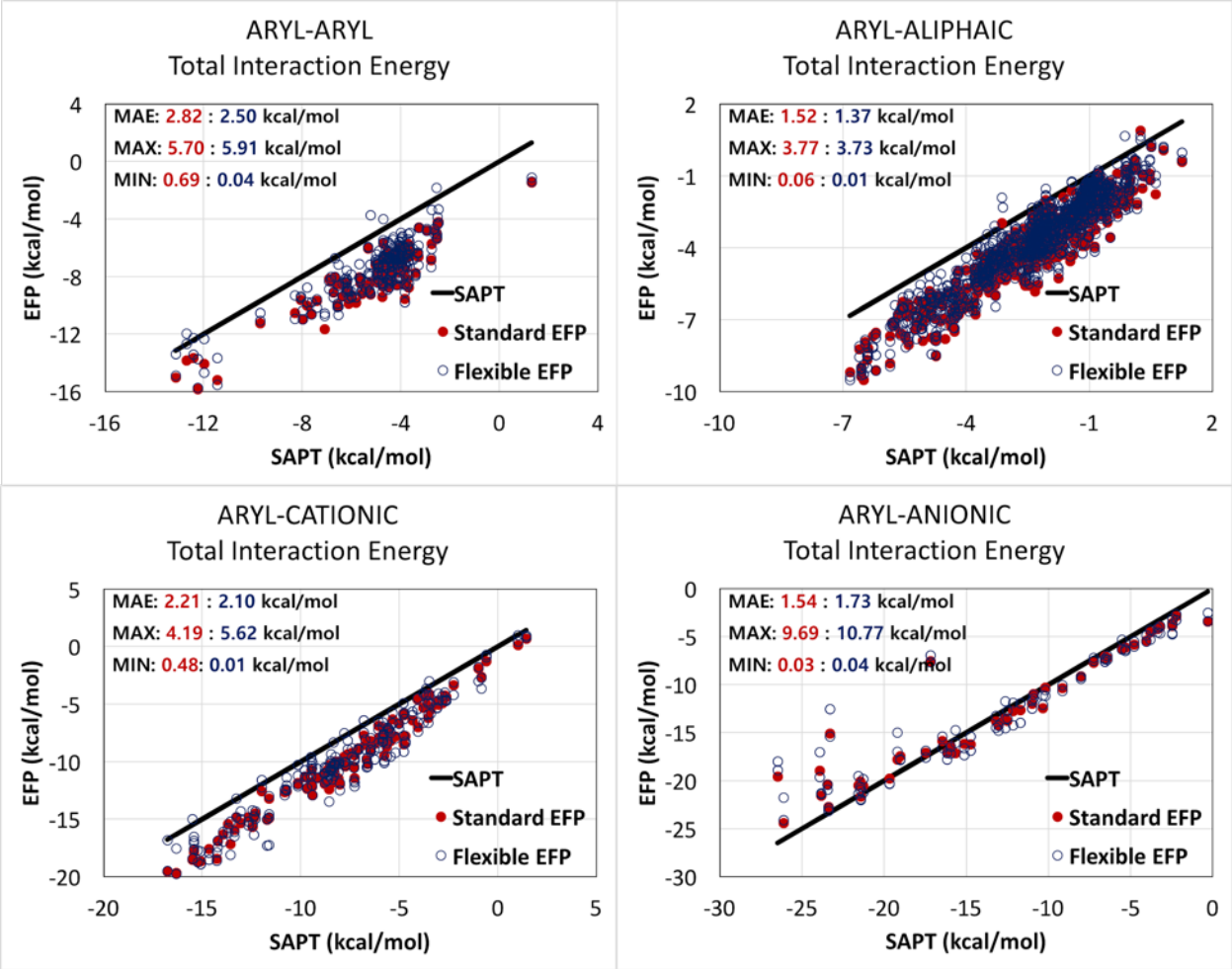

Figure 2.9. Total interaction energies in dimers containing aryl residues computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

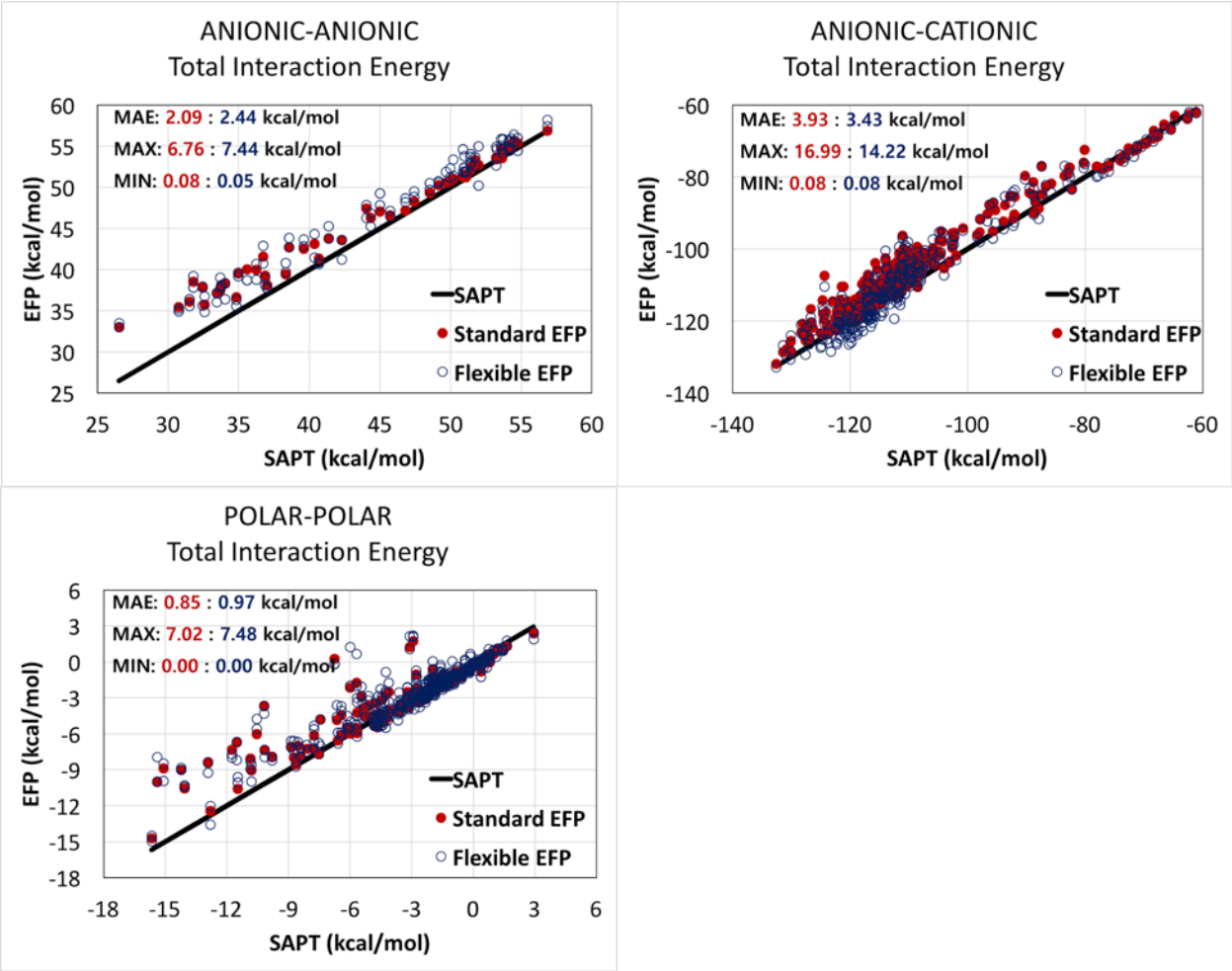

Figure 2.10. Total interaction energies in anionic-anionic, anionic-cationic, and polar-polar dimers computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

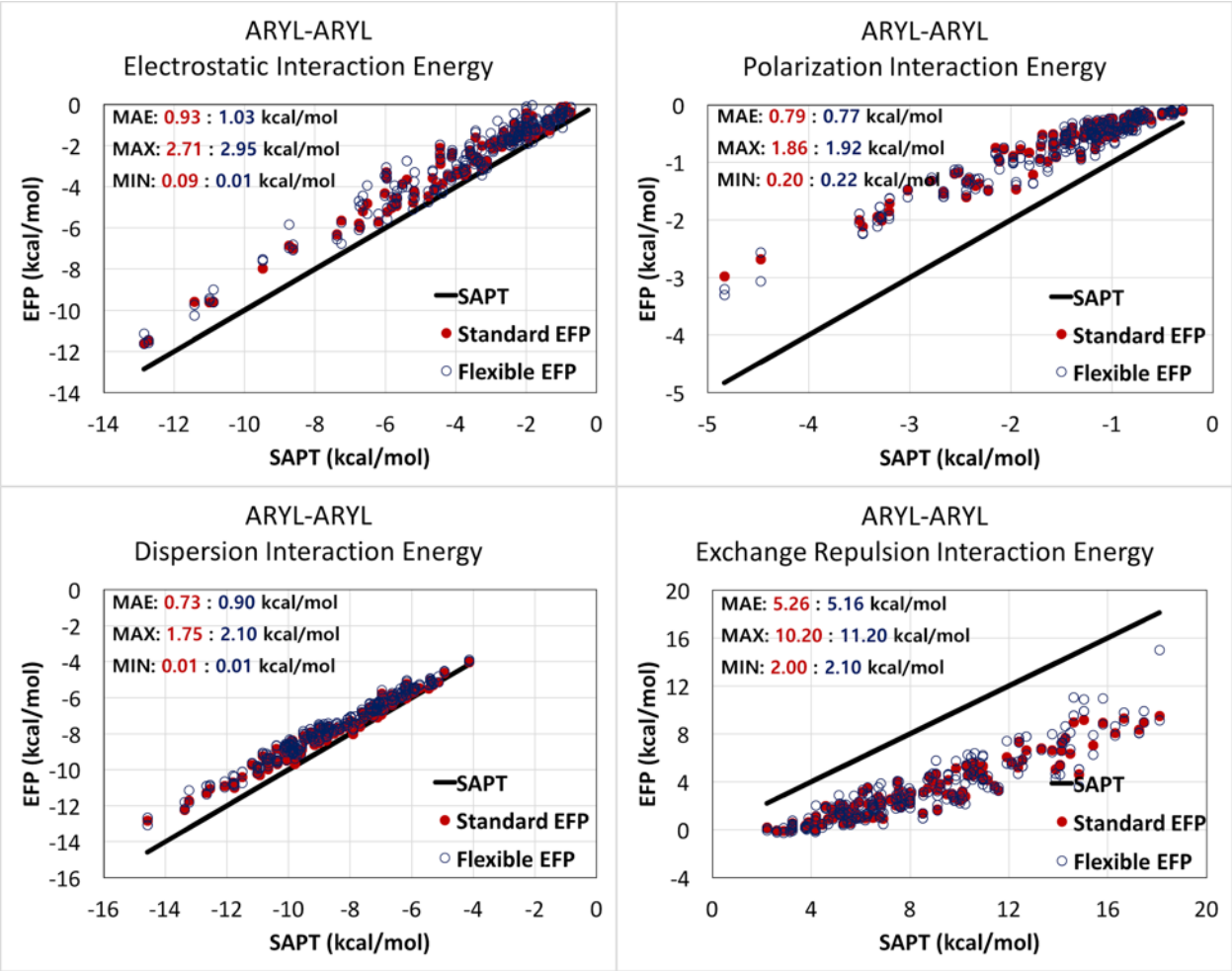

Figure 2.11. Interaction energy components in aryl-aryl complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

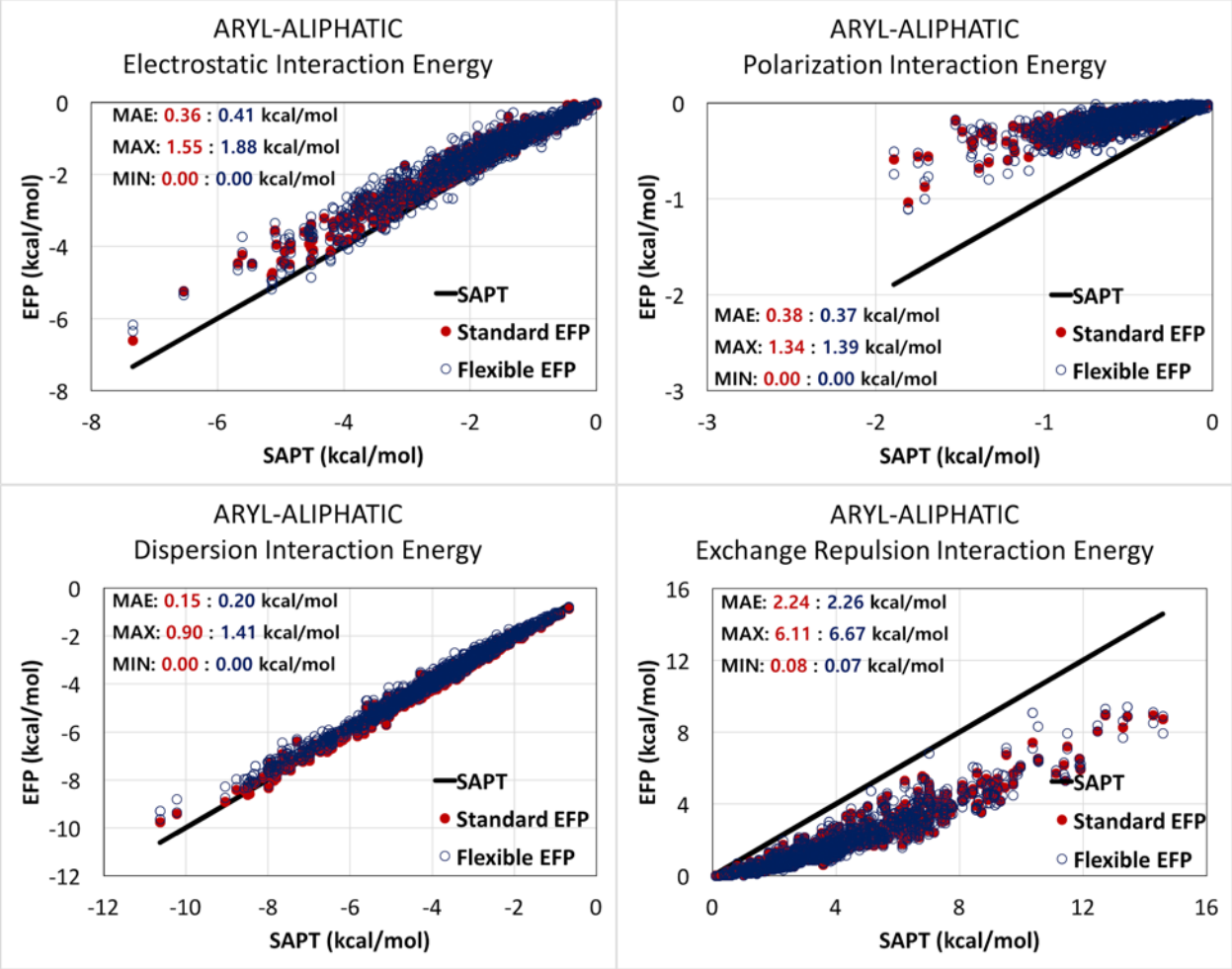

Figure 2.12. Interaction energy components in aryl-aliphatic complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

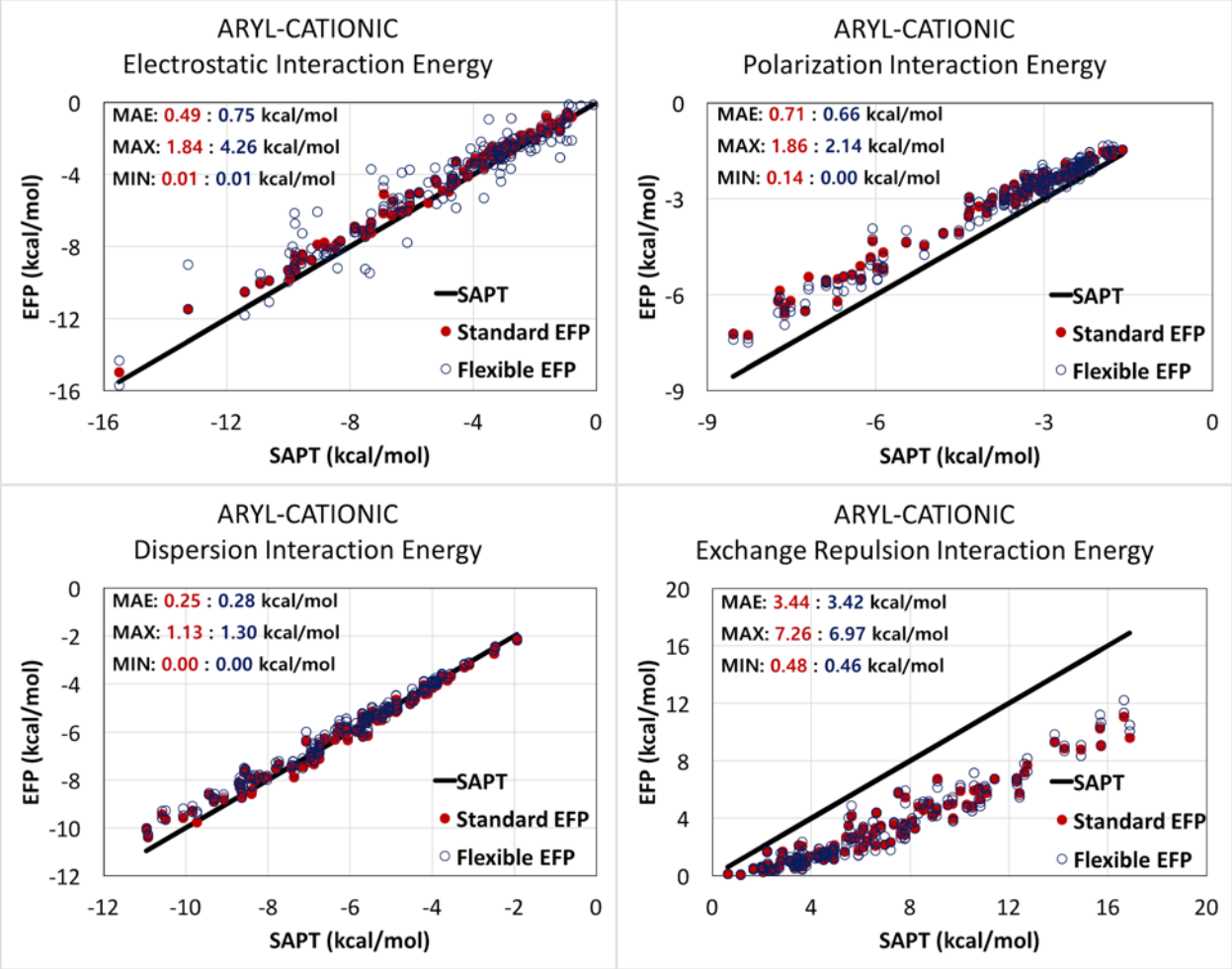

Figure 2.13. Interaction energy components in aryl-cationic complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

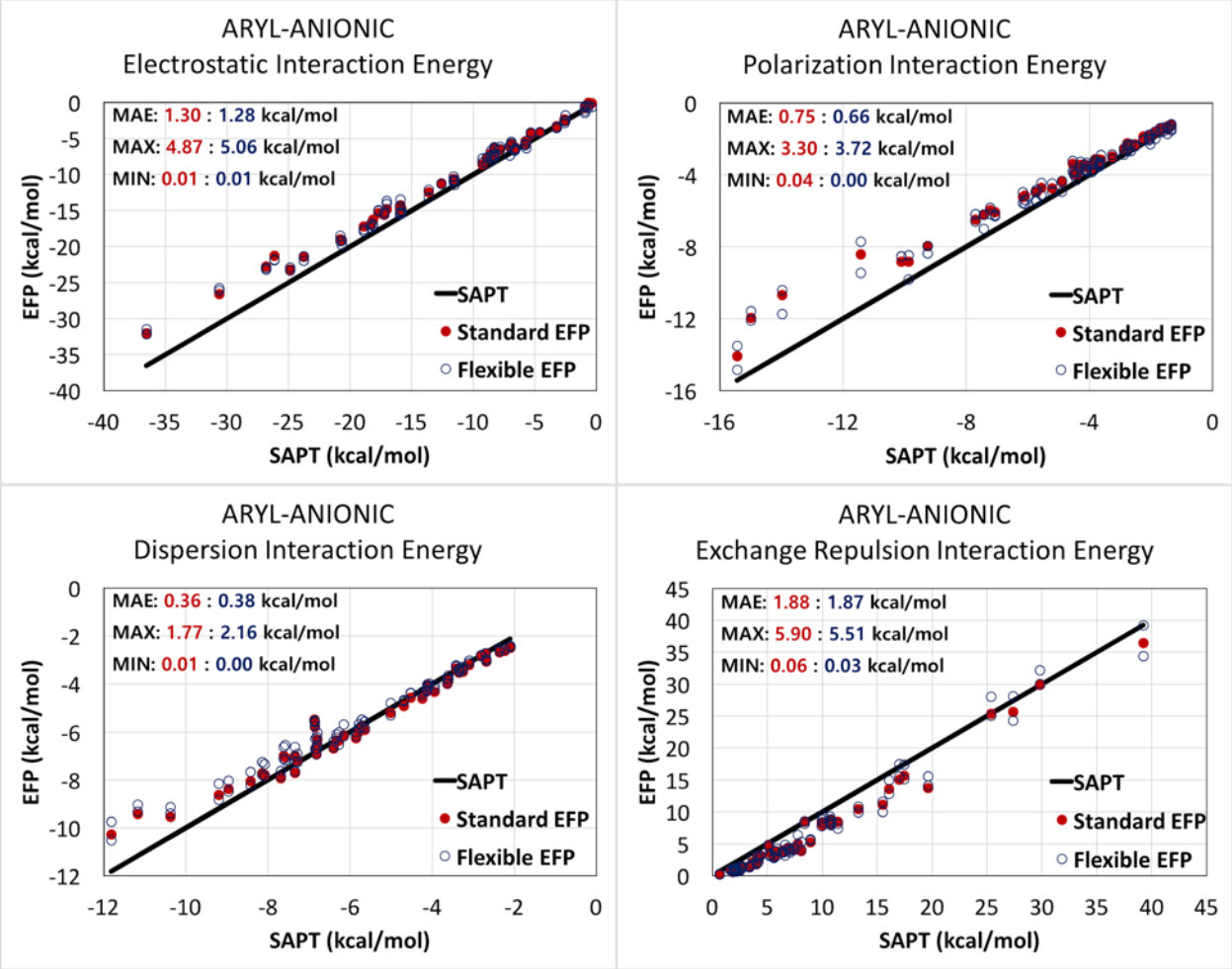

Figure 2.14. Interaction energy components in aryl-anionic complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

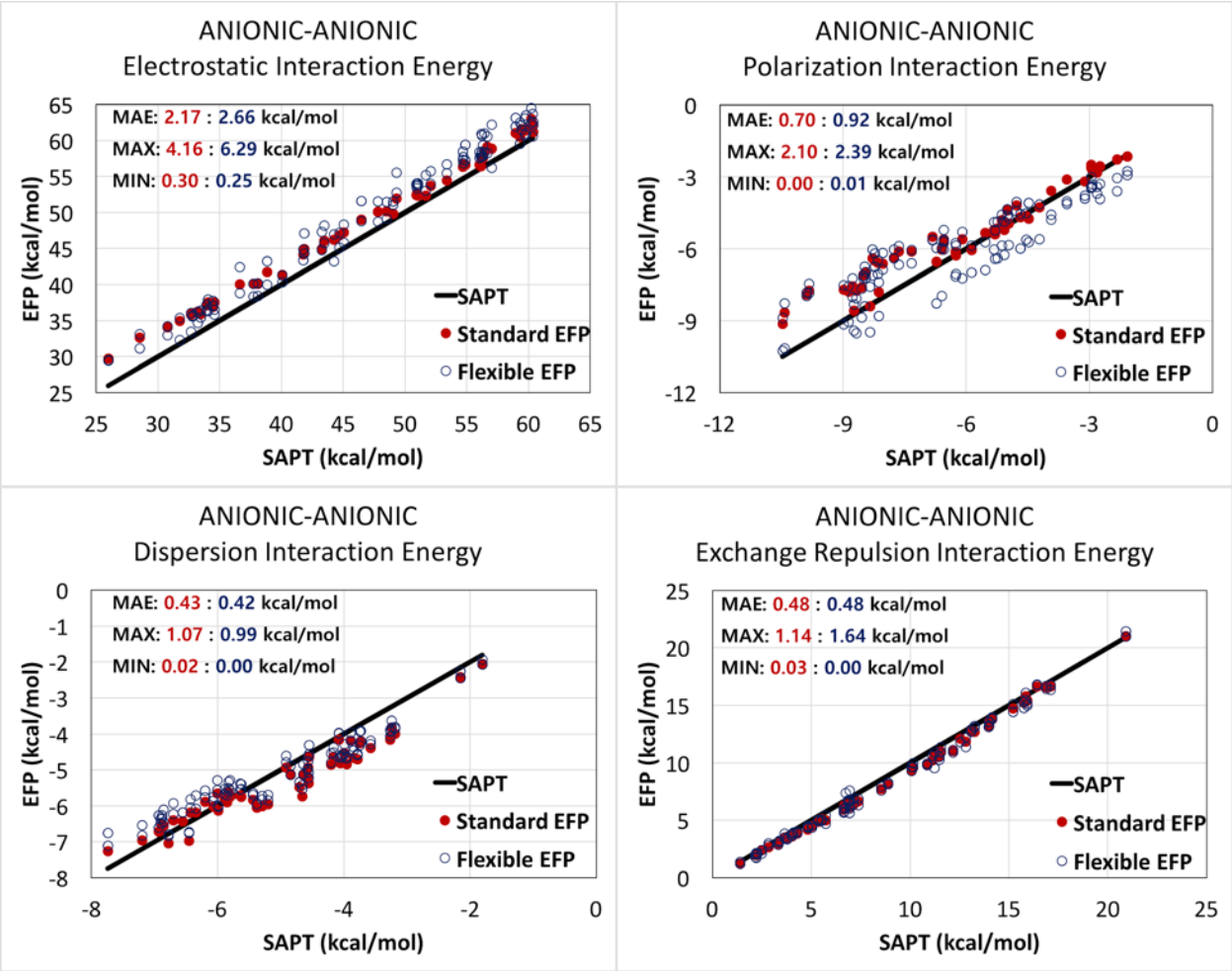

Figure 2.15. Interaction energy components in anionic-anionic complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

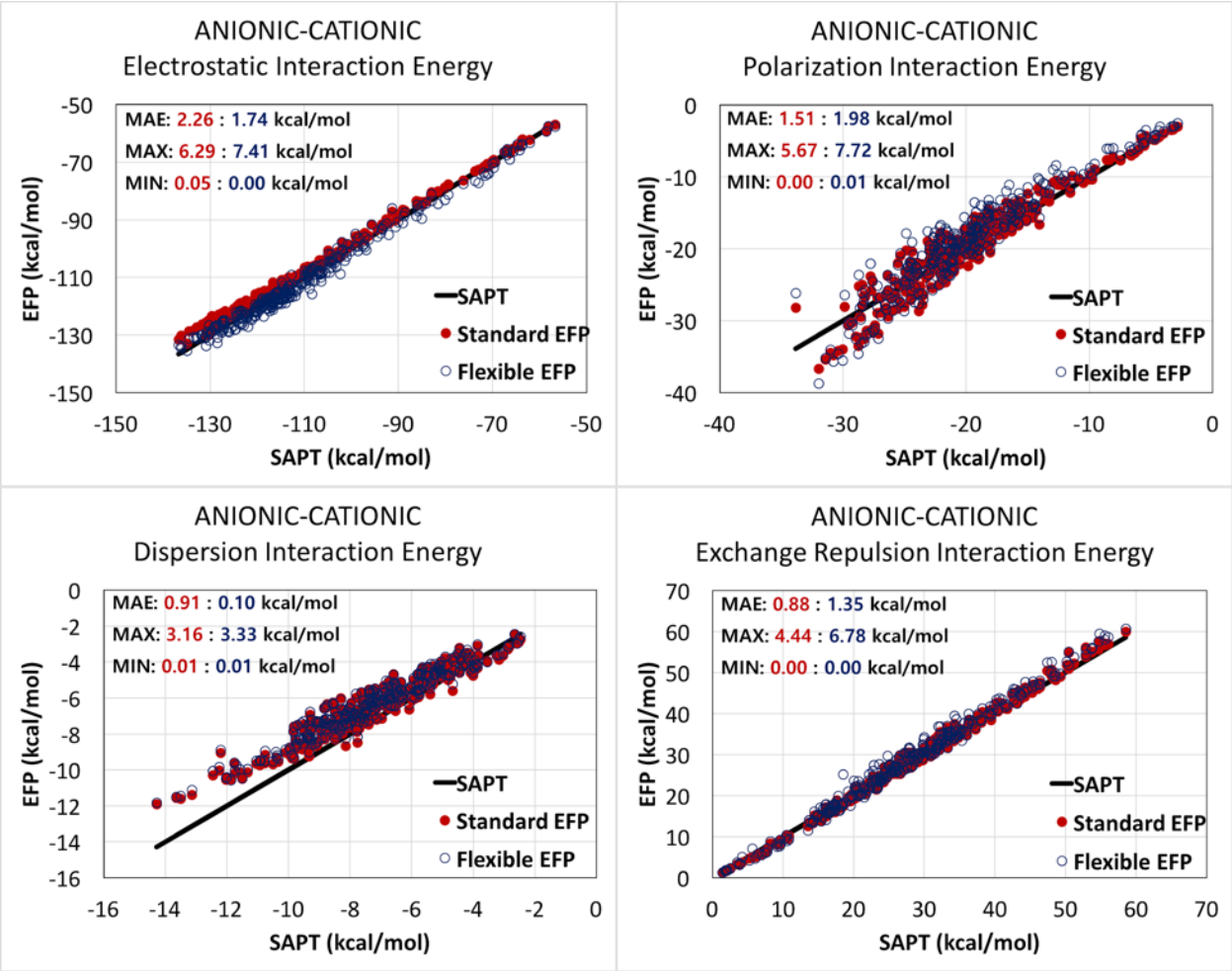

Figure 2.16. Interaction energy components in anionic-cationic complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

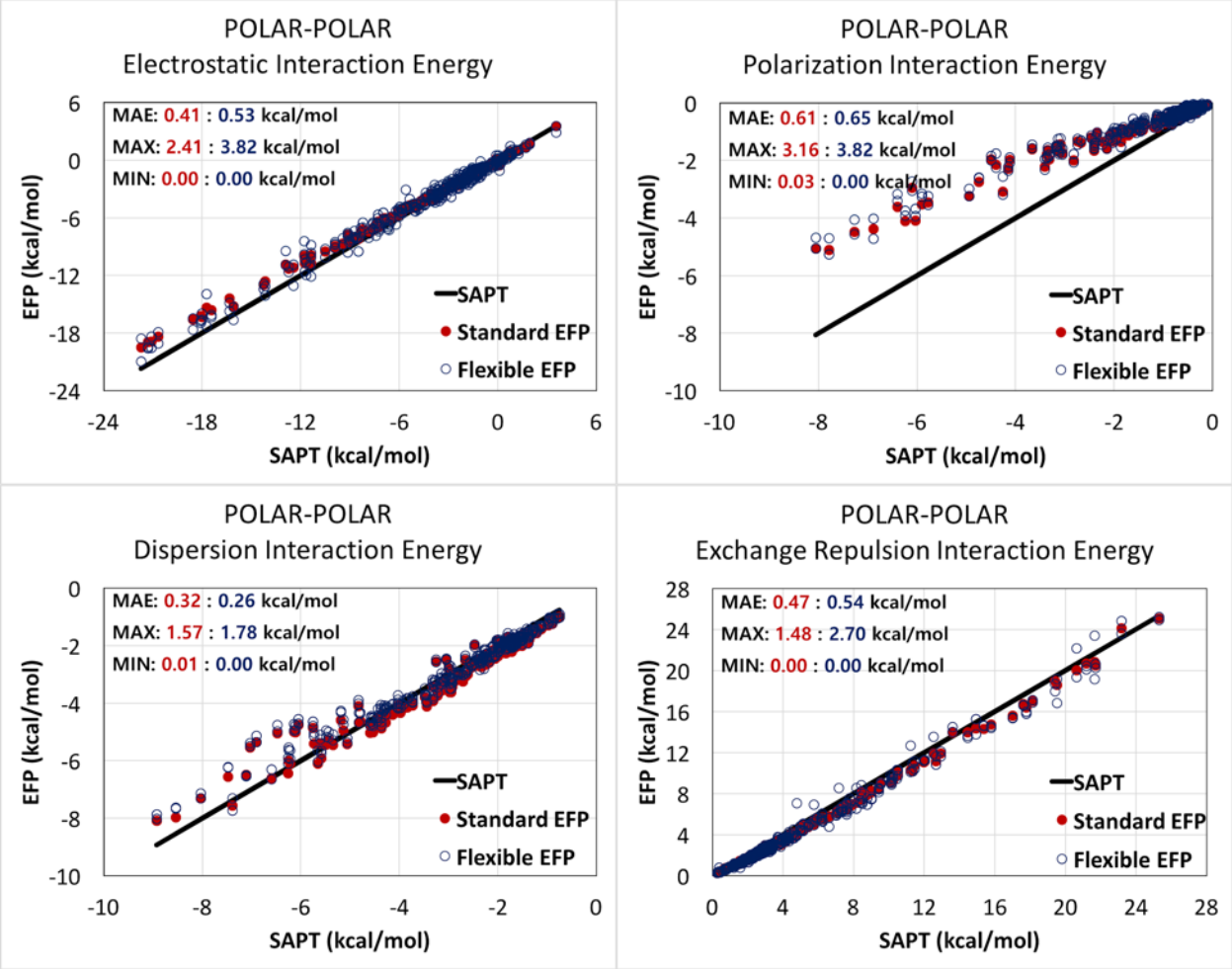

Figure 2.17. Interaction energy components in polar-polar complexes computed with Standard EFP (red circles) and Flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

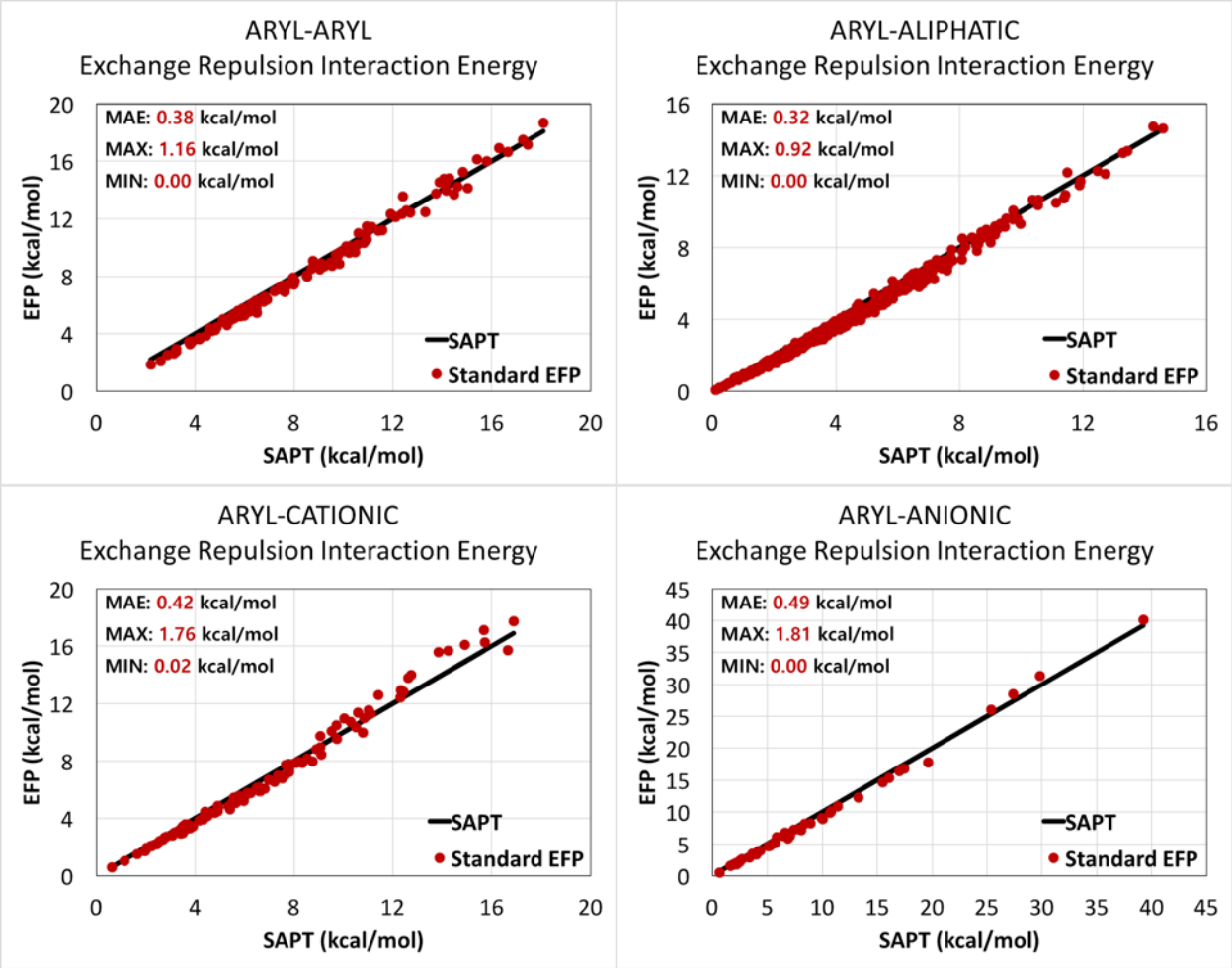

Figure 2.18. Exchange-repulsion energies in aryl-containing complexes computed with Standard EFP (red circles) against sSAPT0/jun-cc-pVDZ.

2.5.2 Binding energy of lumiflavin

A power of EFP for applications in biological chemistry is its ability to compute pairwise interactions in a fully polarized system. In such a way, for example, binding energies of a ligand in a protein pocket can be efficiently evaluated. To demonstrate this functionality of EFP and validate Flexible EFP for large-scale applications, we computed the interaction energy of lumiflavin (a chromophore responsible for redox activity of Cry1At protein) with 483 AAs of the protein excluding interactions with terminal AAs. Standard EFP and Flexible EFP calculations are performed at 25 MD snapshots of the protein.

The current version of the Flexible EFP parameter database does not contain non-standard molecules as well as terminal AAs due to the lack of sampling for the preparation of unique

geometries. Thus, the N-terminal and C-terminal AAs of the Cry1At protein were not considered in this binding energy benchmark study. Except for the terminal AAs, all other parameters were generated using Flexible EFP protocol as described in 2.2.1. and 2.3.2. In the case of lumiflavin molecule, the EFP potentials were prepared with hybrid 6-31G(d)/6-311++G(3df,2p) basis using Standard EFP protocol.

Comparison of the Standard EFP and Flexible EFP total interaction energies and energy components between lumiflavin and surrounding AAs are presented in Figure 2.19. Energies at different geometrical snapshots are shown as separate points. As in the case of dimer systems, the electrostatic, polarization, and exchange-repulsion energies computed with the Flexible EFP scheme are in excellent agreement with those obtained using the Standard BioEFP method. However, the dispersion energies in the Flexible EFP are significantly less attractive than those energies in the Standard EFP. As a result, the MAE of the total interaction energies of the Flexible EFP against Standard EFP is about 8.0 kcal/mol due to the poor agreement in the dispersion energies. However, it does not necessarily mean that the accuracy of Flexible EFP is worse than that of Standard BioEFP. For a better estimate of lumiflavin-protein binding energies, additional SAPT computations were carried out (see Figure 2.20). In these SAPT calculations, the lumiflavin-protein interaction energy was computed as a sum of interactions in all lumiflavin-AA dimers, where AAs were capped by H atoms at C and C_α sites (i.e., fragmentation was done analogously to the BioEFP scheme). While such SAPT calculations cannot fully reproduce the polarization component of the total interaction energy, electrostatic, dispersion, and exchange-repulsion terms do not incorporate strong many-body effects and should be well reproduced by this additive scheme [76]. Interestingly, as shown in Figure 2.20, Flexible EFP compares better to SAPT than the Standard EFP, demonstrating almost perfect agreement in dispersion energies. The main difference in dispersion energies between the two EFP schemes is that there is no dynamic polarizability point on a border between the peptide group and AA side chain fragments (i.e., dynamic polarizability of the localized $C_\alpha - C_\beta$ bond) in Flexible EFP, while such points are present in fragments in the Standard EFP scheme. These boundary points containing static and dynamic polarizabilities are removed between all covalently-linked fragments to ensure stability of polarization self-consistent procedure. Indeed, both Standard and Flexible EFP schemes miss such points between each pair of C and C_α atoms in a peptide chain. While the dynamic polarizability points at the boundary between fragments could be kept in one of the participating

fragments (as the dispersion interactions do not have convergence issues), we leave exploring this possibility to future work.

Returning to the analysis of Figure 2.19 and 2.20, it is clear that the Standard and Flexible EFP schemes perform consistently between different geometrical snapshots, both for the energy components and for the total interaction energies. Additionally, with exception of systematic and well understood discrepancies in the dispersion component, the discrepancies between Standard and Flexible EFP are smaller than the errors of EFP against SAPT, justifying utilization of the Flexible EFP model in large-scale biological applications

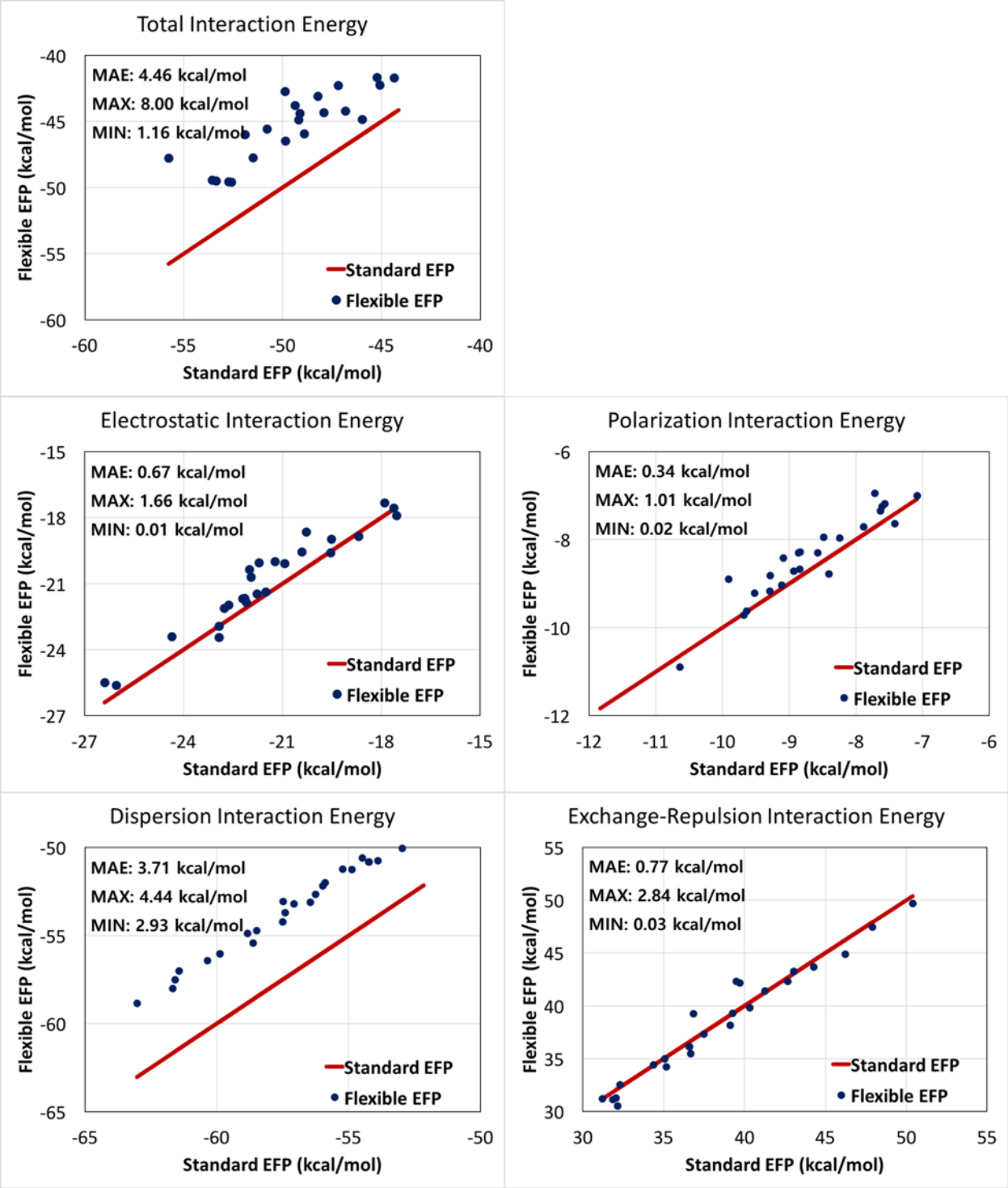

Figure 2.19. Comparison of the total interaction energies and energy components of the lumiflavin chromophore with the protein, computed at 25 MD snapshots by the Standard and Flexible EFP.

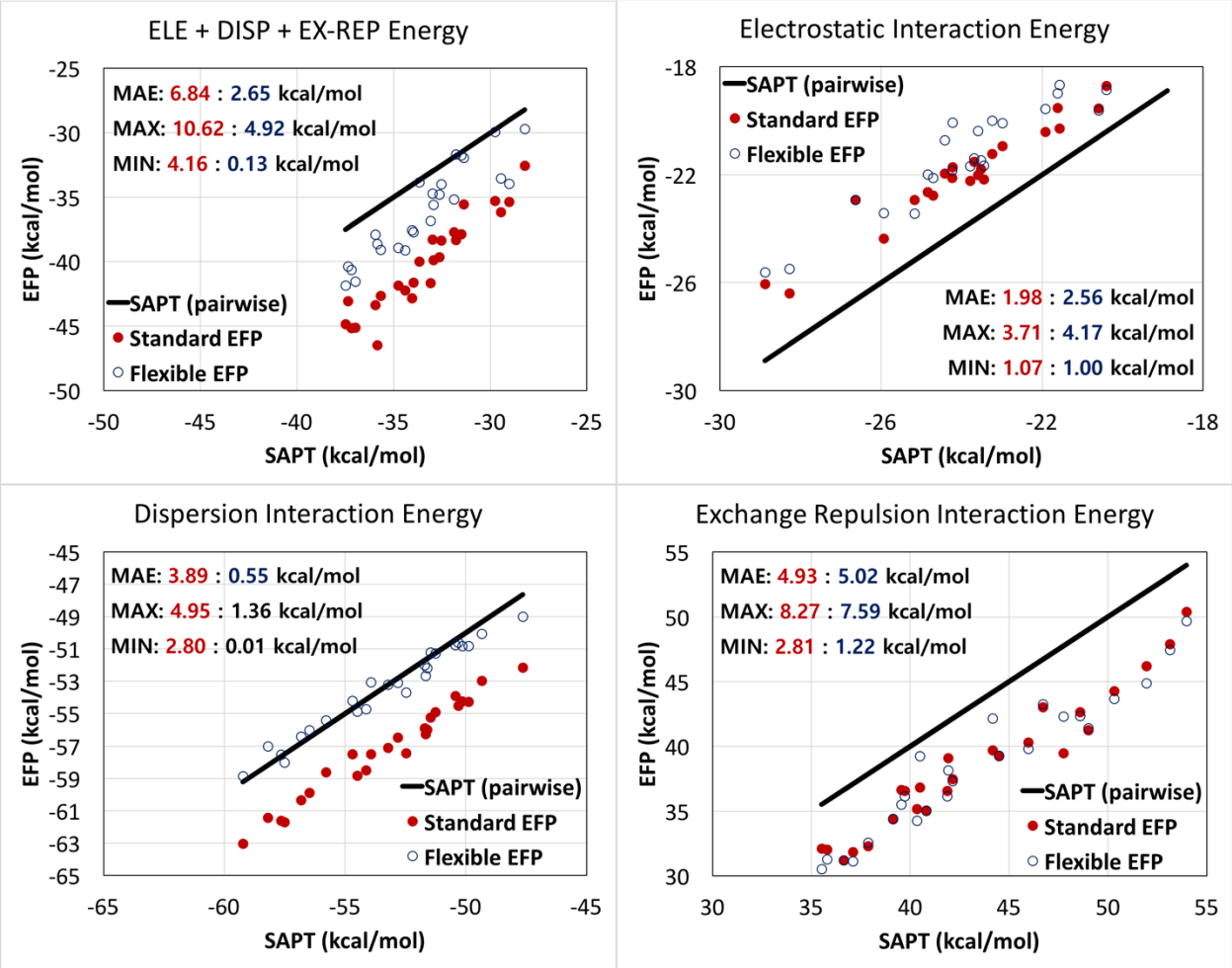

Figure 2.20. Comparison of pairwise interaction energy components in the lumiflavin-protein system. Red circles: Standard EFP; blue circles: Flexible EFP. Comparison is done against sSAPT0/jun-cc-pVDZ.

2.5.3 Application of Flexible EFP in QM/EFP modeling

At last, we examine the applicability of the Flexible EFP scheme in a framework of the polarizable embedding QM/EFP model. Here we examine two electronic properties, ensemble-averaged vertical electron affinity <*VEA*> and ensemble-averaged vertical ionization energy <*VIE*>, which have been recently reported for lumiflavin in Cry1At protein [74]. In these calculations lumiflavin is treated quantum-mechanically, while the interaction between lumiflavin and protein are described with EFP electrostatic and polarization terms (Eqns. (1.2) and (1.3)). The overall setup of the system is identical to the one reported in REF [74], and is pictorially represented in Figure 2.21. Following the Flexible EFP protocol, the EFP parameters of the protein AAs at 25 MD structures were adjusted based on the fragments from the EFPDB. Terminal AAs

were still prepared with the Standard EFP protocol. EFP parameters of solvent (water, Na^+, and Mg^{2+}) and cofactors (FAD^+ and ATP) were taken from REF [74]. We note that the goal of this benchmark is not to obtain the best description of lumiflavin within the protein but rather to demonstrate a close agreement between the Standard and Flexible EFP schemes.

As seen in Table 2.1, the average discrepancies between the Standard and Flexible EFP schemes are ~0.029 eV (0.67 kcal/mol) and ~0.044 eV (1.01 kcal/mol) for <*VEA*> and <*VIE*>, respectively, which is almost an order of magnitude smaller than the uncertainties (standard deviations) in these values. Specific data on each MD snapshot, as well as a brief description of the computational protocol for computing ensemble-averaged VIE and VEA are provided in Table 2.2. Overall, this example demonstrates that the Flexible EFP scheme is reliable in biological simulations and can be safely used for modeling electronic properties of biochromophores.

It should be emphasized that the Flexible EFP scheme provides enormous computational savings compared to Standard EFP as the necessity to recompute parameters for hundreds of fragments at different geometries is avoided. For example, for computing <*VIE*> and <*VEA*> in the cryptochrome protein, using the Flexible EFP protocol and the EFPDB database allowed to avoid generation of 12075 EFP parameters and saved thousands of CPU hours. Additional critical benefit of using Flexible EFP and EFPDB is a significant simplification of setup of EFP simulations and avoiding human errors in preparing, storing and book-keeping parameters of multiple fragments.

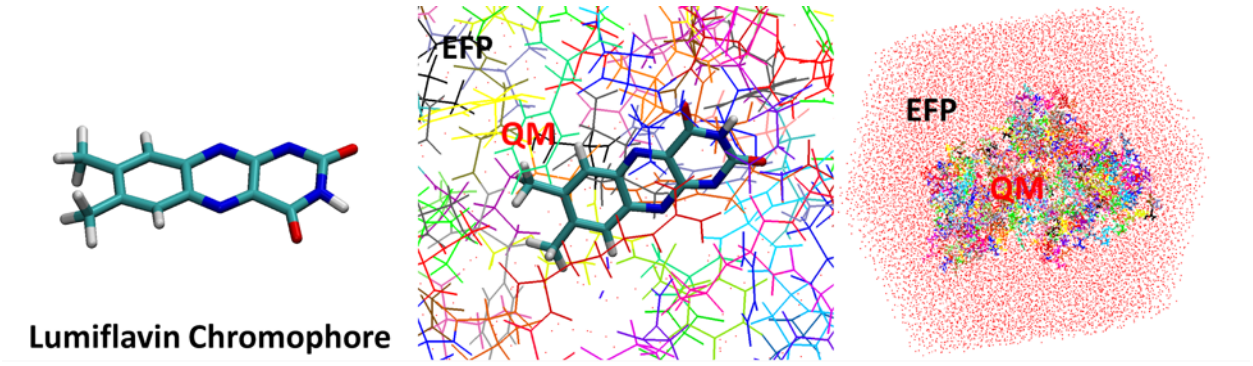

Figure 2.21. Scheme of QM/EFP calculations of Cry1At protein.

Table 2.1. <*VEA*> and <*VIE*> of lumiflavin in Cry1At cryptochrome protein, computed with Standard and Flexible EFP schemes. Ensemble-averaged values with standard deviations, as well as discrepancies between Standard and Flexible EFP scheme are provided

	<*VEA*>, eV		<*VIE*>, eV	
EFP Scheme	Standard EFP	Flexible EFP	Standard EFP	Flexible EFP
Ensemble-averaged values +/- standard deviations	2.407±0.262	2.378±0.261	4.926±0.214	4.882±0.215
Average error	0.029±0.038		0.044±0.037	
Max error	0.130		0.112	
Min error	0.000		0.000	

Table 2.2. VEA and VIE at each MD structure computed with the Standard and Flexible EFP schemes. OX and SEMI correspond to MD simulations for oxidized and semi-reduced forms of lumiflavin, respectively

Ensemble	VEA, eV		Ensemble	VIE, eV	
OX	Standard EFP	Flexible EFP	SEMI	Standard EFP	Flexible EFP
snap_01	2.324	2.280	snap_01	4.993	4.907
snap_02	2.703	2.642	snap_02	4.766	4.764
snap_03	2.768	2.772	snap_03	4.577	4.467
snap_04	2.390	2.383	snap_04	4.882	4.835
snap_05	2.658	2.655	snap_05	4.978	4.896
snap_06	2.431	2.372	snap_06	4.861	4.840
snap_07	2.547	2.503	snap_07	4.893	4.892
snap_08	2.573	2.568	snap_08	5.083	5.036
snap_09	2.056	2.053	snap_09	5.080	5.080
snap_10	2.630	2.603	snap_10	4.858	4.784
snap_11	2.694	2.664	snap_11	5.221	5.208
snap_12	2.480	2.426	snap_12	4.752	4.750
snap_13	2.035	2.008	snap_13	4.651	4.657
snap_14	2.876	2.902	snap_14	5.189	5.171

(Table 2.2 continued)

Ensemble	VEA, eV		Ensemble	VIE, eV	
OX	Standard EFP	Flexible EFP	SEMI	Standard EFP	Flexible EFP
snap_15	2.125	2.099	snap_15	5.080	5.038
snap_16	2.359	2.365	snap_16	5.112	5.030
snap_17	2.449	2.383	snap_17	5.279	5.271
snap_18	2.702	2.573	snap_18	5.261	5.148
snap_19	2.407	2.342	snap_19	4.965	4.880
snap_20	2.297	2.220	snap_20	4.771	4.744
snap_21	1.890	1.913	snap_21	4.468	4.436
snap_22	2.337	2.379	snap_22	4.925	4.929
snap_23	2.207	2.156	snap_23	4.982	4.916
snap_24	1.991	1.935	snap_24	4.892	4.802
snap_25	2.253	2.251	snap_25	4.637	4.577

2.6 Conclusions

In the present work, the concept and implementation of the Flexible EFP model are reported. The Flexible EFP formalism is based on connecting geometries of two chemically identical but structurally different fragments by a series of local coordinate rotations and translations. The distributed parameters associated with atoms, bond midpoints and LMO centroids of one fragment are subjected to the corresponding reorientations, producing a new set of parameters that are suitable for describing the fragment in a new geometry. To utilize this approach in studies of proteins, we developed the EFPDB parameter database that contains EFP parameters of standard AAs. We validated the Flexible EFP scheme by considering interaction energies in the AA dimers and showed that Flexible EFP is at least as accurate as Standard EFP with respect to sSAPT0/jun-cc-pVDZ. Applications of Flexible EFP scheme to binding energies as well as vertical electron attachment and ionization energies in the cryptochrome protein demonstrate substantial benefits of utilizing Flexible EFP in realistic simulations of biological systems. While the computational cost of obtaining EFP parameters becomes negligible in the Flexible EFP scheme, the predicted ensemble-averaged $\langle VEA \rangle$ and $\langle VIE \rangle$ deviate from those obtained with the Standard EFP by only 0.029 and 0.044 eV, respectively. Hence, the Flexible EFP approach represents a

computationally efficient way of conducting large-scale EFP simulations without compromising accuracy.

While the Flexible EFP allows transferring the EFP parameters between different fragment geometries, the utilized model in which the distributed parameters are merely shifted and rotated but not otherwise adjusted to reflect changes in the electronic structure of a fragment at a new geometry might break in the case of significant geometrical rearrangements of a fragment. To address these concerns, the Flexible EFP scheme might be further refined by accounting for changes in parameter magnitudes upon structure rearrangements of the fragment, for example by expanding the parameter elements in a Taylor series around the equilibrium position. These ideas might be explored in future studies.

CHAPTER 3. PREDICTIVE FIRST-PRINCIPLES MODELING OF A PHOTOSYNTHETIC ANTENNA PROTEIN: THE FENNA-MATTHEWS-OLSON COMPLEX

The following chapter is published in a peer reviewed journal.
DOI: 10.2012/acs.jpclett.9b03486

3.1 Abstract

High efficiency of light harvesting in photosynthetic pigment-protein complexes is governed by evolutionary-perfected protein-assisted tuning of individual pigment properties and inter pigment interactions. Due to the large number of spectrally overlapping pigments in a typical photosynthetic complex, experimental methods often fail to unambiguously identify individual chromophore properties. Here, we report a first-principles-based modeling protocol capable of predicting properties of pigments in protein environment to a high precision. The technique was applied to successfully uncover electronic properties of the Fenna-Matthews-Olson (FMO) pigment-protein complex. Each of the three subunits of the FMO complex contains eight strongly coupled bacteriochlorophyll a (BChl a) pigments. The excitonic structure of FMO can be described by an electronic Hamiltonian containing excitation (site) energies of BChl a pigments and electronic couplings between them. Several such Hamiltonians have been developed in the past based on the information from various spectroscopic measurements of FMO; however, fine details of the excitonic structure and energy transfer in FMO, especially assignments of short-lived high-energy sites, remain elusive. Utilizing polarizable embedding quantum mechanics/molecular mechanics with the effective fragment potentials, we computed the electronic Hamiltonian of FMO that is in general agreement with previously reported empirical Hamiltonians and quantitatively reproduces experimental absorption and circular dichroism spectra of the FMO protein. The developed computational protocol is sufficiently simple and can be utilized for predictive modeling of other wild-type and mutated photosynthetic pigment-protein complexes.

3.2 Introduction

Plants, algae, and photosynthetic bacteria capture solar radiation by means of pigment-protein antenna complex. A variety of nature antenna complexes evolved to maximize

photosynthetic efficiency in different environments. The light harvesting and energy transfer efficiency of these complexes approaches 100% and is governed by electronic properties of individual light-absorbing pigments as well as by couplings between the pigments. While molecular structures of many antenna complexes have been determined by X-ray diffraction, the information about electronic energy levels and energy transfer dynamics often lacks desired precision as it is primarily deduced from optical data. Most of the antenna complexes contain large number of pigments (up to ~250,000 in chlorosome antenna [77]) with overlapping optical spectra, leading to spectral congestion that precludes unambiguous identification of properties of individual pigments and, as a result, multiple models can be proposed to fit the same data.

The Fenna-Matthews-Olson (FMO) pigment-protein complex, found in green sulfur bacteria, is one of the most thoroughly studied photosynthetic pigment-protein complex (see Figure 3.1) [78-112]. The primary function of FMO is to transfer the excitation energy from a much larger chlorosome antenna to the intramembrane reaction center complex, where electronic excitation initiates charge transfer process.

The FMO complex is a trimer possessing C_3 symmetry. Each subunit encloses seven bacteriochlorophyll a (BChl a) chromophores and binds the eighth BChl a pigment between the subunits. Close packing of BChl a pigments in each monomer subunit leads to strong excitonic interactions and delocalization of excited states over multiple pigments. A relative structural simplicity combined with intricate excitonic structure makes the FMO complex a favorite object for developing and testing new computational and experimental techniques. For example, FMO was the first pigment-protein complex for which quantum coherences and beatings between excitonic states were observed by Savikhin et al. in 1997 [113]. The study of excitonic structure and coherence in this complex led later to the development of a two-dimensional spectroscopy [79,83,114,115]. To model excitonic interactions and energy transfer in FMO, a number of different electronic Hamiltonians have been proposed [79,83,85,87,89,96,98,110,116], most of which were obtained by a combination of structural data and fits to available experimental measurements. While the off-diagonal elements of a Hamiltonian represent electronic couplings between pigments and can be estimated using dipole-dipole approximation from available X-ray structures, the diagonal energies (individual transition energies) cannot be observed directly in experiment and are inferred from fits to spectroscopic data. This approach results in a significant ambiguity in determination of BChl a site energies and typically does not account for environment-

driven variations in transition dipole moments of individual pigments and inter-pigment couplings. While molecular modeling based on structural data could provide unambiguous assignment of excitonic interactions and energy flow in FMO, internal complexity of the system and necessity to sample protein degrees of freedom and accurately describe electronic structure and couplings between BChl a pigments and vibronic interactions between pigments and the protein environment makes this task challenging [82,85-88,92,93,99,101,103,108,109,117-119]. Here we report a multiscale first-principles modeling that accurately reproduces the absorption and circular dichroism (CD) spectra of the FMO complex based solely on its X-ray structure and uses no other input from experiments.

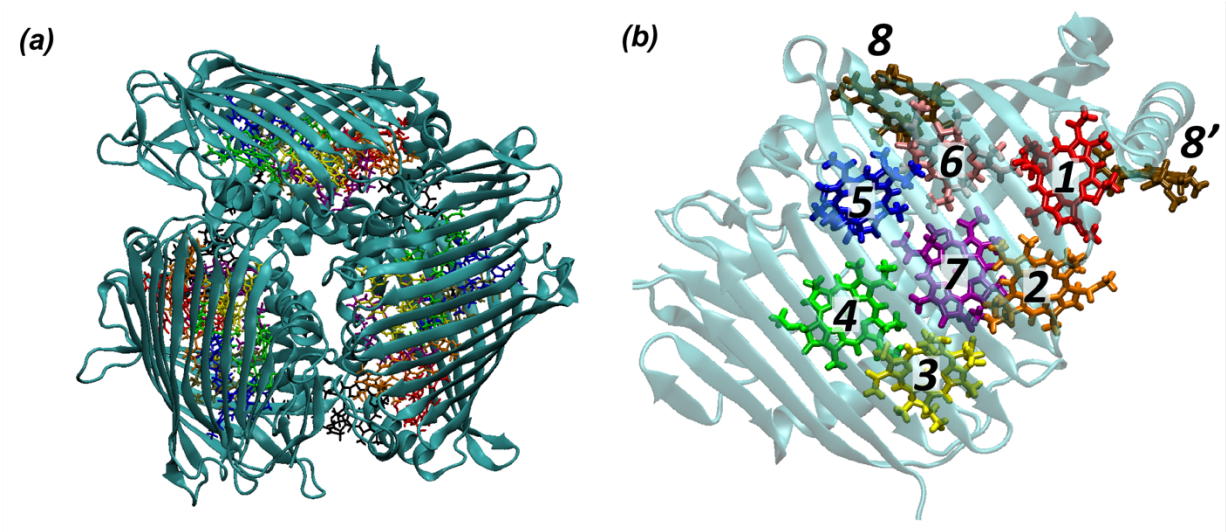

Figure 3.1. (a) FMO trimer of the *Chlorobaculum Tepidum*; (b) FMO monomer with BChl a pigments.

3.3 Computational Details

3.3.1 Preparation of the system and molecular dynamics simulations

The Fenna-Matthews-Olson (FMO) complex in this study is based on the crystal structure of *Chlorobaculum Tepidum* obtained from protein data bank (PDB: 3ENI [109]). Based on the FMO crystal monomer unit, the initial trimeric structure was constructed using C_3 symmetric transformation. The initial trimeric structure of the FMO complex was automatically protonated using the ANTICHAMBER module of AMBER package [120] with neutral pH condition. All histidine residues were protonated at δ position, except for HIS 296 that was protonated at ε

position and HIS 12 protonated at both positions (positively charged) in order to maintain axial binding to MG atoms of BChl a pigments. The prepared crystal structure was immersed in a water cubic box. 6 Cl^- ions were added to neutralize total system charge. Overall, the system consisted of 24 BChl a pigments, 72008 water molecules, and 6 Cl^- counter ions in a simulation box size of approximately $13.5 \times 13.5 \times 13.5$ nm^3 with periodic boundary conditions. The force field for BChl a was obtained from the previous study by Ceccarelli et al [121]. This force field was utilized in previous studies of photosynthetic reaction center [122,123]. AMBER03 force field [124] was used for standard amino acids (AAs), GAFF force field for counter ions and TIP3P model for water [125,126]. After energy minimization, molecular dynamics (MD) equilibrations were performed with the NVT and NPT ensembles for 500 ps each. Then, the production run with the NPT ensemble was carried out with a velocity rescale thermostat [127] for temperature control (300 K) and Parrinello-Rahman barostat [128] for pressure control (1 bar). C-H, O-H and N-H bond lengths were constrained with the LINCS algorithm [129]. A 1 nm cutoff was used to handle Lennard-Jones potentials. Electrostatic long-range interactions were treated with particle mesh Ewald summations (PME) [130,131] with a real-space cutoff of 1 nm. Total 80 ns production run was carried out with 2 fs time step. All classical simulations were performed with the GROMACS package (version 2016.5) [132].

The root mean square deviation (RMSD) of the protein backbone is plotted to verify a proper equilibration of the FMO structure (see Figure 3.2). The RMSD of the protein backbone quickly reached its equilibrium state and is settled at ~0.17 nm (with respect to the crystal structure) after 2 ns. We randomly extracted 100 atomic configurations from the last 30 ns of the MD trajectory without bias. Specifically, 30,000 MD structures were saved to disk (one structure every 1 ps). Out of those, 100 snapshots were selected using a random number generator in the range from 1 to 30,000. These 100 random configurations have been considered for the electronic structure calculations. For each configuration, the water molecules outside 15 Å from the outer protein surface were truncated.

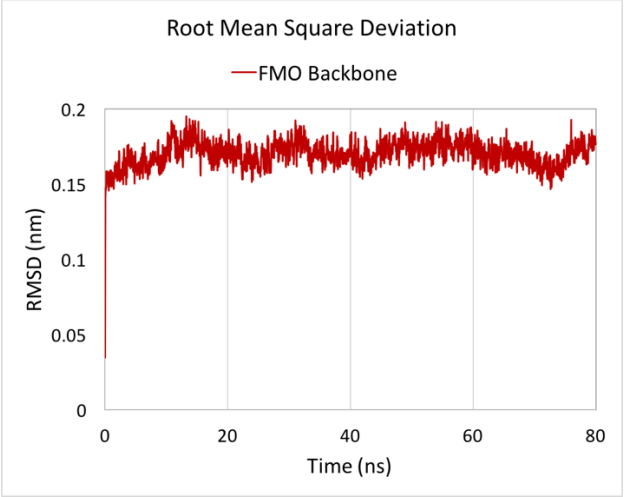

Figure 3.2. RMSD of a protein backbone as a function of time in MD trajectory.

3.3.2 Constrained QM/MM geometry optimizations

Starting from selected snapshots, we performed QM/MM constrained geometry optimizations in which MM regions were kept fixed. Electrostatic embedding QM/MM scheme was employed [133]. Each QM region contained a single BChl a molecule with a phytyl tail truncated between carbons C3 and C5 according to the PDB nomenclature, i.e. on the bond following the double bond. Additionally, few neighboring AAs that significantly influence internal structures of BChl a pigments were included in the corresponding QM regions. QM regions for each BChl a are shown in Figure 3.3. In all cases, the Mg-coordinating AA (or water molecule in case of site #2) was included in the QM region, as well as two or three other AAs strongly interacting (often H-bonding) with BChl a. A choice of AAs included in QM regions was dictated by analysis of solvatochromic shifts of neighboring AAs to each BChl a pigment, as will be discussed later.

Three truncation schemes were employed for the AAs included in the QM region (see Figure 3.4). AAs that interact with BChl a through a side chain were truncated between C_α and C_β carbons, such that only a side chain of the AA was included into the QM region (Figure. 3.4 (b)). For the AAs that interact with BChl a via backbone, as occurs for coordination of sites #2, #3, #5, and #8, truncation was done along $C_\alpha - C$ and $C_\alpha - C_\beta$ bonds (Figure 3.4 (c) - (e)). The truncated atoms were capped with hydrogen link atoms.

All atoms not included in the QM region, i.e, protein AAs, remaining BChl a pigments, water molecules and Cl^- counter ions were described at the MM level with the same force fields

(AMBER03) as used in MD simulations. The MM atoms were held fixed while the QM was optimized with PBE0/6-31G(d) [134-139] level of theory using L-BFGS optimization method (gradient tolerance = 300×10^{-5} a.u). Electrostatic embedding scheme for QM-MM interactions of the whole system was employed. Link hydrogen atoms scheme was used to cap broken bonds. All geometry optimizations were performed using QM/MM interface between GAMESS [140] quantum chemistry package and GROMACS 4.6.5 molecular dynamics package [132].

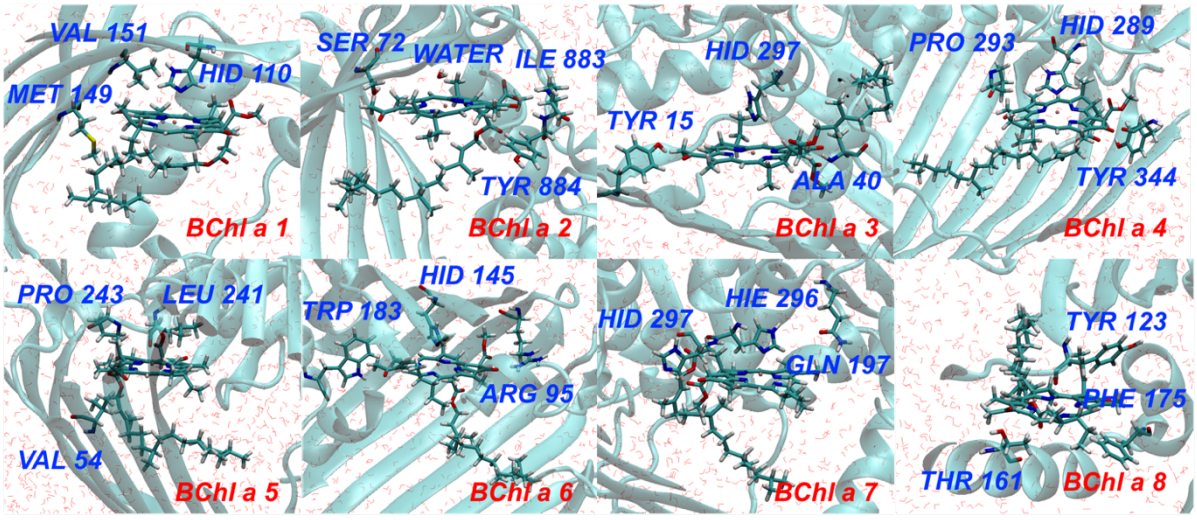

Figure 3.3. QM regions used in QM/MM geometry optimizations of each BChl a site.

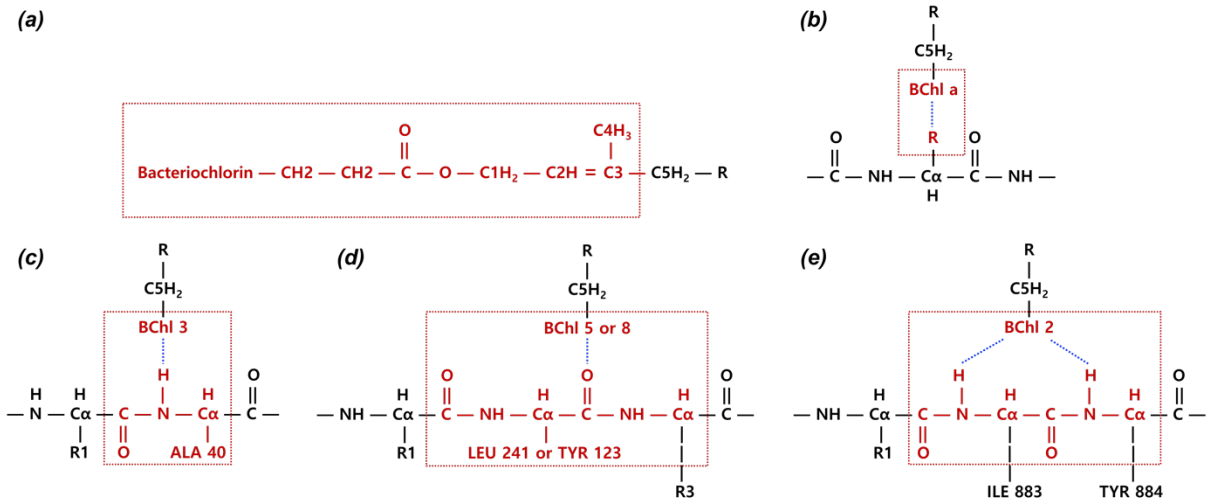

Figure 3.4. Truncation schemes employed in QM/MM geometry optimizations. (a) BChl a, (b) AAs interacting with BChl a via a side chain, and (c) – (e) AAs interacting with BChl a via backbone.

3.3.3 EFP parameters

The interactions between an ab initio region and effective fragments are modeled by the polarization embedding; namely, fragments communicate with the QM Hamiltonian through one-electron Coulomb and polarization terms (Eqns. (1.2) and (1.3)) [29,30,33,48-50,55].

The converged dipoles on the effective fragments are consistent with electronic distributions of all EFP fragments and the QM wave function for the ground electronic state. The induced dipoles are not recomputed for the electronically excited states. In this work, dispersion and exchange-repulsion terms are not included in the QM-EFP coupling Hamiltonian and do not affect excitation energies of chromophores, even though efficient ways of inclusion of both terms were recently developed [61,141].

For EFP calculations, the FMO protein was fragmented into individual AAs following the procedure described in REF [33]. Specifically, the peptide chain was split into fragments along $C_\alpha - C$ bonds; the open valences were capped with H atoms and residual charges were added to the capped C atoms to ensure integer charge of each fragment. EFP parameters for AA fragments, i.e. distributed multipoles and static anisotropic polarizability tensors, were computed with 6-31G(d) basis set. The EFP parameters for a water molecule were computed with 6-31G+G(d) basis set for electrostatic and 6-311++G(3df,2p) basis set for polarization term. Exponential electrostatic damping for charges (SCREEN2) was used to account for charge-penetration effects between EFP fragments [57]. Electrostatic QM-EFP interactions were not screened. Similarly, Gaussian-type polarization damping with default settings (POLAB 0.6) was employed between the fragments but not between the QM and EFP regions. Parameters for AAs were recomputed for all individual snapshots of the protein structure. A more computationally efficient procedure that allows reusing parameters for fragments with similar geometries is reported in Chapter 2 of this thesis. All EFP parameters were computed in the GAMESS electronic structure package [140].

3.3.4 Covalent boundaries in excited state QM/MM and QM/EFP calculations

While BChl a pigments are not covalently linked to a protein scaffold, inclusion of critical AAs into the QM region introduces a covalent boundary between quantum and classical regions. In geometry optimization QM/MM calculations, the covalent boundaries between the QM and MM regions are resolved by using link atoms scheme [142] as implemented in GROMACS 4.6.5

package with removing the first MM atom on the broken bond from a set of MM charges used in QM calculation to avoid over-polarization of the wave function. A treatment of covalent linkage between QM and EFP regions is more complicated because EFP fragments contain both multipoles centered at atoms and bond midpoints and polarizabilities centered at localized molecular orbital (LMO) centroids. Even though Jansen and Kairys proposed a procedure to describe the covalent linkage in the QM/EFP model [143], it requires two steps of EFP parameter generation and is not well suited for multiple covalent bond connections between the QM and EFP regions. Hence, a different approach was utilized in the present study (see Figure 3.5). Cleavage sites in the QM region were capped with H atoms to make the QM region a closed-shell species. Atoms and corresponding EFP parameters directly adjacent to the QM region were removed. Their residual charges were redistributed to the nearest atom to preserve the integer charge of each EFP fragment. The next neighbors were considered as a buffer zone, which only possessed electrostatic multipoles. These buffer regions (in which polarizability tensors were removed) were introduced to stabilize convergence of the QM wave function. To minimize regions with atom deletions, only Mg-coordinating AAs were modeled quantum mechanically in QM/EFP calculations. In the EFP fragments with deleted atoms, residual charges were redistributed to the neighboring atoms as in a standard BioEFP fragmentation scheme [33]. For a direct comparison with QM/EFP, excited state QM/MM calculations followed the same procedure of atom deletion and charge redistribution.

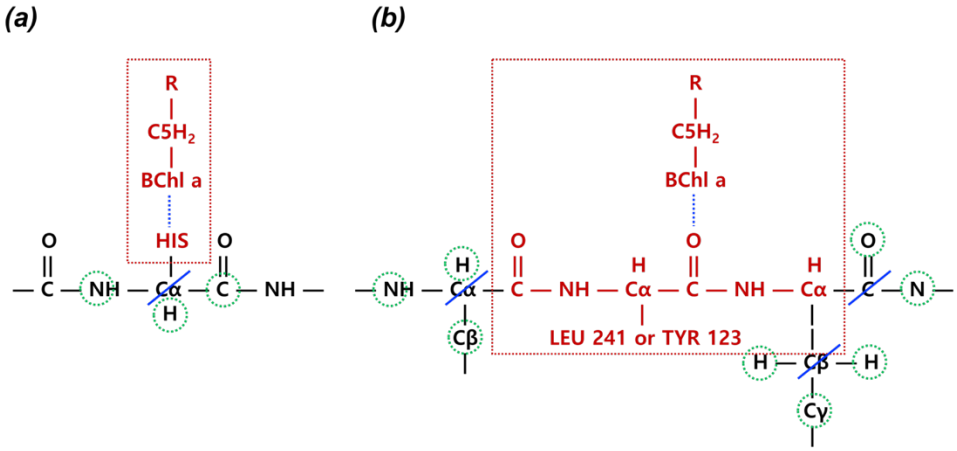

Figure 3.5. Covalent boundaries in excited state QM/MM and QM/EFP computations. (a) a scheme for BChl a site #1, #3, #4, #6, and #7 in which Mg is coordinated with a histidine side chain. (b) a scheme for BChl a site #5 and #8 coordinated with peptide groups. Red: QM regions, blue: removed atoms, and green: buffer zone with removed polarizability tensors.

3.3.5 Excited state QM/MM and QM/EFP calculations

Three embedding schemes for electronic excited states were probed in this work, QM/MM, QM/EFP-noPOL, and QM/EFP. QM/EFP-noPOL refers to non-polarizable QM/EFP, i.e. EFP model in which the polarization interactions are turned off but electrostatic part of the potential is still represented by distributed multipoles up to octopoles. In all three schemes, the QM regions included a single BChl a molecule and the corresponding Mg-coordinating AA, as shown in Figure 3.6. In the case of the QM/EFP model, AAs, remaining BChl a pigments, and solvent molecules within 15 Å from the QM BChl a head were represented as effective fragments. Outside the 15 Å shell, all atoms were described by the MM point charges with the same force fields as used in MD simulations (see Figure 3.6).

TDDFT PBE0/6-31G(d) was used for computing excitation energies, transition dipole moments (TDMs) and transition charges (TrEsp) in all schemes. Transition charges, TrEsp were determined by fitting to a transition density and constrained to reproduce a total transition dipole moment of the given electronic states [144]. The code to compute TrEsp was implemented in a local version of GAMESS. Using transition charges, the electronic couplings between a pair of pigments *i* and *j* are computed as:

$$V_{ij}^{TrEsp} = \sum_{a\in i}\sum_{b\in j}\frac{q_a q_b}{r_{ab}} \tag{3.1}$$

where q_a and q_b are transition charges on atoms a and b of pigments i and j, respectively; r_{ab} is the distance between atoms a and b. For comparison, the electronic couplings were also computed with the point dipole approximation (PDA):

$$V_{ij} = \frac{|\mu_i||\mu_j|}{r_{ij}^3}\left[\vec{\mu}_i \cdot \vec{\mu}_j - 3\left(\vec{\mu}_i \cdot \vec{n}_{ij}\right)\left(\vec{\mu}_j \cdot \vec{n}_{ij}\right)\right] \tag{3.2}$$

where $|\mu_i|$ and $|\mu_j|$ are the strengths of transition dipole moments in i^{th} and j^{th} pigments and $\vec{\mu}_i$ and $\vec{\mu}_j$ represent their corresponding unit vectors; r_{ij} is the distance between i^{th} and j^{th} transition dipole moments (computed as a distance between Mg positions in i^{th} and j^{th} pigments) with the direction $\vec{n}_{ij}$. Direction of computed transition dipole moments was considered to be positive when aligned from ND to NB atoms of BChl a pigment according to the PDB nomenclature.

The TDDFT transition dipole moments and corresponding transition charges were scaled by 0.6 in QM/EFP scheme and 0.74 in QM/MM and QM/EFP-noPOL models to approximately

match the value of the experimental TDM of BChl a in gas phase [145,146]. No additional screening was included in computing the electronic couplings. While calculations of electronic couplings can be further improved by employing screenings due to polarization environment, the present model is found to be sufficient for reproducing major features of the electronic Hamiltonian of FMO [117].

All excited state computations were carried out in the GAMESS electronic structure package [140].

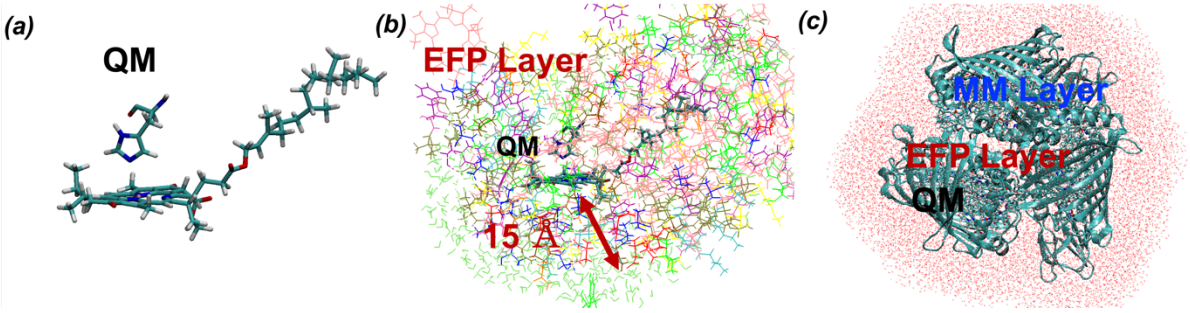

Figure 3.6. Separation into (a) QM, (b) EFP, and (c) MM layers in QM/EFP calculations. The QM region includes one BChl a and a coordinating AA. The EFP layer includes AAs, BChl a fragments, and solvent molecules within 15 Å from the QM BChl a head. All atoms outside 15 Å from the QM BChl a head are described with MM point charges.

3.3.6 Absorption and Circular Dichroism (CD) spectra

Eigenvalues (ε) and eigenvectors (U) of the electronic Hamiltonian can be used to model the absorption and circular dichroism (CD) spectra through computing oscillator (f_k) and rotary (r_k) strengths, respectively:

$$f_k = \sum_{i,j}(\vec{\mu}_i \cdot \vec{\mu}_j)U_{ik}U_{jk} \tag{3.3}$$

$$r_k = \sum_{i,j}\varepsilon_k\left[\vec{R}_{ij}(\vec{\mu}_i \times \vec{\mu}_j)\right]U_{ik}U_{jk} \tag{3.4}$$

where i and j are the indices of chromophores, k is the index of the excitonic state, $\vec{\mu}_i$ and $\vec{\mu}_j$ are the transition dipole moments for chromophore i and j, and $\vec{R}_{ij}$ the distance between the transition dipoles.

In order to compare computed spectra with more structured low-temperature experimental spectra, 60 cm^{-1} full width at half max (FWHM) broadening was used in individual snapshot

spectra, while the spectrum of the average Hamiltonian was broadened by $\sqrt{60^2 + 100^2}$=116.6 cm^{-1} for consistency. For room-temperature spectra, the good agreement is obtained when the excitonic band width is set to 200 cm^{-1}.

3.4 Results

In the described computational details, two steps critically affect the quality of results, as we demonstrate below. These are (i) QM/MM geometry optimizations with sufficiently large QM regions and (ii) excited state calculations in a polarizable protein environment.

3.4.1 Effect of constrained geometry optimization on internal structure of BChl a pigments

Figure 3.7 shows plots of radial distribution functions *g(r)* between Mg atom of BChl a and O or N atom of the coordinating AA or water molecule. Figure 3.7 (a) shows *g(r)* computed from MD snapshots, while Figure 3.7 (b) presents analogues *g(r)* obtained after QM/MM geometry optimizations. Average distances and their standard deviations between Mg and O or N atoms are also shown in Figure 3.7 (c). These plots reveal that classical force field does not describe interaction between Mg atoms and coordinating AAs accurately, which follows from too long Mg-O and Mg-N distances (the corresponding distances in X-ray structure are in 2.0-2.2 Å range). However, QM/MM optimizations correct this shortcoming of classical MD and bring Mg-O/N distances to the appropriate values of 2.2 - 2.3 Å. The only noticeable exception is a coordination of BChl a site #2 by a water molecule. In this case, QM/MM optimizations "repair" coordination in 38 out of 100 snapshots (with Mg-OW distances < 3.0 Å), while in the remaining cases Mg remains 4-coordinated. Further analysis shows that QM/EFP excitation energies (see details below) of 4- and 5- coordinated BChl a site #2 are on average shifted from each other by ~47 cm^{-1}, with 5-coordinated BChl a site #2 being blue-shifted with respect to 4-coordinated one. Thus, while partial missing of water coordination in site #2 might affect the accuracy of those site energies, the expected error should not exceed ~20 cm^{-1} and should not produce any significant effect on spectra. On the other hand, development of a rigorous force field for BChl a molecules is an important direction left for future work.

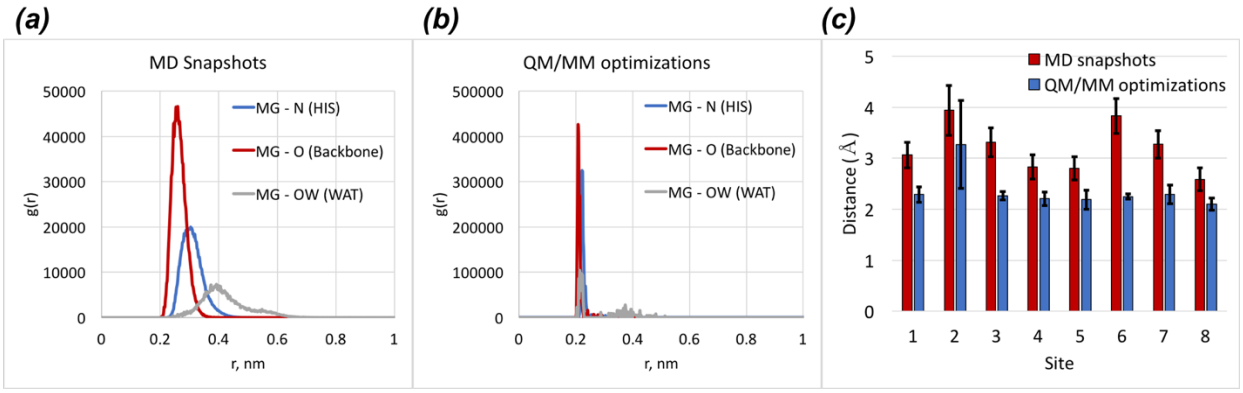

Figure 3.7. (a) and (b) radial distribution functions, *g(r)*, between Mg atoms of BChl a and N atoms of coordinating histidine (blue), O atoms of peptide group (red) and O atom of water (grey). *g(r)* computed in MD snapshots are shown in (a); *g(r)* of QM/MM optimized structures are shown in (b). *g(r)* are not normalized because in each case a single O or N atoms is considered. (c) Average Mg-O or Mg-N distances with standard deviations for MD (red) and QM/MM optimized (blue) structures are shown for each BChl a site.

3.4.2 Effect of a local environment on excitation energies of BChl a pigments

We analyze solvatochromic shifts induced by neighboring AAs on excitation energies of BChl a pigments and benchmark the accuracy of embedding models. Figure 3.8. shows value of excitation energies in BChl a- AA dimers computed with TDDFT PBE0/6-31G(d) (referred to as "full QM"), QM/MM, and QM/EFP schemes as well as excitation energies of the corresponding BChl a pigment in gas phase. In QM/MM and QM/EFP calculations, the QM region contained a BChl a molecule while the AA was described by MM or EFP – see details of QM/MM and QM/EFP excited state calculations above 3.3.6. Mg-coordinating AAs were not included in this analysis. The dimer calculations were performed at a single snapshot produced from a protonated and partially-optimized crystal structure. All dimers with QM/EFP errors in solvatochromic shifts (with respect to full QM dimer calculations) exceeding 20 cm^{-1} are shown. The presented data demonstrate a large impact of several neighboring AAs on excitation energies of BChl a pigments. Indeed, these AAs shift excitation energies by 23-230 cm^{-1}, which is quite substantial for the FMO protein where a spread of site energies of eight pigments is on the order of 500 cm^{-1}. Additionally, these calculations show very large errors (the average relative error is 74%) in the excitation energies when AAs are described by the classical force field (QM/MM scheme). On the other hand, it is very encouraging that the errors in solvatochromic shifts in the QM/EFP model are significantly smaller (the average relative error is 33%) than those in QM/MM. In absolute values,

the largest error of QM/MM model is 102.5, while the largest error of QM/EFP is 47.2 cm^{-1}. While these data do not directly show how the changes in geometries of these AAs influence BChl a excitation energies, we found through a trial-and-error that inclusion of these AAs into QM regions in the constrained QM/MM geometry optimizations significantly improves the quality of the produced electronic Hamiltonian.

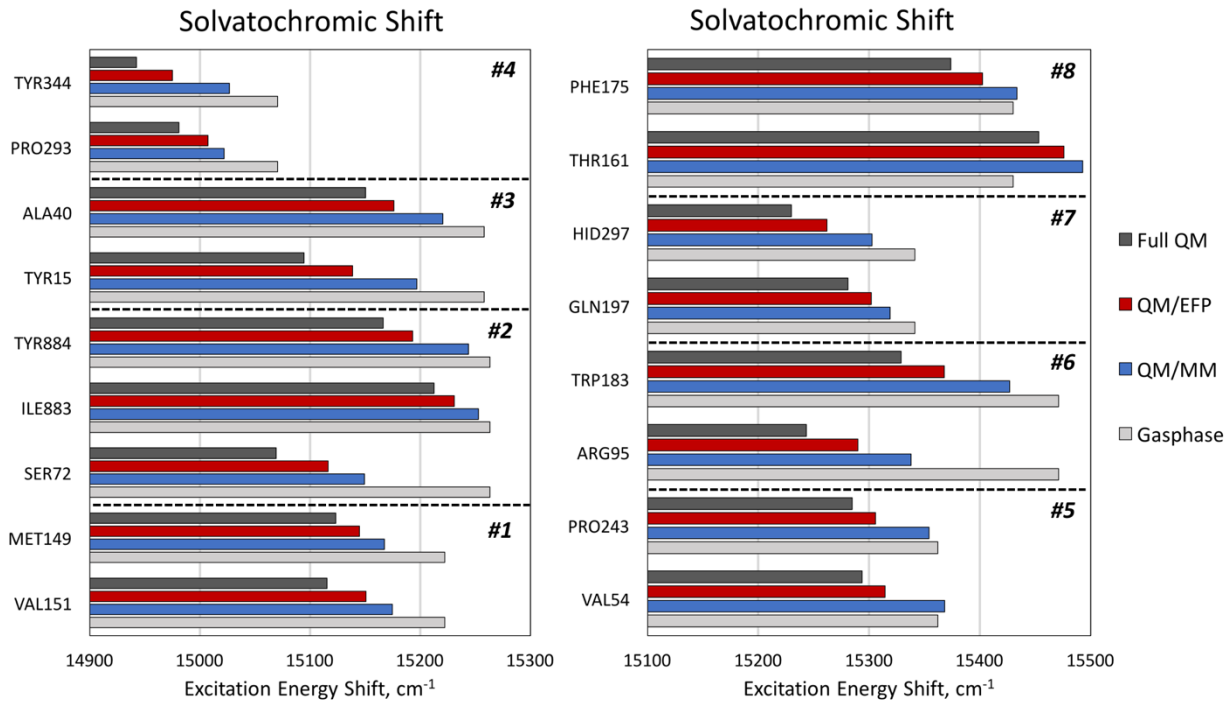

Figure 3.8. Influence of neighboring AAs on excitation energies of BChl a pigments. AAs with the QM/EFP errors in solvatochromic shifts exceeding 20 cm^{-1} are shown.

3.4.3 Electronic properties of the FMO

As shown in Figure 3.9 (a), the electronic excitation energies of BChl a pigments (without inclusion of protein environment, aka "gas phase" site energies) with structures extracted directly from the MD trajectory snapshots are essentially indistinguishable between each other due to large fluctuation of BChl a internal structures during molecular dynamics and general limitations of the classical force field. On the other hand, gas phase site energies computed at the QM/MM optimized structures fluctuate much less, such that energy differences between different sites become more pronounced. The differences in gas phase site energies originate in geometrical constraints imposed by the protein scaffold on each BChl a pigment. Comparison of site energies computed

at the MM-optimized and QM/MM optimized structures, shown in Figure 3.13, further demonstrates importance of correcting the BChl a structures at the QM level.

To demonstrate the importance of polarizable embedding in the description of excited state properties in the FMO complex, we compared the site energies and electronic couplings (averaged over 100 snapshots) computed with electrostatic embedding models QM/MM and QM/EFP-noPOL and with polarizable embedding QM/EFP (see Figure 3.9 (b) and Figure 3.10). As Figure 3.9 (b) and Figure 3.10 demonstrate, the QM/MM and QM/EFP-noPOL schemes produce similar average site energies implying that the inclusion of higher electrostatic multipoles in the description of electronic densities of AAs does not contribute significantly to solvatochromic shifts of BChl a pigments in the FMO complex. Noteworthy, both electrostatic embedding model do not differentiate well between different sites, producing similar solvatochromic shifts of 70-130 cm^{-1} for site #1, #3, #4, #5, #6, and #7 (110-130 cm^{-1} shifts for sites #4, #5, #6, and #7). In contrast to the electrostatic embedding models, the polarizable embedding QM/EFP results in distinct site energies and dramatically stabilizes excitation energy and increase values of transition dipoles (See discussion above regarding the scaling factors applied for computing electronic couplings in Section 3.3.5). Specifically, polarization significantly lower the excitation energy of site #3, which has been proposed to be an exit site of the FMO complex [79,85,87,89,94,99,107,110].

For reference, (unscaled) average values of TDM components and magnitudes are provided in the Figure 3.11. The average values of TDM components were obtained by averaging corresponding components from individual snapshots taken with respect to the system coordinate frame (without accounting for (minor) changes in relative orientations of pigments from snapshot to snapshot). The average TDM magnitude is an average of magnitudes from different snapshots. TDMs of individual snapshots were computed from atomic transition charges as:

$$\vec{\mu}_{tr} = \sum_a q_a \vec{r}_a \tag{3.5}$$

Center of charge of each BChl a pigment is computed as:

$$R_{COCh} = \frac{\sum_a |q_a| \vec{r}_a}{\sum_a |q_a|} \tag{3.5}$$

where q_a is a transition charge on atom a and $\vec{r}_a$ is a position of this atom.

Site #8 possesses the largest TDM, possibly due to its partial exposure to water solvent. Analyzing TDM magnitudes of remaining (internal) pigments, it is quite interesting that the

QM/EFP model shows larger spread of magnitudes of TDMs at different sites compared to QM/MM and QM/EFP-noPOL models. Namely, QM/MM average TDM magnitudes are between 2.75 and 2.84 a.u., i.e. fluctuate by ~3%, while TDM magnitudes are between 3.01 and 3.21 a.u., i.e. ~6% fluctuations, in QM/EFP. Additionally, QM/EFP calculations suggest that the lowest energy (exit) site #3 possesses the second largest TDM (after site #8). The large TDM at site #3 might be an additional factor that stabilizes the lowest energy exciton (by increasing couplings with neighboring site 4) and facilitates a more efficient energy transfer to the reaction center.

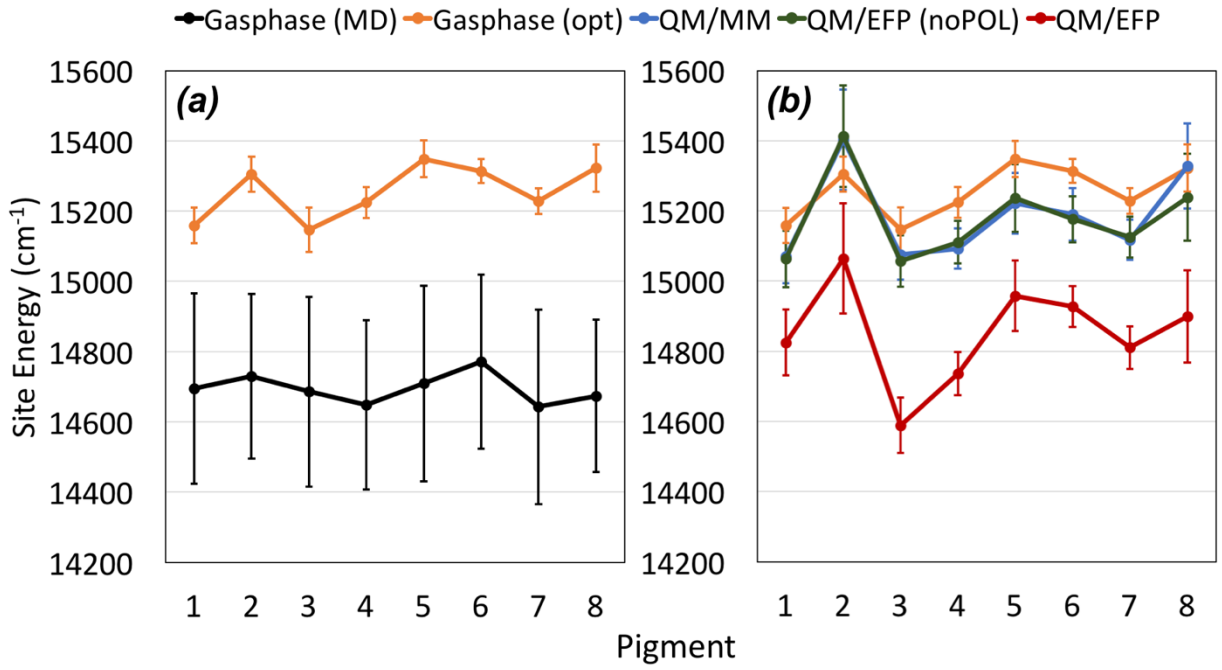

Figure 3.9. BChl a site energies averaged over 100 structures with standard deviations shown as vertical error bars. (a) Gas phase (without protein environment) site energies computed for structures directly extracted from MD snapshots (black) and after QM/MM geometry optimizations (orange). (b) Gas phase (orange), QM/MM (blue), QM/EFP-noPOL (green), and QM/EFP (red) site energies computed for structures after QM/MM geometry optimizations.

Average Hamiltonian

QM/MM	1	2	3	4	5	6	7	8
1	15070.7	-113.3	5.4	-6.8	6.9	-9.3	-2.4	25.1
2	-106.4	15403.5	36.1	9.7	1.5	12.3	7.5	5.6
3	4.8	35.0	15075.8	-51.4	-1.8	-10.1	8.1	1.4
4	-6.3	8.7	-47.9	15093.0	-85.0	-20.0	-50.7	-1.9
5	6.1	1.7	-5.1	-83.7	15223.0	54.1	4.7	5.2
6	-4.8	11.2	-10.9	-19.9	76.0	15190.3	26.1	-10.9
7	-3.5	3.8	17.5	-60.7	-1.3	27.1	15117.6	-14.1
8	33.2	6.4	1.9	-2.1	4.5	-11.0	-13.2	15328.6

QM/EFP-noPOL	1	2	3	4	5	6	7	8
1	15063.3	-113.6	5.3	-6.7	6.8	-8.9	-2.2	24.9
2	-105.9	15413.5	35.9	9.6	1.5	12.1	7.1	5.5
3	4.8	34.8	15057.3	-49.8	-1.9	-10.1	6.6	1.4
4	-6.3	8.6	-45.8	15111.3	-85.5	-19.8	-52.8	-2.0
5	6.1	1.7	-5.2	-83.8	15237.4	53.0	5.9	4.5
6	-4.6	11.0	-10.9	-19.7	75.9	15176.5	27.4	-11.4
7	-3.6	3.9	16.2	-61.3	-0.7	28.1	15125.2	-14.3
8	33.0	6.3	1.9	-2.1	4.5	-11.3	-13.3	15239.2

QM/EFP	1	2	3	4	5	6	7	8
1	14830.4	-113.4	5.3	-6.5	6.4	-8.9	-2.7	24.0
2	-104.8	15058.8	36.8	9.5	1.6	11.7	6.9	5.7
3	4.8	36.0	14585.3	-52.0	-2.0	-10.1	7.2	1.4
4	-6.1	8.7	-47.9	14736.3	-84.0	-19.3	-54.3	-2.0
5	5.8	1.8	-5.2	-82.5	14959.2	51.2	5.0	4.4
6	-4.4	10.9	-10.9	-19.4	72.9	14932.0	26.7	-11.2
7	-3.7	3.9	16.4	-61.4	-0.8	27.7	14815.3	-14.1
8	31.9	6.4	1.9	-2.1	4.4	-11.1	-13.3	14904.9

Standard Deviation

QM/MM	1	2	3	4	5	6	7	8
1	76.6	6.2	0.5	0.5	0.9	3.8	1.8	4.5
2	5.6	142.1	1.8	0.9	0.9	1.4	2.1	1.5
3	0.5	1.8	71.5	7.4	1.1	0.4	5.3	0.4
4	0.5	0.8	10.3	57.8	10.5	1.3	7.9	0.3
5	0.8	1.2	2.1	5.0	85.7	8.9	3.0	7.4
6	3.4	1.4	0.5	1.3	19.2	73.8	4.3	1.4
7	1.8	5.4	11.8	7.1	4.7	4.0	56.8	0.8
8	4.9	1.4	0.3	0.2	0.5	1.2	0.8	120.4

QM/EFP-noPOL	1	2	3	4	5	6	7	8
1	79.8	6.3	0.5	0.5	0.7	3.8	1.3	4.7
2	5.5	143.9	1.8	0.9	0.8	1.4	2.1	1.5
3	0.5	1.8	72.7	7.2	1.2	0.4	4.7	0.4
4	0.5	0.8	10.4	60.1	6.2	1.3	6.5	0.3
5	0.7	1.0	1.3	4.6	96.0	7.7	3.3	0.3
6	3.4	1.4	0.4	1.3	8.8	65.7	4.4	1.5
7	1.4	2.3	7.7	5.9	3.4	4.0	57.6	0.8
8	5.0	1.4	0.3	0.2	0.3	1.2	0.7	123.7

QM/EFP	1	2	3	4	5	6	7	8
1	94.3	6.4	0.5	0.5	0.6	3.7	1.4	4.5
2	5.5	146.0	1.9	0.8	0.8	1.3	2.0	1.5
3	0.5	1.8	79.2	7.4	1.2	0.4	5.0	0.4
4	0.5	0.8	10.5	64.8	6.5	1.3	6.6	0.2
5	0.6	1.0	1.3	4.8	95.4	7.4	3.0	0.3
6	3.3	1.3	0.4	1.2	8.4	54.7	4.3	1.4
7	1.4	2.3	7.8	6.0	3.3	3.9	63.3	0.8
8	4.9	1.4	0.3	0.2	0.3	1.2	0.7	123.9

Figure 3.10. Averaged Hamiltonians from QM/MM, QM/EFP-noPOL and QM/EFP models and corresponding standard deviations of all matrix elements. TrESP couplings are shown in (red) upper triangular sections of each matrix; PDA couplings are shown in (blue) lower triangular sections. All values are in cm^{-1}. The 8^{th} pigment is BChl a 8' as shown in Figure 3.1.

QM/MM	μ_x	μ_y	μ_z	$\lvert\mu\rvert$
1	**2.178**	**1.336**	**-0.450**	**2.745**
2	**2.103**	**-0.779**	**1.312**	**2.779**
3	**0.488**	**-1.160**	**2.357**	**2.839**
4	**2.327**	**1.007**	**-0.701**	**2.775**
5	**1.790**	**-1.467**	**1.212**	**2.773**
6	**0.706**	**0.336**	**-2.548**	**2.808**
7	**1.356**	**2.171**	**-0.538**	**2.766**
8	**1.401**	**-1.806**	**-1.706**	**2.928**

QM/EFP-noPOL	μ_x	μ_y	μ_z	$\lvert\mu\rvert$
1	**2.192**	**1.347**	**-0.432**	**2.761**
2	**2.088**	**-0.781**	**1.304**	**2.761**
3	**0.506**	**-1.157**	**2.353**	**2.837**
4	**2.312**	**1.006**	**-0.722**	**2.763**
5	**1.786**	**-1.476**	**1.217**	**2.768**
6	**0.669**	**0.337**	**-2.570**	**2.818**
7	**1.366**	**2.153**	**-0.541**	**2.752**
8	**1.409**	**-1.811**	**-1.729**	**2.951**

QM/EFP	μ_x	μ_y	μ_z	$\lvert\mu\rvert$
1	**2.401**	**1.463**	**-0.457**	**3.017**
2	**2.342**	**-0.893**	**1.465**	**3.105**
3	**0.558**	**-1.309**	**2.660**	**3.206**
4	**2.586**	**1.132**	**-0.791**	**3.090**
5	**1.944**	**-1.613**	**1.318**	**3.011**
6	**0.744**	**0.374**	**-2.782**	**3.054**
7	**1.516**	**2.381**	**-0.618**	**3.051**
8	**1.530**	**-2.013**	**-1.919**	**3.260**

QM/MM	x	y	z
1	**173.697**	**105.006**	**113.884**
2	**153.072**	**95.513**	**112.827**
3	**129.850**	**90.508**	**133.438**
4	**134.184**	**93.484**	**153.903**
5	**157.432**	**101.699**	**156.761**
6	**165.703**	**113.929**	**141.794**
7	**142.270**	**107.236**	**133.646**
8	**168.070**	**125.429**	**100.222**

QM/EFP-no POL	x	y	z
1	**173.696**	**105.004**	**113.886**
2	**153.042**	**95.473**	**112.760**
3	**129.903**	**90.502**	**133.434**
4	**134.183**	**93.486**	**153.898**
5	**157.147**	**101.328**	**157.373**
6	**165.701**	**113.919**	**141.805**
7	**142.188**	**107.041**	**133.976**
8	**168.059**	**125.439**	**100.222**

QM/EFP	x	y	z
1	**173.679**	**104.978**	**113.827**
2	**152.974**	**95.454**	**112.828**
3	**129.986**	**90.490**	**133.413**
4	**134.186**	**93.491**	**153.939**
5	**157.157**	**101.316**	**157.360**
6	**165.675**	**113.887**	**141.872**
7	**142.189**	**107.055**	**133.974**
8	**168.088**	**125.428**	**100.233**

Figure 3.113.12. Averaged TDM components and magnitudes (in a.u.) from QM/MM, QM/EFP-noPOL, and QM/EFP models, and average positions of center of charge of each BChl a (in a.u.). The 8th pigment is BChl a 8' as shown in Figure 3.1.

3.4.4 Site energy fluctuations of the BChl a pigments

Fluctuations of the QM/EFP site energies, originating in thermal motion of a protein scaffold captured by sampling of 100 protein structures, are visualized in Figure 3.12. Figure 3.12 also compares the average values of the QM/EFP site energies (shown with black vertical lines) with the excitation energies utilized in empirical Hamiltonians by Kell and Brixner (shown with blue and red arrows, respectively [79,89]). For an easier comparison with empirical Hamiltonians and experimental spectra, QM/EFP site energies are shifted by -2430 cm^{-1}, and QM//MM and QM/EFP-noPOL site energies are shifted by -2800 cm^{-1}. A mismatch between absolute values of experimental and computed excitation energies of BChl a pigments originates in well-known intrinsic inaccuracies of a chosen level of theory (PBE0 functional in 6-31G(d) basis set) [147,148]. However, it is expected that the chosen level of theory reasonably describes a potential energy surface of the excited state (i.e., relative energy changes due to vibrational motion of a chromophore), and relative energy changes due to interaction of the excited state with the protein

environment. These relative energy changes (due to slight geometrical distortions and interactions with the heterogeneous protein environment) determine energetic order of site energies and shape of the absorption and CD spectra.

As seen in Figure 3.12, there is a remarkable agreement between the average site energies from the QM/EFP model and Kell's and Brixner's Hamiltonians for the four lowest sites #1, #3, #4, and #7, while the energies of higher-energy sites deviate from each other. The deviations between empirical model Hamiltonians for high-energy sites are not surprising as it is harder to unambiguously deduce these energies from experiment due to short lifetimes and broader spectral bandwidth of high-energy states. On the other hand, the accuracy of the QM/EFP site energies is expected to be similar for all sites. As follows from Figure 3.12, widths of energy distributions of site #3, #4, #6, and #7 are comparable to each other. In general, bacteriochlorin heads of these sites have at least one H-bond with neighboring AAs in addition to coordination with N atoms of histidine (HIS) residues (see Figure 3.3). On the contrary, site #1, #5, and #8 do not have H-bonds to nearby AAs and are somewhat weaker bound to the protein scaffold. As a consequence, the excitation energy distribution of site #7 is wider than that of BChl a pigments that are H-bonded to a protein. Additionally, as site #8 is not fully embedded in a protein envelope, its motion is even less constrained, resulting in a significantly larger spread of excitation energies. In the case of site #2 that also shows a very large spread of energies, Mg is coordinated by a water molecule that moves freely in MD simulations such that BChl a 2 is found four-coordinated in about half of the snapshots. Additionally, several other water molecules that are present in a cavity near site 2 head group change their positions and H-bonding patterns during MD simulations and produce significant fluctuations to the site energy. Hence, we conclude that the width of the site energy distribution is largely governed by the effective H-bonding of BChl a heads with the protein scaffold as well as fluctuations of Mg-coordinating residues.

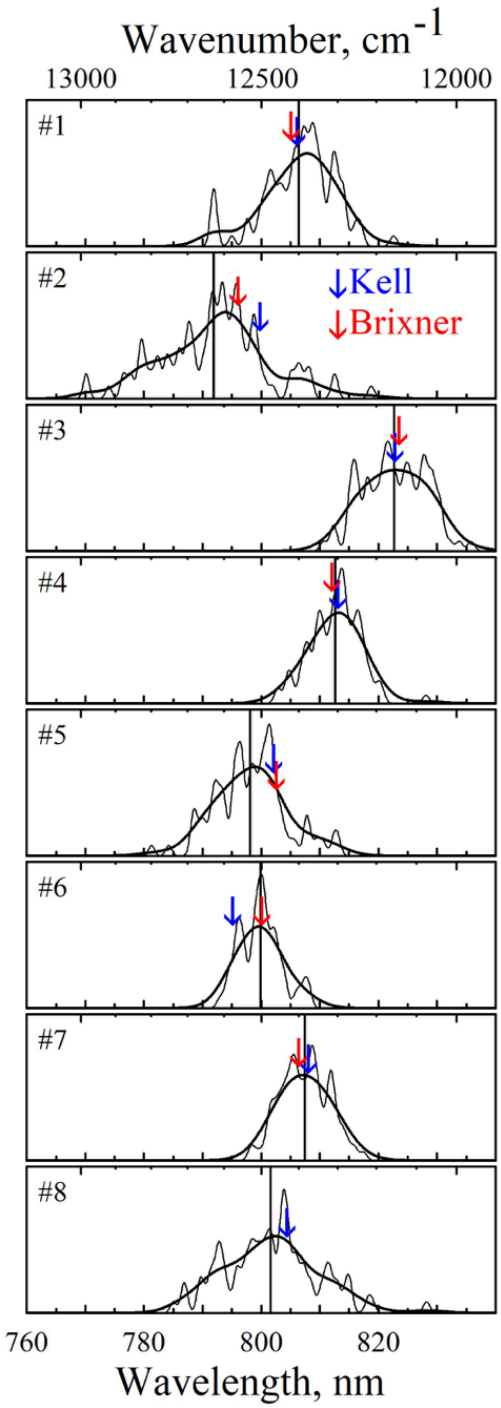

Figure 3.13. QM/EFP site energy fluctuations and comparison to model Hamiltonians by Kell [89] and Brixner [79]. QM/EFP excitation energies are shifted by -2430 cm^{-1} to math experimental absorption and CD spectra. Note that the Brixner Hamiltonian contains only seven pigments. QM/EFP site energy stick spectra of 100 snapshots of MD trajectory are broadened by Gaussians with FWHM of 20 (thinner black lines) and 100 cm^{-1} (thicker black lines) Average QM/EFP site energies are shown with black vertical lines; the site energies from Kell and Brixner empirical Hamiltonians are shown with blue and red arrows, respectively.

3.4.5 Effect of constrained geometry optimization on excitation energies of BChl a pigments

Figure 3.13 shows distribution of the QM/EFP site energies computed at 25 structures extracted from the MD trajectory, as well as those exposed to local constrained MM and QM/MM optimizations. Constrained MM optimizations, in which the optimized regions were identical with those in the QM/MM optimizations, lead to a significant decrease in the excitation energy fluctuations, compared to the energy fluctuations at the MD structures. However, the site energy order obtained at the MM-optimized structures is not correct and a spread of the average energies of different sites is too narrow to produce any reasonable spectra. Thus, correcting the chromophore structures at the QM level is indeed a necessity with the currently utilized force fields

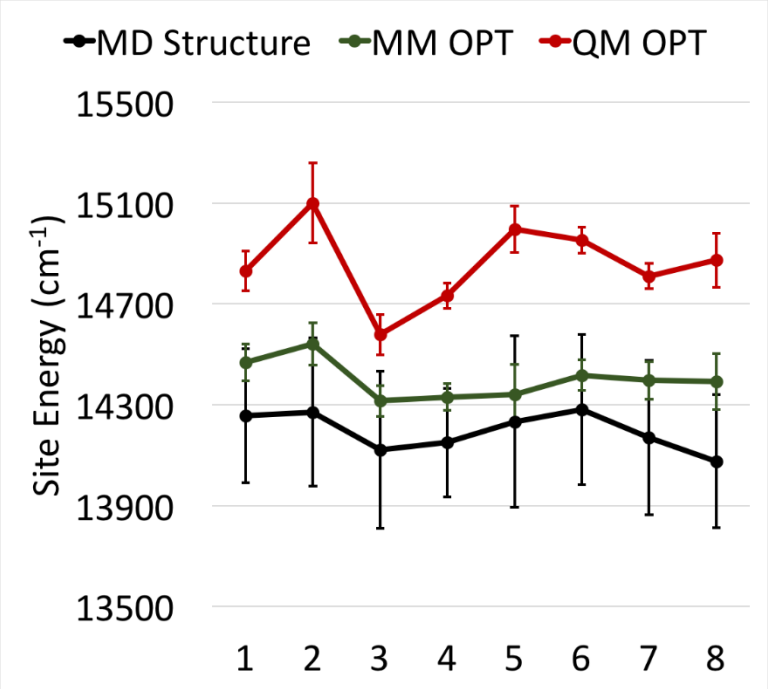

Figure 3.14. QM/EFP site energies computed at 25 snapshots from the MD trajectory (black), as well as at the same snapshots after constrained MM optimizations (green) and constrained QM/MM optimizations (red).

3.4.6 Computed model absorption and Circular Dichroism (CD) spectra

The excitonic Hamiltonians (see Figure 3.10) can be utilized to model absorption and CD spectra. Detailed comparison of spectra built off on the average electronic Hamiltonian and by averaging the spectra of different structural snapshots, as well as comparisons between transition

charges and point dipole approximation models, is provided in the Figure 3.16. First of all, we compare the absorption and CD spectra computed with the three embedding models QM/MM, QM/EFP-noPOL, and QM/EFP. All spectra are computed as averages of spectra of individual structural snapshots, i.e. thermal motion of the protein is directly accounted for. These data are presented in Figure 3.14, together with experimental absorption and CD spectra measured at 295 and 77 K.

It is known from experimental studies [149] that BChl a 8 is only weakly bound in the FMO protein and an occupancy of site #8 depends on a protein preparation procedure. It is estimated that in the procedure utilized in samples that were used for measuring the spectra shown in Figure 3.14 (a) and (b), BChl a 8 is present in ~55% of cases [102,112]. Thus, we mimicked a partial occupancy of site #8 by combining spectra of individual snapshots with a proportion of 45% of 7-site Hamiltonian and 55% of 8-site Hamiltonians.

Analyzing Figure 3.14, it is noteworthy that the quality of the QM/MM, QM/EFP-noPOL, and QM/EFP models cannot be deduced from comparing absorption spectra alone, as all three models produce two intense central peaks and red- and blue- side shoulders. However, the QM/EFP absorption spectrum appears to be somewhat more structured with a more pronounced red-side shoulder that corresponds to the absorption of the lowest energy excitonic state. CD spectra provide a more stringent test for the model accuracy, as the signal from different excitonic transitions can have both positive and negative signs. Out of the three computational models, only QM/EFP manages to reproduce characteristic (starting from red side) down-up-down-up-down sequence of peaks observed both in room-temperature and low-temperature CD experiments, while the QM/MM and QM/EFP-noPOL models fail to reproduce a qualitative shape of the experimental CD spectra. A decomposition of the absorption and CD spectra into contributions of individual excitonic state is provided in the Figure 3.15 and Figure 3.16, from which it follows that the first (right-most, negative) and second (positive) peaks in CD are produced by the lowest energy excitons 1 and 2, while exciton 4 determines the third (negative) peak of CD. Interestingly, exciton 3 has opposite (positive or negative) CD signals in 7-site and 8-site model, due to the contribution of BChl a 8. Overall, we conclude that only the polarizable QM/EFP model faithfully captures excitonic interactions in FMO and is capable of producing quantitatively both absorption and CD spectra.

Comparison of spectra computed with TrEsp and PDA couplings (see Figure 3.16) reveals non-significant but noticeable differences both in absorption and CD. Using TrEsp couplings as compared to PDA couplings results in a more pronounced low-frequency shoulder (corresponding to the first exciton), slight redistribution of intensities of the two main peaks at 800 and 815 nm, and higher intensity of CD signal in 795-805 nm region. Overall, TrEsp couplings bring modeled spectra to a slightly better agreement with experimental spectra.

As discussed above and seen in Figure 3.10 and 3.11, different matrix elements of the QM/EFP average electronic Hamiltonian possess different uncertainties. Thus, even in the absence of correlation between site energies and couplings, it is expected to observe some differences between the spectra from the averaged Hamiltonian and from averaging the spectra of individual Hamiltonians. Such comparisons are also provided in Figure 3.15. Fluctuations of the Hamiltonian matrix elements and of the corresponding spectra due to different protein structures (compared to the spectra from the average Hamiltonian) result in a somewhat broader spectrum range with noisy shoulders extending both to lower and higher wavelengths. These spectral broadenings are more apparent in the absorption than CD spectra, due to partial cancellation of noisy signals in CD. Additionally, averaging the individual spectra results in higher intensity of 815 nm band in absorption

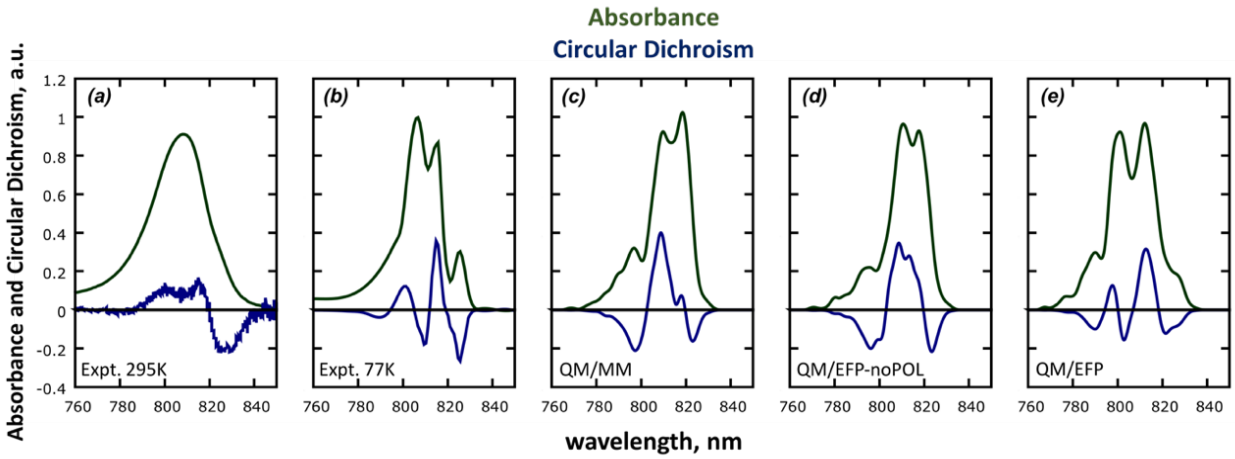

Figure 3.15. Absorption (green) and CD (blue) spectra of the FMO complex. Experimental spectra measured at 295 K (a) and 77 K (b) and computed spectra using QM/MM (c), QM/EFP-noPOL (d), and QM/EFP (e) models. Computed spectra are obtained by combining spectra of individual snapshots with a proportion of 45% of 7-site Hamiltonians and 55% of 8-site Hamiltonians. For a comparison with experimental spectra, QM/MM and QM/EFP-noPOL spectra are shifted by -2800 cm^{-1}, and QM/EFP spectra are shifted by -2430 cm^{-1}.

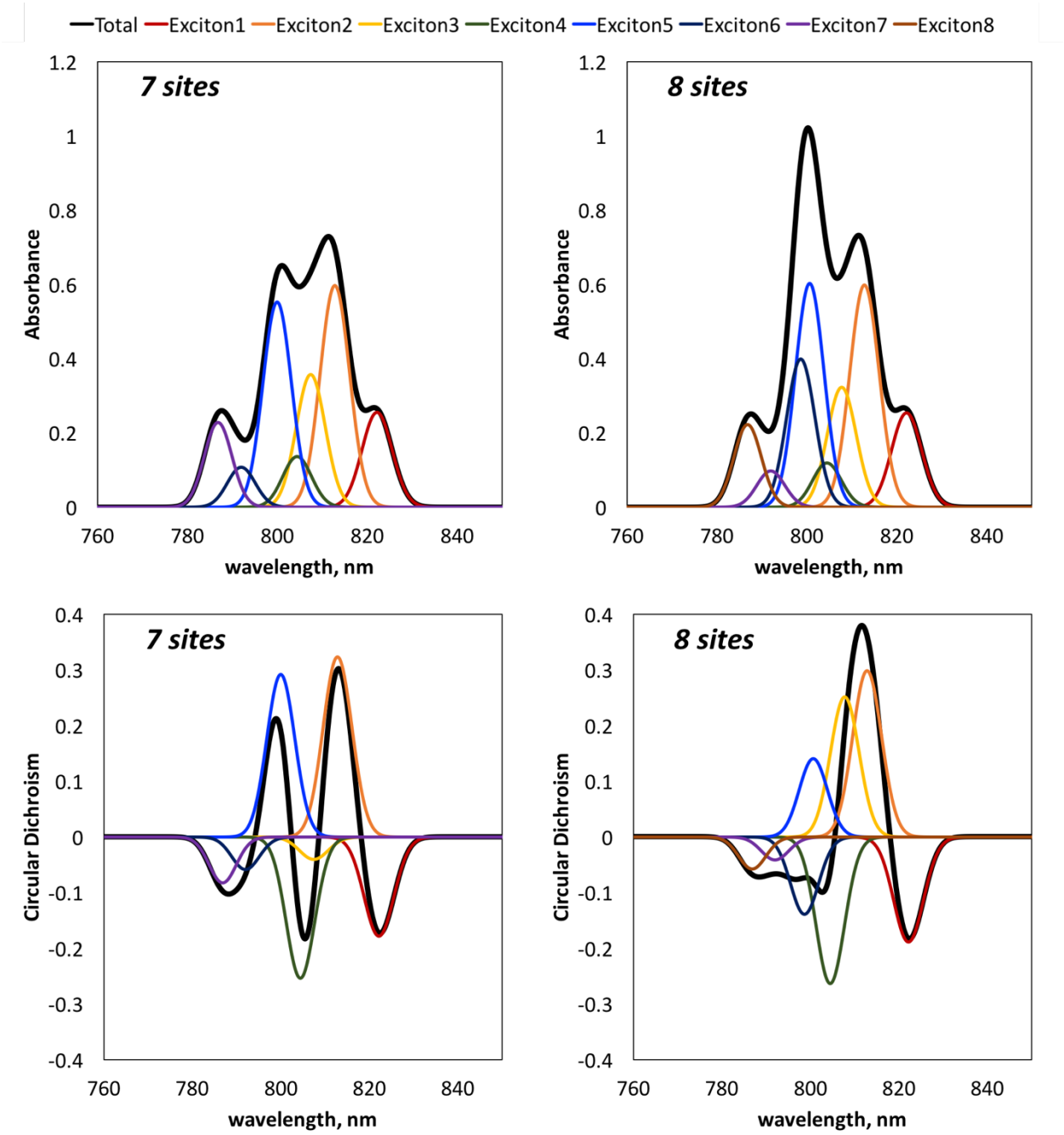

Figure 3.16. Absorption and CD spectra computed as averages of spectra produced from 100 QM/EFP TrEsp Hamiltonians and their decomposition into excitonic contributions. The spectra are shifted by -2430 cm^{-1} to math experimental spectra.

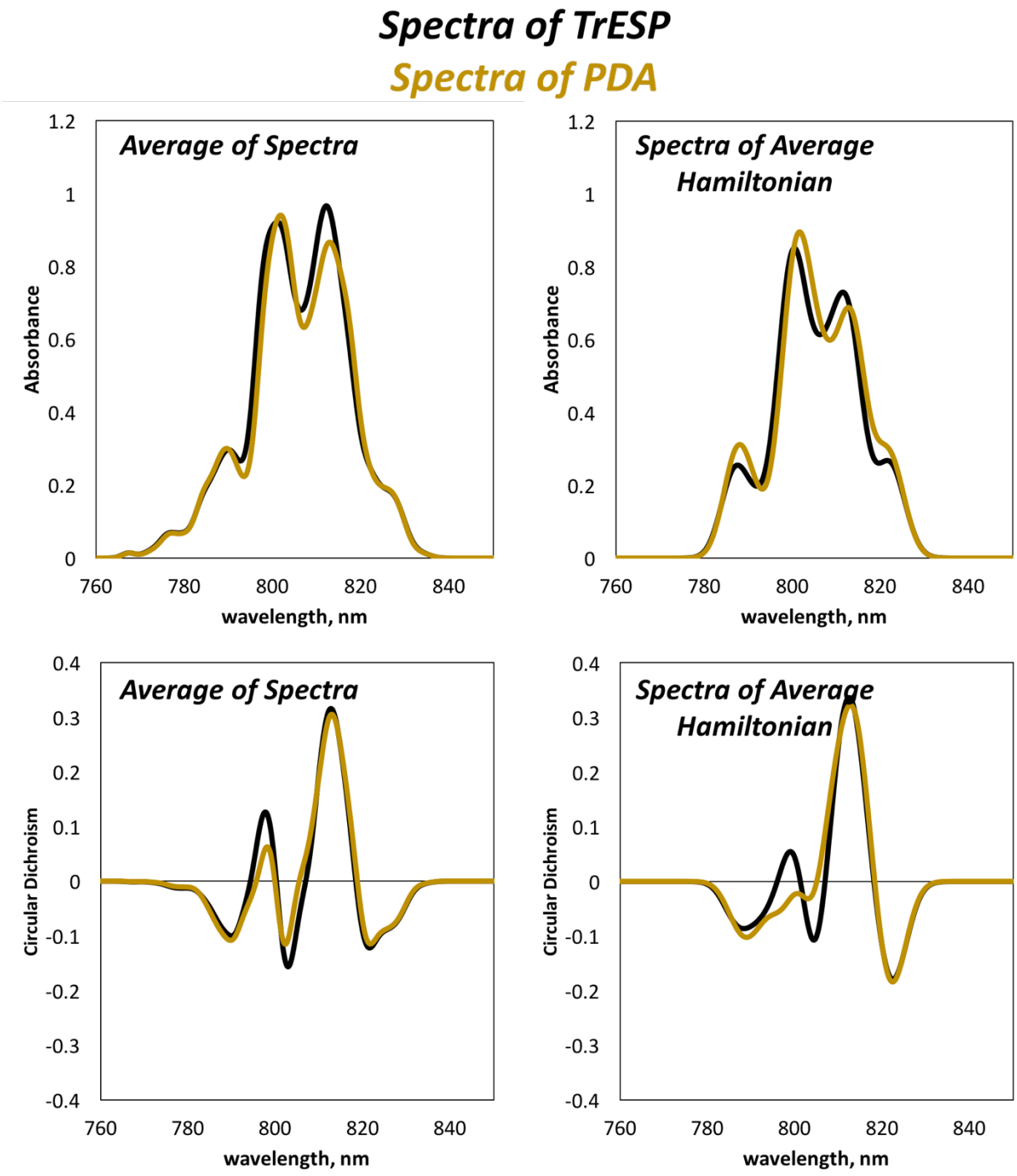

Figure 3.17. Comparison of QM/EFP absorption and CD spectra computed with TrEsp (yellow) and PDA (black) couplings using averaging of individual snapshots (left panels) and average Hamiltonian (right panels). All spectra correspond to mixed model with 55% population of 8-site Hamiltonians.

3.5 Conclusions

We developed a multiscale modeling strategy for describing excitonic properties of pigment-protein complexes. The modeling is based solely on the X-ray structure of the protein complex and does not use other input from experiments. The main steps of the modeling procedure include classical MD simulations, followed by partial correction of pigment structure by constrained QM/MM geometry optimizations. Then, the excited state calculations of electronic states and transition charges of chromophores are conducted with TDDFT PBE0/6-31G(d) in polarizable embedding in which the protein is modeled with the effective fragment potentials. This computational protocol was successfully applied to model the excitonic properties of the Fenna-Matthews-Olson (FMO) photosynthetic pigment-protein complex and resulted in the electronic Hamiltonian that is in an excellent quantitative agreement with previously proposed empirical Hamiltonians. The theory reproduces all major features of absorption and CD spectra of the FMO protein complex. We demonstrate that such an agreement between modeling and experiment becomes possible due to (i) utilizing accurate structures of photosynthetic pigments for computing excitation energies and (ii) representing the protein environment with a polarizable model. Successful first-principles-based modeling of the FMO complex opens exciting avenues for predictive modeling of other wildtype and mutated photosynthetic pigment-protein complexes that will provide mechanistic understanding of interactions in these complex systems.

CHAPTER 4. REVISING THE ELECTRONIC PROPERTIES OF THE FENNA-MATTHEWS-OLSON COMPLEX

4.1 Abstract

As discussed in Chapter 3, the essence of the successful modeling of a photosynthetic antenna protein is (i) utilizing accurate structures of photosynthetic pigments for computing excitation energies and (ii) representing the protein environment with a polarizable model. Due to the large size of BChl a pigments (each BChl a molecule contains 140 atoms), constrained QM/MM geometry optimizations reported in Chapter 3 have included a few critical amino acids and the BChl a head group only. Here, we employ a GPU-based high-performance parallel protocol (GROMACS/TERACHEM interface) capable of accelerating electronic structure computations tremendously. The protocol was successfully applied to constrained QM/MM geometry optimizations of several critical amino acids and entire BChl a pigment. Utilizing advanced pigment structures and polarizable embedding QM/EFP model, we revisited the electronic Hamiltonian of the wild type FMO complex. We found that the new computational strategy resulted in a closer agreement with previously reported empirical Hamiltonians and again quantitatively reproduces experimental absorption and circular dichroism spectra of the FMO complex.

4.2 Introduction

One of the central challenges in the quantum chemistry simulations is how to scale-up a system size while maintaining the computational time and accuracy of simulations. The tremendous efforts of encoding efficient *ab initio* algorithms allow one to explore systems containing ~200 first and second row atoms quantum-mechanically. However, these advances are still insufficient for studying biological systems, which generally consist of thousands of atoms. As mentioned in Chapters 2 and 3, the Effective Fragment Potential (EFP) method [20,29-31,45,150-153] has become a promising means to conduct simulations of extended molecular systems by modeling different parts of the system with different levels of theory. For example, a part of the system where electronic excitations or chemical reactions occur can be represented at the QM level. On the other hand, protein or solvent molecules can be described as model potentials

derived from first principles [152,153]. A combination of QM with EFP results in the so-called QM/EFP method, which was utilized in Chapters 2 and 3.

The work described in Chapter 3 presents a computationally straightforward yet accurate way to predict the electronic properties of the FMO complex. As discussed in Section 3.4.5, the excitation energies of photosynthetic pigments are significantly affected by the quality of their internal structures. Hence, excellent quantitative agreement of theoretical electronic Hamiltonian with previously proposed empirical Hamiltonians becomes possible through partial correction of photosynthetic pigment structures at the QM level. This implies that it might be achievable to describe pigment internal structures even better by utilizing larger QM regions with a decent level of theory. The chosen level of theory in Chapter 3 (i.e., PBE0 DFT functional) was based on the previous study by List et al. [93], which shows that the PBE0 functional predicts the most accurate excitation energies and transition dipole moments of BChl a pigments against DFT-BHLYP/MRCI level of theory. As aforementioned, extending QM regions, however, is not as simple as selecting an appropriate QM method within CPU-based (central processing unit) parallel implementation. Specifically, increasing the size of the QM region for the geometry optimizations, during which multiple gradient computations are followed by multiple SCF cycles, makes this task more challenging than the single point energy or excitation energy computations. As a result, the constrained QM/MM optimization of BChl a pigments in Chapter 3 was performed without including BChl a tail group (see Section 3.3.2 and Figure 3.4). The recent effort of GPU-based (graphics processing unit) high-performance parallel implementation of DFT and TDDFT methods in TERACHEM software [154-156] and development of the interface between GROMACS and TERACHEM makes it possible to include entire BChl a pigment (i.e., both head and tail groups) into the QM region during QM/MM optimizations. In this Chapter, we revisit the electronic properties of the FMO complex and provide improved first-principles modeling of the absorption and circular dichroism (CD) spectra of the FMO complex based on improved BChl a internal structures.

4.3 Computational details

4.3.1 Constrained QM/MM geometry optimizations

Started from the partially optimized structures of BChl a described in Section 3, we performed additional QM/MM constrained geometry optimizations including phytyl tail groups of BChl a pigments into the QM-optimized regions. In the present version, each QM region contains a complete BChl a molecule and few neighboring AAs that significantly influence internal structures of BChl a pigments. The list of AAs and details of truncation schemes for each AA can be found in Section 3.3.2 and Figure 4.1. As discussed in Section 3.3.2, all atoms not included in the QM region, i.e. protein AAs, remaining BChl a pigments, water molecules and Cl^- counter ions, were described at the MM level with the AMBER03 force fields [124]. The MM atoms were held fixed while the QM regions were optimized with PBE0/6-31G(d) [134-139] level of theory using L-BFGS optimization method (gradient tolerance = 300×10^{-5} a.u). Electrostatic embedding scheme for QM-MM interactions of the whole system was employed. Link hydrogen atoms scheme was used to cap broken bonds. All geometry optimizations were performed using QM/MM interface developed by Dr. Dmitry Morozov (University of Jyvaskyla) between TERACHEM [154-156] GPU-accelerated electronic structure package and GROMACS 4.6.5 molecular dynamics package [132].

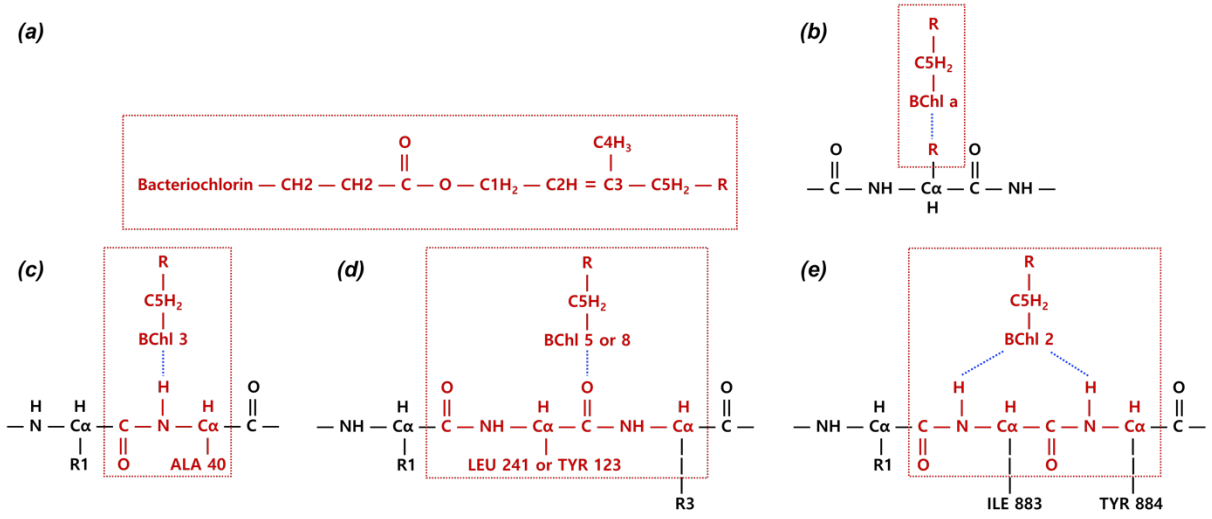

Figure 4.1. Truncation schemes employed in QM/MM geometry optimizations. Red color depicts QM regions; MM regions are shown in black. (a) BChl a including a phytyl tail group, (b) AAs interacting with BChl a via a side chain, and (c) – (e) AAs interacting with BChl a via backbone.

4.3.2 EFP parameters.

For EFP calculations, the FMO protein was fragmented into individual AAs following the procedure described in REF [33]. Specifically, the peptide chain was split into fragments along $C_\alpha - C$ bonds; the open valences were capped with H atoms and residual charges were added to the capped C atoms to ensure integer charge of each fragment. EFP parameters for AA fragments, i.e. distributed multipoles and static anisotropic polarizability tensors, were computed with 6-31G(d) basis set. The EFP parameters for a water molecule were computed with 6-31+G(d) basis set for electrostatic and 6-311++G(3df,2p) basis set for polarization term. Exponential electrostatic damping for charges (SCREEN2) was used to account for charge-penetration effects between EFP fragments [57]. Electrostatic QM-EFP interactions were not screened. Similarly, Gaussian-type polarization damping with default settings (POLAB 0.6) was employed between the fragments but not between the QM and EFP regions. Parameters for only re-optimized AAs were recomputed for all individual snapshots of the protein structure. The remaining AA parameters were taken from the work described in Chapter 3.

All EFP parameters were computed in the GAMESS electronic structure package [140].

4.3.3 Covalent boundaries in excited state QM/MM and QM/EFP calculations.

While BChl a pigments are not covalently linked to a protein scaffold, the incorporation of critical AAs into the QM region introduces a covalent boundary between quantum and classical regions. In geometry optimization QM/MM calculations, the covalent boundaries between the QM and MM regions are resolved by using link atoms scheme [142] as implemented in GROMACS 4.6.5 package with removing the first MM atom on the broken bond from a set of MM charges used in QM calculation to avoid over-polarization of the wave function. Treatment of covalent linkage between QM and EFP regions is more complicated because EFP fragments contain both multipoles centered at atoms and bond midpoints and polarizabilities centered at localized molecular orbital (LMO) centroids. As shown in Chapter 3, we successfully employed atom deletion and buffer atom introductions to avoid over-polarization of the wave function. Hence, the same covalent boundary treatments were utilized to minimize regions with atom deletions. The detailed description of covalent boundary conditions for the QM/EFP scheme is demonstrated in Section 3.3.4.

4.3.4 Excited state QM/EFP calculations.

In Chapter 3, we demonstrated that electrostatic embedding models (QM/MM and QM/EFP-noPOL) did not sufficiently differentiate site energies of different BChl a sites. In contrast to the electrostatic embedding models, the polarizable embedding QM/EFP produced distinct site energies of different pigments. As a result, only polarizable embedding QM/EFP scheme showed quantitative agreement in absorption and circular dichroism (CD) spectra with those obtained from experiments. Hence, only polarizable embedding QM/EFP scheme for electronic excited states was probed in the present work. The QM regions included a single BChl a pigment and corresponding Mg-coordinating AAs, as shown in Figure 3.6. AAs, remaining BChl a pigments, and solvent molecules within 15 Å from the QM BChl a head were represented as effective fragments. Outside the 15 Å shell, all atoms were described by the MM point charges with the AMBER03 force fields (see Figure 3.6). Overall, the only difference in the excited state calculations between the present work and the one described in Chapter 3 is slightly different structures of BChl a pigment and few neighboring AAs.

To ensure consistency with the previous study, TDDFT PBE0/6-31G(d) was used for computing excitation energies. Only transition charges, TrESP were considered for computing electronic couplings between BChl a pigments (See Eqn (3.1)) and were scaled by 0.6 to approximately match the value of the experimental TDM of BChl a in gas phase [145,146].

All excited state computations were carried out in the GAMESS electronic structure package [140].

4.3.5 Absorption and Circular Dichroism (CD) spectra.

In order to compare computed spectra with the previous and low-temperature experimental spectra, the same 60 cm^{-1} FWHM broadening was used in spectra generated for individual snapshots. The detailed scheme for obtaining absorption and circular dichroism (CD) spectra is shown in Section 3.3.6.

4.4 Results

4.4.1 Electronic properties of the FMO

As shown in Figure 4.2 and Table 4.1, the two models (the one from Chapter 3 and the present one) produce average site energies within 100 cm^{-1} from each other. It is also fascinating that the two models result in similar (statistically indistinguishable) average site energies in the two lowest energy sites #3 and #4, with differences of 1.2 and 10.0 cm^{-1}, respectively. In contrast to the two lowest energy sites, BChl a tail group optimizations result in lower site energies in sites #1, #2, #5, #6, #7, and #8 showing a range of 37.7 to 95.2 cm^{-1} of red shifts. A less expected outcome of the present analysis is that the phytyl tail degrees of freedom couple stronger to the excited states in the higher-energy pigments than to those in pigments #3 and #4.

Figure 4.3 compares site energy distributions from the present model with the model of Chapter 3. In the present model (shown with black lines), the widths of energy distributions of sites #1, #3, #4, #5, #6, #7 become comparable to each other. In general, Mg atoms of bacteriochlorin heads of these sites are coordinated with either N atom of histidine groups (HIS) or O atom of peptide groups. Site #2 shows a larger spread of site energies due to the unstable binding of Mg atom of BChl a head to a water molecule. As discussed in Section 3.4.4, Mg-coordinating water molecule moves freely in MD simulations, and not all consequent geometry optimizations recover 5-coordination of Mg. However, the present results show a narrower excitation energy distribution of site #2 compared to the previous study. Specifically, while the complete spectrum is shifted toward red in the present model, the red-most shoulder (which, as was discussed in Section 3.4.4, was ascribed to the snapshots with 4-coordinated Mg) becomes smaller, suggesting that the inclusion of the phytyl tail into the geometry-optimized region contributes to "curing" of site #2 coordination issue. On the other hand, as site #8 is not fully embedded in a protein envelope, its motion is even less constrained when the tail is allowed to move in geometry optimizations, such that pigment #8 might find multiple local minima with the inclusion of the tail group into the QM region in QM/MM optimizations. Hence, site #8 shows the largest energy spread among all pigments and, as expected, the energy fluctuations became more pronounced in the present model.

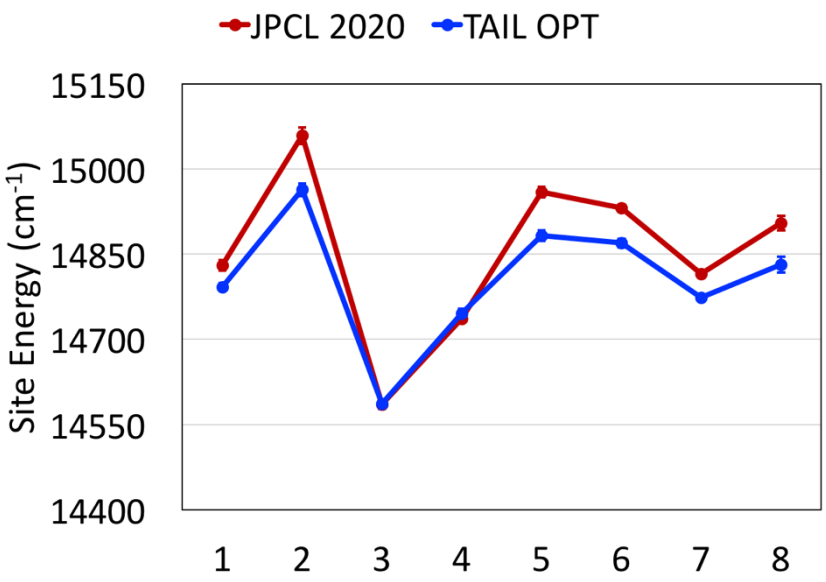

Figure 4.2. BChl a site energies averaged over 100 structures with uncertainties in average values shown as vertical error bars. Red plot represents site energies obtained in Chapter 3 (i.e., after geometry optimizations that included only critical amino acids and BChl a head group). Blue line indicates site energies from the present work, i.e., obtained after geometry optimizations with critical amino acids and whole BChl a pigments.

Table 4.1. Average site energies and their uncertainties as computed in Chapter 3 and obtained in the present work. The differences in average energies of the two models (ΔE) and uncertainty in determining ΔE are also shown. All values are in cm^{-1}.

Site	**QM/EFP (Chapter 3)**	**QM/EFP (Present work)**	ΔE	uncertainty
1	14830.4 $\pm$ 9.5	14792.7 $\pm$ 6.9	-37.7	11.7
2	15058.8 $\pm$ 14.7	14963.6 $\pm$ 11.2	-95.2	18.5
3	14585.3 $\pm$ 8.0	14586.5 $\pm$ 8.6	+1.2	11.8
4	14736.3 $\pm$ 6.5	14746.3 $\pm$ 8.1	+10.0	10.4
5	14959.2 $\pm$ 9.6	14882.5 $\pm$ 9.0	-76.7	13.2
6	14932.0 $\pm$ 5.5	14869.6 $\pm$ 7.4	-62.4	9.2
7	14815.3 $\pm$ 6.4	14774.1 $\pm$ 5.7	-41.2	8.5
8	14904.9 $\pm$ 12.4	14831.5 $\pm$ 14.2	-73.4	18.9

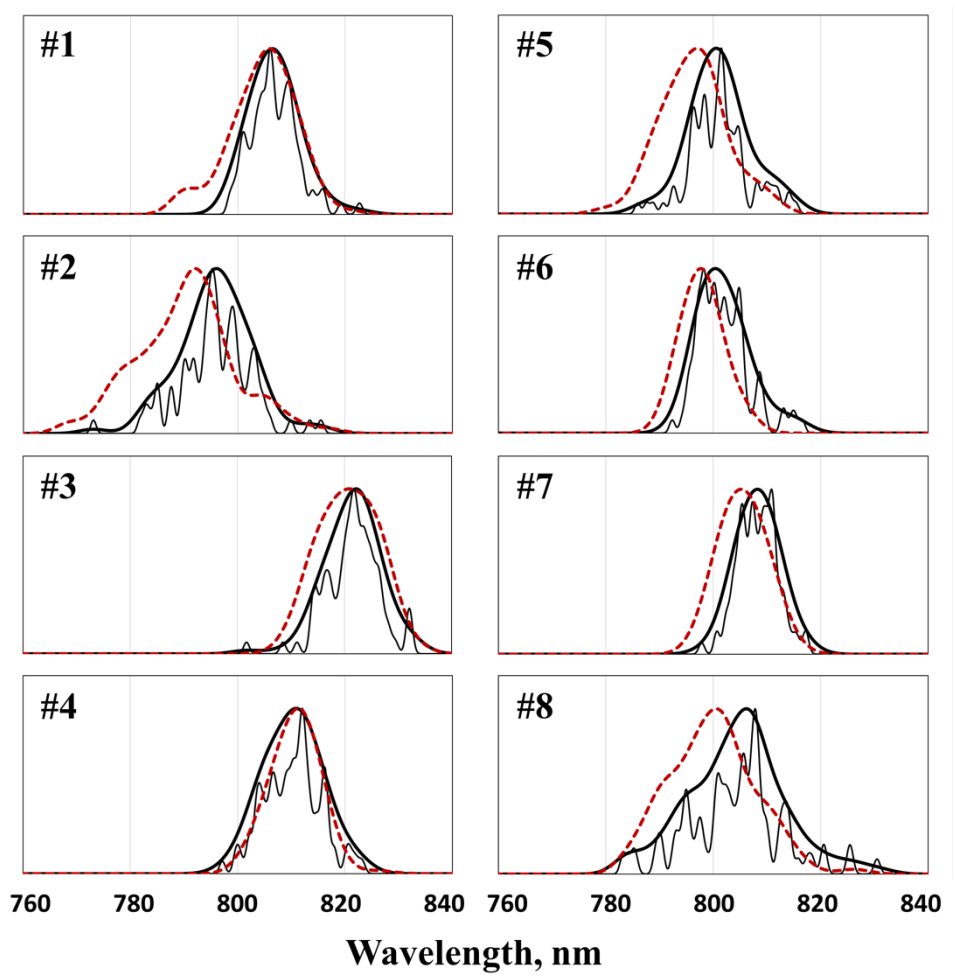

Figure 4.3. QM/EFP site energy fluctuations. Black lines show the present results. Stick spectra of 100 snapshots of MD trajectory are broadened by Gaussians with FWHM of 20 cm^{-1} (thinner black lines) and 100 cm^{-1} (thicker black lines). FWHM 100 cm^{-1} broadened spectra from Chapter 3 are shown with red dashed lines.

4.4.2 Computed absorption and circular dichroism spectra

The excitonic Hamiltonians are constructed to model absorption and CD spectra. Here, we only consider the absorption and CD spectra computed as averages of spectra of individual snapshots, i.e., thermal motion of the protein is directly accounted for. As discussed in Section 3.4.6, BChl #8 is only weakly bound in the FMO protein, and occupancy of site #8 depends on a protein preparation procedure. Thus, a partial occupancy of site #8 was mimicked by combining 45% of the 7-site Hamiltonian and 55% of the 8-site Hamiltonian in Chapter 3. The same strategy is used for generating the spectra shown in Figure 4.4. For an easier comparison with experimental spectra and the previous work, all site energies are shifted by -2420 cm^{-1}.

Analyzing Figure 4.4, it is noteworthy that the quality of the QM/EFP model cannot be deduced from comparing the CD spectra, as both models produce characteristic (starting from the red side) down-up-down-up-down sequence of peaks with similar peak intensities. However, the absorption spectrum from the present work is in a better resemblance to the experimental 77 K spectrum, specifically showing a more pronounced red-side shoulder that corresponds to the absorption by the lowest-energy excitonic states. In addition, the present work produces a better agreement with experimental spectra in terms of positions of intensive peaks in both absorption and CD spectra. Overall, we conclude that BChl a phytyl tail optimizations produce a more compact distribution of the site energies of the lowest energy site #3 (see Figure 4.3), resulting in a better separation of the lowest excitonic state from the mixed excitonic states by other sites. Additionally, overall spread of the site energies is narrower in the present model, which results in spectra that are in a better agreement with experiment.

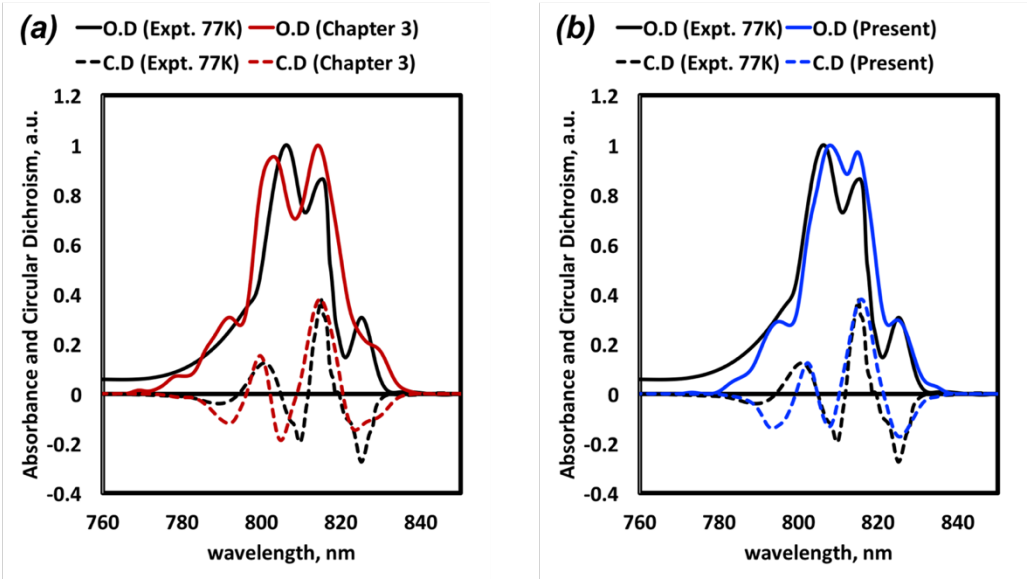

Figure 4.4. Absorption (sold line) and CD (dashed line) spectra of the FMO complex. (a) Experimental spectra measured at 77 K and computed spectra using polarizable embedding QM/EFP model but different QM regions that exclude of phytyl tails (results of Chapter 3), and (c) include phytyl tails (present work). Computed spectra are obtained by summing individual snapshot spectra with a proportion of 45% of 7-site Hamiltonians and 55% of the 8-site Hamiltonians. For a better comparison with experimental spectra, (a) QM/EFP is shifted by -2430 cm^{-1}, and (b) QM/EFP is shifted by -2420 cm^{-1}.

4.5 Conclusions

We implemented a GPU-based high-performance parallel protocol via GROMACS/TERACHEM interface capable of accelerating electronic structure computations. This protocol was successfully applied to extend QM regions of constrained QM/MM optimizations in the FMO pigment-protein complex. The improved pigment structures resulted in the absorption and CD spectra of the FMO complex that quantitatively reproduce experimental spectra. Rigorous calculations of structures of BChl a pigments lead to a more pronounced red-side peak in the absorption spectrum, due to the more compact distribution of excitation energies of the lowest site. It should be emphasized that an accurate shape of the absorption spectra obtained in the present model is an essential prerequisite for reproducing spectral differences between the FMO wild type and mutated species, as will be discussed in Chapter 7.

CHAPTER 5. PHOTOPROTECTION OF THE FMO COMPLEX

5.1 Abstract

The FMO protein plays an important role in transferring excitation energies from chlorosome antenna to the reaction center. In order to function properly, light-harvesting complexes also developed repair and protection mechanisms against excess of solar radiation. It is believed that a typical channel of the excessive excitation energy is an intersystem crossing to a triplet state. Thus, elucidating triplet excited states is essential for understanding the mechanisms of photoprotection in light-harvesting organisms. In the present study, we utilize molecular modeling for exploring a possible route of photoprotection in the FMO protein through protein stabilization of its triplet states.

5.2 Introduction

Photoprotection is a critical survival mechanism evolutionary developed by photosynthetic organisms. Under sunlight excitation a significant fraction of (bacterio)chlophyll-like molecules can undergo intersystem crossing ending in a long-living triplet state. The latter can readily transfer energy to molecular oxygen generating highly reactive singlet oxygen that oxidizes nearby organic molecules eventually leading to oxidative stress and damage of living cells [157-164]. In a typical pigment-protein complex, the harmful effect of excess of solar radiation is avoided by strategically positioning nearby carotenoid molecules that quickly scavenge (B)Chl triplet states and safely dissipate this energy into heat as shown in Figure 5.1. Surprisingly, the Fenna-Matthews-Olson (FMO) pigment-protein complex does not contain any carotenoids yet it is found to be extremely stable in aerobic conditions.

The trimeric FMO complex is a part of photosynthetic apparatus in anaerobic green sulfur bacteria where it functions as an electronic excitation energy transfer channel between the large chlorosomal antenna and the reaction center [78-112]. It has been shown that up to 27 % of singlet excitations in the complex can undergo intersystem crossing and result in the triplet excited state [165]. Based on indirect evidence, it has been proposed that exceptional photostability of the FMO complex may rely on a sufficient shift in the triplet energies of BChl a bringing the latter below

the singlet oxygen energy, effectively blocking the energy transfer process and preventing the formation of reactive oxygen species.

In Chapters 3 and 4, we demonstrated that quantitative agreement of molecular modeling with experiments could be achieved by utilizing accurate structures of photosynthetic pigments and representing the protein environment with a polarizable model. In addition, we proposed a relatively straightforward yet rigorous computational procedures for the study of the FMO protein, showing excellent agreement in absorption and CD spectra with experiments. Here we extend a research scope to modeling triplet states of the FMO complex, with a goal to answer the question whether indeed the energy of triplet states in FMO is sufficiently stabilized in the protein to become below the energy of the singlet oxygen.

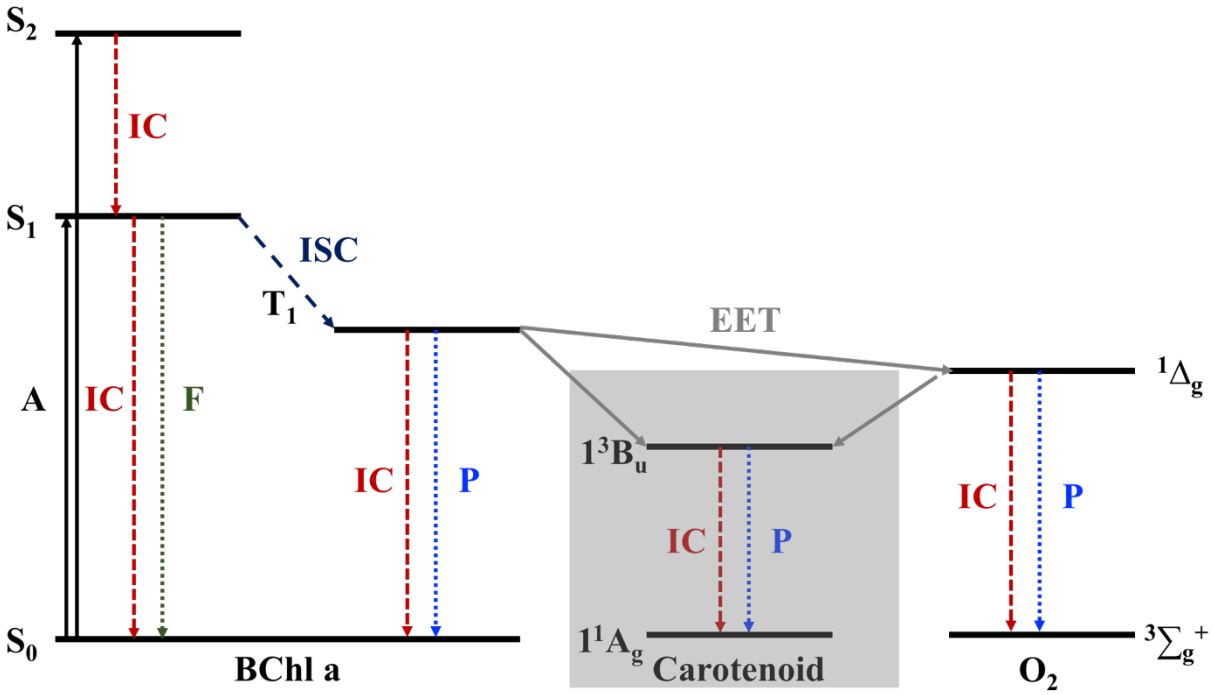

Figure 5.1. Simplified BChl a energy level diagram showing the formation of the BChl a triplet state (T_1) from the singlet excited state (S_1). Because the carotenoids are not present in the FMO complex, BChl a triplet state can either undergo a spin-allowed excitation energy transfer (EET) to an oxygen molecule or decay to the ground state through internal conversion (IC) or phosphorescence (P). Here A stands for the absorption, F – fluorescence, ISC – intersystem crossing [166].

5.3 Computational details

5.3.1 Triplet state constrained QM/MM geometry optimizations

Starting from the 40 optimized structures of the FMO protein from Chapter 4 (i.e., constrained QM/MM optimizations of the singlet ground states of BChl a pigments), constrained QM/MM geometry optimizations were performed for the lowest BChl a triplet states. Each QM region contained a complete BChl a and several neighboring AAs, as shown in Figure 3.3. The list of AAs and details of truncation schemes for each AA can be found in Section 3.3.2 and Figure 4.1. All atoms not included in the QM region, i.e. protein AAs, remaining BChl a pigments, water molecules and Cl^- counter ions were described at the MM level with the AMBER03 force field, as discussed in Section 3.3.2. The MM atoms were held fixed while the QM regions were optimized with the unrestricted PBE0/6-31G(d) [134-139] level of theory using L-BFGS optimization method (gradient tolerance = 300×10^{-5} a.u, multiplicity = 3). The QM/MM geometry optimizations of the triplet states were then conducted as described in Section 4.3.1.

5.3.2 Triplet excited state QM/EFP calculations

EFP parameters of AAs included into optimized QM regions were recomputed for all individual snapshots of the protein structure, using 6-31G(d) basis set. The parameters of other AAs and waters, which geometries were not affected by QM/MM optimizations, were taken from the work described in Chapter 4.

Then, the excited state QM/EFP calculations of the triplet states were conducted analogously to those undertaken in Chapters 3 and 4. Namely, the quantum regions in the excited state computations are the same as described in Chapter 4 and in Figure 3.6. The only difference in the excited state calculations between the present work and the one described in Chapter 4 is that the QM region now corresponds to the triplet rather than singlet optimized structure, and the requested multiplicity of the excited states is set to 3 to ensure transitions to the triplet excited state.

Figure 5.2 shows a schematic representation of the ground and excited potential energy surfaces (PES). As the triplet states in FMO have significantly weaker interactions than the excited singlets (see detailed discussion of triplet-triplet couplings in Chapter 6) and do not excitonically mix, the triplet state quickly localizes on a particular BChl a pigment. Lowering of the triplet state energy upon vibrational relaxation, which generally includes both changes in the internal geometry

of a pigment and reorientation of the solvent, is called the reorganization energy E_r. Thus, values relevant to the spectroscopy of the triplet states are the adiabatic energy differences E_{adiab} between the minimum of the ground state PES and the minimum of the triplet state PES. As shown in Fig. 5.2, the adiabatic energy can be computed either as a sum of the vertical triplet excitation energy at the ground state optimized geometry $E_{S0,vert}$ and the triplet reorganization energy E_r, or as a sum of the vertical triplet excitation energy computed at the triplet optimized geometry $E_{T0,vert}$ and the energy difference at the ground state PES between structures corresponding to the minima of the ground and triplet states ΔE. The latter approach was utilized in this work.

All excited state computations were carried out in the GAMESS electronic structure package [140].

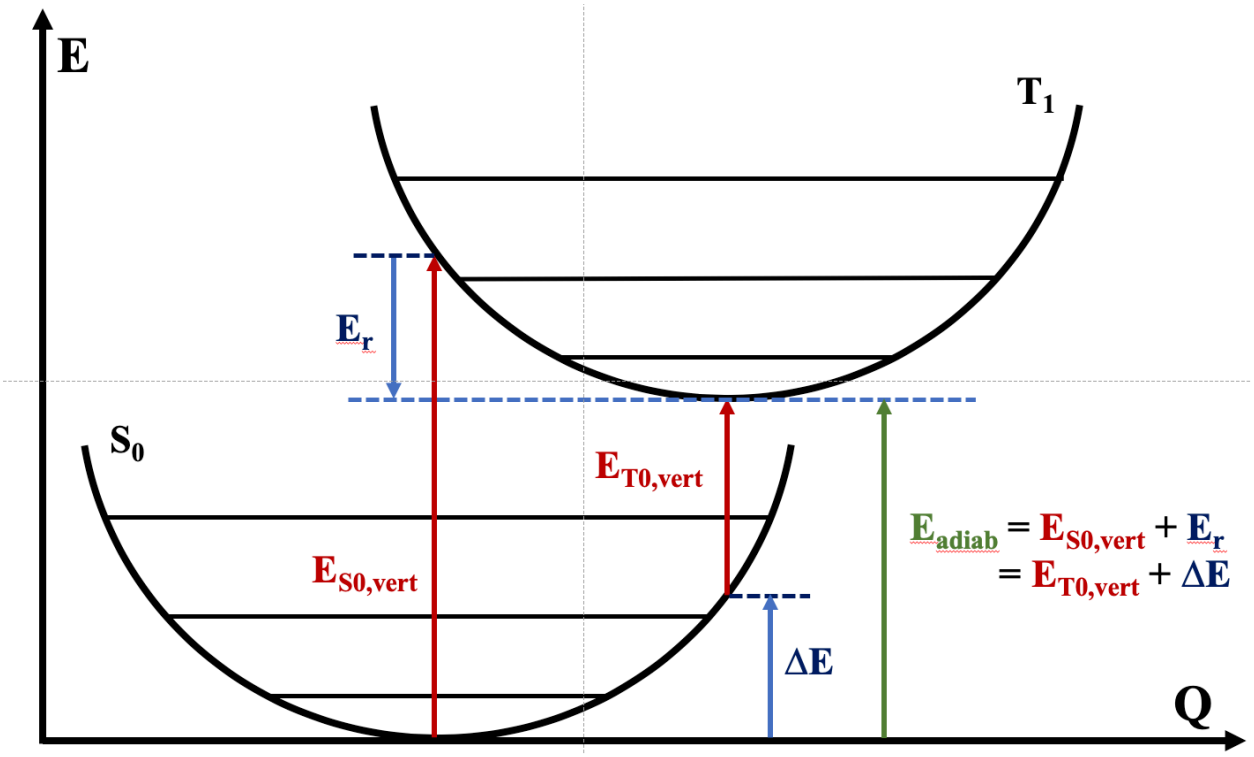

Figure 5.2. Energy diagram of the ground singlet and excited triplet states showing energy computation scheme for the adiabatic transition energies. Singlet-triplet adiabatic energy E_{adiab} can be computed either as a sum of the vertical triplet excitation energy at the ground state optimized geometry $E_{S0,vert}$ and the triplet reorganization energy E_r, or as a sum of the vertical triplet excitation energy computed at the triplet optimized geometry $E_{T0,vert}$ and the energy difference at the ground state PES between structures corresponding to the minima of the ground and triplet states ΔE.

5.3.3 BChl a in pyridine/toluene solution

While DFT and TDDFT computational models are reasonably accurate in predicting relative energy differences when applied to similar species or different configurations of the same species, comparing excitation energies of drastically different species is not very reliable. For example, throughout the thesis, we confidently compare excitation energies of BChl a molecules in different geometries and in different environments, but to compare to the experiment, we had to shift all BChl a excitation energies by ~2400 cm^{-1} (~0.3 eV). Similarly, in case of the triplet transitions, calculations are expected to provide accurate relative energies of triplet excitations, but not their absolute values. However, in order to answer the question of relative energy differences between the S→T transitions in BChl a molecules in FMO and T→S transition in the molecular oxygen, we cannot rely on the calculations, as the absolute excitation energies of both BChl a and molecular oxygen might have significant errors. To resolve this problem, we took the following approach. REF [167] reports experimentally measured phosphorescence energy of BChl a in pyridine/toluene solution. The singlet oxygen transition energy is also provided there. Thus, we estimate the triplet transition energies in FMO by shifting computed triplet energies in FMO by the difference between the experimental and computed phosphorescence energy of a single BChl a molecule in solution. Again, in this approach, we rely on the ability of computational modeling to describe relative changes in excitation energies due to different BChl a geometries and different environments. We also assume that the oxygen transition energy does not change significantly between oxygen condensate and protein. The singlet oxygen emission spectrum was experimentally measured in the cooled and condensed oxygen vapor [168].

In REF [167], it was deduced that BChl a is mainly 6-coordinated in the pyridine/toluene solution. Thus, our simulations of solvated BChl a preserve 6-coordination pattern.

Classical MD simulations: The initial 6-coordinated BChl a structure (see Figure 5.3) was immersed in a pyridine/toluene cubic box. Overall system consisted of 1 BChl a pigment, 6000 toluene and 1500 pyridine molecules (including coordinating pyridines) in a simulation box size of approximately 10.5 × 10.5 × 10.5 nm^3, following toluene/pyridine 4:1 ratio from REF [167]. To preserve 6-coordination of Mg atom of BChl a, we added a constraint between Mg of BChl a and N of pyridine, corresponding to the bond distance of 2.25 Å with a force constant of 543920 kJ/mol. The bonding parameter between N of pyridine and Mg of BChl a was adapted from the previous study [122]. Even though the parameter was applied to between N of histidine and Mg of

BChl a, we assumed that the bond character of histidine N-Mg are similar to that of pyridine N-Mg. In addition, we partially corrected the structures at QM level, the minor artificial effect would be resolved during QM/MM optimizations. The force field for BChl a was obtained from the previous study by Ceccarelli et al [121] and is the same as was used in Chapter 3. This force field was utilized in previous studies of photosynthetic reaction center [122,123]. CHARMM general force field (CGENFF) [169] was used for pyridine and toluene molecules. After energy minimization, molecular dynamics (MD) equilibrations were performed with the NVT and NPT ensembles for 500 ps each. Then, the production run with the NPT ensemble was carried out with a velocity rescale thermostat [127] for temperature control (300 K) and Parrinello-Rahman barostat [128] for pressure control (1 bar). C-H, O-H and N-H bond lengths were constrained with the LINCS algorithm [129]. A 1 nm cutoff was used to handle Lennard-Jones potentials. Electrostatic long-range interactions were treated with particle mesh Ewald summations (PME) [130,131] with a real-space cutoff of 1 nm. Total 10 ns production run was carried out with 2 fs time step. All classical simulations were performed with the GROMACS package (version 2016.5) [132].

We randomly extracted 100 atomic configurations from the last 5 ns of the MD trajectory without bias. Specifically, 5,000 MD structures were saved to disk for each system (one structure every 1 ps). Out of those, 100 snapshots were selected using a random number generator in the range from 1 to 5,000. These 100 random configurations have been considered for the electronic structure calculations.

Constrained QM/MM geometry optimizations: Started from selected atomic configurations, we performed QM/MM constrained geometry optimizations in which MM regions were kept fixed. The electrostatic embedding QM/MM scheme was employed [133]. Each QM region contained a single BChl a molecule with phytyl tail truncated between carbons C3 and C5 according to the PDB nomenclature, i.e., on the bond following the double bond (see Figure 3.4). Additionally, two Mg-coordinating pyridines were included in the QM region. All atoms not included in the QM region, i.e., remaining pyridine and toluene molecules were described at the MM level with the same force fields (CGENFF) as used in MD simulations. The MM atoms were held fixed while the QM was optimized with PBE0/6-31G(d) [134-139] level of theory using L-BFGS optimization method (gradient tolerance = 300×10^{-5} a.u). Followed by ground singlet state geometry optimizations, triplet state geometry optimizations were performed starting from the singlet optimized structures using unrestricted PBE0/6-31G(d) level of theory with multiplicity 3. All

geometry optimizations were performed using QM/MM interface between GAMESS [140] quantum chemistry package and GROMACS 4.6.5 molecular dynamics package [132].

EFP parameters: HF/6-31+G(d) level of theory was used to optimize geometries of the pyridine and toluene molecules. Then, the EFP parameters for the pyridine and toluene molecules were computed with 6-31G(d) basis set for both electrostatic and polarization term. Exponential electrostatic damping for charges (SCREEN2) was used to account for charge-penetration effects between EFP fragments [57]. Electrostatic QM-EFP interactions were not screened. Gaussian-type polarization damping with default settings (POLAB 0.6) was employed between the fragments but not between the QM and EFP regions. The Flexible EFP scheme (see Chapter 2) was utilized to adjust EFP parameters of pyridine and toluene to their structures at MD snapshots. All EFP parameters were computed in the GAMESS electronic structure package [140].

Singlet and triplet excited state QM/EFP calculations: the QM regions included a single BChl a molecule and two Mg-coordinating pyridines. As shown in Figure 5.4, solvent molecules within 15 Å from the QM BChl a head were represented as effective fragments. Outside the 15 Å shell, all atoms were described by the MM point charges with the same force fields (CGENFF) as used in MD simulations. TDDFT PBE0/6-31G(d) was used for computing singlet and triplet excitation energies in all configurations.

All excited state computations were carried out in the GAMESS electronic structure package [140].

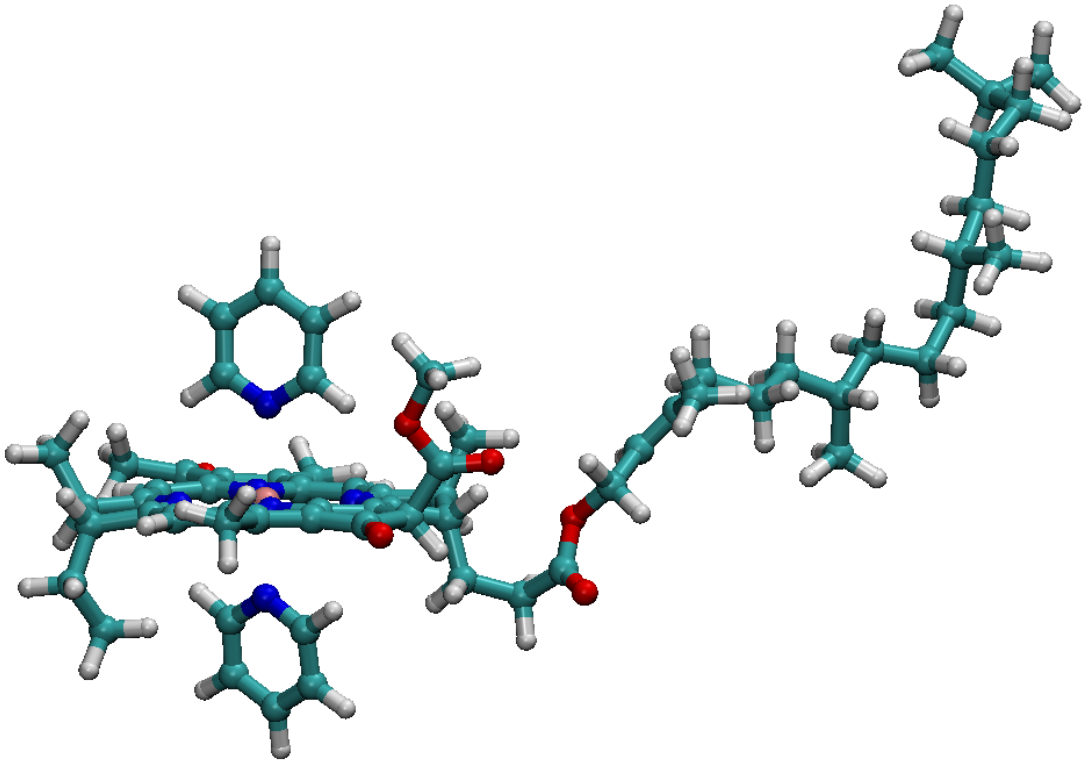

Figure 5.3. Initial atomic configuration of 6-coordinated BChl a.

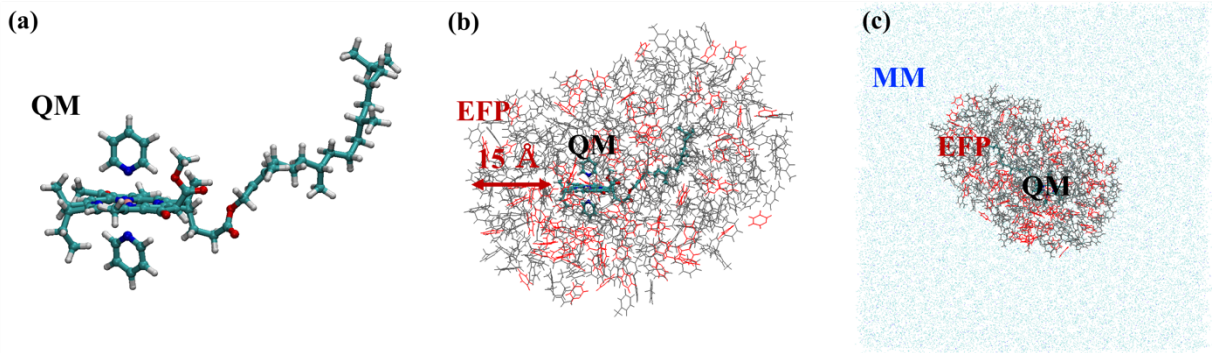

Figure 5.4. Separation into (a) QM, (b) EFP, and (c) MM regions in the QM/EFP calculations. The EFP region includes toluene (gray colored) and pyridine (red colored) effective fragments within 15 Å from the QM BChl a head. All atoms outside 15 Å from the QM BChl a head are described with MM point charges.

5.4 Results and Discussion

Table 5.1 summarizes the main computational results of this work. Figure 5.5 also shows adiabatic triplet excitation energies of BChl a in FMO and in solution.

Discussion: comparison of vertical energies between solvent and FMO sites, comparison of E_r energies – similar relaxation in protein and solvent.

Computed triplet excitation energies in solution can be directly compared to experimental phosphorescence measurements. The experimentally measured phosphorescence of 6-coordinated BChl a in solution is P_{expt} = 8116.88 $\pm$ 60.0 cm^{-1} [167], while the average S-T adiabatic energy from calculations is $P_{calc} = E_{adiab,\ solution}$ = 4932.08 $\pm$ 23.6 cm^{-1}. Thus, we estimate the error of TDDFT PBE0/6-31G(d) in predicting triplet excitation energies as $E_{shift} = P_{expt} - P_{calc}$ = 3184.8 $\pm$ 64.5, where uncertainty of the shift Δ is computed as a $\Delta = \sqrt{\Delta^2_{Expt.} + \Delta^2_{Comp.}}$. Interestingly, the energy shift for the triplet excited state is very different from the shift we adopted for modeling the singlet excitation spectrum of FMO, namely ~ +4900 cm^{-1} for the triplets and ~ -2400 cm^{-1} for the singlets, highlighting limitations of DFT functionals in predicting electronic states of different multiplicities.

It is also worth mentioning that the spread of the triplet site energies (even vertical ones) is significantly wider that the spread of the singlet site energies, namely about ~80 cm^{-1} spread of the triplet states versus ~300 cm^{-1} spread of the singlet excited states. This suggests strong interaction of the protein with the triplet states – even though formally the triplets are less polar than the singlets.

The directly calculated and shifted by E_{shift} adiabatic S-T energies in FMO are plotted in Fig. 5.5. The experimentally measured transition energy of the singlet oxygen molecule (7861 cm^{-1}) is also shown there. In the solution, the experimental S-T energy of the BChl a is about 255 cm^{-1} above the transition energy of the singlet oxygen. However, in FMO, the lowest triplet state at site #3 is found 662 cm^{-1} below the oxygen transition. The other sites are isoenergetic with the oxygen transition energy within the accuracy of our calculations and experiment.

As discussed in Section 5.1, the triplet state (T_1) of BChl a is populated from the singlet excited state (S_1) through the intersystem crossing as shown in Figure 5.1. Because of absence of carotenoids in the FMO complex, triplet state of BChl a can either decay to the ground state or transfer its energy to an oxygen molecule. To prevent the formation of highly reactive singlet

oxygen in the FMO protein, the energy of BChl a triplet state should be below than that of the singlet oxygen, such that the only available pathways for the triplet state decay are the intersystem crossing or phosphorescence to the ground state. Assuming that the triplet-triplet transfer happens reasonably fast (within a μs, as will be discussed in Chapter 6) and most of the triplet population transfers to the lowest site #3, it seems plausible for BChl a pigments in the FMO protein to dissipate most of their triplet excitation energy to the ground state without forming highly reactive singlet oxygen. Thus, exceptionally low energy of the triplet state at the lowest site #3 due to a specific protein environment can be considered as an evolutionary developed mechanism of the photoprotection in the FMO protein.

Table 5.1. Summary of triplet state properties of the FMO complex and BChl a in solution. Average energy values and uncertainties are provided. All values are in cm^{-1}.

	$E_{S0,vert}$	E_r	$E_{T1,vert}$	dE	E_{adiab}
solution	6566.9±17.4	1634.8±16.9	4140.0±29.9	792.1±9.4	4932.1±23.6
FMO site #1	6009.0±18.8	1349.8±27.9	3944.2±47.7	715.1±12.1	4659.3±39.1
FMO site #2	6572.5±34.2	1939.4±41.5	3889.4±42.7	743.7±19.4	4633.1±38.1
FMO site #3	5507.6±20.6	1492.7±32.3	3353.8±45.9	661.1±20.0	4014.9±46.1
FMO site #4	6225.8±25.4	1686.0±35.9	3850.5±43.5	689.3±14.7	4539.8±34.9
FMO site #5	6226.7±27.8	1686.0±54.0	3956.2±42.8	672.7±42.7	4628.8±50.0
FMO site #6	6265.2±22.2	1766.1±40.1	3822.8±40.5	676.2±28.5	4499.1±40.8
FMO site #7	6138.4±16.2	1488.0±57.4	3949.1±38.9	701.3±48.1	4650.4±60.6
FMO site #8	6387.1±37.3	1319.7±91.5	4158.7±41.1	908.7±88.8	5067.4±94.6

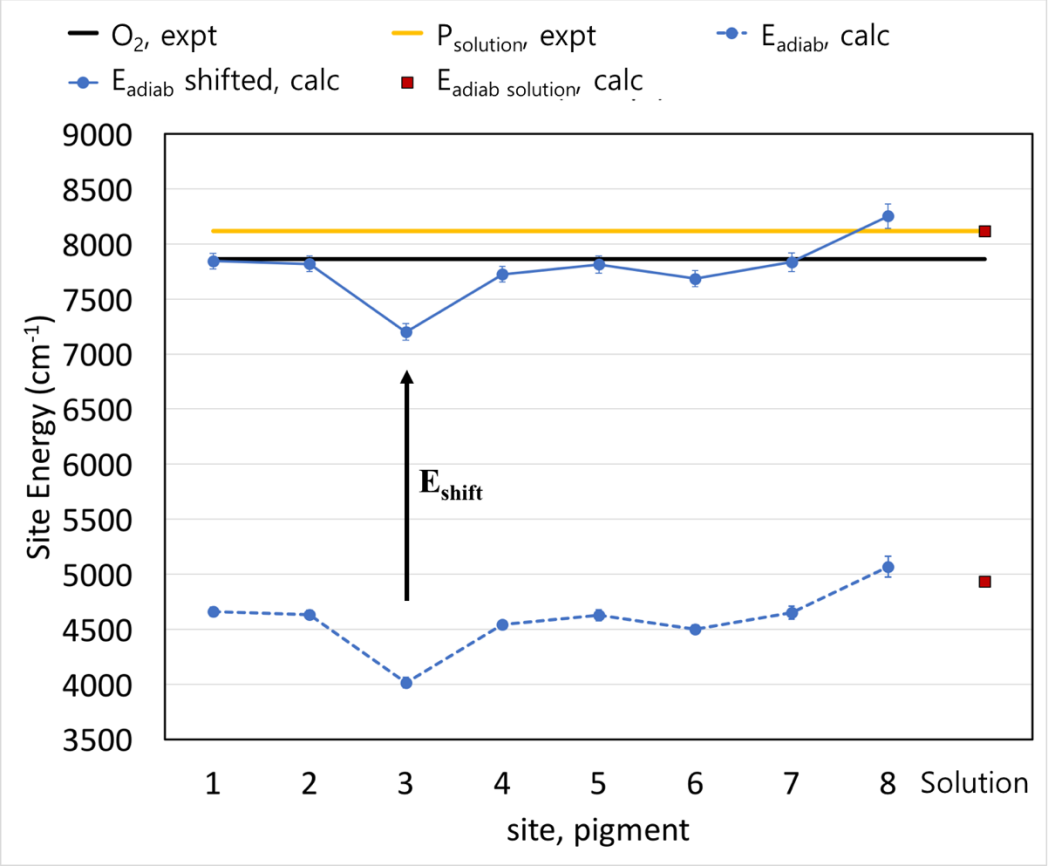

Figure 5.5. Triplet state BChl a energies in FMO and solution. Directly computed (E_{adiab}) and shifted (E_{adiab} shifted) S-T energies of pigments in FMO are compared to the T-S transition in molecular oxygen (O_2, expt) (REF [167]).

5.5 Conclusions

In this study, we have expanded our multiscale modeling strategy for the FMO complex to the triplet excited states. Started from the optimized wild-type structures, we generated a set of triplet-optimized BChl a pigments. In addition, to relate computed triplet transition energies to the experimental values, we modeled BChl a 6-coordination state in toluene/pyridine solution, for which the phosphorescence experimental data are available. Due to the absence of carotenoids, the triplet state of BChl a pigment in the FMO protein should be lower than the singlet oxygen state to protect itself from photodegradation. As shown in triplet BChl a energy analysis, the lowest site triplet energy is below the transition energy in molecular oxygen. This result suggests that EET to molecular oxygen is energetically unfavorable, such that at least one of the photoprotection mechanisms in FMO is based on sufficiently low energy of its triplet states. Further investigations of this topic may include analysis of possible locations of O_2 in the FMO protein through molecular

dynamics simulations, i.e., answering the question whether photoprotection mechanism also includes structural barriers to the oxygen molecule to approach BChl a sites, and investigation of the photostability in FMO mutants, especially mutant #3, where the site #3 energy is significantly destabilized.

CHAPTER 6. TRIPLET DYNAMICS OF THE FMO COMPLEX

6.1 Introduction

Investigation of detailed mechanisms of energy transfer in photosynthetic pigment-protein complexes provides knowledge necessary for the development of biomimetic systems such as bioinspired artificial antenna, light-harvesting materials, and artificial photosynthetic devices [170-174]. The Fenna-Matthews-Olson (FMO) pigment-protein complex, found in green sulfur bacteria, is one of the most thoroughly studied photosynthetic pigment-protein complex (see Figure 3.1) [78-112]. The primary function of FMO is to transfer the excitation energy from a much larger chlorosome antenna to the intramembrane reaction center complex, where electronic excitation initiates charge transfer process. The FMO complex is a trimer possessing C_3 symmetry. Each subunit encloses seven bacteriochlorophyll a (BChl a) chromophores and binds the eighth BChl a pigment between the subunits. Close packing of BChl a pigments in each monomer subunit leads to strong excitonic interactions and delocalization of excited states over multiple pigments.

An appealing way to investigate excitonic interactions in the FMO complex is to introduce controlled changes to the system and study the effect of the mutation on the excitonic spectra. This approach allows decomposing contributions of specific protein parts into the total spectroscopic signals. Chapter 7 presents our work on modeling such effects due to single-point mutations. Another possibility of modifying the excitonic interactions in FMO is by considering so-called "transient mutants", i.e., an excitonic system in which the excited FMO protein undergoes intersystem crossing to a triplet state. In this scenario, the BChl a pigment in the triplet state does not participate in the excitonic interactions with other pigments, which leads to effective removal of this pigment from the excitonic Hamiltonian. Such transient mutants were recently observed and characterized spectroscopically in REF [158]. New transient circular dichroism (TRCD) experiments are also underway in the Savikhin lab [Savikhin, private communications]. The TRCD spectra provide additional structural information and are more sensitive to changes in the excitonic structure compared to the original transient absorption measurements of the excitonic states. The goal of this Chapter is to use molecular modeling to facilitate the understanding and interpretation of triplet state dynamics in the FMO complex.

6.2 Motivation

6.2.1 Transient spectroscopy of FMO

As described in Chapter 3 and Figure 6.1 (a), the excitonic interactions of the FMO complex can be describe with the electronic Hamiltonian H_{ij}:

$$H_{ii} = H_i,\ i=1,n \tag{6.1}$$

$$H_{ij} = V_{ij}, i \neq j,\ i,j=1,n \tag{6.2}$$

where $H_{ii} = H_i$ are the electronic excitation energies of pigment i in the absence of interactions with other pigments (i.e., site energies) and the off-diagonal elements $H_{ij} = V_{ij}$ are interaction energies between pigments i and j. The n eigenvalues found by matrix diagonalization correspond to the energies ε_i of excitonic levels, while the respective eigenvectors c_{ij} represent expansion coefficients of the excitonic wave function, ψ_i, in terms of wave functions of individual pigments U_j as follows:

$$|\psi_i\rangle = \sum_{j=1}^{n} c_{ij}|U_j\rangle \tag{6.3}$$

Combining transition dipole moments of each pigment with eigenvalues and eigenvectors of the Hamiltonian allows computing absorption and CD spectra (see Eqns (3.3) and (3.4) in Chapter 3).

In REF [158], it was demonstrated that the information on the individual pigments in FMO can be accessed using *triplet state* dynamics that occurs on nanosecond-microsecond time scale, as opposed to the femtosecond singlet exciton dynamics. Upon excitation, a fraction of excited complexes (up to 10%) ends up in a triplet excited state. Our exciton simulations show that the triplet excitation is localized on a single pigment, unlike a delocalized singlet excited state (see Figure 6.5 and corresponding discussion in 6.2.3). Thus, the excited state absorption signal for a triplet excitation corresponds to the excitonic absorption spectrum of the remaining, un-excited pigments, with (*n-1)* bands. Essentially, the FMO complex in a triplet state is similar to a mutant missing one pigment, with the rest of the structure largely unaffected (see Figure 6.1 (b)), and the measured difference spectra reveal the properties of that pigment in the most direct way.

Figure 6.2 shows absorption and CD difference spectra (ΔA and ΔA_{CD}, respectively) kinetics probed at a number of wavelengths at room temperature shown as spectral changes at

different time delays between excitation and probe pulses [Savikhin, private communications]. A global fit to this data reveals 4 major decay components; respective decay associated difference spectra for ΔA and ΔA_{CD} are shown in Figure 6.2 c and d. The data could be described with four kinetic components (1 µs, 11 µs, 55 µs and 0.1 µs). The three slowest components are consistent with those measured in low-temperature experiments in REF [158]. Below, we will concentrate on the analysis of the most intense 55 µs components.

(a) Excitonic Hamiltonian

Pigment	i	j	$\cdots$	k
i	H_i	V_{ij}		V_{ik}
j	V_{ji}	H_j		V_{jk}
$\vdots$				
k	V_{ki}	V_{kj}		H_k

(b) Triplet Hamiltonian

Pigment	i	j	$\cdots$	k
i	0	0		0
j	0	H_j		V_{jk}
$\vdots$				
k	0	V_{kj}		H_k

Figure 6.1. Excitonic Hamiltonians (a) of the n-pigment system and (b) the system in which i^{th} pigment is in the triplet state. The triplet pigment does not interact excitonically with other pigments, such that the excitonic Hamiltonian misses i^{th} column and row. In the simplest model, other sites are assumed to be unaffected.

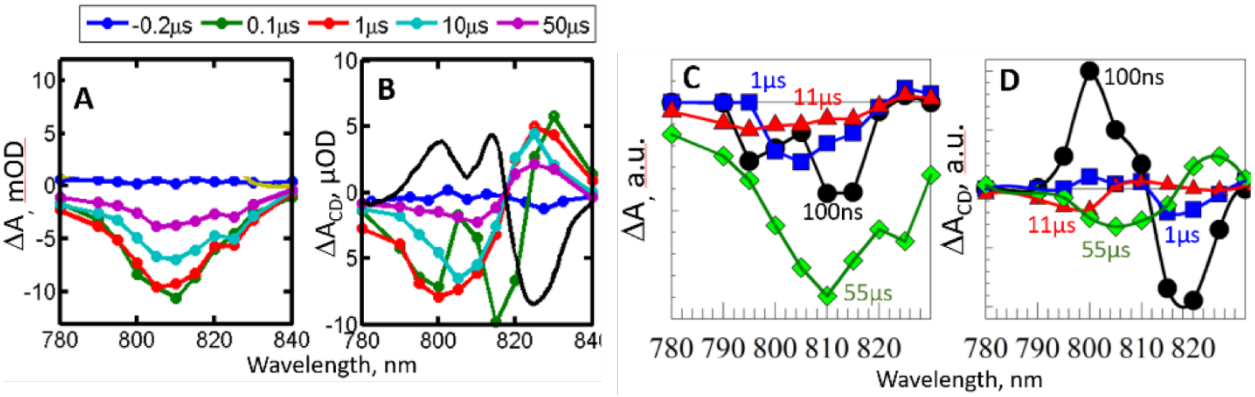

Figure 6.2. **A** and **B**: ΔA and ΔA_{CD} difference spectra measured at fixed times before and after excitation. **C** and **D**: decay associated spectra reveal 4 major decay components reflecting optical properties of individual pigments involved in the triplet energy transfer [Savikhin, private communications].

6.2.2 Triplet excitonic Hamiltonian of the FMO complex

Figure 6.3 and Table 6.1 show distribution of triplet site energies in wild type FMO. Note, since the interaction between triplets is orders of magnitude weaker than that between the singlets (see discussion in Section 6.2.3), the triplet state is expected to localize on a single pigment forcing this pigment to undergo vibrational relaxation to a minimum on the triplet state potential energy surface (PES) (see Figure 5.2). Site energies shown in Figure. 6.3 account for such geometrical relaxations of each pigment in the triplet state. To achieve that, we performed constrained QM/MM geometry optimizations on 40 snapshots in which the pigment in the QM region was described with the triplet state unrestricted DFT PBE0/6-31G(d) (see Computational details in Section 5.3). Surprisingly, the triplet site energies of different pigments are spread (4014 - 5067 cm^{-1}) almost twice wider than the singlet site energies (14564~14984 cm^{-1}), suggesting a stronger interaction (i.e., larger electrochromic shifts) of the protein with the pigments in the triplet state.

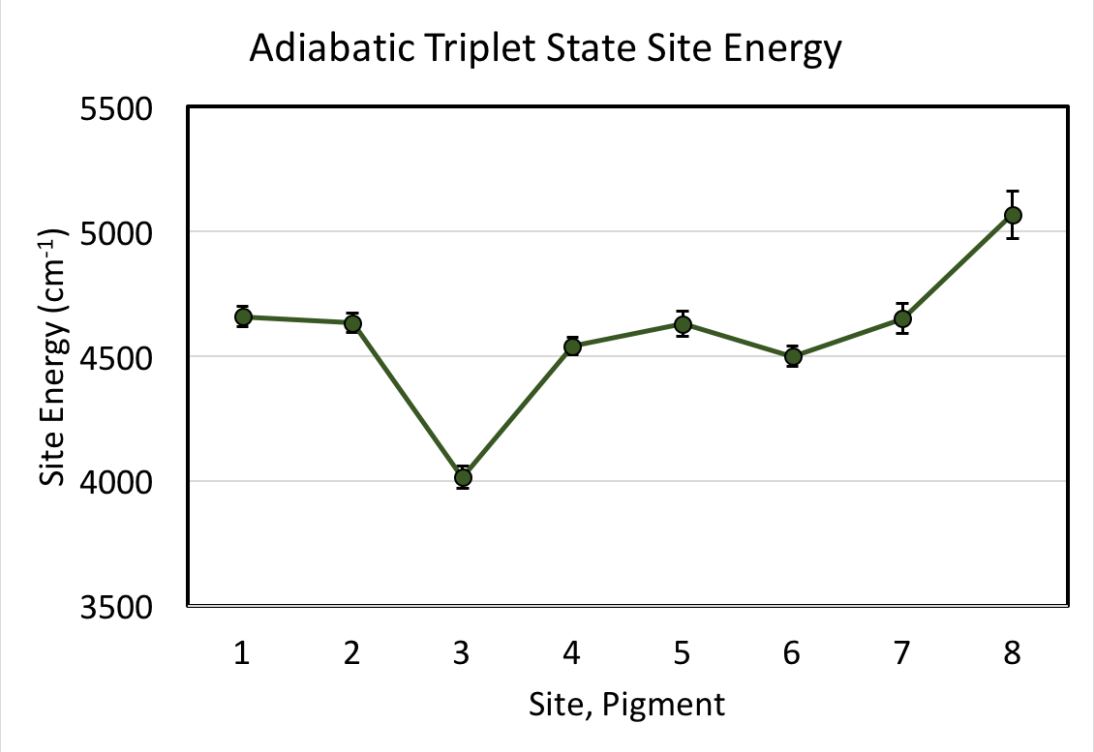

Figure 6.3. Average triplet site energies computed with QM/EFP at 40 structural snapshots after constrained QM/MM optimizations of the lowest triplet state with unrestricted DFT PBE0/6-31G(d) level of theory. Uncertainties of the average values are shown with vertical lines.

Table 6.1. Average, standard deviation, and uncertainty (cm^{-1}) of QM/EFP singlet (vertical) and triplet (adiabatic) site energies computed at 40 snapshots.

Site	Singlet	σ	uncertainty	Triplet	σ	uncertainty
1	14798.0	62.1	9.8	4659.3	247.4	39.1
2	14984.4	105.1	16.6	4633.1	240.9	38.1
3	14564.3	69.5	11.0	4014.9	291.4	46.1
4	14749.9	65.1	10.3	4539.8	220.4	34.9
5	14891.1	66.8	10.6	4628.8	316.1	50.0
6	14899.5	50.5	8.0	4499.1	258.0	40.8
7	14765.9	57.2	9.0	4650.4	383.4	60.6
8	14899.0	127.5	20.2	5067.4	598.1	94.6

6.2.3 Triplet-triplet (T-T) coupling calculations

We also estimated electronic couplings and rate constants for transitions between triplet states of neighboring pigments, as this information can facilitate analysis of the triplet state dynamics. As described in Section 6.3.1, the electronic Hamiltonian is composed of two components: (i) diagonal elements, i.e., site energies, and off-diagonal elements, i.e., electronic couplings between sites. In the case of the (traditional) singlet electronic Hamiltonian, the electronic couplings can be computed by either point-dipole approximation (PDA), transition charges (TrEsp) (see Section 3.3.5), or transition densities [119]. However, triplets are electronically dark states with zero transition dipoles or transition charges. Hence, electronic couplings between the triplets are dominated by the Dexter exchange mechanism, in which the interaction happens due to the overlap of electronic wave functions of the donor and acceptor molecules [175-178]. Here we evaluate T-T couplings between pigments by utilizing the fragment excitation difference (FED) method [176]. In the FED method, the electronic couplings are described from analysis of electron (attachment) and hole (detachment) densities in the dimer:

$$V_{T-T} = \frac{(E_m - E_n)\Delta x_{mn}}{\sqrt{(\Delta x_{nn} - \Delta x_{mm})^2 + 4\Delta x_{mn}^2}} \tag{6.4}$$

where E_m and E_n are excitation energies of the triplet states m and n. Δx_{mn} is an excitation difference operator:

$$\Delta x_{mn} = x_{mn}^D - x_{mn}^A = \int_{r\in D} \rho_{ex}^{(mn)}(r)dr - \int_{r\in A} \rho_{ex}^{(mn)}(r)dr \tag{6.5}$$

where D and A define the donor and acceptor spaces. $\rho_{ex}^{(mn)}(r)$ is is the sum of attachment and detachment densities for transition $|m\rangle \rightarrow |n\rangle$, as they correspond to the electron and hole densities in an excitation.

Triplet-triplet interactions are short-range and decay exponentially with the distance between a donor and acceptor molecules. Thus, only the closest pairs of the pigments might have non-negligible T-T couplings and rate constants. Table 6.2 shows the smallest average distances between BChl a atoms involved in aromatic system (shown in Figure 6.4) for different pairs of BChl a pigments.

We computed triplet-triplet couplings, $V_{T\text{-}T}$, in the closely spaced pigment pairs, namely pairs 1-2, 1-6, 3-4, 3-7, 4-5, 4-7, 5-6, 5-7, and 6 -7, and plotted them as a function of the distance between aromatic carbons of BChl a pigments (see Figure 6.5). Larger distances in pairs of other pigments suggest that the T-T couplings in those pairs will be significantly smaller, and a potential transfer of the triplet state from one pigment to another - significantly slower. As a preliminary investigation, we calculated T-T couplings based on 25 FMO structures. In all calculations, both BChl a geometries are optimized in the triplet state. All calculations were performed using FED method with TDDFT PBE0/6-31+G(d) level of theory in the QCHEM 5.0 electronic structure package [75]. These calculations were performed on BChl a dimers in the gas phase (no protein environment or neighboring AAs). Figure 6.5 reveals that the largest triplet-triplet couplings between neighboring pigments are on the order of 1-10 μeV (0.01 – 0.1 cm^{-1}), which is at least 1000 times smaller than the largest singlet-singlet couplings (see Figure 3.10 for singlet couplings). Such small T-T couplings suggest tha all triplet states in FMO are completely localized on individual pigments. The largest T-T couplings are observed in a pair 5-6, and couplings between pigments 1-2, 3-7 and 4-7 being slightly smaller. While there is a general tendency of the couplings to quickly decay with increase of the distance between carbon atoms in interacting pigments, the coupling-distance correlation is rather weak and is often broken both within each pair of the pigments and among different pigment pairs. One such example is the couplings in similarly spaced pairs 4-7 and 6-7 that differ more than twice. Thus, while a simple distance-based formula provides qualitative estimation of the T-T couplings, quantum-mechanical calculations are necessary for a quantitative analysis.

The triplet-triplet energy transfer rate constant k between a pair of pigments can be estimated using the following equation [179]:

$$k = 0.93 \frac{2\pi}{\hbar} V_{T-T}^2 \tag{6.6}$$

where reduced Planck constant $\hbar$=6.58×10^{-16} eV·s, V_{T-T} is coupling in eV and 0.93 eV^{-1} is the best-case scenario spectral overlap coefficient between the absorption and emission BChl a spectra [179]. Figure 6.6 shows rate constants computed using Eqn (6.6) based on the couplings from Figure 6.5. The main results are also summarized in Table 6.3, which shows the largest and average couplings and the fastest and average rates for each pair of fragments.

Based on the present modeling (see Figure 6.6 and Table 6.3), the fastest triplet-triplet transfer occurs between BChl a #5 and #6. In a timeframe of the pump-probe experiments (~1 μs), triplet energy transfer might also occur in pigment pairs 1-2, 3-7, and 4-7, with slower rates in pairs 3-4, 4-5 and 6-7. The very fast energy transfer between pigments #5 and #6 suggests that the triplet population on these two BChl a is equilibrated. Pigments #1 and #2, even though can exchange triplets with each other, are effectively separated from the remainder of the system, so we do not expect any triplet transfer from/to the pair of pigments #1 and #2. Similarly, pigment #8 is sufficiently separated from other pigments and does not exchange triplets with the remainder of the system.

Even though the data presented in Figures. 6.5-6.6 and Table 6.3 provide a qualitative picture of possible triplet transfer pathways in FMO, there are several effects that need to be accounted for a more quantitative analysis of couplings and rate constants. One minor drawback of the presented calculations is that in the construction of the dimers, both monomers were optimized at the triplet state geometries. A more accurate estimate of the couplings can be obtained if the accepter monomer is in the optimized ground state geometry while the donor monomer is in the optimized triplet geometry. Additionally, environment effects on the couplings can be non-trivial and require investigation. As a preliminary study, we further examined the T-T coupling between sites #3 and #4 with five computation schemes as shown in Figure 6.7. In all calculations, the site #3 geometry is optimized at the singlet ground state and the site #4 geometry is in the triplet state. All calculations were performed using FED method with TDDFT PBE0/6-31+G(d) level of theory in the QCHEM 5.0 electronic structure package [75].

Figure 6.8 provides computed T-T couplings between sites #3 (acceptor) and #4 (donor) depending upon inclusion of neighboring AAs (quantum-mechanically) and polarizable protein environment using the QM/EFP model. Additionally, the couplings were computed when both

pigments were positioned in their triplet state geometries. Comparisons of T-T couplings between T_1-T_1 and S_0-T_1 geometries show that the triplet geometry of the donor pigment contributes to the increase of the couplings. The calculations for sites #3 and #4 also suggest that the T-T coupling is not affected by inclusion of Mg-coordinating AAs, but slightly increases when the H-bonded tyrosine residue is included in the QM region. Based on a visual inspection of the molecular orbitals involved in triplet excitations, this H-bonded tyrosine is the only AA that shares some of the BChl a electronic density in the pigment pair 3-4. On the other hand, once a full protein environment is included at the EFP level, T-T coupling between sites #3 and #4 increases significantly (from 2.5×10^{-6}~3.5×10^{-6} eV in gas phase to ~7.7×10^{-6} eV in protein), suggesting that the protein environment delocalizes triplet states of BChl a pigments. Note that a cumulative increase in the coupling value from ~0.02 cm^{-1} (gas phase dimer) to ~0.06 cm^{-1} (the dimer with H-bonding AA in polarizable environment) results is almost an order of magnitude increase in the rate constant and an order of magnitude faster rate. While these single-point calculations should not be generalized to other structures of the same pigment pair and other pairs, they nevertheless suggest that the results presented in Table 6.2 are qualitative only, and the real rates can easily differ from the computed by an order of magnitude.

Table 6.2. The smallest average distances between aromatic carbons in all pairs of pigments in Å (above diagonal) obtained at 100 snapshots of molecular dynamics trajectory and root mean square (rms) fluctuations from the average values (below diagonal). Pigment numbers are shown in diagonal.

1	5.00	20.27	24.27	17.65	8.52	15.99	8.27
0.14	**2**	9.11	17.59	20.09	16.17	11.26	13.26
0.20	0.23	**3**	5.71	16.23	19.66	5.54	27.32
0.27	0.30	0.15	**4**	6.20	15.42	6.81	31.54
0.25	0.29	0.29	0.20	**5**	4.86	7.65	26.47
0.24	0.25	0.28	0.26	0.18	**6**	7.09	15.02
0.25	0.33	0.15	0.20	0.24	0.24	**7**	19.00
0.29	0.42	0.45	0.36	0.32	0.31	0.34	**8'**

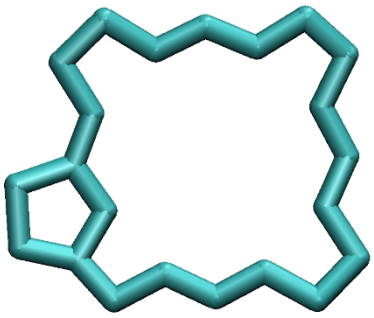

Figure 6.4. Aromatic carbons of BChl a pigment used for edge-to-edge distance calculations between pairs of pigments in FMO.

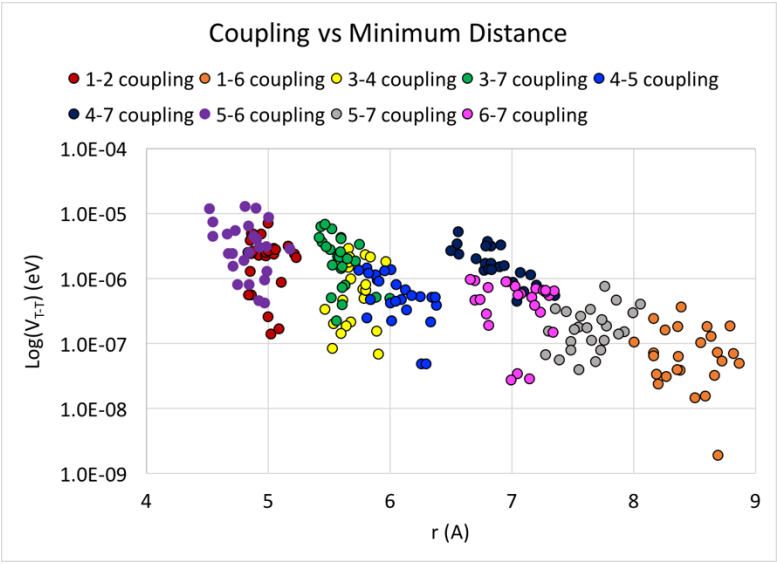

Figure 6.5. Computed T-T couplings (cm^{-1}) between several neighboring BChl a pairs. Calculations are performed at 25 structural snapshots in gas phase using FED method.

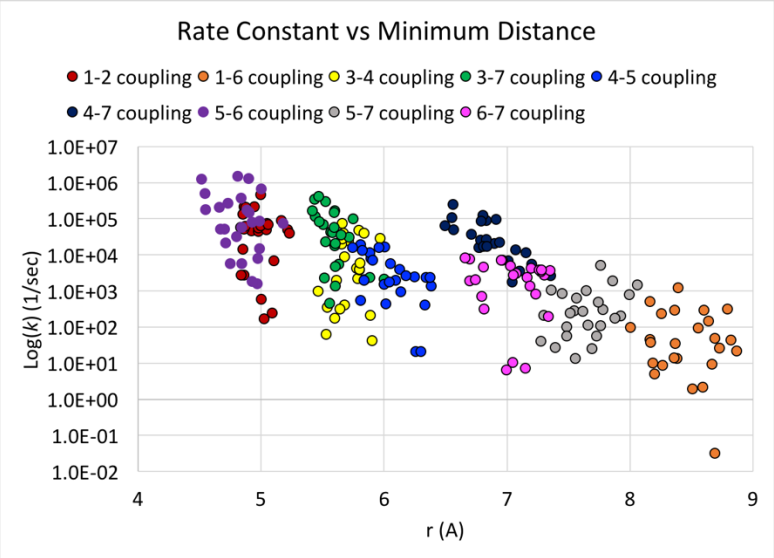

Figure 6.6. Computed rate constants using Eqn (6.6) based on the coupling from Figure 6.5.

Table 6.3. Triplet-triplet couplings $V_{T\text{-}T}$ (eV) and transfer rates (1/k, μs) for selected pairs of BChl a pigments.

Pigment pair	Average distance, Å	Average coupling $V_{T\text{-}T}$, eV	Largest coupling $V_{T\text{-}T}$, eV	Average rate, μs	Fastest rate, μs
1-2	4.98	2.52×10^{-6}	7.25×10^{-6}	5.02×10^{2}	2.14
1-6	8.45	9.44×10^{-8}	3.72×10^{-7}	1.30×10^{6}	815
3-4	5.72	9.89×10^{-7}	2.94×10^{-6}	2.44×10^{3}	13.1
3-7	5.59	2.56×10^{-6}	6.94×10^{-6}	1.97×10^{2}	2.34
4-5	6.06	6.71×10^{-7}	1.48×10^{-6}	4.25×10^{3}	51.6
4-7	6.86	1.95×10^{-6}	5.33×10^{-6}	9.03×10^{1}	3.97
5-6	4.82	4.34×10^{-6}	1.31×10^{-5}	7.57×10^{1}	0.66
5-7	7.64	2.13×10^{-7}	7.60×10^{-7}	1.09×10^{4}	195
6-7	7.04	5.24×10^{-7}	9.78×10^{-7}	1.59×10^{4}	118

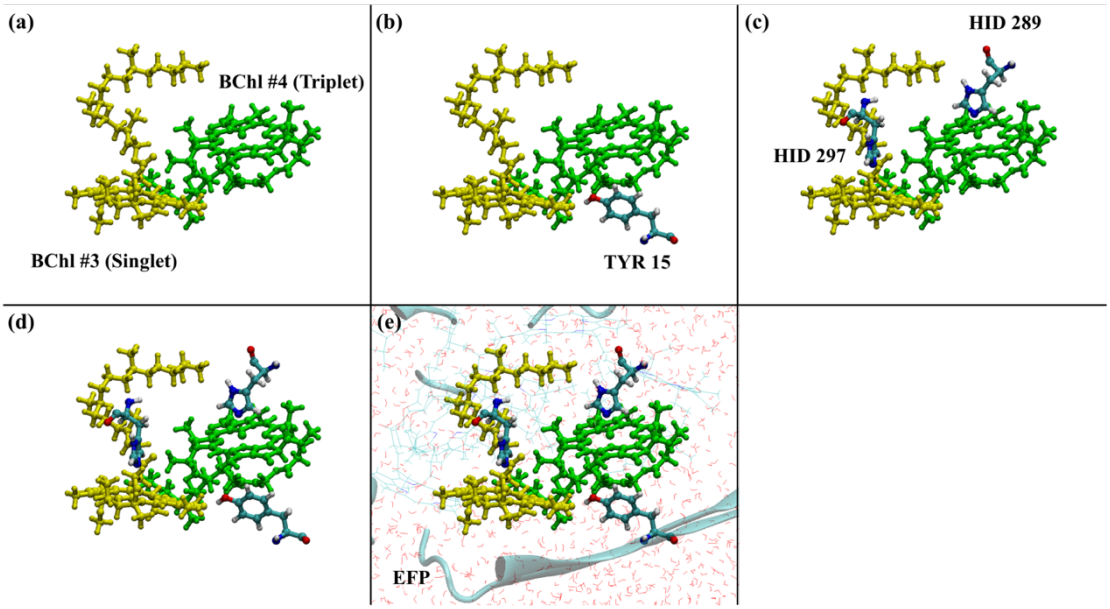

Figure 6.7. Schemes of T-T coupling calculations between sites #3 and #4. (a) site #3 and #4 (dimer) in gas phase, (b) sites #3, #4 and tyrosine residue hydrogen bonded to site #3 (dimer+HB) in gas phase, (c) sites #3, #4 and corresponding Mg-coordinating histidine residues (dimer+HIDs) in gas phase, (d) sites #3, #4, tyrosine, and Mg-coordinating histidine residues (dimer+HB+HIDs) in gas phase, and (e) the scheme (d) considered in the protein environment via QM/EFP calculations.

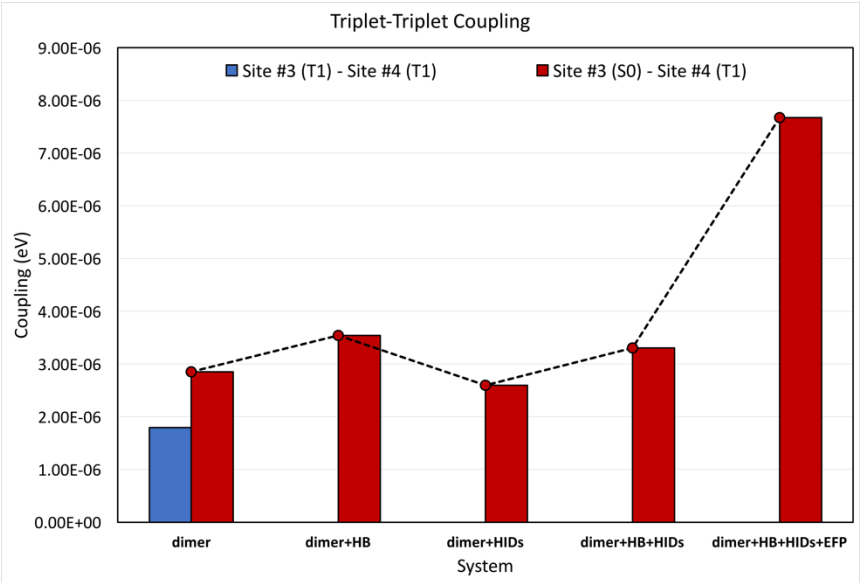

Figure 6.8. Computed T-T couplings (eV) between sites #3 and #4 using FED method. The system descriptions are provided in Figure 6.7. T-T couplings based on the structures where the donor pigment (site #4) are in the triplet geometry and the acceptor pigment (site #3) is in the ground state geometry is shown in red. Coupling computed at the structures where pigments are in their triplet geometries is shown in blue.

6.2.4 Electrochromic shifts due to the triplet BChl a

An interesting and not actively discussed phenomenon that we observed in our modeling is that BChl a in the triplet state possesses a quite different charge distribution compared to that in the ground singlet state. For example, static dipole moments of the ground singlet and excited triplet state are almost perpendicular to each other, as shown in Figure 6.9. Thus, triplet BChl a in the FMO complex, due to its different charge distribution and potentially different geometry, might affect non-negligibly site energies of neighboring pigments.

To verify this hypothesis, we conducted preliminary gas phase computations of several neighboring BChl a pairs on a single snapshot. First, we performed Stone's distributed multipole analysis (DMA) [180-182] using unrestricted DFT PBE0/6-31G(d) [134-139] to compute electrostatic multipoles corresponding to the electronic density of BChl a in the triplet state. Polarization parameters (LMO centroids and polarizability tensors) of BChl a were still adopted from the singlet ground state parameters as utilized in Chapters 3 and 4. Combining electrostatic multipoles (DMA analysis, triplet state) and polarizabilities (singlet ground state adjusted via Flexible EFP), a complete "triplet" BChl a EFP parameters were generated. Then, QM/EFP calculations were performed for a pair of BChl a molecules in which EFP BChl a is described with either ground state singlet or triplet parameters, and QM BChl a is excited to the singlet state (TDDFT PBE0/6-31G(d)).

Table 6.4 shows differences in solvatochromic shifts on the site energies of specific pigments when another pigment undergoes a change from the ground singlet to the triplet state. Thus, in this more elaborate model, the triplet BChl a is not simply excluded from the excitonic Hamiltonian but indeed affects the remaining 7×7 Hamiltonian by providing electrochromic shifts to the neighboring sites. Even though such triplet-induced shifts to the singlet site energies are modest, inclusion of this effect produces a drastic improvement to the modeled 55 μs component of the triplet kinetics. Figure. 6.10 shows comparison of the experimental 55 μs component and two theoretical models, with and without inclusion of the effect of the triplet-induced electrochromic site energy shifts.

The modeling shown in Figure. 6.10 assumes that 55 μs component is governed by a decay of the triplet state population into the ground state. To reproduce a broad band at ~810 nm observed in room-temperature measurements (but not in 77 K experiments, REF [158]), we estimated the population of the triplet states at different pigments using the Boltzmann distribution and triplet

site energies shown in Figure. 6.3. Assuming that the triplet energy transfer occurs sufficiently fast, the average fractional population P_i of pigment i is given by:

$$P_i = \frac{\sum_k e^{-E_k/k_B T} c_{ki}^2}{\sum_k e^{-E_k/k_B T}} \tag{6.7}$$

where c_{ki} is the eigenvector for excitonic state k with the energy E_k, k_B is the Boltzmann coefficient and T is temperature. Based on calculations of the T-T couplings, we hypothesized that pigment #1, #2, and #8 do not exchange triplets with the other pigments, such that only pigments #3, #4, #5, #6, and #7 become equilibrated.

It is noteworthy that inclusion of the triplet-induced electrochromic shifts provides a significant change to the shape of the absorption difference spectra and brings the modelled spectra in an excellent agreement with experiment.

To elaborate computational modeling of electrochromic shifts, we also calculated site energy shifts in the presence of the protein environment at the same 40 snapshots utilized in Section 6.3.2. Because only a few pairs show non-negligible electrochromic shifts (> 5 cm^{-1}) in gas phase, we limited our calculations to 14 pairs of BChl a pigments, as shown in Table 6.5. These Hamiltonians will be used for modeling of 55 μs and other components of the triplet kinetics.

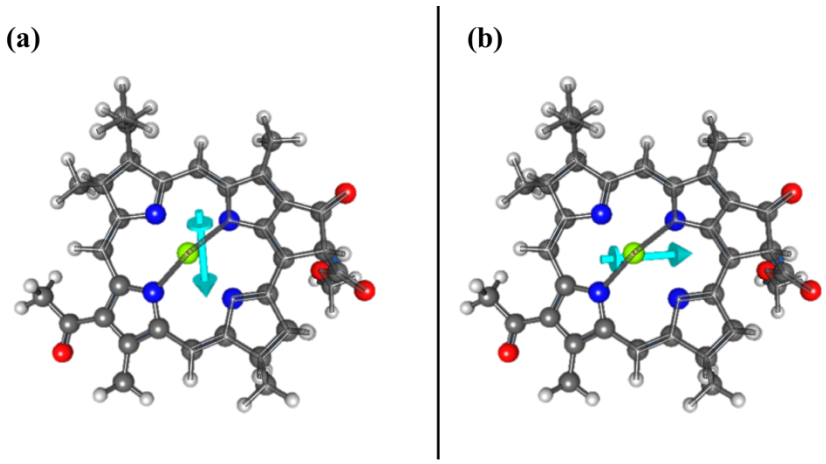

Figure 6.9. Static dipole moments of the (a) ground singlet and (b) excited triplet states of BChl a head group.

Table 6.4. Site energy shifts (cm^{-1}) due to the triplet BChl a. The numbers represent singlet site energy shifts of BChl a (rows) due to the triplet BChl a (columns). Red shifts are shown with red colors; blue shifts - with blue colors.

		Triplet BChl a							
		1	2	3	4	5	6	7	8
Singlet excited BChl a	1		+5.6						+2.2
	2	-36.6		+16.6	+2.6	+1.1	+3.9	+6.6	
	3		-3.8		-18.0	+0.5	-0.9	+2.6	
	4			-25.5		-27.4		+4.0	
	5				-9.3		+42.4	-2.4	
	6					+13.1		-6.3	
	7		+1.0	-3.6	-16.7	+17.7	+11.6		
	8	+2.9							

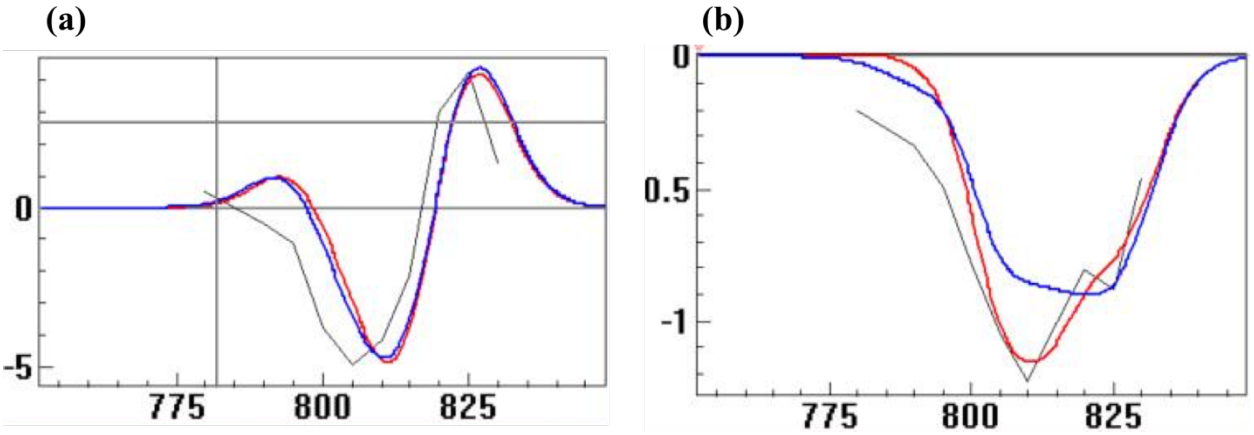

Figure 6.10. The measured (a) absorption and (b) CD signatures of 55 µs component of the triplet kinetics shown in black. Blue and red lines show modeling by assumption of the decay of the Bolzmann-weighted population of triplets on pigments #3, #4, #5, #6, and #7 into the ground state. Blue line is generated by assumption based on missing pigment as discussed in Section 6.2.1 (see Figure 6.1). Red line is produced by using the Hamiltonian that accounts for the triplet electrochromic shifts.

Table 6.5. QM/EFP average site energy shifts and uncertainties (cm^{-1}) due to the triplet BChl a over 40 snapshots of protein structures. The numbers represent singlet site energy shifts of BChl a (rows) due to the triplet BChl a (columns). Red shifts are shown with red colors; blue shifts - with blue colors

		Triplet BChl a						
		1	2	3	4	5	6	7
Singlet excited BChl a	1		**+14.7**±1.0					
	2	**-35.9**±1.3		**+13.8**±0.3				**+2.5**±0.2
	3				**-14.9**±0.5			
	4			**-29.3**±0.6		**-24.3**±0.8		
	5				**-8.2**±0.4		**+39.6**±0.7	
	6					**+12.3**±0.6		**-39.8**±2.1
	7				**-9.8**±0.3	**+14.0**±0.3	**+7.9**±0.3	

6.3 Conclusions

We performed modeling of the triplet states in the FMO protein and managed to explain the origin of the most intense component of the triplet kinetics observed experimentally. Future work involves more accurate calculations of triplet couplings and exploring the effects of electrochromic effects of the excited pigment in other types of pump-probe spectroscopy.

CHAPTER 7. QUANTUM MECHANICAL MODELING OF EXCITONIC STRUCTURE: THE FENNA-MATTHEWS-OLSON MUTANT COMPLEXES

7.1 Abstract

In Chapter 3 we conducted first-principles modeling of the wild type FMO pigment protein complex and demonstrated that the QM/EFP modeled absorption and circular dichroism (CD) spectra are in an excellent agreement with experiments. Building upon this success, we proceed with investigating excitonic interactions and spectroscopy of Fenna-Matthews-Olson (FMO) mutant complexes.

7.2 Introduction

The Fenna-Matthews-Olson (FMO) pigment-protein complex plays a crucial role in green sulfur bacteria as a channel that interfaces the chlorosome antenna to the reaction center. The fundamental function of FMO is a transfer of excitation energy from an antenna complex to the reaction center complex, where electron reactions utilize solar energy for chemical transformations. FMO is one of the actively studied photosynthetic genus because not only it possesses relatively simple structure [78-112], but also FMO is the first crystalized chlorophyll-containing protein structure [95,183]. The FMO structure is an assembly of eight bacteriochlorophyll a (BChl a), and three such moieties associate to form a trimer structure with C_3 symmetry (see Figure 3.1).

Close packing of BChl a pigments in each subunit leads to strong excitonic interactions and delocalization of excited states over multiple pigments. To describe the dynamics of excitonic interactions and energy transfer within the FMO protein, many of electronic Hamiltonians have been proposed [79,83,85,87,89,96,98,110,116]. However, most of them were obtained by a combination of structural information and fits to available empirical measures. While the off-diagonal elements of a Hamiltonian represent electronic couplings between pigments and can be estimated using dipole-dipole approximation from available X-ray structures, the diagonal energies (individual transition energies) cannot be observed directly in experiment and are inferred from fits to spectroscopic data. This approach results in a significant ambiguity in determination of BChl a site energies and typically does not account for environment-driven variations in

transition dipole moments of individual pigments and inter-pigment couplings. While molecular modeling based on structural data could provide unambiguous assignment of excitonic interactions and energy flow in FMO, internal complexity of the system and necessity to sample protein degrees of freedom and accurately describe electronic structure and couplings between BChl a pigments and vibronic interactions between pigments and the protein environment makes this task challenging [82,85-88,92,93,99,101,103,108,109,117-119]. In Chapter 3, we demonstrated that quantitative agreement of molecular modeling with experiments could be achieved by utilizing accurate structures of photosynthetic pigments and representing the protein environment with a polarizable model. In addition, we proposed relatively straightforward, yet rigorous computational procedures for the study of the FMO protein, showing excellent agreement in absorption and CD spectra with experiments. However, in all cases, experimental and molecular modeling based approaches are scoped to the wild-type FMO protein.

Recently, Blankenship and coworkers developed the FMO mutagenesis method and generated a series of FMO mutants, as shown in Figure 7.1 [102,184]. In each mutant, a single amino acid mutation occurs near a particular BChl a pigment, such that the mutants can be referred to by the site number. FMO mutants can provide excellent test sets to both theoreticians and experimentalists for developing and improving new techniques. In addition, studies of the FMO mutants can provide in-depth insight into strategies for engineering artificial photosynthetic devices by unraveling absorbance changes due to mutagenesis.

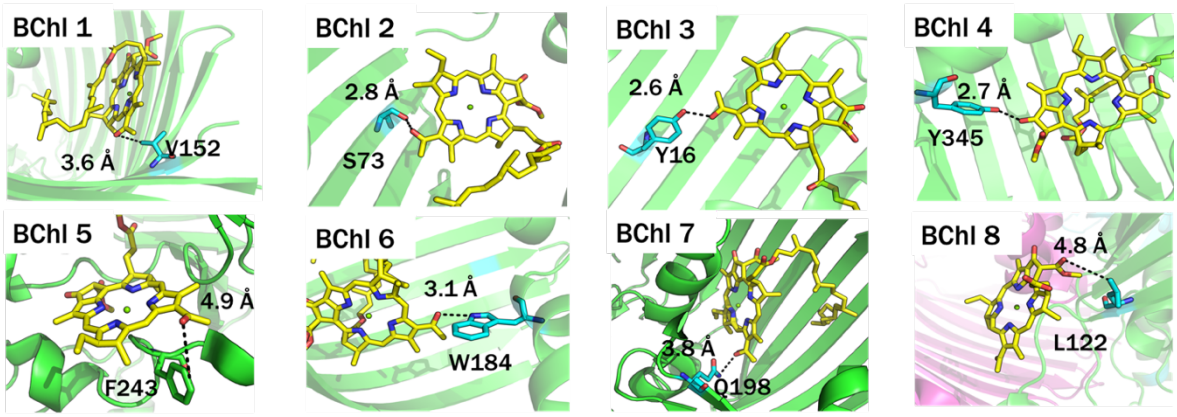

Figure 7.1. Representative snapshots of target amino acid residues for a single point mutagenesis. BChl 1: VAL to ASN (V152N), BChl 2: SER to ALA (S73A), BChl 3: TYR to PHE (Y16F), BChl 4: TYR to PHE (Y345F), BChl 5: PHE to TYR (F243Y), BChl 6: TRP to PHE (W184F), BChl 7: GLN to VAL (Q198V), BChl 8: LEU to GLN (L122Q). Figure is a courtesy of Dr. Rafael G. Saer and Dr. Valentyn Stadnytskyi.

7.3 Preliminary study

7.3.1 Motivation

Blankenship and coworkers analyzed a series of eight site-directed FMO mutants by a combination of absorption and circular dichroism (CD) spectroscopy [102,184]. All FMO mutants showed changes in intensities and shifts in absorption bands. Specifically, relatively large variations of absorbance features were demonstrated by mutagenesis of V152N, S73A, Y16F, Y345F, W184F (i.e., mutants BChl #1, #2, #3, #4, and #6). Even more detailed spectral changes due to the mutagenesis were shown by CD analysis. This is because the CD spectroscopy is more sensitive to the relative energy differences and spatial orientations of photosynthetic pigments than the absorption spectroscopy as was discussed in Sections 3.3.7 and 3.4.6. Specifically, mutants BChl a 2, 3, 4, and 6 displayed noticeable changes in their CD spectra [102]. Blankenship and coworkers also modeled absorption and CD spectra of mutants by modifying empirical Hamiltonians of wild-type FMO (Vuto et al [110], Brixner et al [79], Cho et al [185], and Kell et al [89]). One assumption utilized in this modeling is that the effect of each mutation is short-ranged, such that only the site energy of the closest pigment (i.e., site #1 in the mutant V152N near BChl a #1) is affected. Other site energies were assumed to be unaffected. Table 7.1, adapted from REF [102] shows approximate site energy shifts due to one-site mutagenesis, that were estimated by fitting model spectra to the experimental absorption and CD spectra. The main conclusion from

this modeling was that while it was possible to mimic changes in the absorption spectra by modifying a single element of excitonic wild-type FMO Hamiltonians, the modeled CD spectra did not agree well with the ones measured experimentally (see Figures 7.2 and 7.3). These findings hint to a more complex nature of changes in the structure and excitonic interactions due to mutagenesis. However, before exploring these more complex effects, we decided to conduct electronic structure modeling using the simple assumption that the structure of the protein is not changed upon mutations. Thus, this chapter consists of two parts: the following Section 7.3 describes modeling where only the closest to mutation region is allowed to relax while the majority of the protein is kept fixed, and Section 7.4 in which possible structural changes due to mutation are also accounted for.

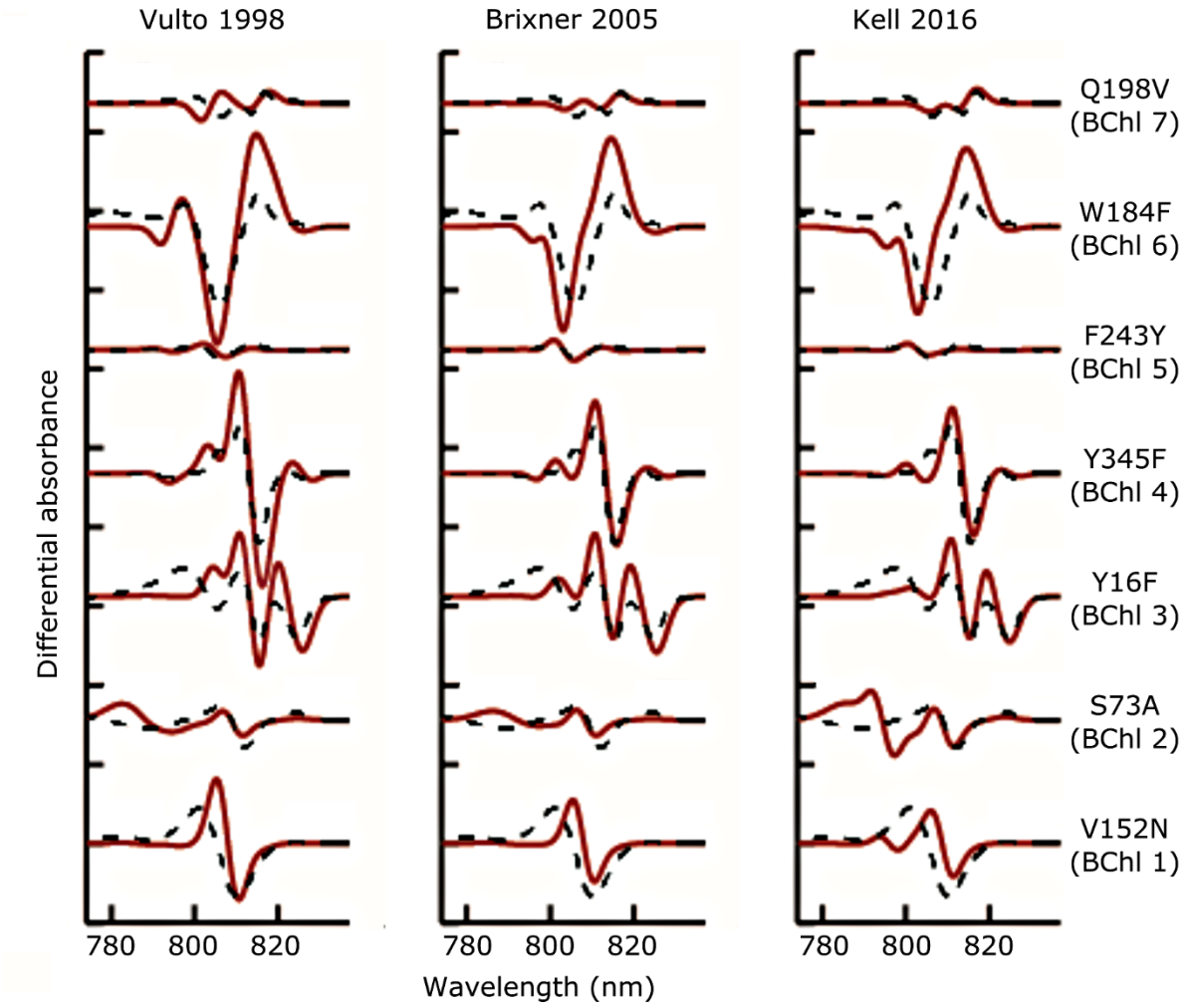

Figure 7.2. Simulated mutant minus wild-type absorption difference spectra calculated with the published Hamiltonians. Simulated difference spectra are shown as red lines, whereas the experimental difference spectra are shown as black dashed lines. This figure is taken from the REF [102].

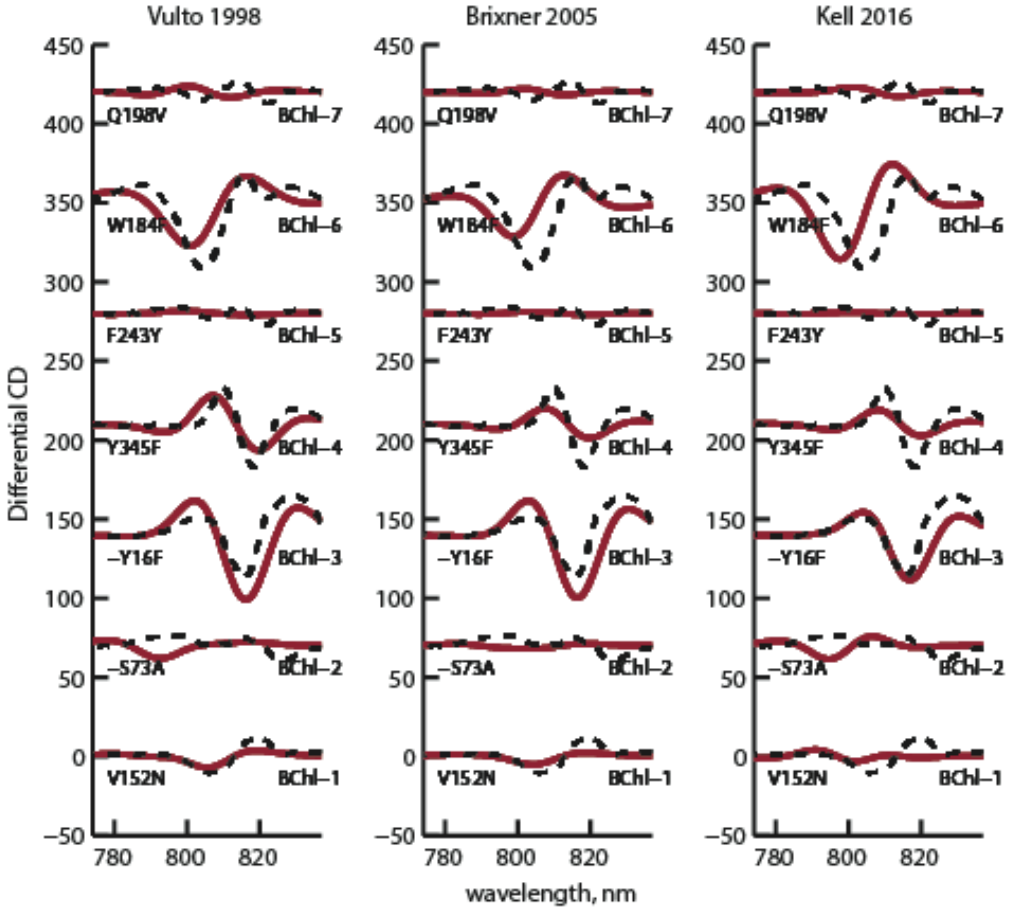

Figure 7.3. Simulated mutant minus wild-type CD difference spectra calculated with the published Hamiltonians. Simulated difference spectra are shown as red lines, whereas the experimental difference spectra are shown as black dashed lines. This figure is taken from the REF [102].

Table 7.1. The wild-type site energy and the best-fit (mutant) site energies for each of the published Hamiltonians. Differences (Diff.) in site energies between wild-type and mutants are shown in units of wavelength and wave numbers. This table is taken from the REF [102].

		Site energy						
		BChl 1 V152N	BChl 2 S73A	BChl 3 Y16F	BChl 4 Y345F	BChl 5 F243Y	BChl 6 W184F	BChl 7 Q198V
Vulto et al.	Wild type	806.5	793.7	823.7	814.3	800	800	804.5
	Mutant	802.6	787.4	811	806.6	799.5	815	806.5
	Diff., nm	+3.9	+6.3	+12.7	+7.7	+0.5	-15	-2
	Diff., cm^{-1}	+60	+101	+190	+117	+8	-230	-31

(Table 7.1 continued)

		Site energy						
		BChl 1 V152N	BChl 2 S73A	BChl 3 Y16F	BChl 4 Y345F	BChl 5 F243Y	BChl 6 W184F	BChl 7 Q198V
Cho et al.	Wild type	805.2	796.2	823.7	812	802.6	800	806.5
	Mutant	802.6	792.5	812.2	808.1	802	815	807.5
	Diff., nm	+2.6	+3.7	+11.5	+3.9	+0.6	-15	-1
	Diff., cm^{-1}	+40	+59	+172	+59	+9	-230	-15
Kell et al.	Wild type	806.1	799.7	823	813	801.9	795.2	808.1
	Mutant	803.2	789	813	809.2	801.6	815	809.7
	Diff., nm	+2.9	+10.7	+10	+3.8	+0.3	-19.8	-1.6
	Diff., cm^{-1}	+45	+170	+149	+58	+5	-306	-24

7.3.2 System preparation and QM/MM geometry optimizations

Here we consider three mutants, Y16F, Y345F, and Q198V (i.e., BChl 3, 4, and 7). Mutations at sites BChl 3 and 4 produce significant changes in both absorption and CD. The observed spectral changes could be reasonably reproduced by shifting a single component of the empirical excitonic Hamiltonians (see Table 7.1 and Figures 7.2 and 7.3). However, while spectral changes due to mutation Q198V at BChl 7 produce less drastic effects on the excitonic spectra, those could not be modeled even qualitatively with this simple approach. Thus, considering these three mutations provide a stringent test to modeling strategies.

Starting from the optimized wild-type structures from Chapter 3 (i.e., QM/MM optimizations that did not include phytyl tails), single-point mutations were conducted using our own Python script (see Figures 7.1 and 7.4). Then, 25 constrained QM/MM geometry optimizations were performed for mutated sites of each of three considered mutants. Each QM region contained a single BChl a with the truncated phytyl tail and corresponding neighboring amino acids, as shown in Figure 3.3. It should be noted that the neighboring AAs considered in the QM/MM optimizations in the wild-type study included target residues for the mutagenesis. The QM/MM geometry optimizations of the mutated sites were conducted as described in Section 3.3.2.

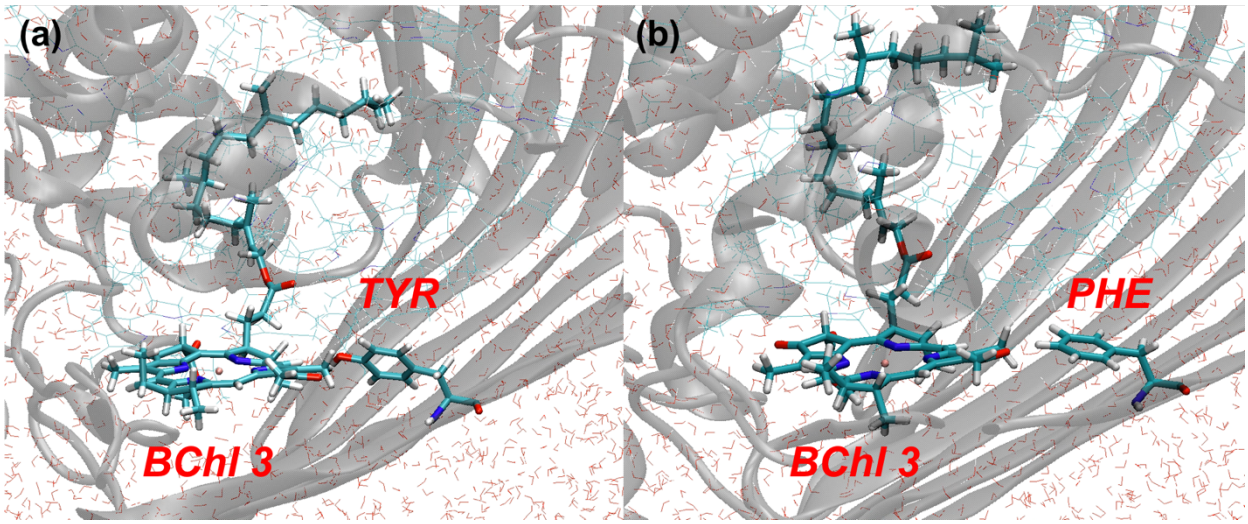

Figure 7.4. Single-point mutagenesis is shown on the example of Y16F (BChl 3). Tyrosine residue H-bonded to BChl 3 (a) is mutated to phenylalanine AA (b).

7.3.3 QM/EFP excited state calculations

In REF [102], only mutated sites were assumed to be affected by mutagenesis. To verify the hypothesis, we computed site energies of the BChl a pigments located within 15 Å of the mutated sites. Other (further separated) sites were assumed to be unaffected; values of their site energies were taken from Chapter 3. Thus, for mutant BChl3-Y16F, we recomputed site energies of pigments #3, #4, and #7; for mutant BChl4-Y345F, sites #3, #4; and for mutant BChl7-Q198V, sites #4, #5, #6, and #7.

The QM/EFP calculations in mutated proteins were conducted analogously to those undertaken in Chapter 3. Namely, regions in the excited state computations are the same as described in Chapter 3 and shown in Figure 3.6 and contain a single BChl a and the corresponding Mg-coordinating residue. Remaining BChl a pigments, AAs, and solvent molecules within 15 Å from the QM BChl a head were represented as effective fragments. Outside the 15 Å shell, all atoms were described by the MM point charges with the AMBER03 force fields (see Figure 3.6). To make consistency with the wild-type study, TDDFT PBE0/6-31G(d) was used for computing excitation energies. Electronic couplings were computed using the point dipole approximation (PDA) (see Eqn. (3.2)) and were scaled by 0.6 to approximately match the value of the experimental TDM of BChl a in gas phase [145,146].

Absorption and CD spectra of mutants and difference ΔA and ΔCD spectra between a mutant and wild-type protein were modeled as an average of 25 excitonic spectra produced at individual snapshots.

All excited state computations were carried out in the GAMESS electronic structure package [140].

7.3.4 Results

As shown in Figure 7.5, single point mutagenesis produces significant energy shifts in mutants Y16F and Y345F. The energies of site #3 in BChl3-Y16F, site #4 in BChl4-Y345F, and site #7 in BChl7-Q198V are blue-shifted by 142 cm^{-1}, 67.1 cm^{-1}, and 17.7 cm^{-1}, respectively, which is comparable to the values estimated in REF [102] (see Table 7.1). Even though the site energy of site #4 in Y16F is also blue-shifted by 22.8 cm^{-1}, the magnitude is much smaller than that of site #3. Mutation Y345F near site #4 does not affect the energies of other sites. It should be noted, however, that the estimated shifts in REF [102] were not able to reproduce CD spectral changes even qualitatively. Interestingly, our modeling suggests that a closely lying BChl a #6 shows a comparable shift (-21.8 cm^{-1}) to the mutated #7 site.

Figure 7.6 shows the experimentally measured and QM/EFP modeled absorption and CD difference spectra. As in the wild-type study (Chapter 3), the site energies are shifted by -2430 cm^{-1}. As we expected from comparing site energies, the mutants of the lowest sites (Y16F and Y345F) qualitatively (but not quantitatively) capture positive and negative spectral differences in both absorption and CD spectra. In case of Q198V, the absorption and CD spectra of the mutant are similar to those of the wild-type protein, resulting in small intensities in absorption and CD difference spectra. QM/EFP results are in a reasonable agreement for the CD difference spectrum (Figure 7.6 (c) lower panel), but fail to reproduce the experimentally measured absorption changes (Figure 7.6 (c) upper panel), implying that the conducted modeling is not sufficient for explaining the properties of the Q198V mutation. Interestingly, Kell's electronic Hamiltonian qualitatively reproduced the absorption difference spectrum, but failed to mimic the CD difference spectrum (see Figures. 7.2 and 7.3) [102].

In order to understand the failure of theoretical model to reproduce experimental difference spectra of mutant BChl7-Q198V, we conducted analysis of structural changes created by this

mutation. Specifically, we performed molecular dynamics simulations explicitly for Q198V, and compared MD trajectories with the ones obtained from the wild-type MD simulations. As shown in Figure 7.7 (a), the wild-type BChl #7 forms several stable H-bonds with a nearby water molecule, glutamine (mutation site), and histidine (which is not coordinated to Mg of BChl a) residues in the MD simulations. However, MD simulations of Q198V reveal a completely broken H-bonding network between BChl #7 and nearby residues (see Figure 7.7 (b)). Further, the orientation of a nearby histidine residue (HID) dramatically changes because there is no stabilizing H-bond formed between the histidine and glutamine residues after mutating glutamine to valine. Thus, the example of Q198V mutant suggests that explicit structural changes of the whole protein need to be accounted for in modeling at least some of the mutants. The present preliminary study is done with local modifications of the wild-type structures, such that we do not capture the detailed structural discrepancies between wild-type and mutants accurately, resulting in a non-sufficient site energy shifts.

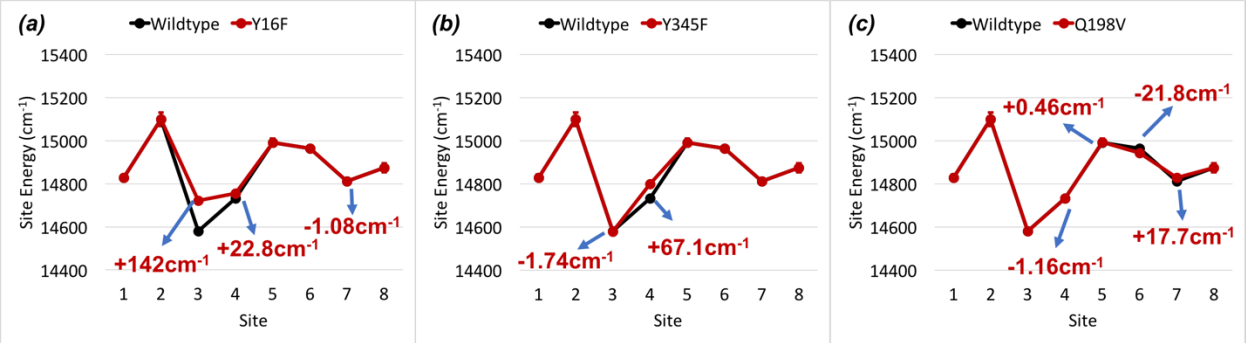

Figure 7.5. BChl a site energies averaged over 25 structures with standard deviations shown as vertical error bars. Site energies of mutants are shown with red lines, site energies of wild-type FMO are shown with black lines. (a) BChl3-Y16F, (b) BChl4-Y345F, and (c) BChl7-Q198V.

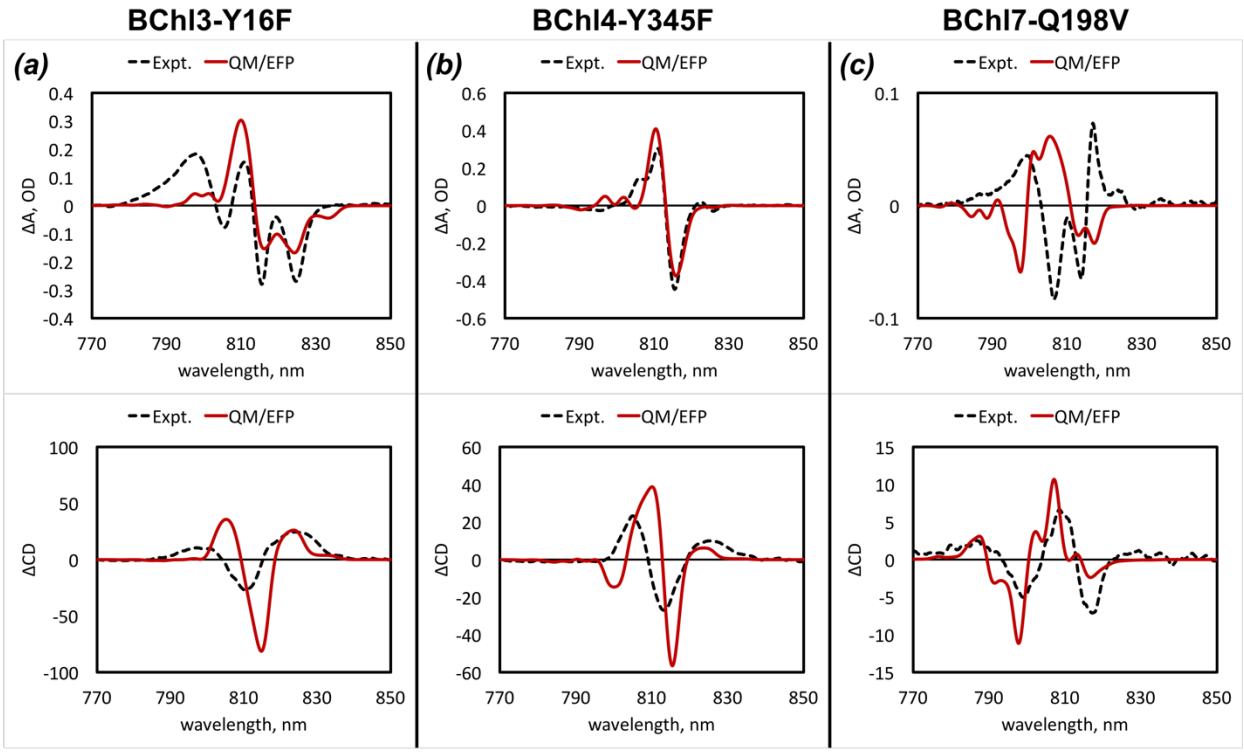

Figure 7.6. The absorption (upper panel) and CD (lower panel) difference spectra of (a) Y16F, (b) Y345F, and (c) Q198V mutant experimentally measured [102] (black dashed line) and computed with QM/EFP method (red stick line).

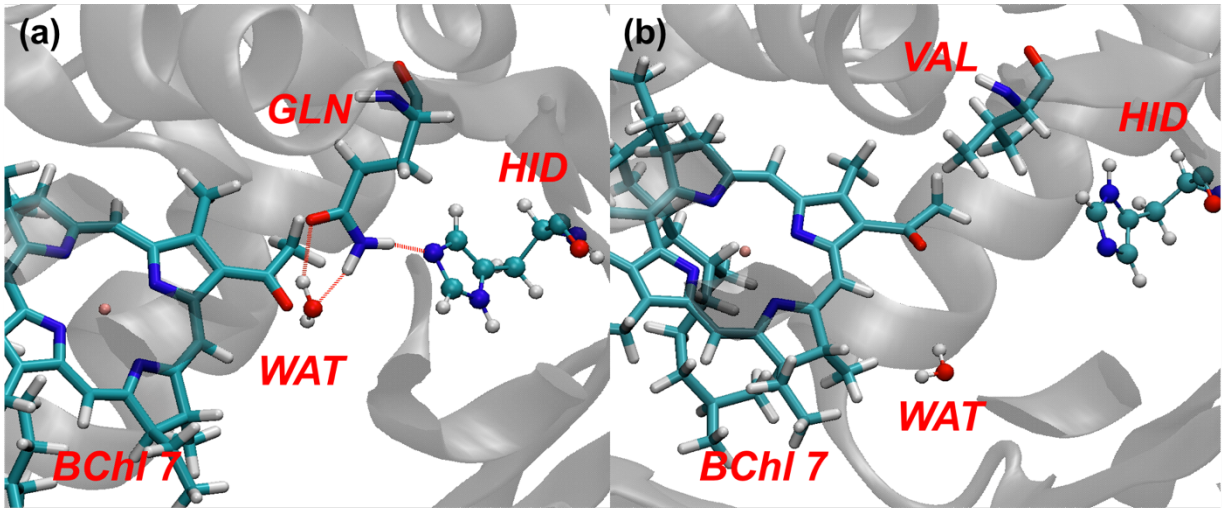

Figure 7.7. Single snapshot from the molecular dynamics simulation trajectories of (a) the wild-type FMO and (b) Q198V BChl7 mutant.

7.3.5 Intermediate conclusion

In this preliminary study, we have investigated effect of mutagenesis on three different sites in the FMO protein. Started from the optimized wild-type structures in Chapter 3, we generated a set of FMO mutants (BChl3-Y16F, BChl4-Y345F, and BChl7-Q198V). Computed spectra of Y16F and Y345F mutants agree well with experimentally measured spectra, while the absorption difference spectra of mutant Q198V are not reproduced by our modeling. As shown in structural analysis of Q198V, these preliminary data emphasize the need for explicit molecular dynamics simulations that can describe structural changes due to mutagenesis. In addition, as the modeled spectra are quite noisy, we conclude that a larger number of structures should be sampled to produce statistically reliable results.

7.4 Revisit FMO mutant complexes

7.4.1 Preparation of the system and molecular dynamics simulations

Here we explicitly perform molecular dynamics simulations to generate the FMO configurational ensembles due to mutagenesis. The Fenna-Matthews-Olson (FMO) mutant complexes in this study are based on the crystal structure of *Chlorobaculum Tepidum* obtained from protein data bank (PDB: 3ENI [109]). Based on the wild-type FMO crystal monomer unit, single-point mutations were conducted using our own Python script (see Figures 7.1 and 7.4). The initial trimeric structures of mutants were constructed using C_3 symmetric transformation. Using ANTICHAMBER module of AMBER package [120], these structures were automatically protonated with neutral pH conditions. The system preparation and equilibration procedures are exactly same as in case of the wild-type protein (see Section 3.3.1). Total 80 ns production run was carried out with 2 fs time step for each mutant system. All classical molecular dynamics simulations were performed with the GROMACS package (version 2016.5) [132]. From the last 30 ns of the MD trajectory, 100 atomic configurations were extracted randomly. For each configuration, the water molecules outside 15 Å from the outer protein surface were truncated.

7.4.2 Constrained QM/MM geometry optimizations

Started from selected structures (100 each), we performed QM/MM constrained geometry optimization in which MM regions were kept fixed. As in wild-type studies, electrostatic

embedding QM/MM scheme was employed [133]. Each QM region contained a complete BChl a molecule including the phytyl tail and few neighboring AAs, as described in Chapter 4. It again should be mentioned that the neighboring AAs included in the QM/MM optimizations of the wild-type FMO already contain AA residues targeted by the mutagenesis. The detailed procedures of the QM/MM geometry optimizations, such as truncation schemes for AAs, level of theory, and gradient tolerance etc. are available in Section 4.3.1.

All geometry optimizations were performed using our own QM/MM interface (developed by Dr. Dmitry Morozov, University of Jyvaskyla) between TERACHEM [154-156] GPU-accelerated electronic structure package and GROMACS 4.6.5 molecular dynamics package [132].

7.4.3 EFP parameters

The number of AA EFP parameters for the FMO mutants is approximately 424,000 (eight mutant structures × 100 atomic configurations each × ~530 amino acid parameters per atomic configuration). The parameter computation time for the smallest AA (glycine) takes about 0.01 CPU hours. This implies that just parameter preparation step takes about 4240.0 CPU hours if the FMO protein consists of only glycine residues. To circumvent the large number of parameter computations, we employed the Flexible EFP scheme described in Chapter 2. The RMSD criterion to select EFP parameters from the amino acid database was 0.2 Å resulting in only 30~40 new parameter calculations per atomic configuration (approximately 7 %). In this study, the AA database was built upon the wild-type FMO EFP parameters. Details of the Flexible EFP procedure and employed levels of theory can be found in Chapter 2 and Section 3.3.3.

7.4.4 Covalent boundaries in excited state QM/MM and QM/EFP calculations

Covalent boundaries between QM and classical regions in QM/MM and QM/EFP calculations were treated as in the wild-type FMO study, as described in detail in Chapter 3.

7.4.5 Excited state QM/EFP calculations

Polarizable embedding QM/EFP scheme for electronic excited states was used in this work, with technical details described in Chapters 3 and 4. Transition charges, TrESP were utilized for

computing electronic couplings between BChl a pigments (See Eqn (3.1)) and were scaled by 0.6 to approximately match the value of the experimental TDM of BChl a in gas phase [145,146].

All excited state computations were carried out in the GAMESS electronic structure package [140].

7.5 Results

In the computational details, we described computational procedure for all eight FMO mutants. However, we are only focusing on BChl 3 (Y16F) system in the present study. The same computational formalism can be applied to other mutants.

7.5.1 Structural changes in mutant BChl3-Y16F

In order to investigate structural differences between wildtype and Y16F mutant, we computed RMSD between 100 atomic configurations from wildtype and mutant MD trajectories. RMSD plots comparing only wildtype snapshots, only mutant snapshots, and those comparing wildtype and mutant structures, are shown in Figure 7.8. All heavy atoms of the FMO protein were included in RMSD computations. For the RMSD calculations between wildtype and mutant configurations, mutated AA was excluded. As shown in Figure 7.8, wildtype and mutant species explore significantly different configurational ensemble during MD simulations, showing near 2.5 Å RMSD between them. However, each configurational ensemble (i.e, wildtype and mutant itself), demonstrate less than 1.5 Å RMSD ranges.

In order to analyze structural changes due to mutagenesis, we computed the closest atomic distances between BChl a head groups (see Figure 6.3) and neighboring AAs (heavy atoms) for all sites in the FMO protein. Figure 7.9 provides such averaged atomic distances in the mutant complexes with respect to the wildtype. Table 7.2 summarizes AAs which show more than 1.0 Å differences. Drastic changes in atomic configurations occur near sites #2 and #8. Specifically, the distance between the BChl a #2 head and the water molecule that is coordinated to Mg atom in the wild type increases dramatically in the mutant, implying that mutagenesis at site #3 leads to closing a water channel at site #2. Hence, water dynamics might be significantly different in mutant species.

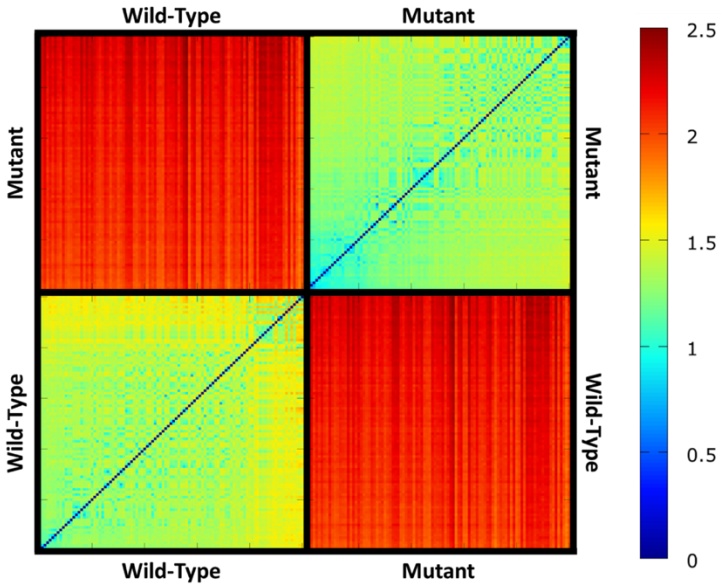

Figure 7.6. RMSDs (Å) between wildtype and mutant Y16F configurational ensembles. RMSDs are computed between 100 atomic configurations of wildtype and mutant. Color scale goes from blue (no structural differences) to red (large differences).

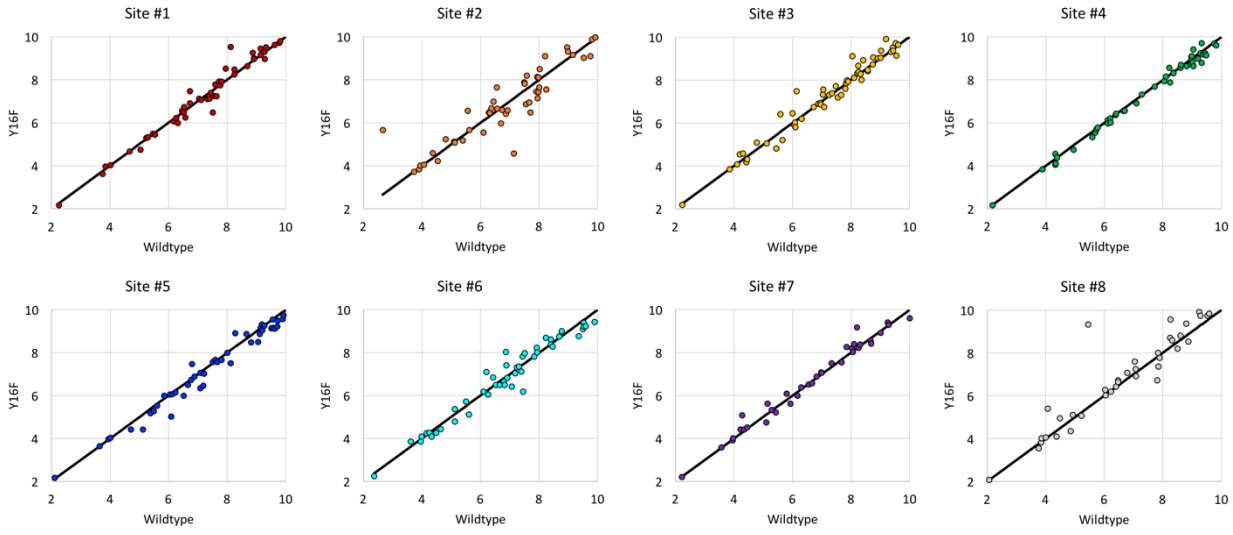

Figure 7.7. Changes in atomic distances between BChl a heads and neighboring AAs due to single point mutagenesis. Black lines indicate distances between BChl a head and AAs in wildtype configurations. Each colored circle represents atomic distance in Y16F mutant. Distances were computed based on selected 100 configurations and averaged.

Table 7.2. Average atomic distances and their RMS deviations between BChl a heads and AAs within 10 Å that show more than 1.0 Å deviations between the wild type and Y16F mutant species.

Site	Amino acid	Wildtype, Å	Y16F, Å	Diff, Å	RMS, Å
1	PHE	8.1	9.5	+1.4	0.05
	SER	7.5	6.5	-1.0	0.07
2	CYS	6.5	7.7	+1.2	0.07
	ILE	5.6	6.6	+1.1	0.10
	GLY	9.3	10.6	+1.3	0.13
	ASN	7.7	6.5	-1.2	0.11
	PRO	7.1	4.6	-2.5	0.10
	WAT	2.7	5.7	+3.0	0.12
3	THR	8.3	10.1	+1.8	0.09
	PHE	6.1	7.5	+1.4	0.03
	VAL	8.0	9.1	+1.1	0.10
4	GLY	9.5	10.8	+1.3	0.08
5	PHE	6.1	5.0	-1.1	0.05
6	MET	6.9	8.0	+1.1	0.05
	GLN	7.4	6.2	-1.2	0.07
7	-	-	-	-	-
8	THR	5.4	9.3	+3.9	0.10
	GLY	9.1	12.1	+3.0	0.07
	ALA	8.1	10.0	+1.9	0.06
	ASN	8.3	9.6	+1.3	0.05
	ASN	9.9	11.0	+1.1	0.05
	PHE	4.1	5.4	+1.3	0.05
	ARG	7.8	6.7	-1.1	0.05

7.5.2 Site energies in Y16F mutant

As shown in Figure 7.10 and Table 7.3, the two species, the wildtype and the Y16F mutant, produce similar average site energies only for sites #1, and #5, which is opposed to the assumption undertaken in REF [102]. Thus, our data imply that a single point mutagenesis can affect most of pigments in the FMO protein. The most drastic changes due to mutation are observed in sites #3 and #6. It is not surprising to observe a large excitation energy shift in site #3 because this site loses H-bond with tyrosine due to mutagenesis (TYR to PHE), resulting in blue shift of the site energy. However, the shift observed in the present study exceeds the values reported in REF [102] and our "frozen structure" analysis from Section 7.3 by additional 80~130 cm^{-1}. Interestingly, mutagenesis near site #3 strongly influences the excitation energy of site #6. Detailed analysis of the origin of these shifts is underway.

Figure 7.11 compares site energy distributions of the Y16F mutant and the wildtype. The widths of the energy distributions are comparable for most sites. Site #2 shows slightly larger spread of site energies in the mutant, possibly due to a missing coordination of Mg atom of BChl a head in the mutant. Excitation energies of sites #3 and #6 are shifted to blue and red, respectively, showing similar energy distributions. Hence, it can be concluded that the site energy shifts are not due to the statistical noise.

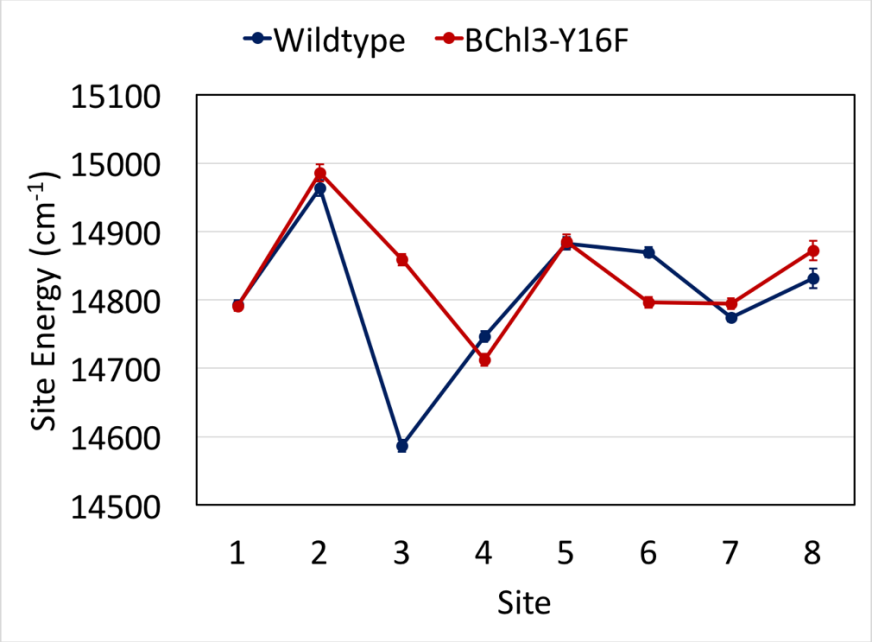

Figure 7.8. Average site energies with uncertainties shown as vertical error bars. Wildtype data (in navy color) are from Chapter 4; Y16F mutant data are shown in red color.

Table 7.3. Average site energies and uncertainties of the average values of wildtype (Chapter 4) and Y16F mutant and the differences in average energies of the two models (ΔE). All values are in cm^{-1}.

Site	**Wildtype**	**Y16F**	ΔE
1	14792.7 ± 6.9	14790.4 ± 6.5	-2.3 ± 9.5
2	14963.6 ± 11.2	14985.5 ± 12.5	+21.9 ± 16.8
3	14586.5 ± 8.6	14859.0 ± 8.3	+272.5 ± 11.9
4	14746.3 ± 8.1	14712.1 ± 8.9	-34.2 ± 12.0
5	14882.5 ± 9.0	14885.2 ± 10.6	+2.7 ± 13.9
6	14869.6 ± 7.4	14706.5 ± 7.7	-73.1 ± 10.6
7	14774.1 ± 5.7	14794.4 ± 7.9	+20.3 ± 9.7
8	14831.5 ± 14.2	14871.9 ± 14.4	+40.4 ± 20.2

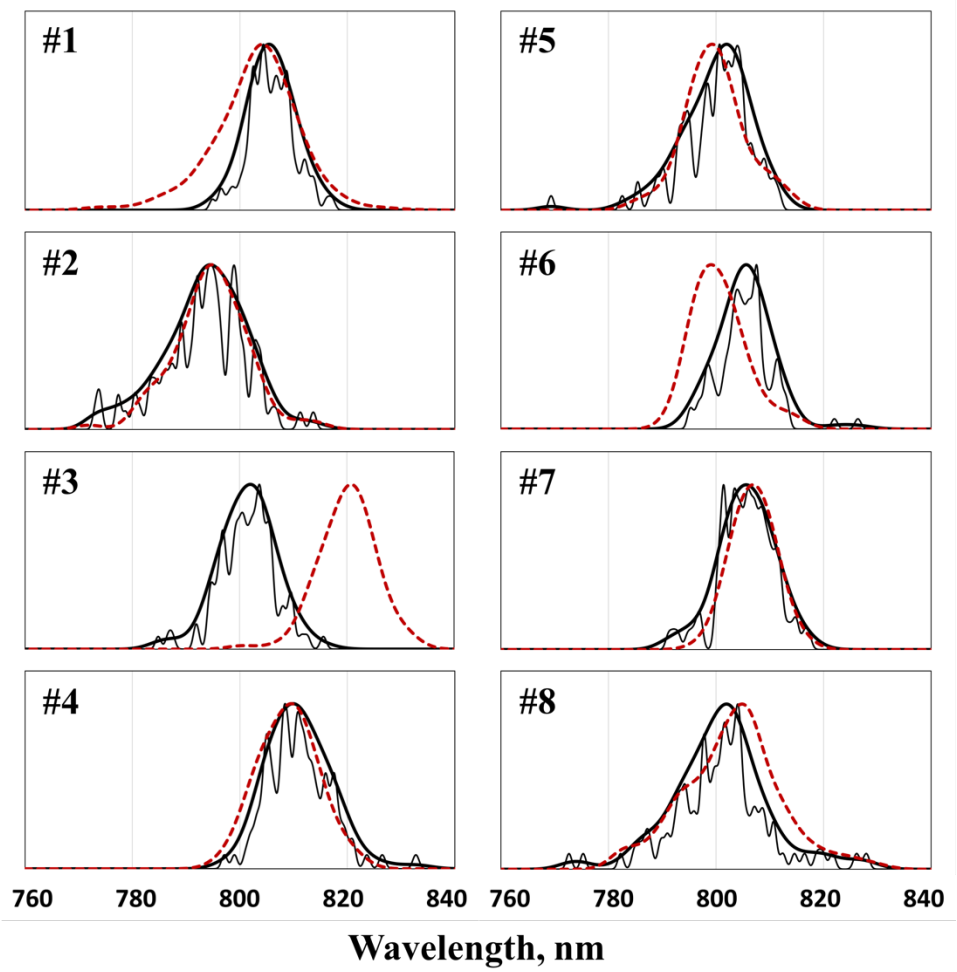

Figure 7.9. Site energy fluctuations. Black lines show the Y16F results. Stick spectra of 100 snapshots of MD trajectory are broadened by Gaussians with FWHM of 20 cm^{-1} (thinner black lines) and 100 cm^{-1} (thicker black lines). FWHM 100 cm^{-1} broadened spectra from wildtype (Chapter 4) are shown with red dashed lines.

7.5.3 Absorption and Circular Dichroism (CD) spectra of Y16F mutant

The excitonic Hamiltonians are constructed to model absorption and CD spectra of the Y16F mutant. Here, we only consider the absorption and CD spectra computed as averages of spectra of individual snapshots, i.e., thermal motion of the protein is directly accounted for. As discussed in Section 3.4.6, BChl #8 is only weakly bound in the FMO protein, and occupancy of site #8 depends on a protein preparation procedure. Indeed, the occupancy of site #8 in the mutant is not known and needs to be investigated separately. However, for consistency, we assumed the same occupancy of site #8 in the mutant as in the wildtype and generated the spectra by mixing 45% of

the 7-site Hamiltonians and 55% of the 8-site Hamiltonians as in Chapter 3. For an easier comparison with experimental spectra, all site energies in the wild type and mutant are shifted by -2380 cm^{-1}. Computed wildtype and Y16F mutant spectra are shown in Figure 7.12. A single point mutagenesis strongly affects the lowest energy range, resulting in missing red-side shoulder in the absorption spectrum. In wildtype complex, the red-side shoulder is due to the lowest exciton mainly shared by sites #3 and #4. However, mutagenesis induces significant blue shift to the site #3 and thus shifts the absorption intensity toward blue. The blue shift of site #3 also induces significant changes in CD. Most noticeably, in the lower energy region, a negative CD signal dominated by the lowest exciton located on site #3 in the wildtype disappears.

We also modelled absorption and CD differences between the mutant and wildtype species, shown in Figure 7.13. It is fascinating that the computed absorption and CD difference spectra show good qualitative agreement with the experimental spectra. Even though there are minor discrepancies in relative intensities, QM/EFP modelling captures main features of both absorption and CD changes due to mutagenesis.

Overall, we conclude that the Y16F mutation produces significant and non-local structural changes that are reflected in site energy shifts in most sites in the FMO complex. The presented modelling is capable of capturing critical changes in optical properties of the FMO complex upon mutagenesis.

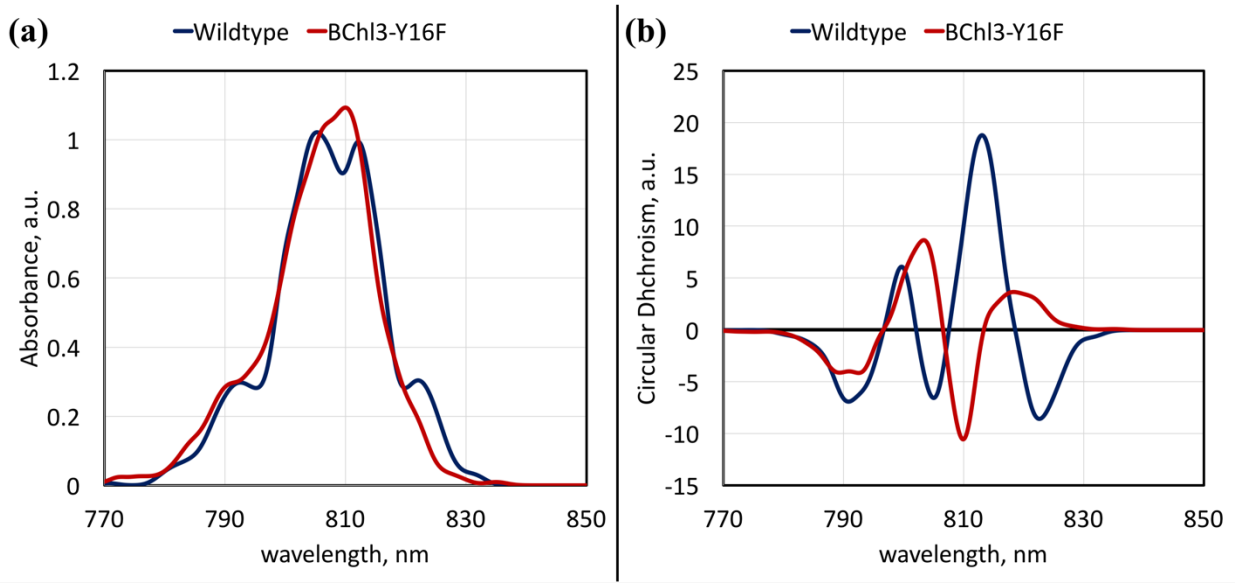

Figure 7.10. Computed absorption and CD spectra of wildtype FMO complex (navy line) and the Y16F mutant (red line). Both wildtype and mutant spectra are shifted by -2380 cm^{-1}.

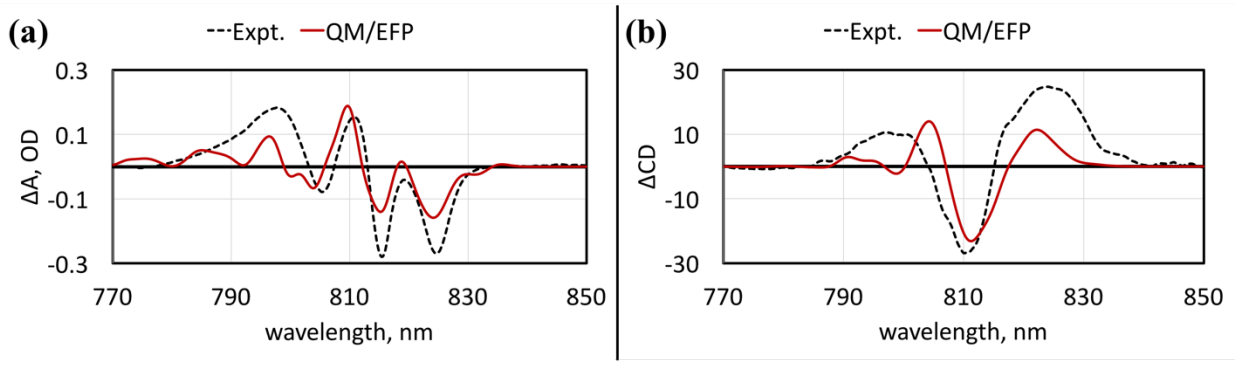

Figure 7.11. (a) Absorption and (b) CD difference spectra. Experimental absorption (77 K) and CD (Room-temperature) difference spectra are shown with black dashed lines [102]; QM/EFP modelled spectra are shown with red solid lines.

7.6 Conclusions

In this work, we have investigated single-point mutations of the FMO protein. The same computational protocol, procedures, and level of theories used in the wild-type study were straightforwardly applied to the FMO mutant genus. We noticed that a single point mutagenesis can introduce significant non-local changes to protein structure and affect optical properties of several pigments. The computed absorption and circular dichroism (CD) difference spectra of the Y16F mutant are in a qualitative agreement with experimental data. Based on the success of

modeling the Y16F mutant system, we are working on Q198V mutant, whose difference spectra cannot be explained by a simple “frozen structure” model or by shifting a single site energy of empirical Hamiltonians.

Analysis of the effects of mutagenesis in the FMO complex can be further strengthened by decomposing electronic properties of BChl a pigments by contributions of individual AAs. This work will be conducted in future studies.

CHAPTER 8. DIMER

8.1 Abstract

In Chapters 3-7 of this thesis it has been demonstrated that polarizable embedding is critical for predicting accurate electronic properties of photosynthetic pigments in biological environments. However, the detailed characterization of the effects of individual amino acids on the electronic excited states of BChl a pigments was missing. Understanding of specific interactions between BChl a pigments and neighboring amino acids is essential for engineering artificial systems with desired optical properties. In this study, we suggest computational schemes for evaluating contributions of individual amino acids on excitation energies of BChl a pigments, by using a mixed quantum-classical approach.

8.2 Introduction

As discussed in the previous chapters, the EFP method decomposes non-covalent interactions into Coulomb (electrostatic), polarization, dispersion, and exchange-repulsion terms. Specifically, the polarization term is computed in a self-consistent manner because the induced dipoles on a fragment are determined by the electrostatic field due to static multipoles as well as induced dipoles on the other fragments (see Figure 1.1). The QM/EFP models also treat polarization interactions self-consistently between the QM and EFP regions. In this regard, EFP inherently provides insights into many-body interactions of the system. Recently, decomposition of the EFP energy of a many-fragment system into pairwise fragment-fragment interactions, so-called pairwise EFP, has been implemented (see Section 2.3.2). However, the pairwise decomposition of the interaction energy is so far limited to the system containing only EFP fragments, i.e., to the cases when the *ab initio* region is not present. In this chapter, we utilize the pairwise EFP scheme to elucidate the individual AA contributions to the solvatochromic shifts of BChl a pigments in the FMO protein. Because the pairwise EFP scheme is not available for QM/EFP models, we reach this goal by representing the ground and excited states of the BChl a pigments with EFP fragments, i.e., by utilizing the fully classical approach. By numerically validating this approach, we suggest new computational strategy that can reproduce optical and electronic properties of pigments by using computationally cheaper classical description. Future

applications of this strategy will span investigations of exciton dynamics in FMO proteins including mutants as well as other more complex photosynthetic systems such as PS I and PS II.

8.3 Computational details

8.3.1 Theory

All the details of conventional and pairwise EFP are described in Chapters 1 and Section 2.3.2. Electrostatics, dispersion, and exchange-repulsion interactions are two-body, so the corresponding system energies are sums of all pairwise fragment-fragment interactions. Many-body polarization energy is decomposed into fragment-fragment interactions in the following way: (i) induced dipoles are self-consistently converged for the whole system; (ii) using converged dipoles, polarization energies for each pair of fragments are computed (see Eq (2.1)). The total polarization energy is a sum of all pairwise energies, but each dimer energy implicitly incorporates many-body effects through the induced dipoles being self-consistently converged for the whole system. Therefore, the pairwise scheme can conveniently decompose the total non-covalent interaction energy into contributions of individual fragment pairs, or extract interactions of a particular fragment with all other fragments of the system. Because our goal is to mimic the QM/EFP scheme, only electrostatic and polarization interactions between fragments (BChl a pigments and AAs) are considered in this study.

8.3.2 System preparation.

Differently from Chapters 3 and 4, here we describe all BChl a pigments as EFP fragments, such that the entire system is described fully classically. We refer to an EFP fragment that represents a QM region (i.e., BChl a pigment in either its ground or excited electronic state) as a "QM-like EFP fragment". Parameters of the QM-like EFP fragments are computed in the following way: (i) Stone distributed multipole analysis (DMA) [180-182] is performed on the entire QM-like EFP fragment at the PBE0/6-31G(d) level of theory. Two separate DMAs are conducted for each BChl a, namely, one to represent its ground state properties and obtained from the ground state PBE0 calculations, and another one to represent BChl a in the singlet excited state, obtained from the electronic density from TDDFT/PBE0. (ii) Polarization parameters (LMO centroids and polarizability tensors) of BChl a are obtained from a MAKEFP computation at the HF/6-31G(d)

level of theory. For polarizability computations, the BChl a molecule is separated into the head and tail groups, which were separately optimized with HF/6-31G(d). Computed distributed polarizabilities of the optimized head and tail groups were adjusted to MD structures via Flexible EFP scheme (see details in Chapter 2). Note that in the present model we ignore changes of BChl a polarizability upon electronic excitation. Combining electrostatic multipoles (DMA analysis) and polarizability tensors (MAKEFP calculations), the parameters of QM-like EFP fragments were generated. The parameter computation scheme is shown in Figure 8.1. The parameters were computed in the GAMESS [140] quantum chemistry package. The parameters for surrounding solvent and AAs are taken from the Chapters 3 and 4.

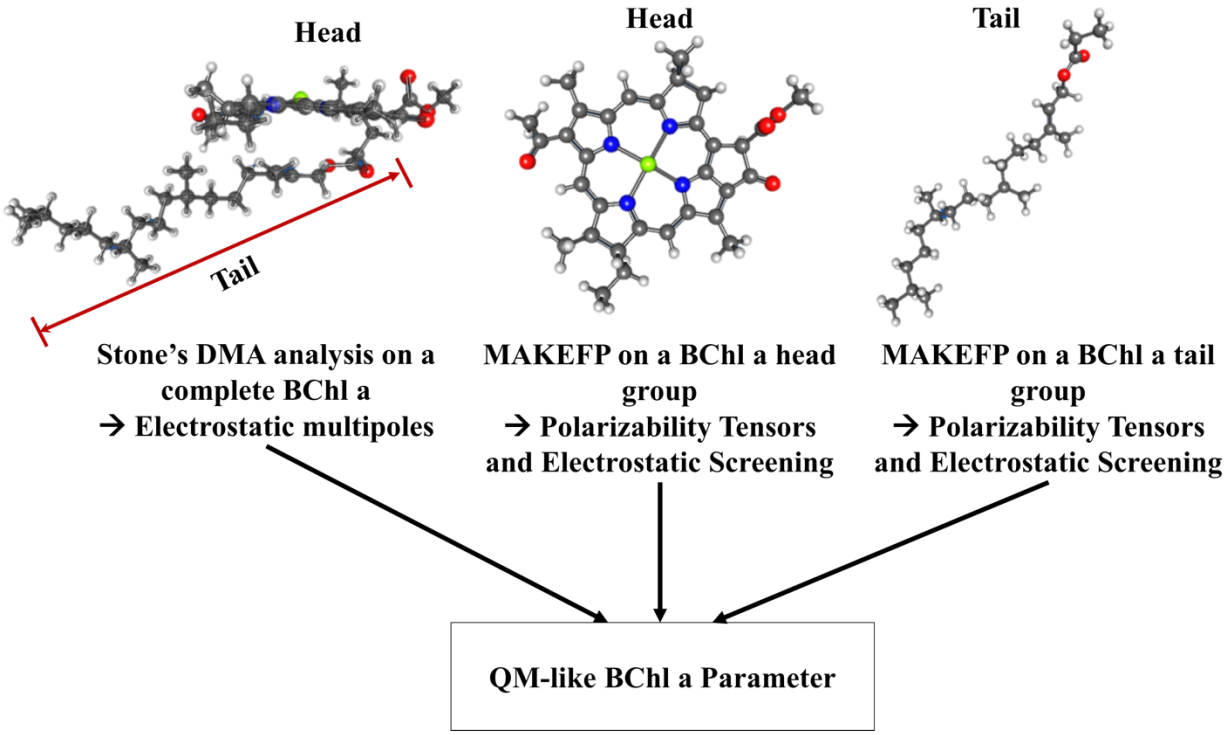

Figure 8.1. Schematic procedure of parameter preparation for a QM-like EFP fragment.

8.3.3 Modeling solvatochromic shifts with EFP calculations

The solvatochromic (or electrochromic) shift is a measure of differential solvation of the ground and electronically excited states of a chromophore. In other words, the solvatochromic shift is the difference of the excitation energies of a chromophore in an environment and in gas phase. In the QM/EFP models, the solvatochromic shift is computed as an excitation energy difference

between QM/EFP and gas-phase excited state computations. In QM/EFP model, surrounding EFP fragments interact with the ground and excited state wave functions of the QM region differentially. However, the same computation scheme cannot be applied to the purely EFP model because in the EFP formalism only the interaction energy between effective fragments is computed. Hence, the computation schemes in the EFP model can be written as:

$$E_{shift} = E_{int}^{ex} - E_{int}^{gs} \tag{8.1}$$

where, E_{int}^{ex} and E_{int}^{gs} represent interaction energies of the excited and ground state QM-like EFP fragment with surrounding EFP fragments. The schematic computation procedure of the solvatochromic shift in the EFP model is shown in Figure 8.2. In this computation formalism, which is similar in spirit to computing differences in binding energy, one can estimate effects of surrounding EFP fragments to the QM-like EFP fragment at different quantum states. In addition, contribution of each fragment can be separated using the pairwise EFP formalism.

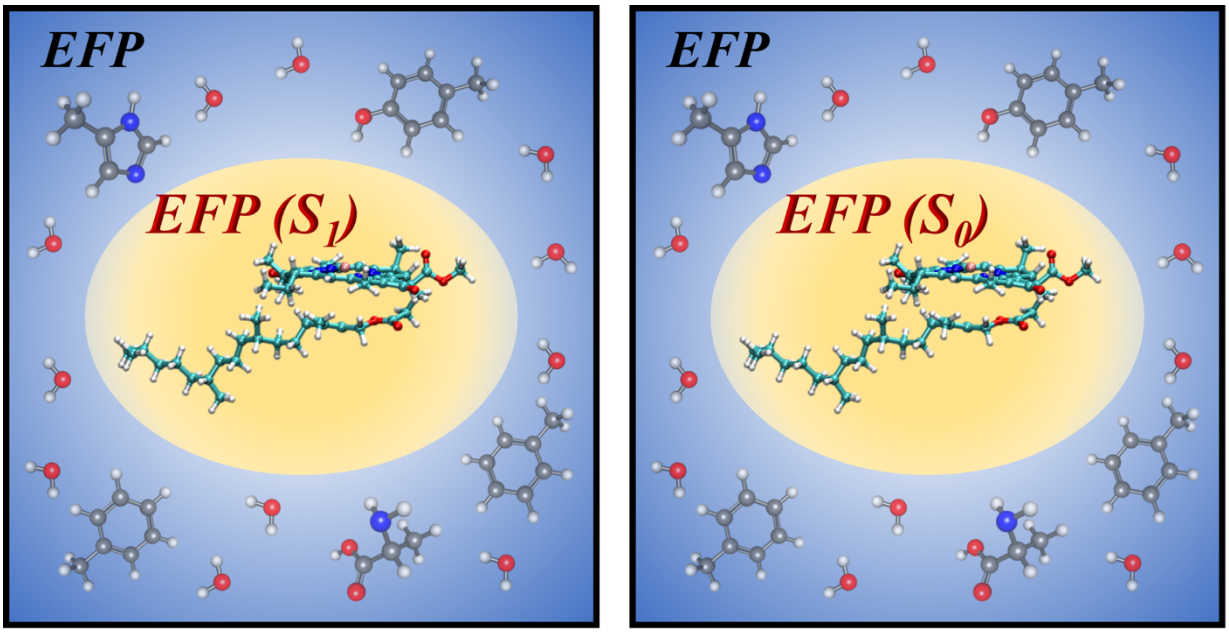

Figure 8.2. Modeling of solvatochromic shifts with EFP calculations. EFP (S_1) and EFP (S_0) represent excited and ground state electrostatic potentials from the Stone DMA computed at the TDDFT and DFT PBE0/6-31G(d) levels, respectively. The solvatochromic shift due to a particular fragment is a difference in pairwise interaction energies of this fragment with the S_1 and S_0 QM-like EFP fragments.

8.3.4 Dimer calculations

To test the quality of the pure EFP scheme in predicting solvatochromic shifts, we considered a set of dimers, in which the system is composed of a specific BChl a pigment and an AA residue or another BChl a molecules within 15 Å from the pigment head group, extracted from 25 snapshots of the MD trajectory, as shown in Figure 3.6. For each snapshot, this procedure resulted in 154, 178, 177, 195, 180, 175, 213, and 137 dimer systems (plus a number of BChl a – water dimers) for each BChl a pigment. The solvatochromic shifts and its energy components (i.e. electrostatic and polarization contributions) in these dimers are compared between the pure EFP and QM/EFP schemes.

QM/EFP calculations: A BChl a pigment is represented as the QM region at the TDDFT PBE0/6-31G(d) level of theory. The excitation energy of this BChl a is computed with including a single AA residue or another BChl a fragment represented by an EFP fragment. QM/EFP excitation energies are computed in the GAMESS quantum chemistry package [140].

Pure EFP calculations: Two sets of pure EFP interaction energy computations are performed for each dimer. In one, the ground state electrostatic potentials of the QM-like BChl a, i.e., EFP (S_0), are used, and in the other, the electrostatic potentials of the electronically excited QM-like BChl a fragment, i.e., EFP (S_1), are utilized. The EFP parameters of the AA residue or another BChl a are identical between QM/EFP and pure EFP calculations. All EFP interaction energies are computed with the EFPMD module of the LibEFP software library [72,73].

8.3.5 Solvatochromic shifts with the EFP scheme in to the full protein environment

Even though the dimer computations provide the majority of solvatochromic shifts due to particular residues, many-body polarization effects are missing in these calculations. In order to account for the many-body polarization effects and mimic the QM/EFP system setup from Chapters 3 and 4, we computed the interaction energies of the QM-like BChl a effective fragments with the reminder of the protein represented as a combination of EFP fragments, including water molecules and MM electrostatic point charges. Similar to the dimer calculations, in these full-system calculations the QM-like BChl a pigment is represented by two different fragments representing S_0 and S_1 electronic states of BChl a. EFP schemes for water molecules and MM point charges are shown in Sections 3.3.3 and 3.3.5.

Then, solvatochromic shifts (Eqn (8.1)) for each QM-like BChl a fragment were computed based on the same 25 snapshots that were utilized in the dimer calculations. The QM/EFP solvatochromic shifts were taken from Chapters 3 and 4 results (gas phase: single BChl a, but the QM region in QM/EFP contains BChl a + Mg coordinating AA). All EFP interaction energies were computed with the EFPMD module of the LibEFP software library [72,73].

8.3.6 Systematic improvement of EFP solvatochromic shifts

As shown in Figure 3.6, QM regions in the QM/EFP excitation energy calculations were composed of BChl a and Mg-coordinating AA residues. This implies that a more precise apples-to-apples comparison of the pure EFP scheme against QM/EFP can be achieved by including Mg-coordinating AA residues into the QM-like fragment regions. A schematic representation of different computational approaches is shown in Figure 8.3. The QM/EFP computations (see Figure 8.3 (a)) were already conducted in Chapter 4. Additional gas phase computations (Figure 8.3 (b)) and EFP parameter calculations for the QM-like EFP fragments with the inclusion of Mg-coordinating AAs were performed for a single structural snapshot. Gas phase excitation energies and EFP parameters were computed with the GAMESS quantum chemistry package [140]. All EFP interaction energies were computed with the EFPMD module of the LibEFP software library [72,73].

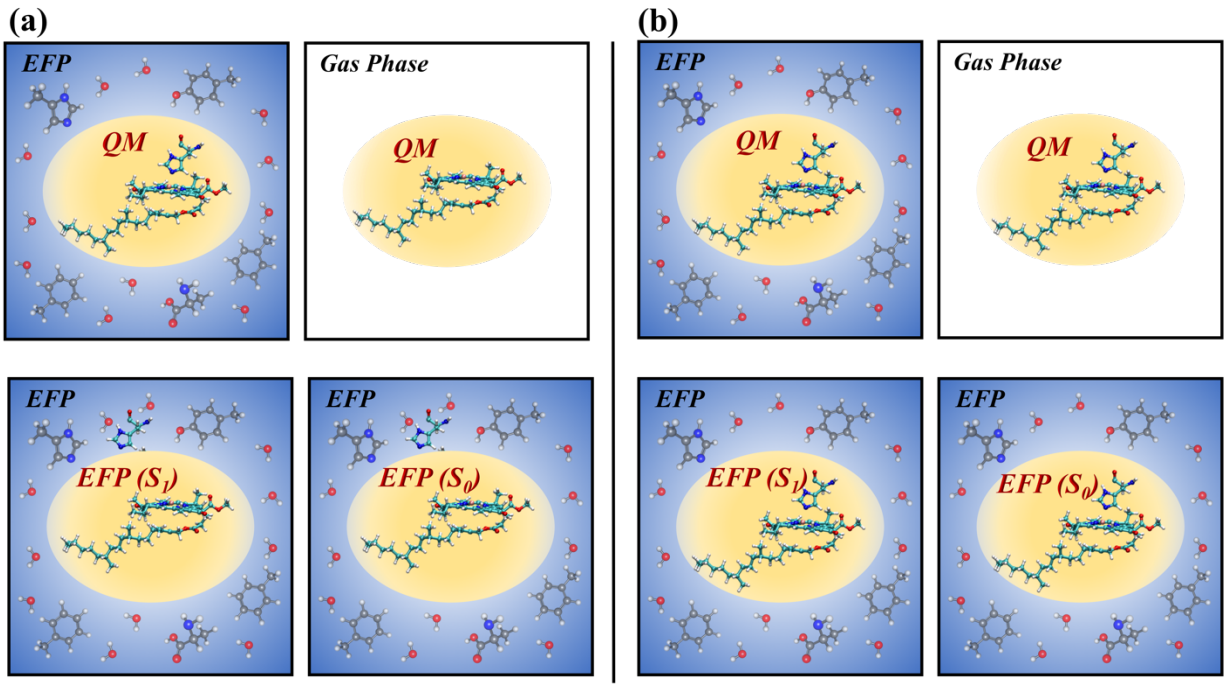

Figure 8.3. Schematic representation of the chromophore and protein regions for calculations of solvatochromic shifts with the QM/EFP and pure EFP schemes. (a) Computation scheme utilized in the preliminary study, and (b) the improved procedure.

8.4 Results

8.4.1 Solvatochromic shifts in dimers

Figures 8.4 – 8.9 provide solvatochromic shifts in QM/EFP (QM: BChl a pigment, EFP: AA residue) and pure EFP (EFP: QM-like BChl a, AA residue) dimer calculations and errors of the pure EFP scheme against QM/EFP for all neighboring pairs in the FMO protein. The color scheme represents the excited pigment. The solvatochromic shifts computed with the pure EFP approach exhibit similar values and trends compared with those computed with QM/EFP. As follows from Figures 8.4 – 8.9, residue contributions increase as the distance between the specific residue and the BChl a pigment decreases. The distances between QM-like BChl a and others were measured as the closest atomic distance (between heavy atoms) of the pigment head group and the other fragment. Because of the larger energy contributions, errors of the pure EFP approach also increase for the residues that are found closer to the BChl a pigments.

The mean absolute errors of the pure EFP versus QM/EFP solvatochromic shifts due to neighboring fragments are shown for each site separately in Figures 8.10 and 8.11. It is noteworthy that the pure EFP scheme exhibits small errors (< 10 cm^{-1}) in the most cases. Several residues near

each BChl a site show errors exceeding 10 cm^{-1}. Those originate from Mg-coordinating AAs, a nearby BChl a molecule, and critical AAs that we discussed in Section 3.4.2. On the other hand, the standard deviations of the errors due to Mg-coordinating AAs, nearby BChl a, and critical AAs are relatively small, suggesting that these errors are systematic and can be corrected with QM calculations. One exception to this trend is the solvatochromic shifts due to the water molecule coordinating Mg of site #2. As discussed in Section 3.4.1, water molecules move freely during MD simulations, such that the water coordination to BChl a #2 is only partially recovered with the QM/MM optimizations. Therefore, the solvatochromic shift due to this water molecule on BChl #2 shows large fluctuations, since position of this water is not well preserved.

Overall, analysis of Figures 8.4 - 8.11 suggests that the pure EFP model qualitatively captures interactions in the FMO protein system. Considering the low computational time of the EFP scheme, this approach is promising for a more detailed investigation, as discussed below.

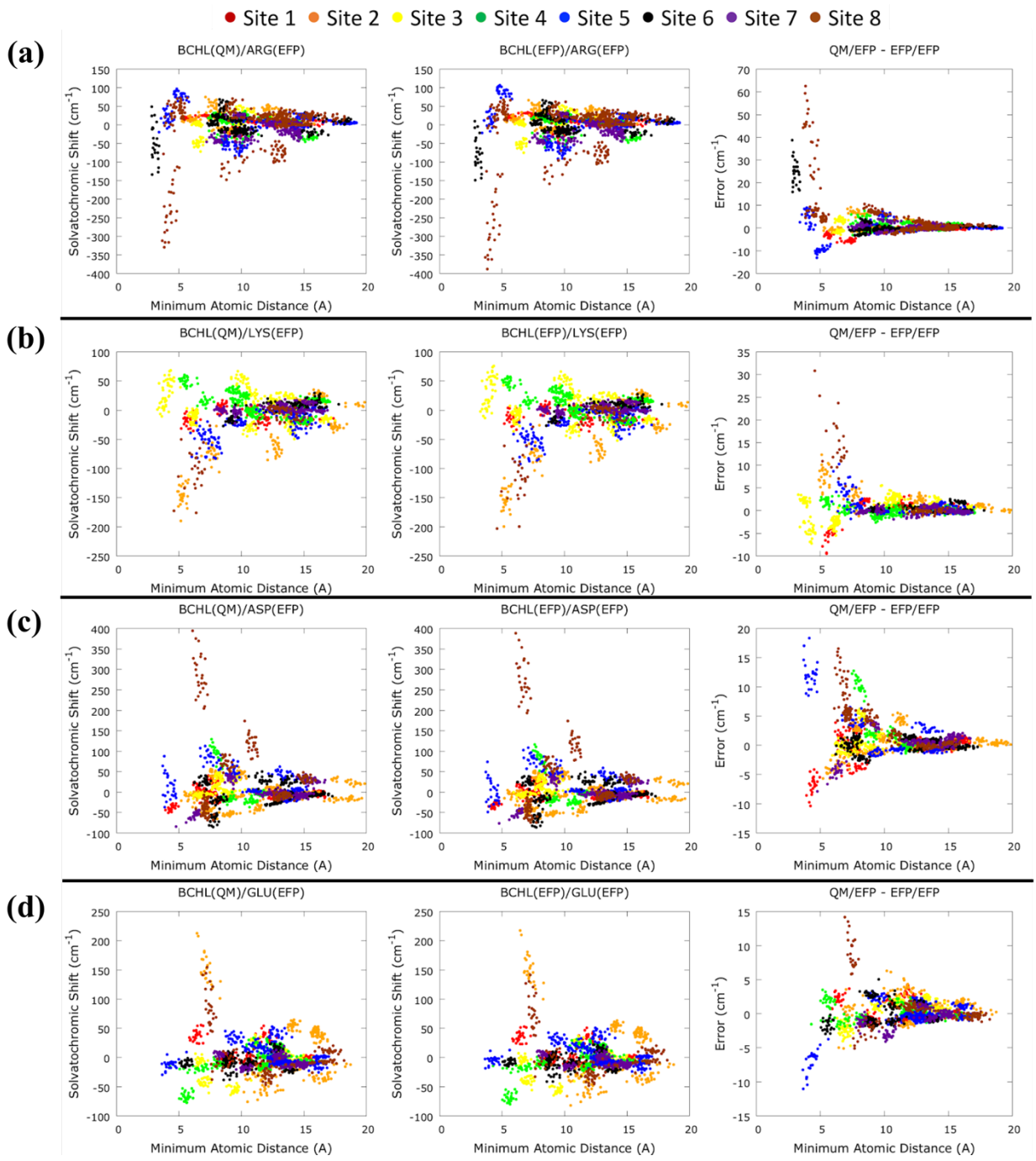

Figure 8.4. Solvatochromic shifts in the dimers containing charged amino acids (a) ARG, (b) LYS, (c) ASP, and (d) GLU. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

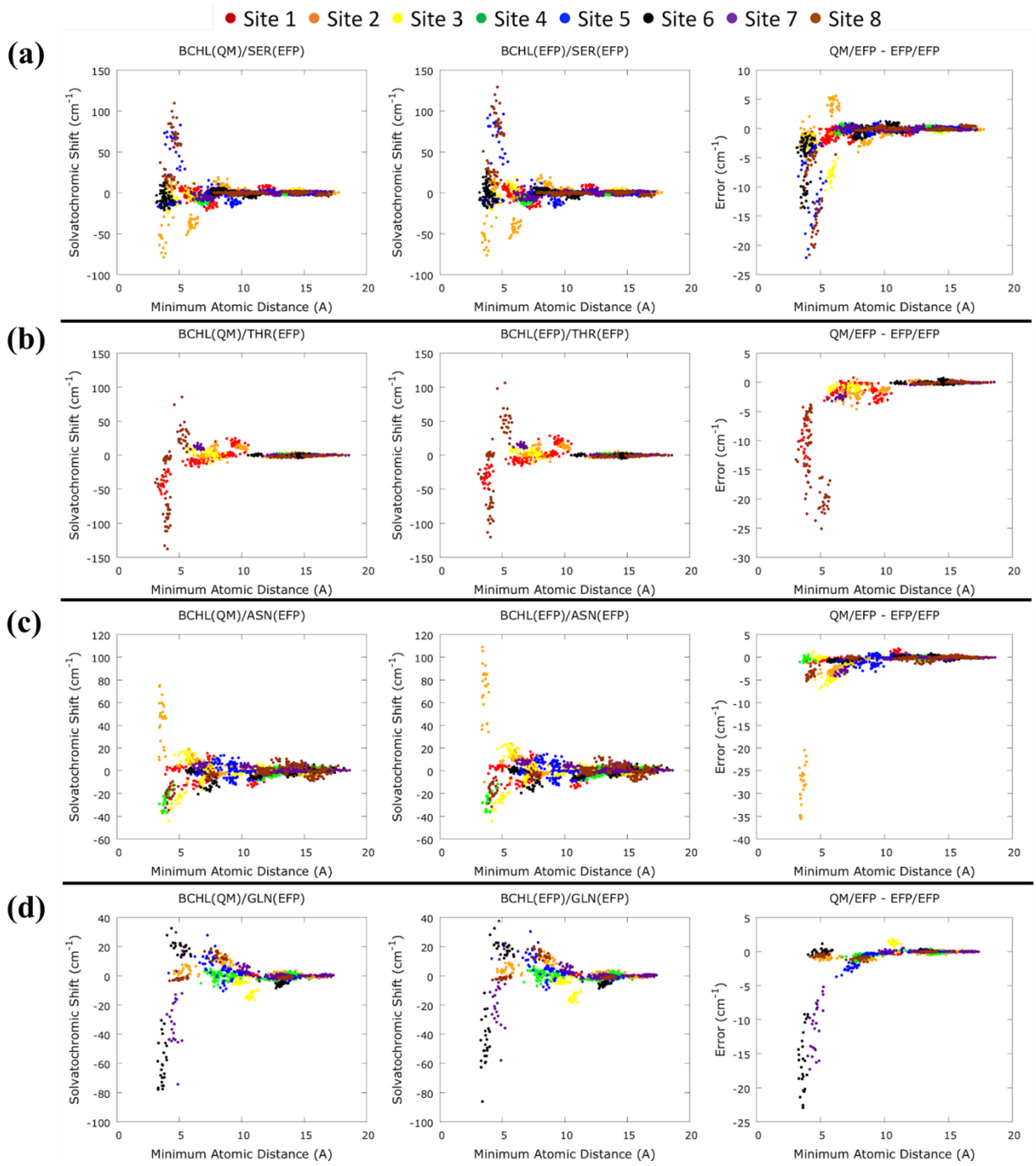

Figure 8.5. Solvatochromic shifts in the dimers containing polar amino acids (a) SER, (b) THR, (c) ASN, and (d) GLN. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

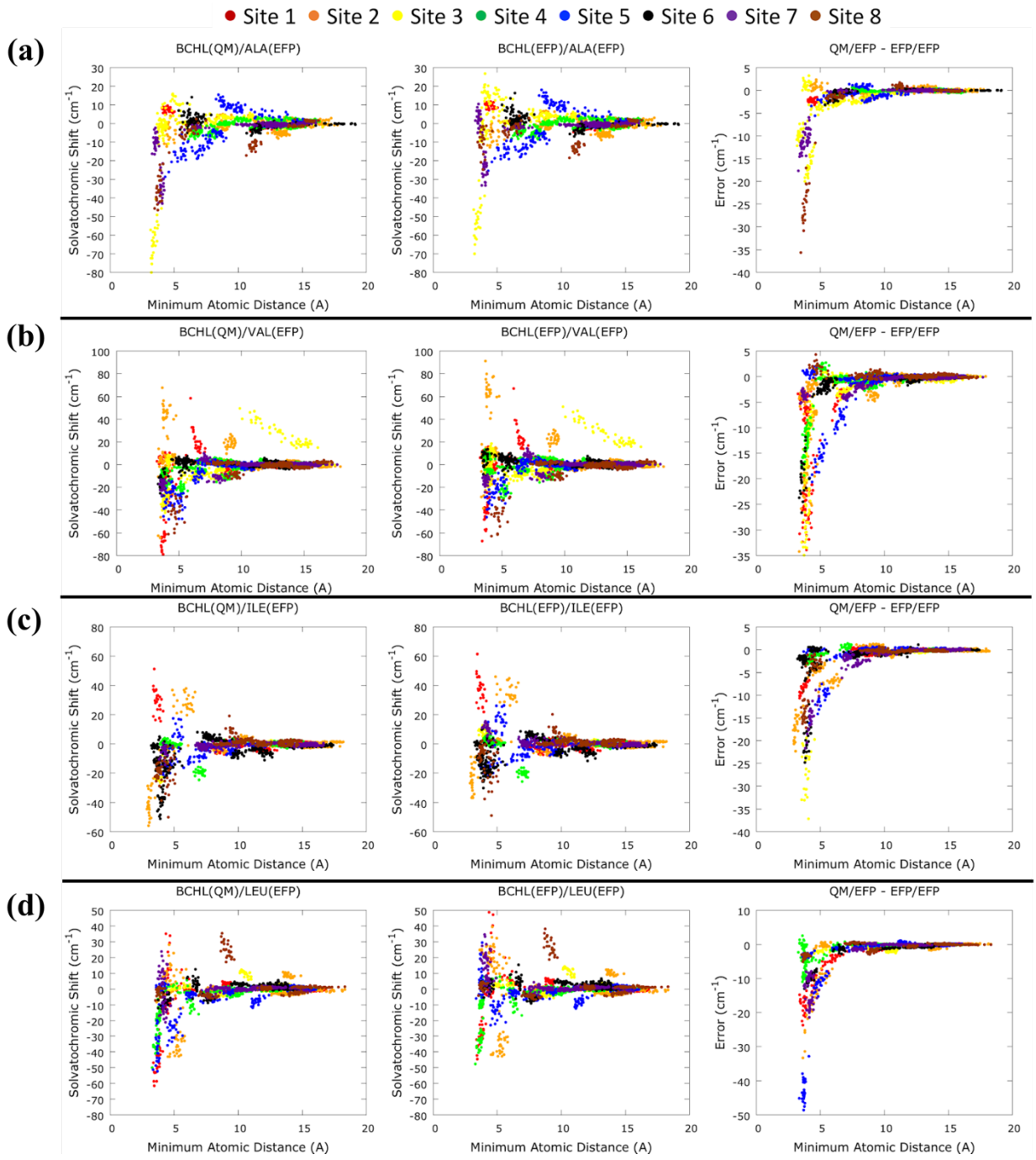

Figure 8.6. Solvatochromic shifts in the dimers containing hydrophobic amino acids (a) ALA, (b) VAL, (c) ILE, and (d) LEU. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

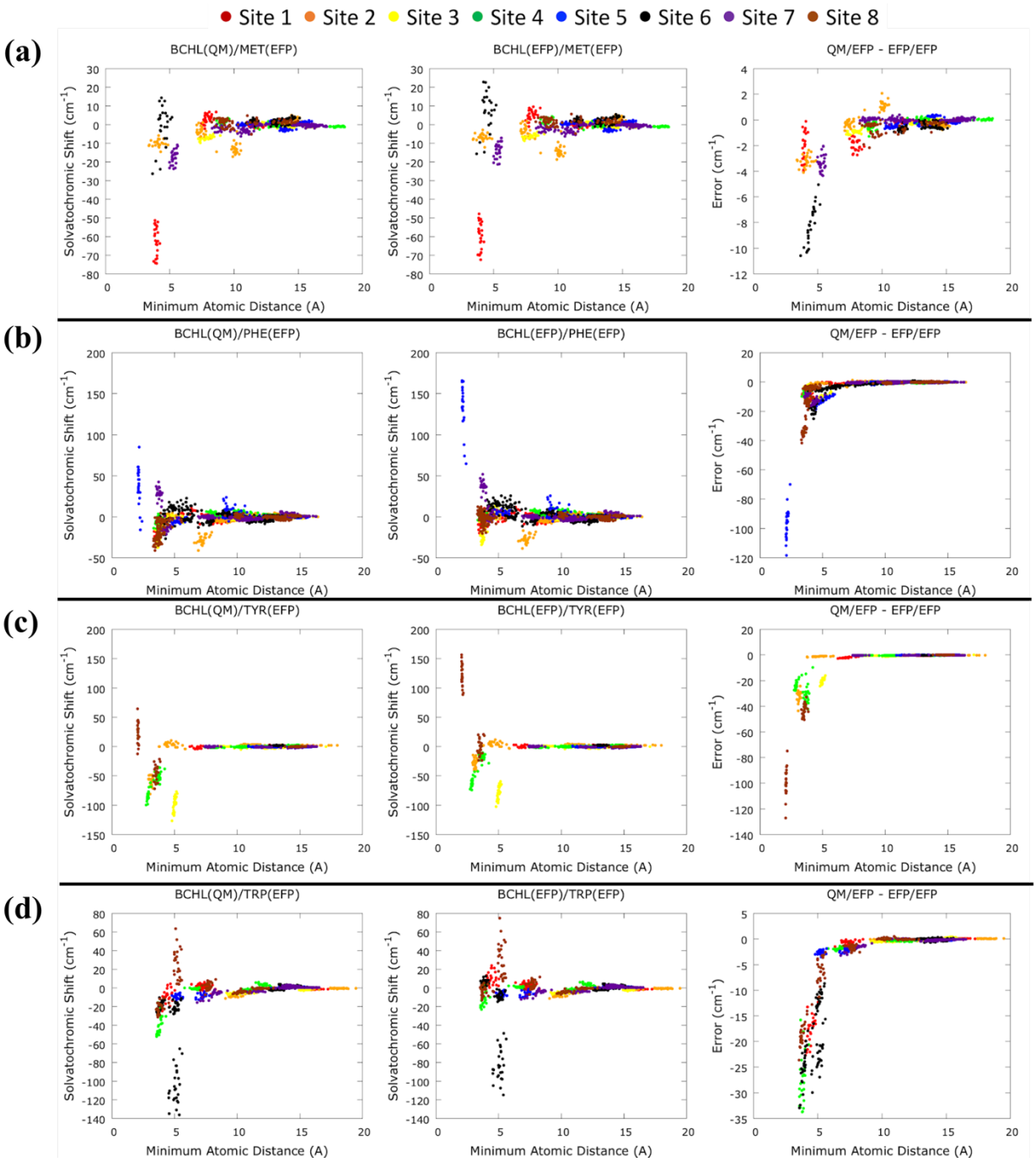

Figure 8.7. Solvatochromic shifts in the dimers containing hydrophobic amino acids (a) MET, (b) PHE, (c) TYR, and (d) TRP. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

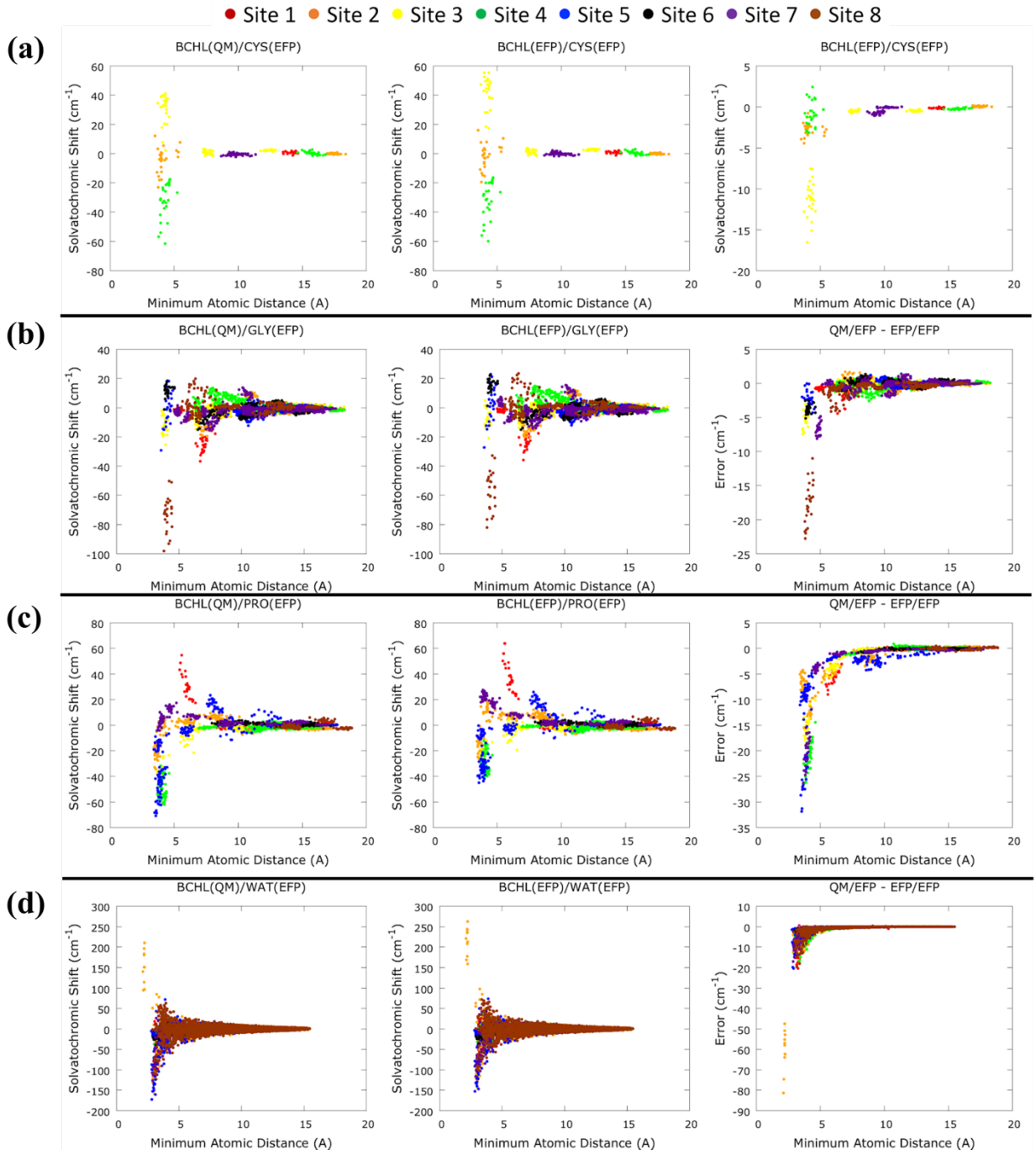

Figure 8.8. Solvatochromic shifts in the dimers containing special cases of (a) CYS, (b) GLY, (c) PRO, and (d) water. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

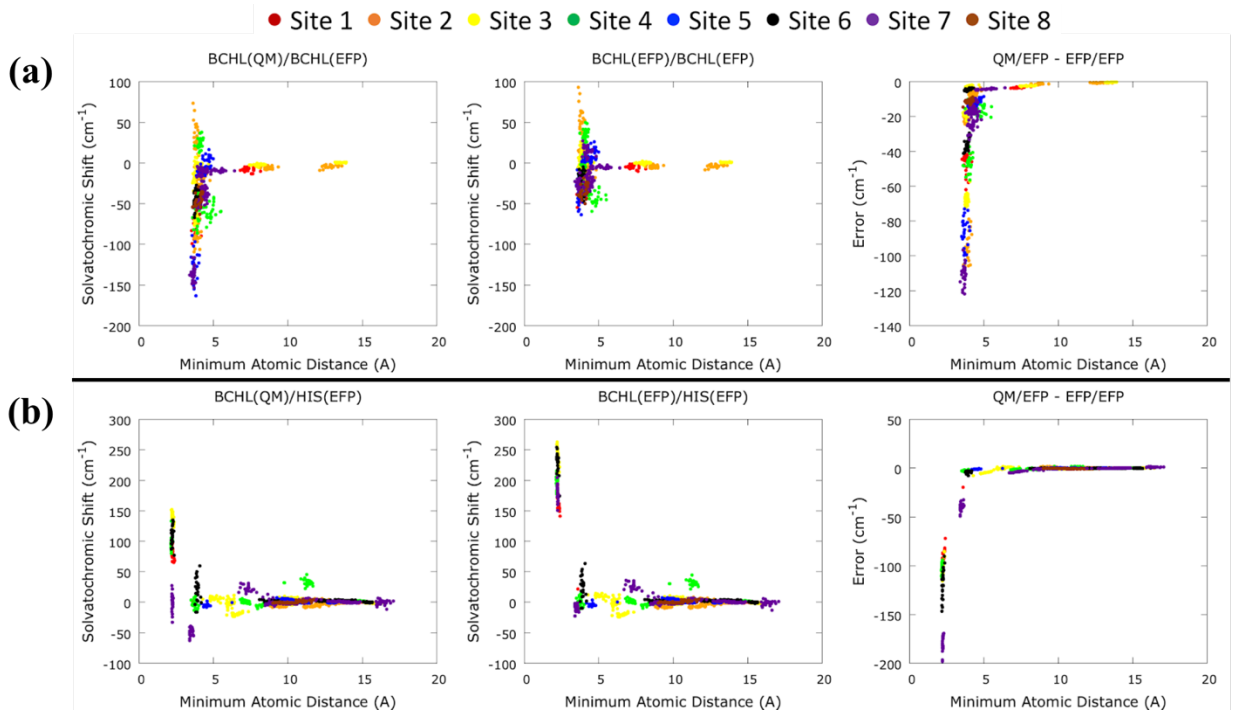

Figure 8.9. Solvatochromic shifts in the dimers containing special cases of (a) BChl a and (b) HIS. The left panels represent solvatochromic shifts computed with QM/EFP, the middle panels show the shifts from the pure EFP scheme, and the right panels indicate errors of the pure EFP scheme against the QM/EFP calculations.

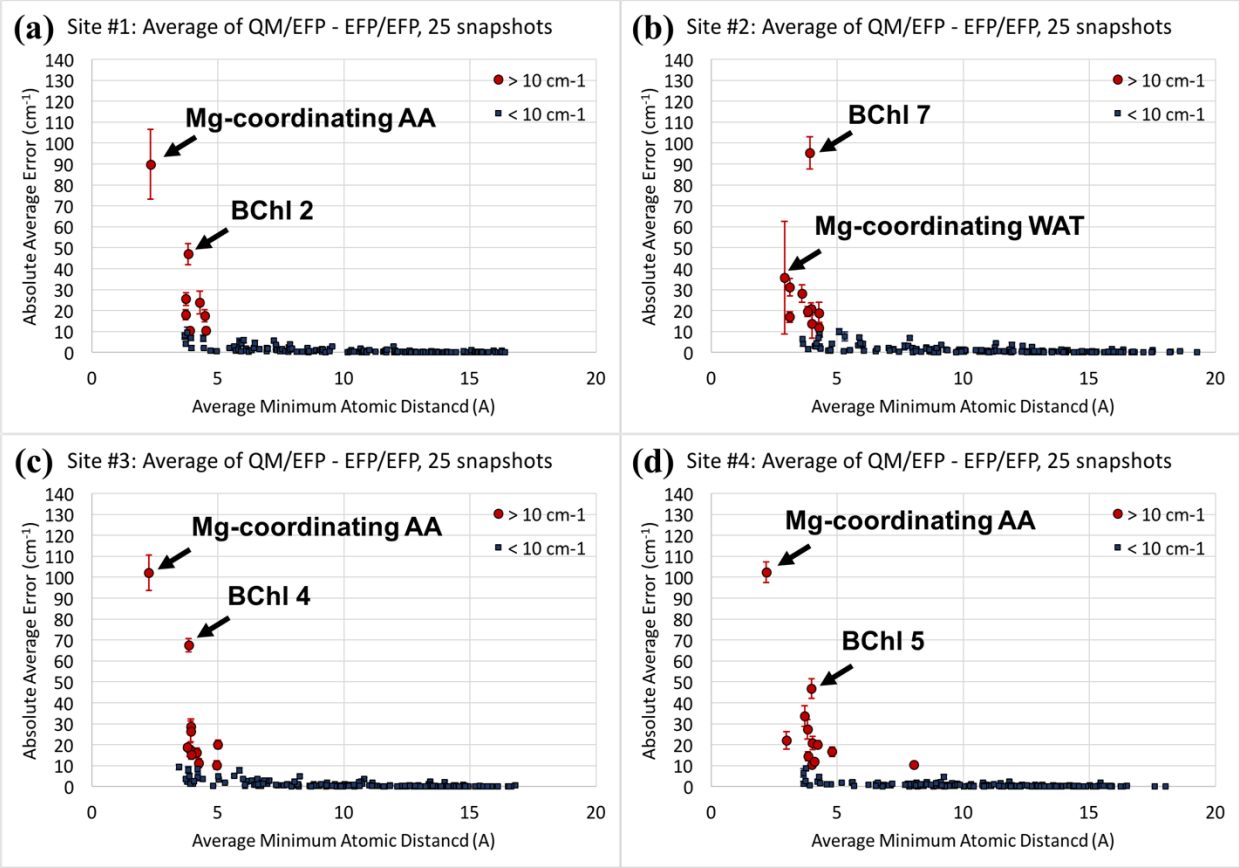

Figure 8.10. Mean absolute errors and standard deviations of the pure EFP solvatochromic shifts versus the QM/EFP shifts. Errors are plotted as a function of the average minimum atomic distance between the QM-like BChl a and neighboring residues. (a) BChl a #1, (b) BChl a #2, (c) BChl a #3, and (d) BChl a #4. Errors exceeding 10 cm^{-1} are colored in red. Residues producing errors of more than 40 cm^{-1} are specifically shown.

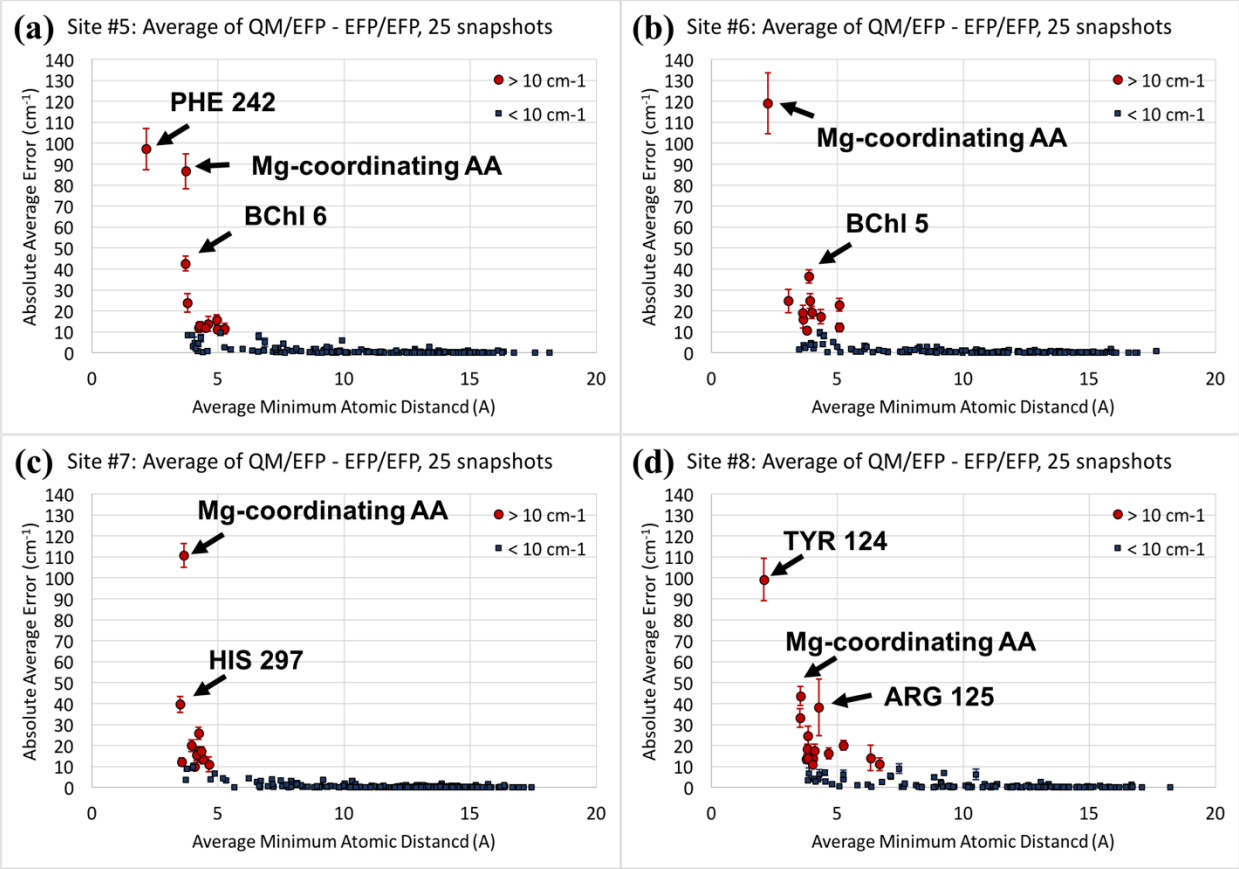

Figure 8.11. Average errors and their standard deviations of the pure EFP solvatochromic shifts versus the QM/EFP shifts. Errors are plotted as a function of the average minimum atomic distance between the QM-like BChl a and neighboring residues. (a) BChl a #5, (b) BChl a #6, (c) BChl a #7, and (d) BChl a #8. Errors exceeding 10 cm^{-1} are colored in red. Residues producing errors of more than 40 cm^{-1} are specifically shown.

8.4.2 Solvatochromic shifts in protein

Using a similar formalism to computation of the binding energy between a ligand and a protein (see Section 2.5.2, 8.3.5, and 8.3.6), protein-induced solvatochromic shifts of BChl a pigments can be effectively evaluated. Here we computed the interaction energies of S_0 and S_1 forms of the QM-like BChl a fragments with the EFP fragments representing AAs and other BChl a molecules. Far-separated residues and solvent molecules were represented as MM point charges, using the same separation into EFP and MM regions as in Chapters 3 and 4. Figure 8.12 presents a comparison of average solvatochromic shifts for each site computed at 25 snapshots computed with three models, electrostatic embedding QM/MM, polarizable QM/EFP, and polarizable but classical pure EFP. Interestingly, the pure EFP model seems to reproduce the solvatochromic shifts

of electrostatic embedding QM/MM model even though EFP fragments polarize each other in the EFP formalism. Most noticeably, the pure EFP model was not able to sufficiently stabilize sites #3 and #4, as the polarizable embedding QM/EFP model did. Two things might explain this discrepancy.

(i) As shown in the dimer calculations, the pure EFP description does not quantitatively reproduce solvatochromic shifts due to Mg-coordinating AAs, nearby BChl a, and several critical AAs (see also Section 3.4.2 and Figure 3.8), suggesting that the simplified classical potentials cannot model the QM wave function rigorously for closely-spaced and strongly interacting molecules.

(ii) As we discussed in Section 8.3.6, Mg-coordinating AAs in the QM/EFP calculations of the shifts are included in the quantum region. However, in the pure EFP scheme, the Mg-coordinating AAs are treated at the EFP level. In other words, the Mg-coordinated AAs are treated quantum-mechanically in QM/EFP but classically in pure EFP, which might induce a non-negligible difference. Thus, QM-like BChl a and corresponding Mg-coordinating AAs should be treated simultaneously as a QM-like fragment in the pure EFP model, for the exact comparison to QM/EFP. In addition, this approach, if utilized, will effectively eliminate the largest errors of the pure EFP scheme arising from the Mg-coordinating AAs, such that the system description will be systematically improved.

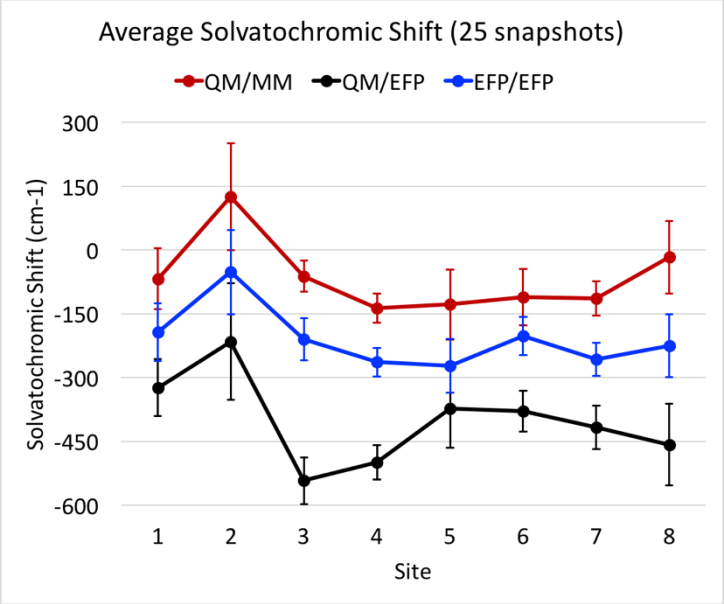

Figure 8.12. Average protein solvatochromic shifts computed at 25 snapshots and their standard deviations shown as vertical bars using the QM/MM (red), QM/EFP (black), and pure EFP (blue) models.

8.4.3 Systematic improvement to the pure EFP calculations

As described in Section 8.3.6, each QM-like BChl a and corresponding AA can be combined in a single effective fragment. In this treatment, the interactions between the QM-like BChl a and corresponding Mg-coordinating AAs are completely ignored. For the direct comparison, we recomputed "gas phase" excitation energies with the inclusion of the Mg-coordinating AAs (see Figure 8.3), and obtained solvatochromic shifts by subtracting the recomputed gas phase excitation energies from the QM/EFP values. Thus, this approach produces solvatochromic shifts of the whole protein with exception of the Mg-coordinating AAs, which effect can be considered separately if needed.

Figure 8.13 provides comparison of solvatochromic shifts computed with pure EFP, electrostatic embedding QM/MM and polarizable embedding QM/EFP. As expected from the previous preliminary study, pure EFP (blue and green lines) schemes produce comparable solvatochromic shifts to the QM/MM model (red line). It is fascinating that the pure EFP model gets closer to the shifts of the polarizable embedding QM/EFP model and are in good agreement with QM/EFP in terms of site to site tendencies, showing stabilization in site #3.

As we saw in the dimer calculations, a few AAs and neighboring BChl a pigments produced significant errors in the pure EFP scheme. In principle, it is possible to correct these errors using the following scheme:

$$E_{correction} = E_{QM/EFP}^{dimer} + \left(E_{EFP/EFP}^{protein} - E_{EFP/EFP}^{dimer}\right) \tag{8.2}$$

where $E_{QM/EFP}^{dimer}$ and $E_{EFP/EFP}^{dimer}$ represent the solvatochromic shifts due to a particular residue from the dimer QM/EFP and pure EFP calculations, respectively, and $E_{EFP/EFP}^{protein}$ indicates solvatochromic shifts of the same residue in the presence of the protein environment. Therefore, $\left(E_{EFP/EFP}^{protein} - E_{EFP/EFP}^{dimer}\right)$ term can be interpreted as a polarizable correction to the dimer solvatochromic shift. The QM/EFP corrected solvatochromic shifts are also shown in Figure 8.13. The corrected solvatochromic shifts become closer to the shifts of the polarizable embedding QM/EFP model and are in excellent agreement with QM/EFP in terms of site to site tendencies.

It should be emphasized that the QM/EFP scheme can provide details of energy component contributions but so far cannot decipher specific effects of individual fragments. Because we show that the corrected pure EFP scheme quantitively reproduces QM/EFP solvatochromic shifts, the total protein solvatochromic effects on each BChl a site can be decomposes into contributions of individual fragments. Thus, analysis of specific AAs that dominate energy lowering or increase on particular sites will be done in the near future to further understand the FMO protein systems.

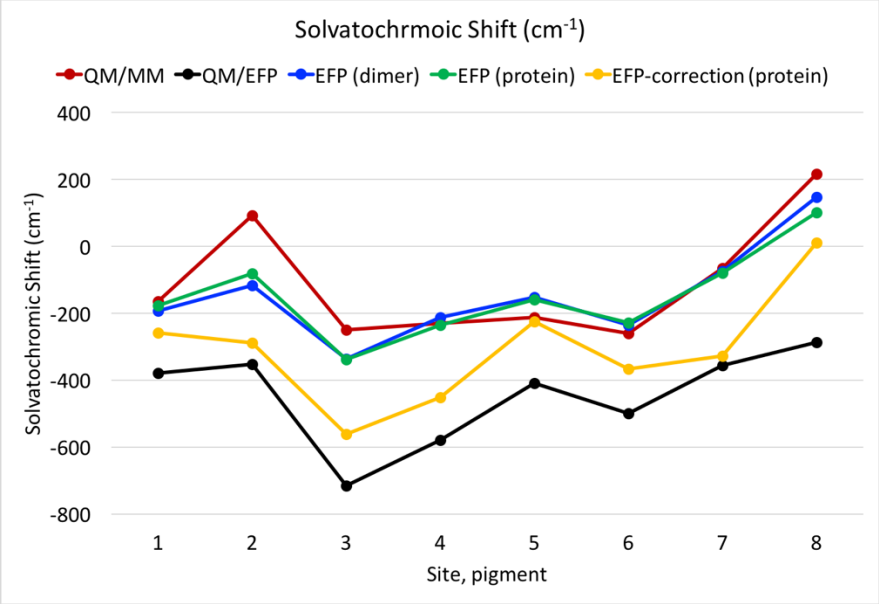

Figure 8.13. Solvatochrmoic shifts computed at a single snapshot with QM/MM (red), QM/EFP (black), and pure EFP models (blue – dimer calculations, green – protein calculations, yellow – protein calculations with the correction (Eqn 8.2)).

8.5 Conclusions

In the present work, the concept and application of the pure EFP model to the FMO protein system are reported. The pure EFP formalism computes solvatochromic shifts due to the protein environment by modeling BChl a chromophores with classical polarizable potentials representing their electronic densities of the ground and excited electronic states. The EFP model can provide details of individual fragment contributions to the system of interest. We validated the EFP scheme by considering solvatochromic shifts in the dimers composed of BChl a pigment and neighboring AA residues and other molecules and showed that the pure EFP model approaches the accuracy of QM/EFP for the dimers separated by a certain threshold. Applications of EFP/EFP scheme to solvatochromic shifts in the FMO protein demonstrate substantial benefits of utilizing EFP/EFP in realistic simulations of biological systems. While the computational cost of computing EFP/EFP energies becomes negligible compared to the QM/EFP model, the predicted solvatochromic shifts and tendencies show reasonable agreement with the ones obtained from QM/EFP model. Future work on this topic includes increasing sample size and decomposing contributions of individual AAs to solvatochromic shifts.

CHAPTER 9. EFFECTIVE FRAGMENT POTENTIAL INVESTIGATION OF TERT-BUTANOL WITH WATER ION MIXTURES

9.1 Abstract

Interactions of t-butyl alcohol (TBA) with water-ion mixtures (NaOH and NaI) are investigated with the Effective Fragment Potential (EFP) method. A series of EFP simulations (EFP/MD and QM/EFP) were carried out to obtain structural data, C-H stretch frequencies, and corresponding Raman activities of TBA molecules. Radial distribution function, g(r), between central carbon of TBA and ions were also computed, showing that there is a more significant fraction of iodide ions than fluoride in the first hydration shell of TBA. Computed Raman spectra suggest that C-H stretch of TBA can be perturbed with ions located in the first hydration shell. Those effects are significant in NaI aqueous solution of TBA due to the higher probabilities of finding Na^+ and I^- ions in a proximity of TBA.

9.2 Computational details

9.2.1 Preparation of the systems and molecular dynamics simulations

The three model systems investigated in this study consist of TBA-water, TBA-water-NaOH, and TBA-water-NaI mixtures as shown in Figure 9.1.

Classical MD simulations: The CHARMM27 force field [6,186] was used for the TBA molecule and Na^+ and OH^- ions. OPLSAA [187,188] force field was used for I^-. For water, the transferable intermolecular potential with 3 points (TIP3P) model was used for all the simulations [125]. The TBA-water system consists of 4 TBAs and 416 water molecules in a simulation box of approximately $24 \times 24 \times 24$ $Å^3$. The water-ion mixtures consist of 4 TBAs, 400 water molecules, 8 Na^+ ions, and 8 negatively charged ions (I^- or OH^-) corresponding to 0.5 M concentration of TBA and 1.0 M concentration of each ion. After energy minimization, molecular dynamics (MD) equilibrations were performed with the NVT and NPT ensembles for 1 ns each. Then, the production run with the NPT ensemble was carried out with a velocity rescale thermostat [127] for temperature control (300 K) and Parrinello-Rahman barostat [128] for pressure control (1 bar). C-H, O-H and N-H bond lengths were constrained with the LINCS algorithm [129]. A 1 nm cutoff was used to handle Lennard-Jones potentials. Electrostatic long-range interactions were treated

with particle mesh Ewald summations (PME) [130,131] with a real-space cutoff of 1 nm. Total 2 ns production run was carried out with 2 fs time step. All classical simulations were performed with the GROMACS package (version 2016.5) [132].

EFP MD simulations: Staring from the final atomic configuration of classical MD simulations, we performed additional molecular dynamics simulations at the EFP level, by representing each molecule as an EFP fragment. The reason for performing EFP MD simulations is to rigorously capture the essential components of noncovalent interactions such as polarization and short-range charge-penetration effects. The EFP parameters in this study were obtained from previous work [42,43]. The TBA geometry for the EFP parameter computations was optimized with MP2/cc-pVTZ level of theory. The EFP parameters for all molecules were computed with hybrid basis sets, namely 6-31+G(d)/6-311++G(3df,2p) for TBA, water, and OH^-, 6-311G(d)/6-311++G(3df,2p) for Na^+, and 6-311G(d)/aug-cc-pVQZ for I^-. After energy minimization, molecular dynamics (MD) equilibrations were performed with the NVT and NPT ensembles for 100 ps each. 900 ps NPT production run was carried out with 2 fs time step at 400 K and 1 bar. All EFP MD simulations were performed with the EFPMD module of the LibEFP software library [72,73]. For each system, we randomly extracted 100 atomic configurations from the EFPMD trajectory without bias. Specifically, a total of 90,000 atomic configurations were saved to disk (one structure every 10 fs). Out of those, 100 snapshots were selected using a random number generator in the range from 1 to 90,000. At these selected snapshots, one (out of four) TBA molecule was positioned at the center of the simulation box using the image-centering technique. The image-centering technique was applied to all four TBAs in each atomic configuration, resulting in 400 different configurational snapshots. These 400 configurational snapshots have been considered for the frequency and Raman intensity computations for each TBA-water-ion system.

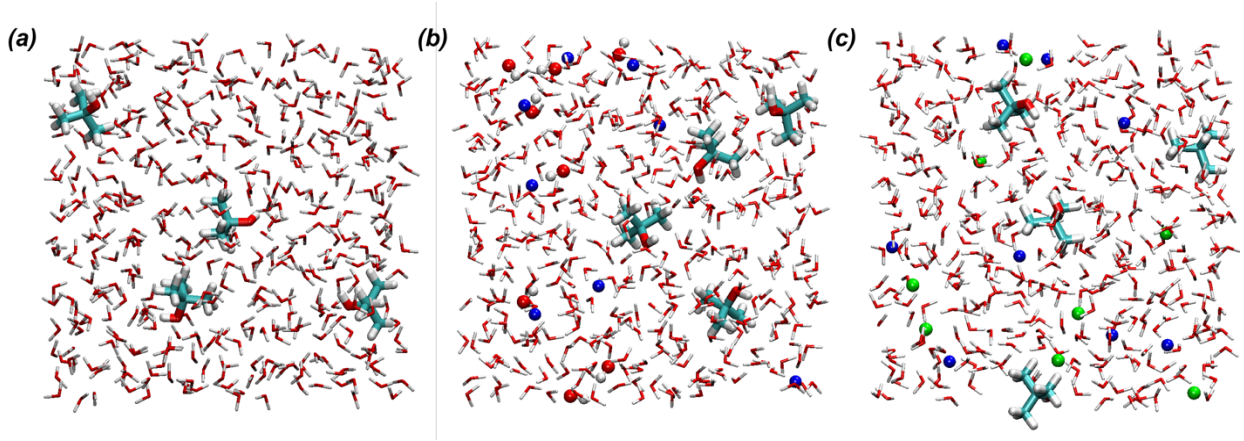

Figure 9.1. Snapshots of (a) TBA-water, (b) TBA-water-NaOH, and (c) TBA-water-NaI systems. The Na^+ ions are represented by blue spheres, OH^- is shown with red and white spheres, and I^- with green spheres.

9.2.2 Frequency and Raman activity computations of C-H stretching vibrations of TBA

Selected 400 TBA configurational snapshots of each water-ion mixture system were used for evaluating C-H frequency shifts using QM/EFP and the partial Hessian technique [189,190]. The central TBA molecule was considered as the QM region. In the case of the TBA-water-NaI system, all sodium and iodide ions were also described at the QM level. The quantum TBA was described with the second-order Møller-Plesset perturbation theory using 6-311++G(d,p) basis set, and the Hay-Wadt effective core potential for Na^+ and I^- ions. The remaining TBA and water molecules were described by EFP parameters as used in the EFP MD simulations. Na^+ and OH^- ions in TBA-water-NaOH system were treated at the EFP level. The exchange-repulsion QM-EFP term developed in REF [141] was included for TBA and water fragments. The EFP fragments and QM ions (in case of the TBA-water-NaI system) were kept fixed, and only vibrational frequencies of the quantum TBA molecule were obtained through diagonalization of a partial Hessian matrix.

Raman intensities were computed for the gas-phase TBA molecule (since the second-order energy derivatives are not implemented for EFP solvation model). The same level of theory (MP2/6-311++G(d,p)) was applied to the Raman computations, while the gradient, Hessian, and dipole derivative tensor matrices were extracted from the QM/EFP frequency computations.

9.3 Results

9.3.1 Radial distribution function, *g(r)*

Specific interactions of TBA with solvent can be identified via the radial distribution functions *g(r)* for specific pair of atoms. Figure 9.2 plots the radial distribution functions between the central carbon (CC) of TBA and (non-hydrogen) atoms of solvent molecules in TBA-water, TBA-water-NaOH, and TBA-water-NaI systems. The dominant peaks at approximately 0.35 nm correlate with Na^+ ions interacting with TBA. Based on *g(r)* between hydroxyl oxygen of TBA (OT) and Na^+ (yellow line), Na^+ specifically interacts with the hydroxyl oxygen of TBA, resulting in high probabilities of finding Na^+ near TBA. This tendency is thrice more pronounced in the TBA-water-NaI mixture than in the TBA-water-NaOH system. Interestingly, the negatively charged ions exhibit different behavior. While the OH^- ions are barely found near the TBA molecule, I^- ions show broad and relatively intensive peaks at 0.4-0.6 nm from the central carbon of TBA.

In order to elaborate the findings from analysis of *g(r)*, we computed and plotted probabilities of finding ions near some distance threshold from TBA. As shown in Figure 9.3, the probabilities of finding ions near the TBA molecule gradually increase as the distance from CC increases. The largest discrepancies between the TBA-water-NaOH and TBA-water-NaI systems are observed for the negatively charged ions. As we observed from analysis of *g(r)*, OH^- has a low probability to be found near a TBA molecule (see Figure 9.2). On the other hand, the probability to find I^- near TBA dramatically increases from the 5 Å radial distance, which is consistent with *g(r)* findings. As a result, the probability of finding both positively and negatively charged ions near TBA continuously increases as we increase the radial distance threshold in the case of the TBA-water-NaI system. In the case of the TBA-water-NaOH mixture, almost half of the TBA molecules do have any nearby ions within a 7 Å sphere around their central carbons.

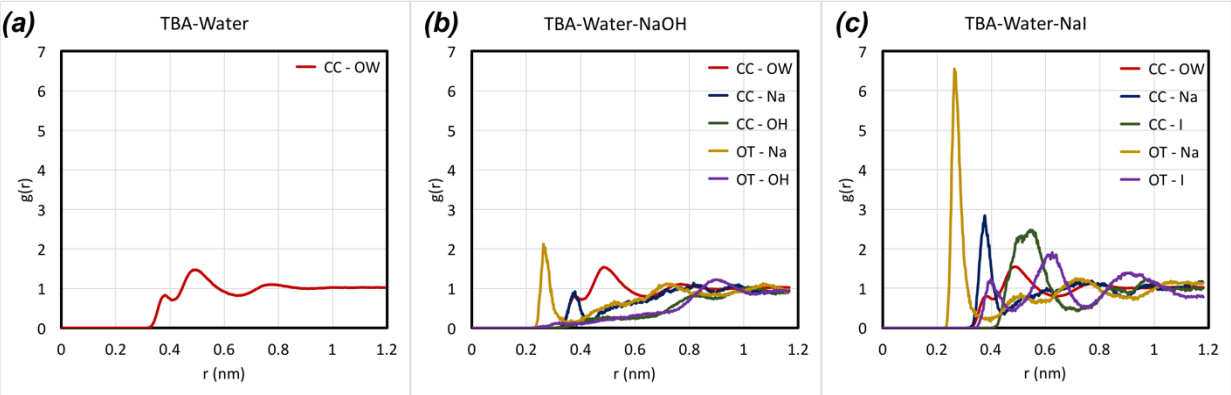

Figure 9.2. Radial distribution functions g(r) between the central carbon atom of TBA (CC) and water oxygen (OW) and ions (Na, OH, I). The radial distribution functions between the hydroxyl oxygen (OT) of TBA and the ions are also shown.

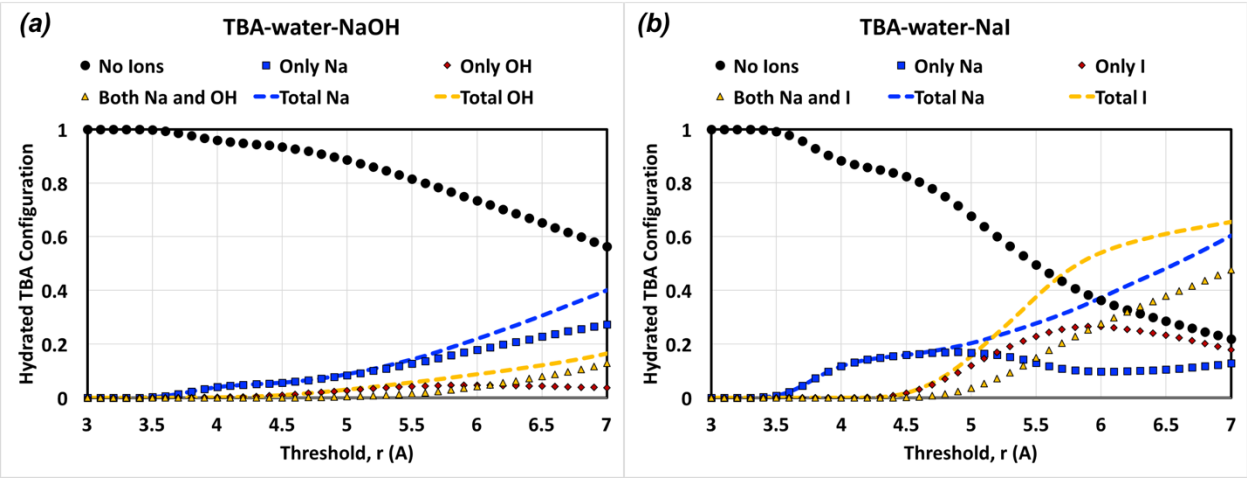

Figure 9.3. Average number of ions found near TBA as a function of the distance cut-off from the central carbon atom of TBA. (a) TBA-water-NaOH system, (b) TBA-water-NaI system.

9.3.2 Influence of ions on C-H vibrational frequencies

The frequencies of the C-H stretch vibrations of TBA were computed using QM/EFP scheme. The frequencies averaged over 400 configurational ensembles and standard deviations are shown in Figure 9.4. As seen in Figure 9.4 (a), the average frequency shifts due to the ions are not significant. On the other hand, we observe significant fluctuations of Raman intensities which suggests that comparing shapes of the overall Raman spectra is more meaningful than comparing shifts of individual frequencies.

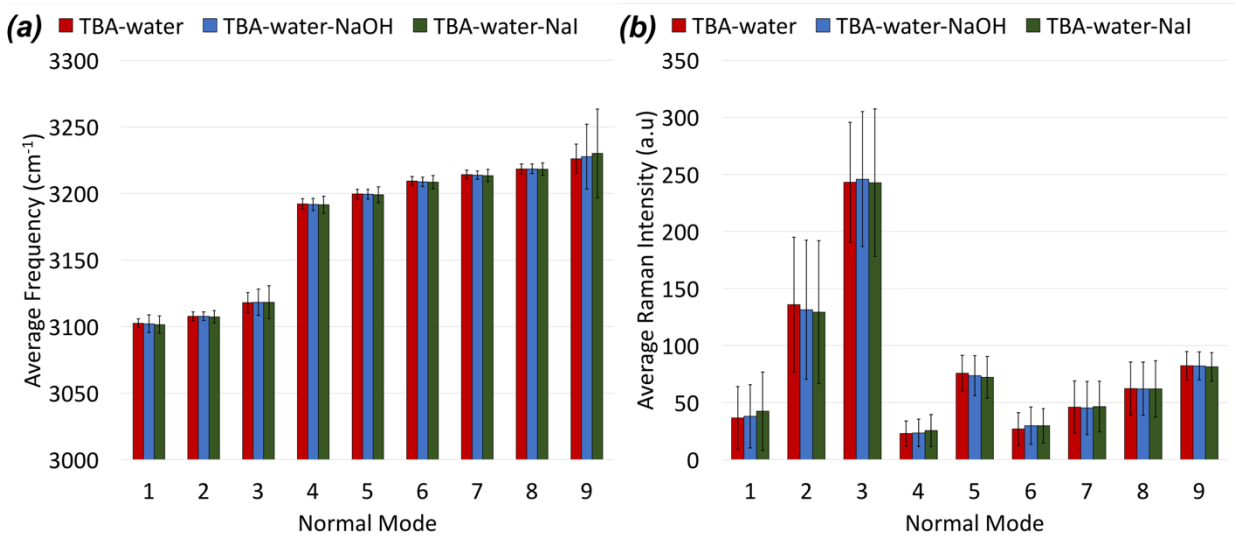

Figure 9.4. Average TBA C-H stretch frequencies and (b) corresponding Raman intensities, and standard deviations of these values shown as vertical error bars. The values are obtained by averaging over 400 configurational snapshots.

9.3.3 Raman spectra of C-H vibrational modes of TBA

In order to compare the frequency shifts due to the presence of ions more directly, we plotted Raman intensity spectra of TBA-water-ion mixtures and compared them with the TBA-water system. In addition, we separate the TBA configurational snapshots based on positions of ions within a particular distance from the central carbon of TBA. Namely, we separated contributions of configurations in which (i) positively charged ions (Na^+), (ii) negatively charged ions (OH^- or I^-), and (iii) both positively and negatively charged ions are present within a specific cutoff distance from CC. Note that the first hydration shell of TBA defined by a minimum in TBA-water g(r) (see Fig. 8.2a) is 6.5 Angstroms. Spectra are constructed by averaging of Raman spectra of individual configurational snapshots. The $10cm^{-1}$ FHWM Gaussian broadening was used in individual snapshot spectra. The averaged Raman spectra were fitted with three Gaussian functions whose maxima were taken as the peak positions for comparison between spectra.

Comparison of Raman spectra based on the TBA hydration conditions is provided in Figures 9.5 – 9.7. In all figures, the black lines indicate the Raman spectra of the TBA-water mixture (reference), and the red lines represent the average spectra of 400 snapshots (without separations). As shown in Figures 9.5 – 9.7, the TBA C-H stretch modes have three major Raman active peaks. Analysis of Figure 9.5 shows that the three Raman active peaks are shifted to the red-

side due to presence of Na^+ ions within the hydration shell. An intriguing feature in the spectra in Figure 9.5 is that the red shifts are more pronounced in the TBA-water-NaI than in the TBA-water-NaOH system even though the probabilities of finding Na^+ ions near TBA are comparable in the two systems. On the other hand, the tendencies are different in the case of negatively charged ions (see Figure 9.6). In the presence of OH^- ions near TBA, two out of three Raman active peaks are shifted to the blue side (except for the first peak), which is opposite for the case of I^- which shifts all peaks to the red side. It should be noted that the probability of finding OH^- ions near TBA is quite low, such that the spectra with OH- in a proximity of TBA are built off on a few snapshots and the observed shifts do not have high statistical certainty. Figure 9.7 shows the case of both positive and negative ions present within the hydration shell. In the case of the TBA-water-NaOH, the Raman active peaks are shifted to the red side but the sampling size is small. In the case of the TBA-water-NaI system, both Na^+ and I^- contribute to the red shifts of C-H frequencies, as shown in Figure 9.5 (b) and 9.6 (b). Hence, Figure 9.7 (b) shows red shifts, which are more pronounced than in the TBA-water-NaOH mixture. Interestingly, both NaOH and NaI cases show more pronounced red shifts when both ions are present within hydration shell than when only one species is present. In this regard, we conclude that all ions shift the spectra to the red side when found in the proximity of TBA molecules, but the relative amount of the red shift is governed by the propensity of both positive and negative ions to TBA.

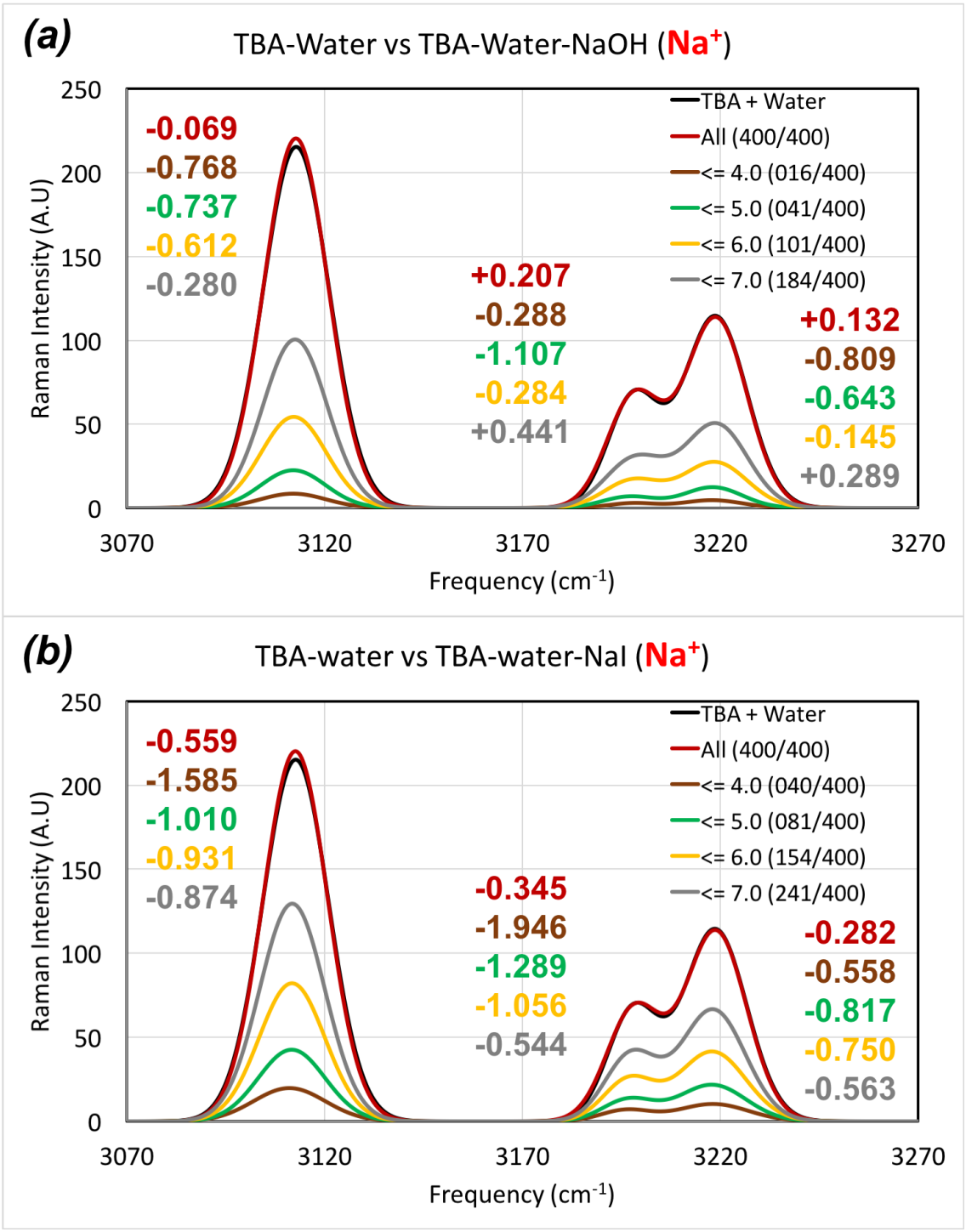

Figure 9.5. Modeled Raman spectra of (a) TBA-water-NaOH and (b) TBA-water-NaI systems. Decomposition of spectra based on the presence of Na^+ within 4.0, 5.0, 6.0 and 7.0 Angstroms from the central carbon of TBA. The number of corresponding snapshots is given in parenthesis. Frequency shifts are in cm^{-1}. The TBA-water spectrum is shown with black lines. The complete TBA-water-ion spectra are shown with red lines.

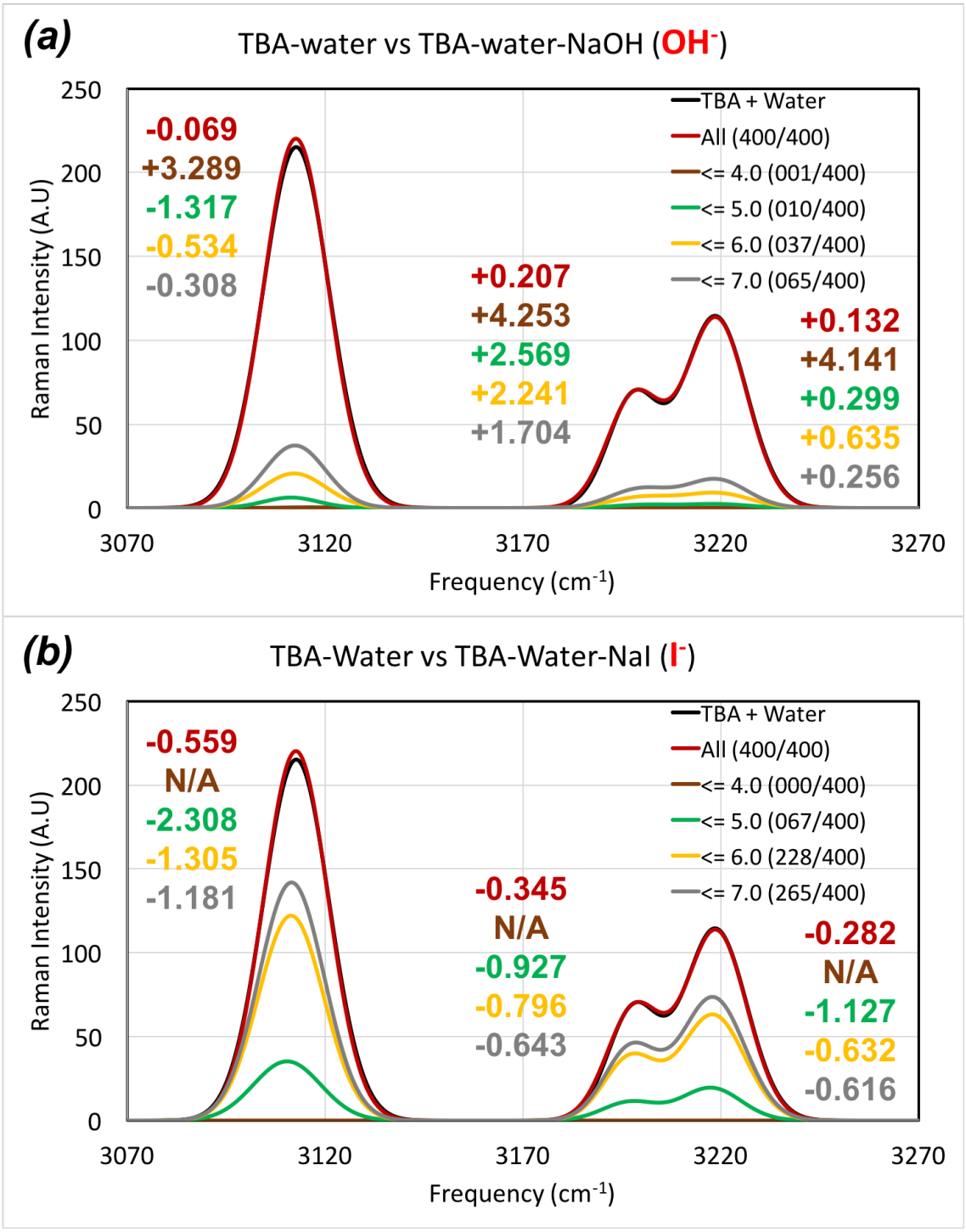

Figure 9.6. Modeled Raman spectra of (a) TBA-water-NaOH and (b) TBA-water-NaI systems. Decomposition of spectra based on the presence of negative ions (OH^- and I^-) within 4.0, 5.0, 6.0 and 7.0 Angstroms from the central carbon of TBA. The number of corresponding snapshots is given in parenthesis. Frequency shifts are in cm^{-1}. The TBA-water spectrum is shown with black lines. The complete TBA-water-ion spectra are shown with red lines.

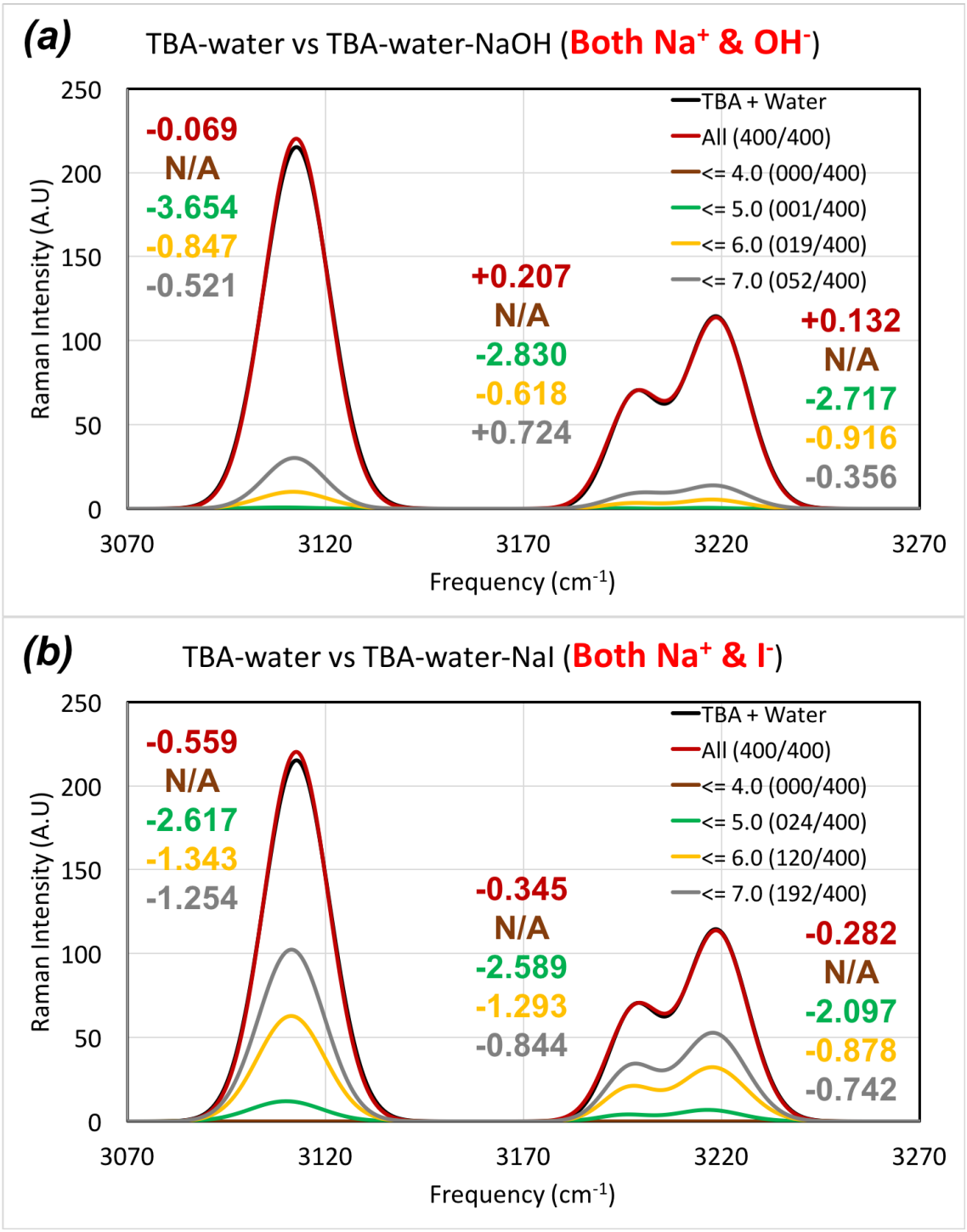

Figure 9.7. Modeled Raman spectra of (a) TBA-water-NaOH and (b) TBA-water-NaI systems. Decomposition of spectra based on the simultaneous presence of positive and negative ions (Na^+ and OH^- pair in (a), Na^+ and I^- pair in (b)) within 4.0, 5.0, 6.0 and 7.0 Angstroms from the central carbon of TBA. The number of corresponding snapshots is given in parenthesis. Frequency shifts are in cm^{-1}. The TBA-water spectrum is shown with black lines. The complete TBA-water-ion spectra are shown with red lines.

9.4 Conclusions

In this work, we have investigated the effect of ions on the C-H vibrational stretch frequencies of the TBA molecule. The simulation techniques used in this study are a combination of classical molecular dynamics and polarizable EFP simulations and the hybrid polarizable embedding QM/EFP model. Structural data from polarizable EFP MD simulations indicate that while OH^- ions are excluded from the TBA hydration shell, the local concentration of Na^+ and I^- ions near the TBA is nearly as large as the bulk ion concentrations. QM/EFP frequency calculations suggest that all ions, when found in a proximity of TBA, shift Raman-active frequencies to red. Because of the high concentration of Na^+ and I^- ions, the TBA molecules are strongly perturbed, resulting in noticeable red-shifted C-H stretch Raman frequencies relative to those in the aqueous TBA system. However, in the TBA-water-NaOH system, a depletion of ions from the first hydration shell of TBA results only in a small red shift, with magnitude within statistical uncertainty of our calculations. Overall, the presented results are in a qualitative agreement with experimental measurements of Raman spectra of aqueous solutions of TBA.

REFERENCES

[1] R. J. Bartlett, and M. Musiał, "Coupled-cluster theory in quantum chemistry," *Rev. Mod. Phys.,* vol. 79, no. 1, pp. 291-352, 2007.

[2] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "Reprint of: A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.,* vol. 589, pp. 37-40, 2013.

[3] M. E. Casida, and M. Huix-Rotllant, "Progress in Time-Dependent Density-Functional Theory," *Annu. Rev. Phys. Chem.,* vol. 63, no. 1, pp. 287-323, 2012.

[4] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Challenges for Density Functional Theory," *Chem. Rev.,* vol. 112, no. 1, pp. 289-320, 2012.

[5] J. W. Ponder, and D. A. Case, "Force Fields for Protein Simulations," *Adv. Protein Chem,.* vol. 66, pp. 27-85, 2003.

[6] N. Foloppe, and J. MacKerell, Alexander D., "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data," *J. Comput. Chem.,* vol. 21, no. 2, pp. 86-104, 2000.

[7] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, "The ONIOM Method and Its Applications," *Chem. Rev.,* vol. 115, no. 12, pp. 5678-5796, 2015.

[8] F. Maseras, and K. Morokuma, "IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states," *J. Comput. Chem.,* vol. 16, no. 9, pp. 1170-1179, 1995.

[9] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma, and M. J. Frisch, "Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM," *J. Chem. Theory Comput.,* vol. 2, no. 3, pp. 815-826, 2006.

[10] N. Dubinets, and L. V. Slipchenko, "Effective Fragment Potential Method for H-Bonding: How To Obtain Parameters for Nonrigid Fragments," *J. Phys. Chem. A,* vol. 121, no. 28, pp. 5301-5312, 2017.

[11] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry," *J. Phys. Chem. A,* vol. 105, no. 2, pp. 293-307, 2001.

[12] M. S. Gordon, L. V. Slipchenko, H. Li, and J. H. Jensen, "The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions," *Annu. Rep. comput. Chem*., vol 3, pp. 177-193, 2007.

[13] N. Gresh, G. A. Cisneros, T. A. Darden, and J.-P. Piquemal, "Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand−Macromolecule Complexes. A Bottom-Up Strategy," *J. Chem. Theory Comput.,* vol. 3, no. 6, pp. 1960-1986, 2007.

[14] P. Lykos, "Modeling the Hydrogen Bond within Molecular Dynamics," *J. Chem. Educ.,* vol. 81, no. 1, p. 147, 2004.

[15] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, "Current Status of the AMOEBA Polarizable Force Field," *J. Phys. Chem. B,* vol. 114, no. 8, pp. 2549-2564, 2010.

[16] K. Babu, and S. R. Gadre, "Ab initio quality one-electron properties of large molecules: Development and testing of molecular tailoring approach," *J. Comput. Chem.,* vol. 24, no. 4, pp. 484-495, 2003.

[17] S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, and R. Walsh, "Additivity rules for the estimation of thermochemical properties," *Chem. Rev.,* vol. 69, no. 3, pp. 279-324, 1969.

[18] M. A. Collins, and V. A. Deev, "Accuracy and efficiency of electronic energies from systematic molecular fragmentation," *J. Chem. Phys.,* vol. 125, no. 10, p. 104104, 2006.

[19] V. Deev, and M. A. Collins, "Approximate ab initio energies by systematic molecular fragmentation," *J. Chem. Phys.,* vol. 122, no. 15, p. 154102, 2005.

[20] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation Methods: A Route to Accurate Calculations on Large Systems," *Chem. Rev.,* vol. 112, no. 1, pp. 632-672, 2012.

[21] W. Li, Y. Li, R. Lin, and S. Li, "Generalized Energy-Based Fragmentation Approach for Localized Excited States of Large Systems," *J. Phys. Chem. A,* vol. 120, no. 48, pp. 9667-9677, 2016.

[22] D. W. Zhang, Y. Xiang, and J. Z. H. Zhang, "New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin−Biotin Interaction Energy," *J. Phys. Chem. B,* vol. 107, no. 44, pp. 12039-12041, 2003.

[23] Z. Bikadi, and E. Hazai, "Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock," *J Cheminformatics,* vol. 1, p.15, 2009.

[24] S. L. Dixon, and K. M. Merz, "Semiempirical molecular orbital calculations with linear system size scaling," *J. Chem. Phys.,* vol. 104, no. 17, pp. 6643-6649, 1996.

[25] S. L. Dixon, and K. M. Merz, "Fast, accurate semiempirical molecular orbital calculations for macromolecules," *J. Chem. Phys.,* vol. 107, no. 3, pp. 879-893, 1997.

[26] E. Nikitina, V. Sulimov, V. Zayets, and N. Zaitseva, “Semiempirical calculations of binding enthalpy for protein-ligand complexes,” *Int. J. Quantum Chem.,* vol. 97, no. 2, pp. 747-763, 2004.

[27] J. J. P. Stewart, “Application of the PM6 method to modeling the solid state,” *J. Mol. Model.,* vol. 14, no. 6, pp. 499-535, 2008.

[28] J. J. P. Stewart, “Application of the PM6 method to modeling proteins,” *J. Mol. Model.,* vol. 15, no. 7, pp. 765-805, 2009.

[29] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, “An effective fragment method for modeling solvent effects in quantum mechanical calculations,” *J. Chem. Phys.,* vol. 105, no. 5, pp. 1968-1986, 1996.

[30] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, “The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry,” *J. Phys. Chem. A,* vol. 105, no. 2, pp. 293-307, 2001.

[31] M. S. Gordon, L. V. Slipchenko, H. Li, and J. H. Jensen, "The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions," *Annu. Rep. comput. Chem*., vol 3, pp. 177-193, 2007.

[32] M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, “Accurate First Principles Model Potentials for Intermolecular Interactions,” *Annu. Rev. Phys. Chem.,* vol. 64, pp. 553-578, 2013.

[33] P. K. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. H. Shao, A. I. Krylov, and L. V. Slipchenko, “Extension of the Effective Fragment Potential Method to Macromolecules,” *J. Phys. Chem. B,* vol. 120, no. 27, pp. 6562-6574, 2016.

[34] L. V. Slipchenko, and P. K. Gurunathan, “Effective fragment potential method: Past, present, and future,” Wiley-Blackwell, 2017, ch. 6, pp. 183-208.

[35] L. A. Burns, M. S. Marshall, and C. D. Sherrill, “Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches,” *J. Chem. Phys.,* vol. 141, no. 23, p. 234111, 2014.

[36] B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz, and H. L. Williams, “SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies,” *Methods and Techniques in Computational Chemistry: METECC,* vol. 94, pp. 79-129, 1993.

[37] B. Jeziorski, R. Moszynski, and K. Szalewicz, “Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes,” *Chem. Rev.,* vol. 94, no. 7, pp. 1887-1930, 1994.

[38] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," *J. Chem. Phys.,* vol. 140, no. 9, p. 094106, 2014.

[39] D. D. Kemp, and M. S. Gordon, "An Interpretation of the Enhancement of the Water Dipole Moment Due to the Presence of Other Water Molecules," *J. Phys. Chem. A,* vol. 112, no. 22, pp. 4885-4894, 2008.

[40] D. D. Kemp, and M. S. Gordon, "Theoretical Study of the Solvation of Fluorine and Chlorine Anions by Water," *J. Phys. Chem. A,* vol. 109, no. 34, pp. 7688-7699, 2005.

[41] Q. A. Smith, M. S. Gordon, and L. V. Slipchenko, "Benzene−Pyridine Interactions Predicted by the Effective Fragment Potential Method," *J. Phys. Chem. A,* vol. 115, no. 18, pp. 4598-4609, 2011.

[42] M. D. Hands, and L. V. Slipchenko, "Intermolecular Interactions in Complex Liquids: Effective Fragment Potential Investigation of Water–tert-Butanol Mixtures," *J. Phys. Chem. B,* vol. 116, no. 9, pp. 2775-2786, 2012.

[43] B. M. Rankin, M. D. Hands, D. S. Wilcox, K. R. Fega, L. V. Slipchenko, and D. Ben-Amotz, "Interactions between halide anions and a molecular hydrophobic interface," *Faraday Discuss.,* vol. 160, no. 0, pp. 255-270, 2013.

[44] L. V. Slipchenko, and M. S. Gordon, "Water−Benzene Interactions: An Effective Fragment Potential and Correlated Quantum Chemistry Study," *J. Phys. Chem. A,* vol. 113, no. 10, pp. 2092-2102, 2009.

[45] D. Ghosh, D. Kosenkov, V. Vanovschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, and A. I. Krylov, "Noncovalent Interactions in Extended Systems Described by the Effective Fragment Potential Method: Theory and Application to Nucleobase Oligomers," *J. Phys. Chem. A,* vol. 114, no. 48, pp. 12739-12754, 2010.

[46] Q. A. Smith, M. S. Gordon, and L. V. Slipchenko, "Effective Fragment Potential Study of the Interaction of DNA Bases," *J. Phys. Chem. A,* vol. 115, no. 41, pp. 11269-11276, 2011.

[47] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, "Accurate Prediction of Noncovalent Interaction Energies with the Effective Fragment Potential Method: Comparison of Energy Components to Symmetry-Adapted Perturbation Theory for the S22 Test Set," *J. Chem. Theory Comput.,* vol. 8, no. 8, pp. 2835-2843, 2012.

[48] A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, "Modeling Solvent Effects on Electronic Excited States," *J. Phys. Chem. Lett.,* vol. 2, no. 17, pp. 2184-2192, 2011.

[49] D. Ghosh, O. Isayev, L. V. Slipchenko, and A. I. Krylov, "Effect of Solvation on the Vertical Ionization Energy of Thymine: From Microhydration to Bulk," *J. Phys. Chem. A,* vol. 115, no. 23, pp. 6028-6038, 2011.

[50] D. Kosenkov, and L. V. Slipchenko, “Solvent Effects on the Electronic Transitions of p-Nitroaniline: A QM/EFP Study,” *J. Phys. Chem. A,* vol. 115, no. 4, pp. 392-401, 2011.

[51] L. V. Slipchenko, “Solvation of the Excited States of Chromophores in Polarizable Environment: Orbital Relaxation versus Polarization,” *J. Phys. Chem. A,* vol. 114, no. 33, pp. 8824-8830, 2010.

[52] M. S. Gordon, and M. W. Schmidt. "Advances in electronic structure theory: GAMESS a decade later," In *Theory and Applications of Computational Chemistry*. Elsevier, 2005, pp. 1167-1189.

[53] H. Li, M. S. Gordon, and J. H. Jensen, “Charge transfer interaction in the effective fragment potential method,” *J. Chem. Phys.,* vol. 124, no. 21, p. 214108, 2006.

[54] H. Li, M. S. Gordon, and J. H. Jensen, “Charge transfer interaction in the effective fragment potential method,” *J. Chem. Phys.,* vol. 124, no. 21, pp. 214108, 2006.

[55] Y. Kim, D. Morozov, V. Stadnytskyi, S. Savikhin, and L. V. Slipchenko, “Predictive First-Principles Modeling of a Photosynthetic Antenna Protein: The Fenna-Matthews-Olson Complex,” *J. Phys. Chem. Lett.,* vol. 11, no. 5, pp. 1636-1643, 2020.

[56] L. V. Slipchenko, and M. S. Gordon, “Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer,” *J. Comput. Chem.,* vol. 28, no. 1, pp. 276-291, 2007.

[57] L. V. Slipchenko, and M. S. Gordon, “Damping functions in the effective fragment potential method,” *Mol. Phys.,* vol. 107, no. 8-12, pp. 999-1016, 2009.

[58] I. Adamovic, and M. S. Gordon, “Dynamic polarizability, dispersion coefficient C-6 and dispersion energy in the effective fragment potential method,” *Mol. Phys.,* vol. 103, no. 2-3, pp. 379-387, 2005.

[59] R. Amos, N. Handy, P. Knowles, J. Rice, and A. Stone, “AB-initio prediction of properties of carbon dioxide, ammonia, and carbon dioxide... ammonia,” *J. Phys. Chem.,* vol. 89, no. 11, pp. 2186-2192, 1985.

[60] L. V. Slipchenko, M. S. Gordon, and K. Ruedenberg, “Dispersion Interactions in QM/EFP,” *J. Phys. Chem. A,* vol. 121, no. 49, pp. 9495-9507, 2017.

[61] D. D. Kemp, J. M. Rintelman, M. S. Gordon, and J. H. Jensen, “Exchange repulsion between effective fragment potentials and ab initio molecules,” *Theor. Chem. Acc.,* vol. 125, no. 3-6, pp. 481-491, 2010.

[62] C. I. V. Rojas, J. Fine, and L. V. Slipchenko, “Exchange-repulsion energy in QM/EFP,” *J. Chem. Phys.,* vol. 149, no. 9, p. 094103, 2018.

[63] C. Bertoni, and M. S. Gordon, “Analytic Gradients for the Effective Fragment Molecular Orbital Method,” *J. Chem. Theory Comput.,* vol. 12, no. 10, pp. 4743-4767, 2016.

[64] S. R. Pruitt, C. Steinmann, J. H. Jensen, and M. S. Gordon, "Fully Integrated Effective Fragment Molecular Orbital Method," *J. Chem. Theory Comput.,* vol. 9, no. 5, pp. 2235-2249, 2013.

[65] C. Steinmann, D. G. Fedorov, and J. H. Jensen, "Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods," *J. Phys. Chem. A,* vol. 114, no. 33, pp. 8705-8712, 2010.

[66] T. Nagata, D. G. Fedorov, K. Kitaura, and M. S. Gordon, "A combined effective fragment potential–fragment molecular orbital method. I. The energy expression and initial applications," *J. Chem. Phys.,* vol. 131, no. 2, p. 024101, 2009.

[67] T. Nagata, D. G. Fedorov, T. Sawada, and K. Kitaura, "Analysis of Solute-Solvent Interactions in the Fragment Molecular Orbital Method Interfaced with Effective Fragment Potentials: Theory and Application to a Solvated Griffithsin-Carbohydrate Complex," *J. Phys. Chem. A,* vol. 116, no. 36, pp. 9088-9099, 2012.

[68] C. Steinmann, D. G. Fedorov, and J. H. Jensen, "Mapping Enzymatic Catalysis Using the Effective Fragment Molecular Orbital Method: Towards all ab initio Biochemistry," *Plos One,* vol. 8, no. 4, p. e60602, 2013.

[69] F. Zahariev, and M. S. Gordon, "Development of a combined quantum monte carlo-effective fragment molecular orbital method," *Mol. Phys.,* vol. 117, no. 9-12, pp. 1532-1540, 2019.

[70] C. A. Brautigam, B. S. Smith, Z. Ma, M. Palnitkar, D. R. Tomchick, M. Machius, and J. Deisenhofer, "Structure of the photolyase-like domain of cryptochrome 1 from Arabidopsis thaliana," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 101, no. 33, pp. 12142-12147, 2004.

[71] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, "Psi4: an open-source ab initio electronic structure program," *WIREs Computational Molecular Science,* vol. 2, no. 4, pp. 556-565, 2012.

[72] I. A. Kaliman, and L. V. Slipchenko, "LIBEFP: A new parallel implementation of the effective fragment potential method as a portable software library," *J. Comput. Chem.,* vol. 34, no. 26, pp. 2284-2292, 2013.

[73] I. A. Kaliman, and L. V. Slipchenko, "Hybrid MPI/OpenMP Parallelization of the Effective Fragment Potential Method in the libefp Software Library," *J. Comput. Chem.,* vol. 36, no. 2, pp. 129-135, 2015.

[74] R. N. Tazhigulov, P. K. Gurunathan, Y. Kim, L. V. Slipchenko, and K. B. Bravaya, "Polarizable embedding for simulating redox potentials of biomolecules," *Phys. Chem. Chem. Phys.,* vol. 21, no. 22, pp. 11642-11650, 2019.

[75] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," *Mol. Phys.,* vol. 113, no. 2, pp. 184-215, 2015.

[76] R. M. Parrish, D. F. Sitkoff, D. L. Cheney, and C. D. Sherrill, "The Surprising Importance of Peptide Bond Contacts in Drug–Protein Interactions," *Chemistry – A European Journal,* vol. 23, no. 33, pp. 7887-7890, 2017.

[77] G. S. Orf, and R. E. Blankenship, "Chlorosome antenna complexes from green photosynthetic bacteria," *Photosynth. Res.,* vol. 116, no. 2, pp. 315-331, 2013.

[78] D. L. Andrews, and A. A. Demidov. (1999). *Resonance energy transfer*: Wiley.

[79] T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming, "Two-dimensional spectroscopy of electronic couplings in photosynthesis," *Nature,* vol. 434, no. 7033, pp. 625-628, 2005.

[80] D. R. Buck, S. Savikhin, and W. S. Struve, "Ultrafast absorption difference spectra of the Fenna-Matthews-Olson protein at 19 K: Experiment and simulations," *Biophys. J.,* vol. 72, no. 1, pp. 24-36, 1997.

[81] D. R. Buck, S. Savikhin, and W. S. Struve, "Effect of Diagonal Energy Disorder on Circular Dichroism Spectra of Fenna−Matthews−Olson Trimers," *J. Phys. Chem. B,* vol. 101, no. 42, pp. 8395-8397, 1997.

[82] S. Chandrasekaran, M. Aghtar, S. Valleau, A. Aspuru-Guzik, and U. Kleinekathöfer, "Influence of Force Fields and Quantum Chemistry Approach on Spectral Densities of BChl a in Solution and in FMO Proteins," *J. Phys. Chem. B,* vol. 119, no. 31, pp. 9995-10004, 2015.

[83] A. F. Fidler, J. R. Caram, D. Hayes, and G. S. Engel, "Towards a coherent picture of excitonic coherence in the Fenna–Matthews–Olson complex," *J Phys B-At Mol Opt,* vol. 45, no. 15, p. 154013, 2012.

[84] K. A. Fransted, J. R. Caram, D. Hayes, and G. S. Engel, "Two-dimensional electronic spectroscopy of bacteriochlorophyll a in solution: Elucidating the coherence dynamics of the Fenna-Matthews-Olson complex using its chromophore as a control," *J. Chem. Phys.,* vol. 137, no. 12, p. 125101, 2012.

[85] J. Gao, W.-J. Shi, J. Ye, X. Wang, H. Hirao, and Y. Zhao, "QM/MM modeling of environmental effects on electronic transitions of the FMO complex," *J. Phys. Chem. B,* vol. 117, no. 13, pp. 3488-3495, 2013.

[86] M. Higashi, and S. Saito, "Quantitative Evaluation of Site Energies and Their Fluctuations of Pigments in the Fenna–Matthews–Olson Complex with an Efficient Method for Generating a Potential Energy Surface," *J. Chem. Theory Comput.,* vol. 12, no. 8, pp. 4128-4137, 2016.

[87] X. Jia, Y. Mei, J. Z. Zhang, and Y. Mo, "Hybrid QM/MM study of FMO complex with polarized protein-specific charge," *Sci. Rep.,* vol. 5, no. 1, pp. 1-10, 2015.

[88] S. Jurinovich, C. Curutchet, and B. Mennucci, "The Fenna–Matthews–Olson protein revisited: A fully polarizable (TD) DFT/MM description," *ChemPhysChem,* vol. 15, no. 15, pp. 3194-3204, 2014.

[89] A. Kell, R. E. Blankenship, and R. Jankowiak, "Effect of spectral density shapes on the excitonic structure and dynamics of the Fenna–Matthews–Olson trimer from Chlorobaculum tepidum," *J. Phys. Chem. A,* vol. 120, no. 31, pp. 6146-6154, 2016.

[90] A. Khmelnitskiy, T. Reinot, and R. Jankowiak, "Impact of Single-Point Mutations on the Excitonic Structure and Dynamics in a Fenna–Matthews–Olson Complex," *J. Phys. Chem. Lett.,* vol. 9, no. 12, pp. 3378-3386, 2018.

[91] S. Kihara, D. Hartzler, G. S. Orf, R. E. Blankenship, and S. Savikhin, "The Fate of the Triplet Excitations in the Fenna-Matthews-Olson Complex and Stability of the Complex," *Biophys. J.,* vol. 106, no. 2, p. 182a, 2014.

[92] C. W. Kim, B. Choi, and Y. M. Rhee, "Excited state energy fluctuations in the Fenna–Matthews–Olson complex from molecular dynamics simulations with interpolated chromophore potentials," *Phys. Chem. Chem. Phys.,* vol. 20, no. 5, pp. 3310-3319, 2018.

[93] N. H. List, C. Curutchet, S. Knecht, B. Mennucci, and J. Kongsted, "Toward reliable prediction of the energy ladder in multichromophoric systems: A benchmark study on the FMO light-harvesting complex," *J. Chem. Theory Comput.,* vol. 9, no. 11, pp. 4928-4938, 2013.

[94] R. Louwe, J. Vrieze, A. Hoff, and T. Aartsma, "Toward an integral interpretation of the optical steady-state spectra of the FMO-complex of Prosthecochloris aestuarii. 2. Exciton simulations," *J. Phys. Chem. B,* vol. 101, no. 51, pp. 11280-11287, 1997.

[95] B. W. Matthews, R. E. Fenna, M. Bolognesi, M. Schmid, and J. M. Olson, "Structure of a bacteriochlorophyll a-protein from the green photosynthetic bacterium Prosthecochloris aestuarii," *J. Mol. Biol.,* vol. 131, no. 2, pp. 259-285, 1979.

[96] M. T. Milder, B. Brüggemann, R. van Grondelle, and J. L. Herek, "Revisiting the optical properties of the FMO protein," *Photosynth. Res.,* vol. 104, no. 2-3, pp. 257-274, 2010.

[97] J. Moix, J. Wu, P. Huo, D. Coker, and J. Cao, "Efficient energy transfer in light-harvesting systems, III: The influence of the eighth bacteriochlorophyll on the dynamics and efficiency in FMO," *J. Phys. Chem. Lett.,* vol. 2, no. 24, pp. 3045-3052, 2011.

[98] C. Olbrich, T. L. Jansen, J. r. Liebers, M. Aghtar, J. Strümpfer, K. Schulten, J. Knoester, and U. Kleinekathöfer, "From atomistic modeling to excitation transfer and two-dimensional spectra of the FMO light-harvesting complex," *J. Phys. Chem. B,* vol. 115, no. 26, pp. 8609-8621, 2011.

[99] C. Olbrich, J. Strümpfer, K. Schulten, and U. Kleinekathöfer, "Quest for spatially correlated fluctuations in the FMO light-harvesting complex," *J. Phys. Chem. B,* vol. 115, no. 4, pp. 758-764, 2011.

[100] J. M. Olson. (2005). The FMO protein. In Govindjee, J. T. Beatty, H. Gest, and J. F. Allen (Eds.), *Discoveries in Photosynthesis* (pp. 421-427). Dordrecht: Springer Netherlands.

[101] D. Padula, M. H. Lee, K. Claridge, and A. Troisi, "Chromophore-Dependent Intramolecular Exciton–Vibrational Coupling in the FMO Complex: Quantification and Importance for Exciton Dynamics," *J. Phys. Chem. B,* vol. 121, no. 43, pp. 10026-10035, 2017.

[102] R. G. Saer, V. Stadnytskyi, N. C. Magdaong, C. Goodson, S. Savikhin, and R. E. Blankenship, "Probing the excitonic landscape of the Chlorobaculum tepidum Fenna-Matthews-Olson (FMO) complex: a mutagenesis approach," *Biochimica et Biophysica Acta (BBA)-Bioenergetics,* vol. 1858, no. 4, pp. 288-296, 2017.

[103] S. Saito, M. Higashi, and G. R. Fleming, "Site-Dependent Fluctuations Optimize Electronic Energy Transfer in the Fenna–Matthews–Olson Protein," *J. Phys. Chem. B,* vol. 123, no. 46, pp. 9762-9772, 2019.

[104] S. Savikhin, D. R. Buck, and W. S. Struve, "Toward Level-to-Level Energy Transfers in Photosynthesis: The Fenna−Matthews−Olson Protein," *J. Phys. Chem. B,* vol. 102, no. 29, pp. 5556-5565, 1998.

[105] S. Savikhin, and W. S. Struve, "Ultrafast energy transfer in FMO trimers from the green bacterium Chlorobium tepidum," *Biochemistry,* vol. 33, no. 37, pp. 11200-11208, 1994.

[106] S. Savikhin, and W. S. Struve, "Low-temperature energy transfer in FMO trimers from the green photosynthetic bacterium Chlorobium tepidum," *Photosynth. Res.,* vol. 48, no. 1-2, pp. 271-276, 1996.

[107] M. Schmidt am Busch, F. Müh, M. El-Amine Madjet, and T. Renger, "The eighth bacteriochlorophyll completes the excitation energy funnel in the FMO protein," *J. Phys. Chem. Lett.,* vol. 2, no. 2, pp. 93-98, 2011.

[108] J. Schulze, M. F. Shibl, M. J. Al-Marri, and O. Kühn, "Multi-layer multi-configuration time-dependent Hartree (ML-MCTDH) approach to the correlated exciton-vibrational dynamics in the FMO complex," *J. Chem. Phys.,* vol. 144, no. 18, p. 185101, 2016.

[109] D. E. Tronrud, J. Wen, L. Gay, and R. E. Blankenship, "The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria," *Photosynth. Res.,* vol. 100, no. 2, pp. 79-87, 2009.

[110] S. I. Vulto, M. A. de Baat, R. J. Louwe, H. P. Permentier, T. Neef, M. Miller, H. van Amerongen, and T. J. Aartsma, "Exciton simulations of optical spectra of the FMO complex from the green sulfur bacterium Chlorobium tepidum at 6 K," *J. Phys. Chem. B,* vol. 102, no. 47, pp. 9577-9582, 1998.

[111] S. I. Vulto, M. A. de Baat, S. Neerken, F. R. Nowak, H. van Amerongen, J. Amesz, and T. J. Aartsma, "Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: a kinetic model," *J. Phys. Chem. B,* vol. 103, no. 38, pp. 8153-8161, 1999.

[112] J. Wen, H. Zhang, M. L. Gross, and R. E. Blankenship, "Native electrospray mass spectrometry reveals the nature and stoichiometry of pigments in the FMO photosynthetic antenna protein," *Biochemistry,* vol. 50, no. 17, pp. 3502-3511, 2011.

[113] S. Savikhin, D. R. Buck, and W. S. Struve, "Oscillating anisotropies in a bacteriochlorophyll protein: Evidence for quantum beating between exciton levels," *Chem. Phys.,* vol. 223, no. 2, pp. 303-312, 1997.

[114] M. A. Allodi, J. P. Otto, S. H. Sohail, R. G. Saer, R. E. Wood, B. S. Rolczynski, S. C. Massey, P.-C. Ting, R. E. Blankenship, and G. S. Engel, "Redox Conditions Affect Ultrafast Exciton Transport in Photosynthetic Pigment–Protein Complexes," *J. Phys. Chem. Lett.,* vol. 9, no. 1, pp. 89-95, 2018.

[115] G. S. Engel, "Quantum coherence in photosynthesis," *Procedia Chem.,* vol. 3, no. 1, pp. 222-231, 2011.

[116] J. Adolphs, F. Müh, M. E.-A. Madjet, and T. Renger, "Calculation of pigment transition energies in the FMO protein," *Photosynth. Res.,* vol. 95, no. 2, p. 197, 2007.

[117] C. Curutchet, and B. Mennucci, "Quantum Chemical Studies of Light Harvesting," *Chem. Rev.,* vol. 117, no. 2, pp. 294-343, 2017.

[118] G. Groenhof, V. Modi, and D. Morozov, "Observe while it happens: catching photoactive proteins in the act with non-adiabatic molecular dynamics simulations," *Curr Opin Struc Biol,* vol. 61, pp. 106-112, 2020.

[119] D. S. Kaliakin, H. Nakata, Y. Kim, Q. Chen, D. G. Fedorov, and L. V. Slipchenko, "FMOxFMO: Elucidating Excitonic Interactions in the Fenna–Matthews–Olson Complex with the Fragment Molecular Orbital Method," *J. Chem. Theory Comput.,* vol. 16, no. 2, pp. 1175-1187, 2020.

[120] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *J. Mol. Graphics Modell.,* vol. 25, no. 2, pp. 247-260, 2006.

[121] M. Ceccarelli, P. Procacci, and M. Marchi, "An ab initio force field for the cofactors of bacterial photosynthesis," *J. Comput. Chem.,* vol. 24, no. 2, pp. 129-142, 2003.

[122] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjær, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. Liang, A. V. Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjöhamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. H. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman, and R. Neutze, "Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser," *Nature Methods,* vol. 11, no. 9, pp. 923-926, 2014.

[123] A. B. Wöhri, G. Katona, L. C. Johansson, E. Fritz, E. Malmerberg, M. Andersson, J. Vincent, M. Eklund, M. Cammarata, M. Wulff, J. Davidsson, G. Groenhof, and R. Neutze, "Light-Induced Structural Changes in a Photosynthetic Reaction Center Caught by Laue Diffraction," *Science,* vol. 328, no. 5978, p. 630, 2010.

[124] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *J. Comput. Chem.,* vol. 24, no. 16, pp. 1999-2012, 2003.

[125] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.,* vol. 79, no. 2, pp. 926-935, 1983.

[126] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.,* vol. 25, no. 9, pp. 1157-1174, 2004.

[127] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.,* vol. 126, no. 1, p. 014101, 2007.

[128] M. Parrinello, and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.,* vol. 52, no. 12, pp. 7182-7190, 1981.

[129] A. D. Mackerell Jr., M. Feig, and C. L. Brooks III, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations," *J. Comput. Chem.,* vol. 25, no. 11, pp. 1400-1415, 2004.

[130] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems," *J. Chem. Phys.,* vol. 98, no. 12, pp. 10089-10092, 1993.

[131] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.,* vol. 103, no. 19, pp. 8577-8593, 1995.

[132] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX,* vol. 1-2, pp. 19-25, 2015.

[133] H. Lin, and D. G. Truhlar, "QM/MM: what have we learned, where are we, and where do we go from here?," *Theor. Chem. Acc.,* vol. 117, no. 2, p. 185, 2006.

[134] C. Adamo, and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," *J. Chem. Phys.,* vol. 110, no. 13, pp. 6158-6170, 1999.

[135] R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules," *J. Chem. Phys.,* vol. 54, no. 2, pp. 724-728, 1971.

[136] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, "Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements," *J. Chem. Phys.,* vol. 77, no. 7, pp. 3654-3665, 1982.

[137] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, "Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements," *J. Am. Chem. Soc.,* vol. 104, no. 10, pp. 2797-2803, 1982.

[138] P. C. Hariharan, and J. A. Pople, "The influence of polarization functions on molecular orbital hydrogenation energies," *Theor. Chim. Acta,* vol. 28, no. 3, pp. 213-222, 1973.

[139] W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules," *J. Chem. Phys.,* vol. 56, no. 5, pp. 2257-2261, 1972.

[140] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery Jr, "General atomic and molecular electronic structure system," *J. Comput. Chem.,* vol. 14, no. 11, pp. 1347-1363, 1993.

[141] C. I. V. Rojas, J. Fine, and L. V. Slipchenko, "Exchange-repulsion energy in QM/EFP," *J. Chem. Phys.,* vol. 149, no. 9, p. 094103, 2018.

[142] H. M. Senn, and W. Thiel, "QM/MM Methods for Biomolecular Systems," *Angew. Chem. Int. Ed.,* vol. 48, no. 7, pp. 1198-1229, 2009.

[143] V. Kairys, and J. H. Jensen, "QM/MM Boundaries Across Covalent Bonds: A Frozen Localized Molecular Orbital-Based Approach for the Effective Fragment Potential Method," *J. Phys. Chem. A,* vol. 104, no. 28, pp. 6656-6665, 2000.

[144] M. E. Madjet, A. Abdurahman, and T. Renger, "Intermolecular Coulomb Couplings from Ab Initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers," *J. Phys. Chem. B,* vol. 110, no. 34, pp. 17268-17281, 2006.

[145] R. S. Knox, "Dipole and Oscillator Strengths of Chromophores in Solution," *Photochem. Photobiol.,* vol. 77, no. 5, pp. 492-496, 2003.

[146] R. S. Knox, and B. Q. Spring, "Dipole Strengths in the Chlorophylls," *Photochem. Photobiol.,* vol. 77, no. 5, pp. 497-501, 2003.

[147] F. Segatta, L. Cupellini, M. Garavelli, and B. Mennucci, "Quantum Chemical Modeling of the Photoinduced Activity of Multichromophoric Biosystems," *Chem. Rev.,* vol. 119, no. 16, pp. 9361-9380, 2019.

[148] A. Sisto, D. R. Glowacki, and T. J. Martinez, "Ab Initio Nonadiabatic Dynamics of Multichromophore Complexes: A Scalable Graphical-Processing-Unit-Accelerated Exciton Framework," *Acc. Chem. Res.,* vol. 47, no. 9, pp. 2857-2866, 2014.

[149] A. Ben-Shem, F. Frolow, and N. Nelson, "Evolution of photosystem I – from symmetry through pseudosymmetry to asymmetry," *FEBS Lett.,* vol. 564, no. 3, pp. 274-280, 2004.

[150] I. Adamovic, M. A. Freitag, and M. S. Gordon, "Density functional theory based effective fragment potential method," *J. Chem. Phys.,* vol. 118, no. 15, pp. 6725-6732, 2003.

[151] D. Ghosh, D. Kosenkov, V. Vanovschi, J. Flick, I. Kaliman, Y. Shao, A. T. B. Gilbert, A. I. Krylov, and L. V. Slipchenko, "Effective fragment potential method in Q-CHEM: A guide for users and developers," *J. Comput. Chem.,* vol. 34, no. 12, pp. 1060-1070, 2013.

[152] H. Li, A. W. Hains, J. E. Everts, A. D. Robertson, and J. H. Jensen, "The Prediction of Protein pKa's Using QM/MM: The pKa of Lysine 55 in Turkey Ovomucoid Third Domain," *J. Phys. Chem. B,* vol. 106, no. 13, pp. 3486-3494, 2002.

[153] A. V. Nemukhin, B. L. Grigorenko, A. V. Bochenkova, I. A. Topol, and S. K. Burt, "A QM/MM approach with effective fragment potentials applied to the dipeptide–water structures," *J MOL STRUC-THEOCHEM,* vol. 581, no. 1, pp. 167-175, 2002.

[154] I. S. Ufimtsev, and T. J. Martínez, "Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation," *J. Chem. Theory Comput.,* vol. 4, no. 2, pp. 222-231, 2008.

[155] I. S. Ufimtsev, and T. J. Martinez, "Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation," *J. Chem. Theory Comput.,* vol. 5, no. 4, pp. 1004-1015, 2009.

[156] I. S. Ufimtsev, and T. J. Martinez, "Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics," *J. Chem. Theory Comput.,* vol. 5, no. 10, pp. 2619-2628, 2009.

[157] J. B. Arellano, T. Bernt Melø, C. M. Borrego, J. Garcia-Gil, and K. R. Naqvi, "Nanosecond Laser Photolysis Studies of Chlorosomes and Artificial Aggregates Containing Bacteriochlorophyll e: Evidence for the Proximity of Carotenoids and Bacteriochlorophyll a in Chlorosomes from Chlorobium phaeobacteroides strain CL1401," *Photochem. Photobiol.,* vol. 72, no. 5, pp. 669-675, 2000.

[158] S. Kihara, D. A. Hartzler, G. S. Orf, R. E. Blankenship, and S. Savikhin, "The Fate of the Triplet Excitations in the Fenna–Matthews–Olson Complex," *J. Phys. Chem. B,* vol. 119, no. 18, pp. 5765-5772, 2015.

[159] S. Kihara, D. A. Hartzler, and S. Savikhin, "Oxygen Concentration Inside a Functioning Photosynthetic Cell," *Biophys. J.,* vol. 106, no. 9, pp. 1882-1889, 2014.

[160] H. Kim, H. Li, J. A. Maresca, D. A. Bryant, and S. Savikhin, "Triplet Exciton Formation as a Novel Photoprotection Mechanism in Chlorosomes of Chlorobium tepidum," *Biophys. J.,* vol. 93, no. 1, pp. 192-201, 2007.

[161] A. Krieger-Liszkay, "Singlet oxygen production in photosynthesis," *J. Exp. Bot.,* vol. 56, no. 411, pp. 337-346, 2004.

[162] A. Michaeli, and J. Feitelson, "REACTIVITY OF SINGLET OXYGEN TOWARD AMINO ACIDS AND PEPTIDES," *Photochem. Photobiol.,* vol. 59, no. 3, pp. 284-289, 1994.

[163] A. Telfer, "Too much light? How β-carotene protects the photosystem II reaction centre," *Photoch Photobio Sci,* vol. 4, no. 12, pp. 950-956, 2005.

[164] E. C. Ziegelhoffer, and T. J. Donohue, "Bacterial responses to photo-oxidative stress," *Nat. Rev. Microbiol.,* vol. 7, no. 12, pp. 856-863, 2009.

[165] G. S. Orf, D. M. Niedzwiedzki, and R. E. Blankenship, "Intensity Dependence of the Excited State Lifetimes and Triplet Conversion Yield in the Fenna–Matthews–Olson Antenna Protein," *J. Phys. Chem. B,* vol. 118, no. 8, pp. 2058-2069, 2014.

[166] F. Wilkinson, W. P. Helman, and A. B. Ross, "Quantum Yields for the Photosensitized Formation of the Lowest Electronically Excited Singlet State of Molecular Oxygen in Solution," *J. Phys. Chem. Ref. Data,* vol. 22, no. 1, pp. 113-262, 1993.

[167] D. A. Hartzler, D. M. Niedzwiedzki, D. A. Bryant, R. E. Blankenship, Y. Pushkar, and S. Savikhin, "Triplet Excited State Energies and Phosphorescence Spectra of (Bacterio)Chlorophylls," *J. Phys. Chem. B,* vol. 118, no. 26, pp. 7221-7232, 2014.

[168] S. Jockusch, N. J. Turro, E. K. Thompson, M. Gouterman, J. B. Callis, and G. E. Khalil, "Singlet molecular oxygen by direct excitation," *Photoch Photobio Sci,* vol. 7, no. 2, pp. 235-239, 2008.

[169] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell Jr., "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J. Comput. Chem.,* vol. 31, no. 4, pp. 671-690, 2010.

[170] T. Brixner, R. Hildner, J. Köhler, C. Lambert, and F. Würthner, "Exciton Transport in Molecular Aggregates – From Natural Antennas to Synthetic Chromophore Systems," *Adv. Energy Mater.,* vol. 7, no. 16, p. 1700236, 2017.

[171] M. J. Llansola-Portoles, D. Gust, T. A. Moore, and A. L. Moore, "Artificial photosynthetic antennas and reaction centers," *Comptes Rendus Chimie,* vol. 20, no. 3, pp. 296-313, 2017.

[172] E. Romero, V. I. Novoderezhkin, and R. van Grondelle, "Quantum design of photosynthesis for bio-inspired solar-energy conversion," *Nature,* vol. 543, no. 7645, pp. 355-365, 2017.

[173] G. S. Schlau-Cohen, "Principles of light harvesting from single photosynthetic complexes," *Interface Focus,* vol. 5, no. 3, p. 20140088, 2015.

[174] H. Zhou, J. Xu, X. Liu, H. Zhang, D. Wang, Z. Chen, D. Zhang, and T. Fan, "Bio-Inspired Photonic Materials: Prototypes and Structural Effect Designs for Applications in Solar Energy Manipulation," *Adv. Funct. Mater.,* vol. 28, no. 24, p. 1705309, 2018.

[175] D. A. Hartzler, L. V. Slipchenko, and S. Savikhin, "Triplet–Triplet Coupling in Chromophore Dimers: Theory and Experiment," *J. Phys. Chem. A,* vol. 122, no. 33, pp. 6713-6723, 2018.

[176] C.-P. Hsu, Z.-Q. You, and H.-C. Chen, "Characterization of the Short-Range Couplings in Excitation Energy Transfer," *J. Phys. Chem. C.,* vol. 112, no. 4, pp. 1204-1212, 2008.

[177] Z.-Q. You, and C.-P. Hsu, "The fragment spin difference scheme for triplet-triplet energy transfer coupling," *J. Chem. Phys.,* vol. 133, no. 7, p. 074105, 2010.

[178] Z.-Q. You, C.-P. Hsu, and G. R. Fleming, "Triplet-triplet energy-transfer coupling: Theory and calculation," *J. Chem. Phys.,* vol. 124, no. 4, p. 044506, 2006.

[179] Z.-Q. You, and C.-P. Hsu, "Ab Inito Study on Triplet Excitation Energy Transfer in Photosynthetic Light-Harvesting Complexes," *J. Phys. Chem. A,* vol. 115, no. 16, pp. 4092-4100, 2011.

[180] A. J. Stone, "Distributed multipole analysis, or how to describe a molecular charge distribution," *Chem. Phys. Lett.,* vol. 83, no. 2, pp. 233-239, 1981.

[181] A. J. Stone, "Distributed Multipole Analysis: Stability for Large Basis Sets," *J. Chem. Theory Comput.,* vol. 1, no. 6, pp. 1128-1132, 2005.

[182] A. J. Stone, and M. Alderton, "Distributed multipole analysis," *Mol. Phys.,* vol. 56, no. 5, pp. 1047-1064, 1985.

[183] R. E. Fenna, and B. W. Matthews, "Chlorophyll arrangement in a bacteriochlorophyll protein from Chlorobium limicola," *Nature,* vol. 258, no. 5536, pp. 573-577, 1975.

[184] R. Saer, G. S. Orf, X. Lu, H. Zhang, M. J. Cuneo, D. A. A. Myles, and R. E. Blankenship, "Perturbation of bacteriochlorophyll molecules in Fenna–Matthews–Olson protein complexes through mutagenesis of cysteine residues," *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* vol. 1857, no. 9, pp. 1455-1463, 2016.

[185] M. Cho, H. M. Vaswani, T. Brixner, J. Stenger, and G. R. Fleming, "Exciton Analysis in 2D Electronic Spectroscopy," *J. Phys. Chem. B,* vol. 109, no. 21, pp. 10542-10556, 2005.

[186] A. D. MacKerell Jr., and N. K. Banavali, "All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution," *J. Comput. Chem.,* vol. 21, no. 2, pp. 105-120, 2000.

[187] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.,* vol. 118, no. 45, pp. 11225-11236, 1996.

[188] W. L. Jorgensen, and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *J. Am. Chem. Soc.,* vol. 110, no. 6, pp. 1657-1666, 1988.

[189] J. D. Head, "Computation of vibrational frequencies for adsorbates on surfaces," *Int. J. Quantum Chem.,* vol. 65, no. 5, pp. 827-838, 1997.

[190] H. Li, and J. H. Jensen, "Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy," *Theor. Chem. Acc.,* vol. 107, no. 4, pp. 211-219, 2002.