Multiscale Simulations of Excited States and Redox Properties of Biomolecular Systems

A Dissertation submitted to the Doctoral Committee
of the Faculty of Chemistry at Universität Duisburg-Essen
by

Sinjini Bhattacharjee

In partial fulfilment of the requirements for the Degree of
Doctor in Natural Sciences (***Dr. rer. nat.***)

MAX-PLANCK-INSTITUT
FÜR KOHLENFORSCHUNG

RECHARGE

UNIVERSITÄT
DUISBURG
ESSEN
Offen im Denken

Mülheim an der Ruhr, Germany
August 2024

“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less…” –Marie Curie

DOI: 10.17185/duepublico/82566
URN: urn:nbn:de:hbz:465-20241204-113415-3

The research work in my PhD thesis was carried out from October 2020 to August 2024 in the Department of Molecular Theory and Spectroscopy at the Max-Planck-Institut für Kohlenforschung (Mülheim an der Ruhr, Germany) under the supervision of Prof. Dr. Frank Neese and Dr. Dimitrios A. Pantazis.

This dissertation is approved by the following members of the Final Doctoral Committee:

1st Reviewer: Prof. Dr. Frank Neese, Max-Planck-Institut für Kohlenforschung, Germany.
2nd Reviewer: Dr. Dimitrios A. Pantazis, Max-Planck-Institut für Kohlenforschung, Germany.

Chair of Examination: Prof. Dr. Eckhard Spohr, Universität Duisburg-Essen, Germany.

Date of Doctoral Examination: 29/10/2024

Acknowledgement

I extend my heartfelt thanks to everyone, who has played a vital role in supporting me throughout my PhD journey over the past four years. First and foremost, I am deeply grateful to my supervisors, Prof. Dr. Frank Neese and Dr. Dimitrios A. Pantazis, for giving me the opportunity to work in their research groups at the Max-Planck-Institut für Kohlenforschung in Mülheim, Germany.

I would especially like to acknowledge Dr. Pantazis for his exemplary leadership and unwavering guidance throughout my PhD. His mentorship fueled my passion for exploring various biomolecular systems and allowed me to gain expertise in multiple aspects of Computational Chemistry, Biophysics, and Photosynthesis. His exceptional support has helped me develop independent thinking skills, be innovative in research and acquire experience in several aspects of scientific writing and presentation, which will remain invaluable assets to my future academic career. I am thankful for the leadership of Prof. Neese, Dr. Frank Wennmohs, and the entire ORCA development team for providing me with a positive work environment and all the necessary computational resources that enabled me to conduct my research and delve into the world of quantum chemistry and spectroscopy. I owe particular thanks to Dr. Christoph Riplinger at FAccTs for his constant technical support in multiscale modeling within ORCA.

I would like to specially thank my former co-worker, mentor and collaborator Dr. Abhishek Sirohiwal, for all our productive discussions on multiscale modelling of membrane proteins and photosynthesis. I also thank my former colleagues Dr. Kalishankar Bhattacharyya, Dr. Soumen Ghosh, Dr. Georgi Stoychev and Dr. Ahmet Altun, with whom I engaged in many insightful project discussions. I thank my fellow group members particularly Dr. Mursaleem Ansari and Dr. Maria Drosou, with whom I co-authored two publications during my PhD. I acknowledge my external collaborators, Prof. Dr. Isabella Daidone, Dr. Srilatha Arra, Dr. Miquel Garcia-Ratés, Adam Šrut, Prof. Dr. Vera Krewald, Igor Gordiy, Olivia McCubbin Stepanic and Prof. Serena DeBeer for the exciting research topics we explored together. A special thank you to Penelope Pesara, Markella Mermigki, Dr. Daniel SantaLucia, Dr. Shalini Yadav and Dr. Maria Drosou for their assistance in proofreading this thesis.

I gratefully acknowledge the International Max Planck Research School on Reactive Structure Analysis for Chemical Reactions (IMPRS-RECHARGE) for financial support and development of scientific skills throughout my PhD.

Beginning a PhD during the COVID-19 pandemic came with its unique challenges, and I am particularly thankful to everyone, who stood by me through those difficult times. I owe a special debt of gratitude to Dan for being the most loving and supportive friend. I specially thank my friends Srija Tewari, Haritha Cheraparambil, Dr. Rajat Maji, Dr. Atanu Rana, Dr. Sarmistha Bhunia, Dr. Kushal Sengupta, and members of the Neese and DeBeer departments for all the wonderful memories we shared in Mülheim an der Ruhr.

My deepest thanks go to my family in India, especially my parents and grandparents, whose encouragement has been a constant source of motivation. Finally, I dedicate this thesis to Rabbie and my parents, Ashoke Bhattacharjee and Chaiti Bhattacharjee, for their unconditional love, support, and encouragement, which have helped me grow into an independent, free-thinking person.

Abstract

Modelling biological catalysis is challenging for contemporary computational methods, requiring accurate characterization of chemical properties within complex environments. Recent advancements in hybrid multiscale approaches offer a balance of chemical accuracy and computational efficiency, making them invaluable tools for studying protein complexes. This thesis aims to establish an atomistic and electronic level understanding of biomolecular systems using multiscale quantum mechanics/molecular mechanics (QM/MM) approaches, integrating accurate QM methods with classical molecular dynamics (MD) simulations.

The Introduction provides an overview of QM-based multiscale modelling methods and outlines the structure and functional principles of the Photosystem II (PSII) protein complex. We first illustrate the multilayer approach using the domain-based pair natural orbital implementation of coupled cluster theory on simple QM cluster models. This method is employed to estimate redox potentials and solvation free energies of hydrated transition metal ions at the coupled-cluster level, maintaining high-level treatment of the metal complex and explicitly coordinated water molecules. This protocol represents a promising route for applying accurate QM methods to complex open-shell systems.

The core of this thesis explores mechanisms relevant to excitation energy transfer (EET), primary charge separation and photoprotection in natural photosynthesis. Photosynthesis encompasses photochemical light harvesting by chlorophyll pigments, water splitting to release molecular oxygen, and the fixation of carbon dioxide into organic compounds. The initial processes involve intricate mechanisms of light harvesting and charge separation in the PSII pigment-protein complex. The harvested light energy from the core antennae initiates electron transfer at the reaction center (RC). Photosynthetic organisms also adapt to oxidative stress and environmental changes through photoprotective strategies like energy dissipation, quenching and cyclic electron flow in PSII. Despite extensive studies on primary charge separation processes in the RC, the mechanisms of EET, quenching of chlorophyll triplets, and effects of gene expression on primary processes are not fully deciphered. Theoretical and computational analysis of these systems require investigating chemical reactions and their coupling to molecular processes across various scales. Accurately incorporating solvent and protein environments is crucial, and describing spectroscopic and redox properties requires high-level electronic structure methods. In this thesis, we describe the complete low-energy excitation spectrum, including singlet-triplet excitations and charge transfer (CT) states, using long-range corrected time-dependent density functional theory (TD-DFT) and QM/MM. We also calculate electron paramagnetic resonance (EPR) properties of chlorophyll triplet states and describe the electrostatic modulation of excited state energetics of RC pigments by specific redox-active cofactors. Furthermore, we investigate how three genetic variants of a crucial RC protein differentially tune the optical and redox properties of the pigments involved in the primary processes of photosynthesis. We focus on the critical $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair, identifying residues responsible for specific matrix-induced adjustments of CT states and redox potentials. Finally, we investigate the low-energy excitation manifold of the CP43 core antenna, utilizing TD-DFT/MM combined with large-scale perturbed matrix method (PMM) calculations. This work provides a refined basis for interpreting spectroscopic observations and understanding EET within the PSII core complex.

Overall, the work in this thesis highlights two major themes: (i) methodological aspects for QM based multiscale modelling of proteins, and (ii) the role of solvation and protein environment in obtaining biologically relevant properties, aiming to refine our fundamental understanding of oxygenic photosynthesis.

Publications

A total of six research projects were undertaken in the framework of this PhD thesis. Five research projects are already published in peer-reviewed journals, while one manuscript is in preparation at the time of submission of this thesis.

1. **Bhattacharjee, S.**; Isegawa, M.; Garcia-Ratés, M.; Neese, F.; Pantazis, D. A., Ionization Energies and Redox Potentials of Hydrated Transition Metal Ions: Evaluation of Domain-Based Local Pair Natural Orbital Coupled Cluster Approaches. *J. Chem. Theory Comput.* **2022,** *18*, 1619-1632. https://doi.org/10.1021/acs.jctc.1c01267

2. **Bhattacharjee, S.**; Neese, F.; Pantazis, D. A., Triplet states in the reaction center of Photosystem II. *Chem. Sci.* **2023,** *14*, 9503-9516. https://doi.org/10.1039/D3SC02985A

3. **Bhattacharjee, S.**; Arra, S.; Daidone, I.; Pantazis, D. A., Excitation landscape of the CP43 photosynthetic antenna complex from multiscale simulations. *Chem. Sci.* **2024,** *15*, 7269-7284. https://doi.org/10.1039/D3SC06714A

4. Drosou, M.; **Bhattacharjee, S.**; Pantazis, D. A., A Combined Multireference–Multiscale Approach to the Description of Photosynthetic Reaction Centers, *J. Chem. Theory Comput.* **2024**. https://doi.org/10.1021/acs.jctc.4c00578

5. **Bhattacharjee, S.**; Gordiy, I.; Sirohiwal, A.; Pantazis, D. A., Microscopic Basis of Reaction Center Modulation in PsbA Variants of Photosystem II, **2024**, *submitted*.

6. Ansari, M.; **Bhattacharjee, S.**; Pantazis, D. A., Correlating Structure with Spectroscopy in Ascorbate Peroxidase Compound II. *J. Am. Chem. Soc.* **2024,** *146*, 9640-9656. https://doi.org/10.1021/jacs.3c13169

List of Abbreviations

2DEV: Two-Dimensional Electronic-Vibration

AMBER: Assisted Model Building and Energy Refinement

CASPT2: Complete Active Space Second Order Perturbation Theory

CASSCF: Complete Active Space Self-Consistent Field

CCSD(T): Coupled Cluster with Singles, Doubles and perturbative Triples

CHARMM: Chemistry at Harvard Macromolecular Mechanics

COSMO-RS: Conductor-like Screening Model for Realistic Solvents

CPCM: Conductor-like Polarizable Continuum Model

CP-SCF: Coupled-Perturbed Self-Consistent Field

DFT: Density Functional Theory

DFTB: Density Functional Tight Binding

DLPNO: Domain-Based Local Pair Natural Orbital

DKH: Douglas-Kroll-Hess

EET: Excitation Energy Transfer

ENDOR: Electron–Nucleus Double Resonance

EPR: Electron Paramagnetic Resonance

EVB: Empirical Valence Bond

FARLiP: Far-Red Light Photo Acclimation

FMO: Fenna-Matthews-Olson

FT-IR: Fourier-Transformed InfraRed

GAFF/GAFF2: General Amber Force Field

GIAO: Gauge Independent Atomic Orbitals

GROMOS: Groningen Molecular Simulation System

HFC: Hyperfine Coupling

HOMO: Highest Occupied Molecular Orbital

LED: Local Energy Decomposition

LHC: Light Harvesting Complex

LUMO: Lowest Unoccupied Molecular Orbital

MD: Molecular Dynamics

MK-RESP: Merz-Kollman Restrained Electrostatic Potential

MMPBSA: Molecular Mechanics Poisson-Boltzmann Surface Area

NHI: Non-Heme Iron

NEVPT2: N-Electron Valence State Second Order Perturbation Theory

NMR: Nuclear Magnetic Resonance

NTO: Natural Transition Orbital

ODMR: Optically Detected Magnetic Resonance

OEC: Oxygen Evolving Complex

ONIOM: Own N-layer Integrated molecular Orbital Molecular mechanics

OPM: Orientation of Protein in Membranes

PDA: Point Dipole Approximation

PDB: Protein Data Bank

PCM: Polarizable Continuum Model

PES: Potential Energy Surface

PME: Particle Mesh Ewald

PMM: Perturbed Matrix Method

POPC: 1-Palmitoyl-2-Oleoyl-sn-glycero-3-Phosphocholine

PPC: Pigment Protein Complex

PSII-cc: Photosystem II Core Complex

PT/ET: Proton Transfer/Electron Transfer

PTE(S): Perturbation Theory Energy scheme with Singles

QM/MM: Quantum Mechanics/Molecular Mechanics

QRO: Quasi-Restricted Orbitals

RMSD: Root Mean Square Deviation

SHE: Standard Hydrogen Electrode

SMD: Solvation Model based on Density

SOMF: Spin–Orbit Mean-Field

SOMO: Singly Occupied Molecular Orbital

STEOM-CCSD: Similarity Transformed Equation Of Motion Coupled Cluster Theory

TDA: Tamm–Dancoff Approximation

TD-DFT: Time-Dependent Density Functional Theory

TIP3P: Transferable Intermolecular Potential with Three Points

T-TET: Triplet-Triplet Energy Transfer

VEE: Vertical Excitation Energies

VMD: Visual Molecular Dynamics

XANES: X-ray Absorption Near Edge Structure

XRD: X-Ray Diffraction

Table of Contents

Acknowledgement ii

Abstract iv

Publications vi

List of Abbreviations vii

Chapter 1 Introduction 1

1.1 Multiscale Modelling of Biomolecular Systems 1

- 1.1.1 QM Cluster Models 1
- 1.1.2 QM/QM Multilevel Approaches 2

1.2 Hybrid QM/MM Methods 3

- 1.2.1 MM Force Fields 3
- 1.2.2 Evaluating the QM/MM Energy 4
- 1.2.3 QM/MM Interaction Schemes 5
- 1.2.4 Treatment of the QM/MM Boundary 6

1.3 Oxygenic Photosynthesis 8

- 1.3.1 Structure of Photosystem II 8
- 1.3.2 Charge Separation at the Reaction Center (RC) 11
- 1.3.3 Plastoquinone (PQ) Reduction 13
- 1.3.4 Water oxidation at the Oxygen Evolving Complex (OEC) 14
- 1.3.5 The CP43 and CP47 light harvesting proteins 15
- 1.3.6 The extrinsic proteins of PSII 17

1.4 Spectroscopic Properties of Photosynthetic Pigments 18

1.5 Scope of this Thesis 20

1.6 REFERENCES 22

Chapter 2 Ionization Energies and Redox Potentials of Hydrated Transition Metal Ions: Evaluation of DLPNO Coupled Cluster Approaches 29

2.1 Introduction 29

2.2 Theory and Methodology 32

- 2.2.1 Explicitly Hydrated Models 32
- 2.2.2 Electronic configurations 33
- 2.2.3 Geometry Optimizations 33
- 2.2.4 Electronic Structure Calculations 34
- 2.2.5 Calculation of Ionization Energies and Redox Potentials 35
- 2.2.6 Multilevel QM/QM Scheme for Truncation Thresholds 36

2.3 Results and Discussion 37
  2.3.1 Geometries 37
  2.3.2 Ionization Energies for M-W6 Models 38
  2.3.3 Effect of the Second Solvation Sphere 41
  2.3.4 Standard Reduction Potentials using a Cluster-Continuum Approach 44
  2.3.5 Evaluation of Multilayer DLPNO-based Approaches 46

2.4 Conclusions 49

2.5 REFERENCES 51

Chapter 3 Multiscale Modelling of Photosynthetic Systems 56

3.1 Introduction 56

3.2 System Preparation 56
  3.2.1 Setup of the PSII monomer 56
  3.2.2 Construction of the PsbA (D1) protein variants 58
  3.2.3 Models of the PSII “extrinsic” proteins‡ 59
  3.2.4 Explicitly solvated model of Cytochrome c_{550} (PsbV)‡ 60

3.3 Definition of Force Field Parameters 61

3.4 Molecular Dynamics of the PSII assembly 63
  3.4.1 Energy minimization 63
  3.4.2 Heating and Equilibration 63
  3.4.3 Production simulations 64

3.5 MD simulations of PsbV 65
  3.5.1 System minimization 65
  3.5.2 Equilibration Dynamics 65

3.6 QM/MM Computations 66
  3.6.1 Definition of the QM regions 67
  3.6.2 Geometry optimizations 68
  3.6.3 Computation of Excitation Energies 70

3.7 REFERENCES 72

Chapter 4 Triplet States in the Reaction Center of Photosystem II 76

4.1 Introduction 76

4.2 Methodology 79
  4.2.1 QM/MM Setup 79
  4.2.2 Geometry optimization 80
  4.2.3 Excitation energies 80

4.2.4 EPR parameters 81
4.3 Results and Discussion 82
4.3.1 Singlet–triplet excitations in individual pigments 82
4.3.2 Singlet–triplet excitations in pigment assemblies 85
4.3.3 Relaxed triplet states 88
4.3.4 EPR parameters of triplet chlorophylls 90
4.3.5 Electrostatic effects by plastoquinone Q_A and the OEC 94
4.4 Conclusion 99
4.5 Supplementary Tables 100
4.6 REFERENCES 102
Chapter 5 Microscopic Basis of Reaction Center Modulation in PsbA Variants of Photosystem II **107**
5.1 Introduction 107
5.2 Methodology 109
5.2.1 Preparation of MM models 109
5.2.2 Molecular Dynamics (MD) Simulations 109
5.2.3 Binding Free Energy Calculations 110
5.2.4 QM/MM Geometry optimizations 110
5.2.5 Excited State Calculations 111
5.3 Results 112
5.3.1 Structural analysis of PsbA variants. 112
5.3.2 Functional Dynamics around Pheo_{D1} 115
5.3.3 Redox tuning of the Chl_{D1}–Pheo_{D1} pair 117
5.3.4 Global tuning of Chl_{D1}–Pheo_{D1} excited states 118
5.3.5 Electrostatic effects of D1 substitutions 120
5.4 Discussion 123
5.4.1 Structural changes around Chl_{D1}–Pheo_{D1} 123
5.4.2 Influence on Redox Properties 124
5.4.3 Spectral tuning of Chl_{D1}–Pheo_{D1} excited states 125
5.5 Conclusion 127
5.6 Supplementary Tables and Figures 128
5.7 REFERENCES 131

Chapter 6 Excitation Landscape of the CP43 Photosynthetic Antenna Complex from Multiscale Simulations 134

6.1 Introduction ... 134

6.2 Methodology ... 139

 6.2.1 Molecular Dynamics (MD) simulations ... 139
 6.2.2 QM/MM geometries ... 140
 6.2.3 Calculation of excited states ... 141
 6.2.4 Perturbed Matrix Method (PMM) Calculations ... 141
 6.2.5 Calculation of the absorption spectrum with MD-PMM ... 144

6.3 Results and discussion ... 146

 6.3.1 Chlorophyll site energies ... 146
 6.3.2 Excitonic couplings from PMM ... 153
 6.3.3 Excited states of Chl dimers ... 154
 6.3.4 Global CP43 excitonic states from the MD-PMM approach ... 159

6.4 Conclusion ... 161

6.5 Supplementary Tables and Figures ... 163

6.6 REFERENCES ... 171

Chapter 7 Multiscale Modelling of the Soluble Methane Monooxygenase Hydroxylase Complex 176

7.1 Introduction ... 176

7.2 Computational Methodology ... 179

 7.2.1 System Preparation ... 179
 7.2.2 Force Field Parameters ... 180
 7.2.3 Molecular Dynamics Simulations ... 181
 7.2.4 QM/MM Calculations ... 184

7.3 REFERENCES ... 185

Chapter 8 Summary and Perspective 187

1

Introduction

1.1 Multiscale Modelling of Biomolecular Systems

Proteins and enzymes are crucial to chemical and biological processes and represent a challenging and vital subfield in theoretical and computational chemistry.[1-3] While experimental data, including crystal and solution structures, mutagenesis, spectroscopy, and kinetics, provide significant mechanistic insights, biomolecular systems particularly, enzymes remain a complex topic requiring further exploration.[4] Modern computational chemistry tools are invaluable for addressing these challenges, interpreting and supplementing experimental data to propose mechanistic details, provided the quantitative properties align with experimental observations.[5] Over the past decades, various computational methods have been employed to model systems across length and timescales, such as cluster modeling with quantum mechanics (QM),[6-8] multiscale quantum mechanics/molecular mechanics (QM/MM),[2, 9, 10] and QM/MM combined with molecular dynamics (QM/MM/MD) methods.[11] The QM methodologies range from ab-initio theories to density functional theory (DFT) and various semi-empirical approaches. Additionally, the empirical valence bond (EVB) method,[12] rooted in QM but operationally similar to a force field, has also been used. Molecular mechanics (MM) is typically implemented with fixed-charge force fields,[13] though studies suggest that polarization of the MM region may be important in some cases.[14] The combination of ab-initio QM and MM multiscale methods opens a wide range of applications in describing spectroscopy,[15] thermodynamics and biocatalysis.[16-18] When combined with MD, these methods even allow for a time-dependent description of the entire system.

1.1.1 QM Cluster Models

The initial approach to elucidate spectroscopic properties and gain mechanistic insights for biomolecular systems is the “cluster” or “QM-only” approach which employs simplified chemical models.[6-8] This model treats the significant part of the enzyme at the QM level; that is, the active site and the surrounding residues relevant for the biochemical process. A QM-cluster model is typically derived from an existing crystal structure or nuclear magnetic resonance (NMR) data. A complete model should include all components participating in the biochemical reaction, as well as the residues responsible for stabilization and binding of the substrate, maintaining its configuration, and those with significant short-range, long-range electrostatic and non-covalent interactions. Since a sizable part of the enzyme is excluded in a QM-cluster model, the steric and polarization effects exerted by the protein environment must be compensated for.

The conventional strategy to model the electrostatic influence of the protein environment involves assuming a homogeneous polarizable medium using a dielectric constant i.e., implicit solvation methods (e.g., CPCM, SMD).[19, 20] Steric effects are addressed by fixing several "boundary" atoms at their crystallographic coordinates. However, explicitly solvated systems are difficult to model and might be impractical with expensive electronic structure methods.[21] Therefore, including explicit solvent molecules or using a combination of implicit and explicit solvation approaches is the most feasible way to obtain chemically relevant properties. Nevertheless, the size of the QM cluster models remains restricted to a few hundred atoms, due to limitations in computational power and the QM method employed, necessitating the use of hybrid QM/QM and QM/MM approaches.

1.1.2 QM/QM Multilevel Approaches

A more practical approach to compensate for the environmental effects in QM cluster models and simultaneously treat a larger part of the complex system with accuracy is to use multilevel QM embedding approaches. The main principle of hybrid QM/QM methods is to treat the less critical parts of a large system (e.g., solvent and protein environment) at a lower accuracy, while the chemically relevant part is still treated at the highest level of sophistication. The hybrid QM1/QM2 approach allows the combination of different levels of theory for each part of the system using multiple "layers". The individual QM regions can be described using a single QM method but with different accuracies (e.g., PNO thresholds in DLPNO-CCSD(T)),[22-29] or entirely different QM methods (e.g., DFT/xTB,[30] DLPNO-CCSD(T)/HF-3c). Most QM1/QM2 embedding approaches are based on the Own N-layer integrated molecular orbital molecular mechanics (ONIOM) method,[31] and can be combined with MM force fields (QM1/QM2/MM) and continuum solvation.

The work described in **Chapter 2** follows the QM cluster with explicit solvation and the QM/QM multilevel approaches. Moreover, obtaining accurate energetics for spin-states and redox changes in transition metal clusters and metal active sites is challenging.[32-34] The treatment of the electronic structure problem places heavy demands both on the definition of the computational model in terms of the appropriate representation of the coordination environment and on the QM method. We recently demonstrated a potential application of the domain-based local pair natural approach to coupled cluster theory with singles, doubles, and perturbative triples, DLPNO-CCSD(T),[22-29] paving the way for more affordable and accurate calculations of chemical properties in complex systems.

1.2 Hybrid QM/MM Methods

Hybrid QM/MM methods combine the use of the quantum mechanics and molecular mechanics together to capture the electronic structure, dynamics, and energetics of biomolecules across length and time scales. The whole biomolecular system is divided into two regions (see **Fig. 1.1**) where the QM region usually represents the region of interest (active site or chromophore) and is treated by computationally demanding electronic structure methods, whereas the MM subsystem (i.e., rest of protein environment, solvents and counter ions) is described using classical atomic force-fields.

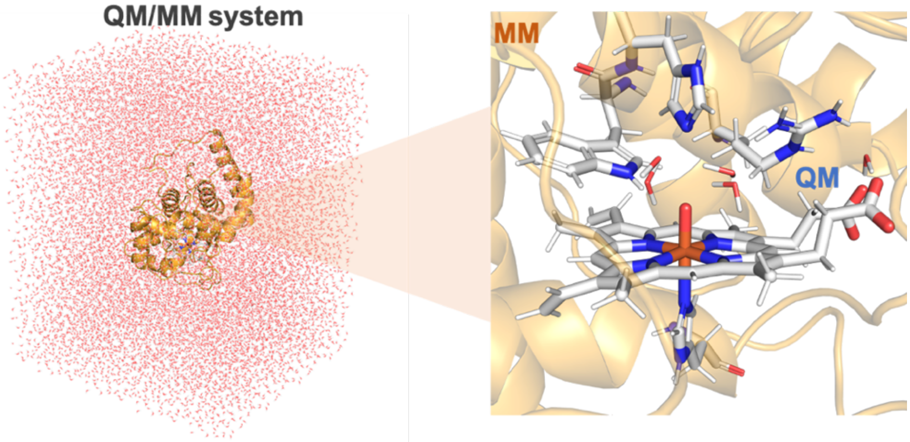


Fig. 1.1 An example illustrating the partitioning of the **QM** (heme, ligands) and **MM** (surrounding protein, solvent) subsystems within a solvated Ascorbate Peroxidase protein (atomic coordinates from 7BI1.pdb).[35]

Additionally, there are four essential aspects to be considered in all QM/MM models: (i) how to describe the MM force field (i) how to evaluate the total energy (ii) treating the interaction between QM and MM regions, and (iii) treating covalent bonds at the QM/MM boundary. In the following sections, we will discuss each of these aspects in further detail.

1.2.1 MM Force Fields

The classical potential energy functions used to describe the interactions among the MM atoms in the protein or biomolecular system are known as Force Fields (FF).[13] The most popular protein-based force-fields are Assisted Model Building and Energy Refinement (AMBER),[36,

[37] Chemistry at Harvard Macromolecular Mechanics (CHARMM),[38, 39] Groningen Molecular Simulation System (GROMOS)[40, 41] and Optimized Potentials for Liquid Simulations-all atom (OPLS-AA)[42, 43] while more advanced polarizable force-fields like AMOEBA[44] may also be required in certain cases. Quantitatively, the FF description of a system can be categorized into bonded (bond stretching, bending angle, torsions) and non-bonded (electrostatics and van der Waals) terms, expressed as:

$$E_{MM} = \sum_{bonds} k_r(r-r_i)^2 + \sum_{angles} k_\theta(\theta-\theta_i) + \sum_{dihedrals} k_\phi[1+\cos(n\omega+\phi)] - \sum_{i,j}\frac{q_iq_j}{4\pi\epsilon_0 r_{i,j}} - \sum_{i,j} 4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right]$$

Where, r, θ and ϕ represent the bond length, angle, and torsion respectively; r_i and θ_i are the equilibrium distance and angle; n and ω are the torsion multiplicity and phase; k_r, k_θ and k_ϕ are the respective bonded force-constants; σ (vdW radii) and ϵ (well-depth) are the Lennard-Jones (LJ) parameters; ϵ_0 represents the dielectric constant; q_i and q_j represents the atomic partial charges and $r_{i,j}$ represents the distance between the non-bonded atoms.

The QM/MM setup for any biomolecular system necessitates an appropriate FF description of the entire system. While optimized FF parameters for the protein, water and counter-ions are generally available, most biomolecular systems contain non-canonical protein residues, organic co-factors and metal-sites, making the overall parameterization process non-trivial. For organic cofactors, the General Amber Force-Field (GAFF2)[45] and CHARMM General Force-Field (CGenFF)[46] are commonly used with their respective protein force-fields, AMBER and CHARMM. However, there is no universal force-field for metal sites due to varying coordination environments, spin-states, oxidation states, and covalency requiring manual derivation of partial charges and the force-field to describe the bonded and non-bonded terms in such cases.

1.2.2 Evaluating the QM/MM Energy

The total energy of the QM/MM system is composed of the individual contribution of the QM and MM regions, and the contribution from their mutual interaction. There exist two different coupling schemes for computing the total QM/MM energy: additive and subtractive. In the additive coupling scheme, the QM/MM energy is evaluated as follows:

$$E_{QM/MM} = E_{QM}(\mathbf{QM}) + E_{QM/MM}(\mathbf{QM+MM}) + E_{MM}(\mathbf{MM})$$

where the QM system is embedded within the larger MM system, and the energy of the whole system is the sum of QM, MM, and QM/MM coupling terms. The middle term represents the interaction between the QM and MM subsystems and contains the bonded interactions:

$$E_{QM/MM}(\mathbf{QM+MM}) = E_{bonding} + E_{vdW} + E_{electrostatic}$$

In the additive scheme, the bonded interactions between the QM and MM atoms are explicitly described by the MM force field whereas the vdW interactions at the boundary are described by LJ potentials.

In the subtractive scheme, the energy is obtained as:

$$E_{QM/MM} = E_{QM}(\mathbf{QM}) + E_{MM}(\mathbf{QM+MM}) - E_{MM}(\mathbf{QM})$$

The subtractive QM/MM scheme involves three separate energy calculations: (i) energy of the total system at the MM level, (ii) energy of the QM sub-system at the QM (high) level of theory, and (iii) energy of the QM sub-system at the MM (low) level of theory. This scheme is relatively straightforward to implement although less accurate compared to the additive scheme due to its simplified nature and because the QM/MM interaction is also computed at the MM level of theory. The ONIOM is a classic example of the subtractive QM/MM approach.

1.2.3 QM/MM Interaction Schemes

The electrostatic interactions between QM and MM parts are treated using the following embedding schemes: (i) electrostatic, (ii) mechanical and (iii) polarizable. In mechanical embedding, the QM atoms are represented by point charges, bond dipoles, or higher multipoles and the electrostatic interaction between the QM and MM parts is completely treated at the MM level. The electrostatic embedding approach enables the polarization of the QM part since the charge distribution of the MM part is incorporated in the QM calculation. The effective Hamiltonian is represented as:

$$H_{electrostatic} = H_{QM} - \sum_{k}^{m} \frac{q_k}{|r_i - R_k|} + \sum_{k}^{m} \sum_{l}^{n} \frac{q_k Z_l}{|r_l - R_k|}$$

where, *m* is the number of MM atoms which has atomic partial charge of q_k; n denotes the number of QM nucleus having atomic number Z_l. The terms r_i, r_l and R_k represents the position of the electron *i*, position of QM nucleus *l* and MM atom *k*, respectively.

In the polarizable embedding scheme,[47-49] the polarization of the MM part in response to the QM charge distribution and the MM charges is also taken into account such that the QM and MM atoms can mutually polarize each other. The mechanical embedding approach is the least accurate compared to electrostatic and polarizable embedding since the charge density of the QM part is not polarized by the MM part and vice-versa. On the other hand, polarizable embedding schemes are computationally demanding and requires specialized polarizable force fields and inclusion of atomic polarizabilities into the effective Hamiltonian. The electrostatic embedding scheme is the most widely used in QM/MM calculations of biomolecular systems.

1.2.4 Treatment of the QM/MM Boundary

QM/MM methods based on the electrostatic embedding technique requires special treatment of the atoms at the QM/MM boundary to avoid "over-polarization" of the QM electron density. In most biomolecular systems the active-sites in proteins are bound by amino acid side chains which are part of the overall polypeptide architecture. The partitioning of the system into the QM and MM regions therefore results in the QM/MM boundary to pass through covalent bonds. In such a case, there are two important aspects to be considered: (a) the dangling QM atom should be capped and (b) the QM density should not be over-polarized by the MM atoms at the QM/MM boundary. There are several techniques to handle the capping of the QM atom, one of which is the Link-Atom (LA) approach.[3] In the LA approach, the bond is cut and replaced with a hydrogen atom covalently linked to the QM atom to saturate its valency (**Fig. 1.2**). The final QM calculation is thus performed on the QM sub-system with the LA.[50]

However, because the LA introduces new degrees of freedom to the system, its proximity to the MM atoms, may lead to over-polarization of the QM density. In order to bypass this problem, the "Charge-Shift" scheme is used where the partial charge of the immediate MM atom is set to zero and equally re-distributed to the neighbouring covalently bound MM atoms to conserve the overall charge of the QM/MM system. The link-atom and charge shift scheme to treat QM/MM boundaries is depicted in **Fig. 1.2**.

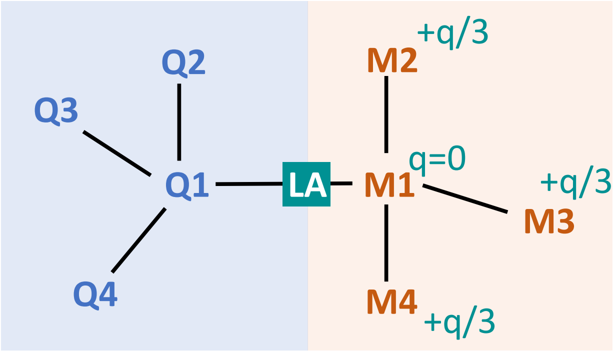


Fig. 1.2 The interaction of the **QM** (Q1–Q4) and **MM** (M1–M4) atoms at the **QM/MM** boundary. The Link Atom (**LA**) is placed at the covalent bond which is cut. The partial atomic charges depicts the Charge-Shift scheme to prevent over-polarization of the QM density by the M1 atom.

The accurate description of complex biochemical or biophysical processes using computational tools presents significant challenges. When investigating phenomena that involve changes in electronic structure, such as the making and breaking of chemical bonds, transition states, spin-states, electron and proton transfer and spectroscopy, the choice of (i) an appropriate molecular model and (ii) a sufficiently accurate QM method is paramount.[1, 51, 52] For instance, in enzyme catalysis, it is important to identify which protein residues, explicit solvation, and counter ions are necessary to represent the active site and which potential reaction mechanisms to consider for each catalytic step. Protein crystallography and cryo-EM provide "frozen" snapshots of individual steps along the catalytic process, serving as starting points for molecular models. However, modelling proteins in biological systems involves chemistry beyond 0 K, the standard starting point for exploring the potential-energy surface (PES). While QM only cluster models can yield significant insights into catalytic mechanisms, the effects of the environment (protein matrix, explicit water molecules and counterions), finite temperature, and entropy must also be considered through a MD equilibration.

The size of the system and available computing resources dictate the selection of a suitable QM method, which must be carefully benchmarked for accuracy. Selecting adequate methodologies and establishing a computational protocol for protein biochemistry and biophysics is daunting and details of suitable QM methodologies and protocols vary depending on the protein and the specific biochemical or biophysical process being studied. Various QM methods such as DFT, wave function-based ab-initio methods, or semi-empirical methods are often employed determined by the nature of the problem. This thesis demonstrates the application of the QM-based multiscale modelling methodologies to various biomolecular systems with particular emphasis on photosynthetic pigment-protein complexes.

1.3 Oxygenic Photosynthesis

Oxygenic photosynthesis is crucial for sustaining all life on Earth through the biological conversion of solar energy into chemical energy.[53-55] The fundamental biological functions are conserved across all photosynthetic organisms (plants, algae and cyanobacteria): light-induced charge separation (CS) that drives electron transfer[56-61] leading to the production of energy-rich reduced chemical compounds and the release of molecular O_2 as a by-product. This process transformed Earth's atmosphere to be O_2-rich and contributed to the formation of the ozone layer, while the evolution of respiration harnessed oxygen's oxidizing power, supporting more complex life forms. Thus, photosynthetic water splitting is one of the most important biochemical reactions on earth.[62] The light-driven oxidation of water[63, 64] and primary CS occur in Photosystem II (PSII),[65-70] the first pigment-protein complex in the photosynthetic chain. The electrons are then transported by mobile electron carriers (plastoquinones and plastocyanins) through other components of the photosynthetic chain such as Cytochrome b_6f (Cyt b_6f)[71-75] and Photosystem I (PSI)[76-80] to be utilized ultimately in the synthesis of NADPH (see **Fig. 1.3**). NADPH facilitates the light-independent carbon fixation reactions that convert CO_2 into glucose. The ET process is driven by the relative redox potentials of the embedded cofactors and coupled with proton translocation, eventually generating a transmembrane potential gradient that powers the chemiosmotic synthesis of ATP, the primary energy source for all metabolic processes.

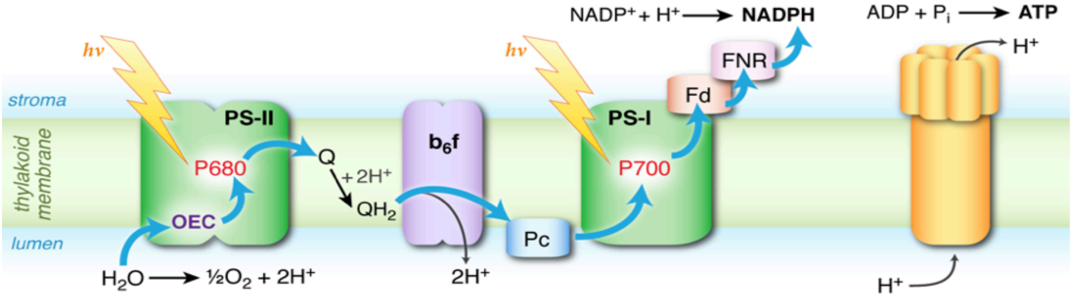


Fig. 1.3 Schematic representation of the linear electron transfer (**LET**) pathway in oxygenic photosynthesis depicting reactions catalysed by each membrane embedded protein complex: Photosystem II (**PSII**), Cytochrome b_6f, Photosystem I (PSI) and ATP synthase. The figure is reproduced in part with permission from Principles of natural photosynthesis, *Solar energy for fuels*, **2016**, pp. 23–48.[54]

1.3.1 Structure of Photosystem II

Photosystem II (PSII) is a membrane-protein complex found in the thylakoid membranes of photosynthetic organisms, from cyanobacteria to higher plants. The cyanobacterial PSII is a

700 kDa dimeric protein comprising 20 subunits per monomer, including 17 transmembrane proteins and 3 membrane-extrinsic subunits (**Fig. 1.4**).[65, 81-84] Among the transmembrane subunits, the D1 and D2 proteins, encoded by the psbA and psbD genes respectively, each contain five transmembrane helices and host the reaction center (RC) of PSII, along with all the cofactors involved in electron transfer and water splitting.[68, 84] The D1 and D2 polypeptides are flanked by two chlorophyll-binding proteins, CP47 and CP43, encoded by the psbB and psbC genes respectively.[85-88] CP47 and CP43 each have six transmembrane helices and bind multiple chlorophyll (Chl) molecules, functioning as intrinsic light-harvesting antennas to the PSII-RC. In addition to these four major transmembrane subunits, there are 12 low-molecular-weight proteins: PsbE, PsbF, PsbH, PsbI, PsbJ, PsbK, PsbL, PsbM, PsbT, PsbX, PsbY and PsbZ.[89] These subunits have molecular masses of less than 10 kDa and typically contain one transmembrane helix each, except for PsbZ, which has two transmembrane helices.

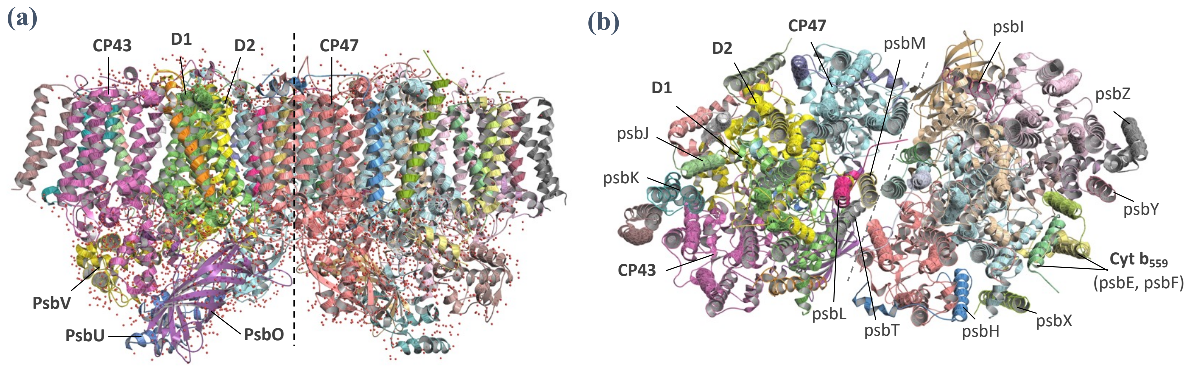


Fig. 1.4 The overall structure of the Photosystem II (**PSII**) dimer and its major subunits at 1.9 Å resolution (**3WU2.pdb**) **(a)** viewed perpendicular to the membrane normal; the red dots indicate positions of the crystallographic waters. **(b)** The arrangement of transmembrane helices in the PSII dimer in a top view from the stromal side. The dashed line in the center divides the two monomers, and the PSII subunits are labeled in one of the monomers.

Fig. 1.4 represents the overall structure of the PSII dimer analyzed at 1.9 Å resolution by Umena et al.,[68] featuring 16 transmembrane and 3 membrane-peripheral subunits. D1 and D2 form the reaction center core complex of PSII, with CP47 and CP43 located on either side of the D1/D2 core binding 16 and 13 Chls, respectively. The three extrinsic, hydrophilic subunits—PsbO, PsbU, and PsbV—are situated on the lumenal side of the thylakoid membrane.[90-93] Together with the extrinsic loops of D1, D2, CP43, and CP47 that extend into the lumenal side, these soluble proteins cap the oxygen evolving complex (OEC), shielding it from the lumenal bulk solution. In addition to the protein subunits, the PSII monomer includes 35 Chls, 2 pheophytins (Pheo), 11 β-carotenes, 2 plastoquinones (Q_A and Q_B), 1 bicarbonate ion, 1 b-type and 1 c-type cytochrome, a nonheme iron (NHI), more than 20 special lipids, at least 2 chlorides, a Mn_4CaO_x cluster, and other ions. The overall arrangement of the PSII cofactors is depicted in **Fig. 1.5**.

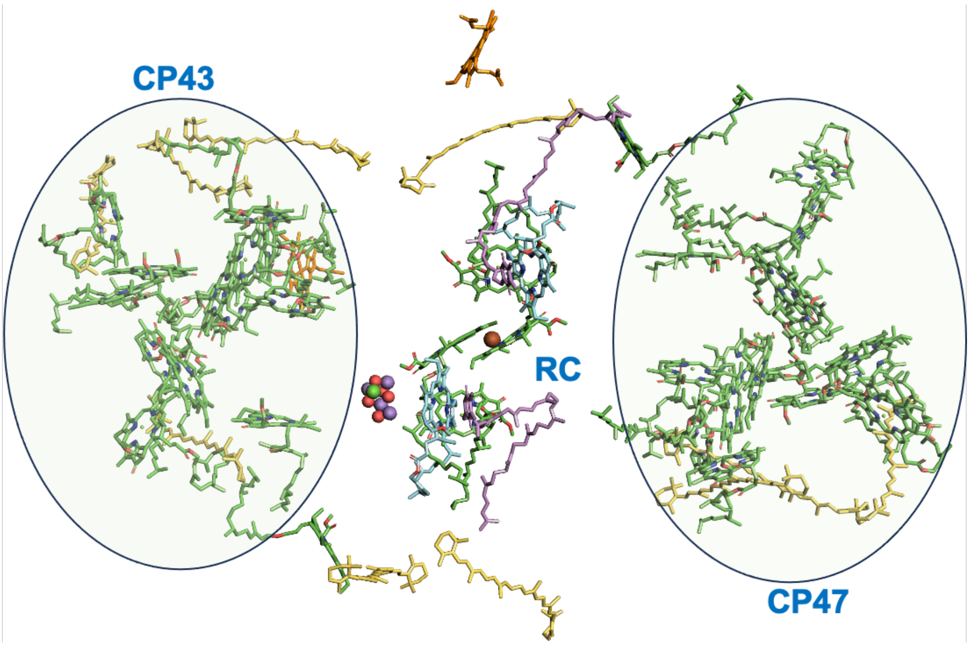


Fig. 1.5 Schematic representation of the various cofactors arranged within the PSII core complex (**PSII-cc**) viewed from the stromal side (protein not shown). The 35 Chls are depicted in green, 2 Pheo (cyan), 2 hemes (orange), β-carotenes (yellow), 2 plastoquinones (pink), 1 non-heme Fe (brown) and the OEC.

One of the most notable features of the high-resolution PSII dimeric structure is the presence of approximately 2,800 water molecules. These water molecules are distributed in two layers: some on the surface of the cytoplasmic (stromal) side of the thylakoid membrane but predominantly on the surface of the lumenal side (**Fig. 1.4a**), with very few water molecules in the transmembrane region. This solvent distribution pattern is typical for membrane-protein complexes. The few water molecules present in the transmembrane region serve as axial ligands or H-bonding partners to Chls that are not ligated by protein residues. Additionally, two extra water molecules are usually found near the chlorin ring, likely H-bonded to the carbonyl groups of the chlorin ring as well as the direct water ligand. These water molecules may play a crucial role in stabilizing the chlorin ring not directly ligated to the protein matrix. In the following sections, we will discuss primarily about (a) light-harvesting complexes (LHCs) and excitation energy transfer (EET) (b) charge-separation (CS) and electron transfer (ET) in the PSII-RC (c) water oxidation and plastoquinone reduction and (d) the role of the PSII extrinsic proteins.

1.3.2 Charge Separation at the Reaction Center (RC)

In photosynthesis sunlight is captured and funneled as excitation energy[88, 94-98] into the reaction center (RC) of PSII, the site of primary charge separation (CS) that initiates the photosynthetic electron transfer (ET) chain.[99] The PSII-RC comprises of six pigments arranged pseudo-symmetrically along the D1 and D2 polypeptides (**Fig. 1.6a**). The central pair of Chls (P_{D1} and P_{D2}) are flanked by two accessory Chls (Chl_{D1} and Chl_{D2}) and two pheophytins (Pheo_{D1} and Pheo_{D2}). The D1 pigments constitute the "active" branch of primary ET,[56-59, 100-107] while those on the D2 side are thought to engage in secondary processes such as regulation and photoprotection.[108-112] Besides the RC pigments, the D1 and D2 proteins include two additional Chls (Chlz_{D1} and Chlz_{D2}) located on the periphery of the RC on each side (see **Fig. 1.6a**).

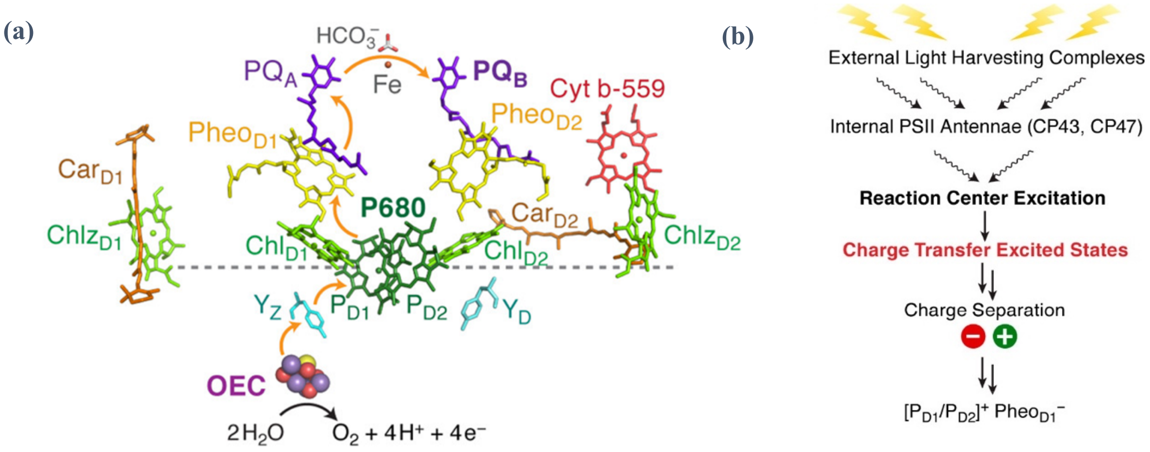


Fig. 1.6 (a) Reaction center (**RC**) pigments and other important cofactors, with a schematic representation of electron flow along the D1 branch; figure adapted from *ACS Catal.* **2018**, *8*, 9477–9507. **(b)** Sequence of major events leading to charge separation (**CS**) in the PSII-RC; figure adapted from *Acc. Chem. Res.* **2023**, *56*, 2921-2932.

The peripheral Chls are not involved in primary ET however, Chlz_{D1} presumably participates in EET between CP43 and the RC, whereas Chlz_{D2}, along with Car_{D2} has been proposed to act as intermediates in the cyclic electron flow between cytochrome b_{559} (Cyt b_{559}) and the D2 branch of the RC.[108-112]

The photoexcitation of the RC triggers primary ET events exclusively along the D1 branch, leading to the formation of the radical cation species ($P_{680}^{+\cdot}$) localized in the central $P_{D1}P_{D2}$ pair.[57, 113, 114] With an estimated redox potential of approximately 1.3 V, $P_{680}^{+\cdot}$ is the strongest known biological oxidizing agent. It facilitates the oxidation of the OEC and drives water splitting at the lumenal side of PSII via the intermediary redox-active tyrosine D1-Y161 (Y_Z).

On the acceptor side of PSII, electrons following primary CS are transferred to plastoquinone Q_A and eventually to the terminal acceptor plastoquinone Q_B.[115-117]

Over the past few decades, various theories have emerged about the identity of the pigments involved in primary CS.[58, 61, 101, 105, 118-124] The most widely accepted mechanism is the "Chl_{D1} pathway"[125-127] where Chl_{D1} serves as the primary electron donor and $Pheo_{D1}$ as the primary acceptor forming the $Chl_{D1}^{+}Pheo_{D1}^{-}$ radical pair (RP), prior to hole stabilization on $P_{D1}P_{D2}$:

$$[Chl_{D1}Pheo_{D1}]^{*} \rightarrow Chl_{D1}^{+}Pheo_{D1}^{-} \rightarrow [P_{D1}P_{D2}]^{+}\ Pheo_{D1}^{-}$$

Some studies propose the "P_{D1} pathway" analogous to purple bacterial RCs (bRC), where Chl_{D1} acts as an intermediary electron acceptor from P_{D1}, forming a $P_{D1}^{+}Chl_{D1}^{-}$ pair.[128] An initial charge-separated state within the central $P_{D1}P_{D2}$ pair has also been considered:

$$[P_{D1}P_{D2}]^{*} \rightarrow P_{D1}^{-}P_{D2}^{+} \rightarrow [P_{D1}P_{D2}]^{+}\ Chl_{D1}^{-} \rightarrow [P_{D1}P_{D2}]^{+}\ Pheo_{D1}^{-}$$

Other hypotheses include initial CS within the $P_{D1}P_{D2}$ pair, direct "long-range" electron transfer from $P_{D1}P_{D2}$ to $Pheo_{D1}$ without forming Chl_{D1}^{-},[129] and concerted action of Chl_{D1} and P_{D1} acting as primary electron donors to $Pheo_{D1}$ i.e., "parallel pathways".[58, 100, 130] There is also an existing debate about the latter mechanism, especially in the context of far-red light photo acclimation (FaRLiP) in certain cyanobacteria.[118, 129, 131-133]

Deciphering the functional principles of photosynthetic RCs[134, 135] involves addressing key questions with regards to the spectroscopic and redox properties of the individual pigments,[136-139] the nature of excitations[127] particularly charge-transfer (CT) states, the functional asymmetry between D1 and D2 pigments,[107, 140] the localization and/or delocalization of excitation energy,[105] identification of electron donor-acceptor pairs, and the influence of the protein environment and excitation wavelength.[81-83] In the past, theoretical[85, 122, 141-143] and computational simulations offered complementary insights into the nature and function of LHCs.[144-149] Recent studies by Sirohiwal et al.[103, 129, 150] demonstrated the use of multiscale computational approaches on a "near-native" state of PSII, to obtain an atomistic description of the excited states of the RC pigments. The studies provided valuable insights into the electronic and structural factors that generate functional asymmetry and determine the directionality of CS within the RC. The influence of the PSII protein electrostatics, redox states and its variants on the low-energy excitation profile of the RC pigments, are discussed in great detail in **Chapters 4** and **5**.

1.3.3 Plastoquinone (PQ) Reduction

The PSII-RC bind two plastoquinones (PQ) at its “acceptor side” (Q_A and Q_B), that are arranged symmetrically around a non-heme iron (NHI).[117, 151, 152] The metal center is a hexacoordinate Fe^{II} (**Fig. 1.7**) with a distorted octahedral geometry. Four vertices of the octahedron are occupied by the N_ε-atoms of histidines, two from each of the D1 and D2 subunits. One His on each side is hydrogen bonded with the respective PQ. The remaining two ligand positions are occupied by the oxygen atoms of a bidentate ligand, which is bicarbonate in PSII-RC.

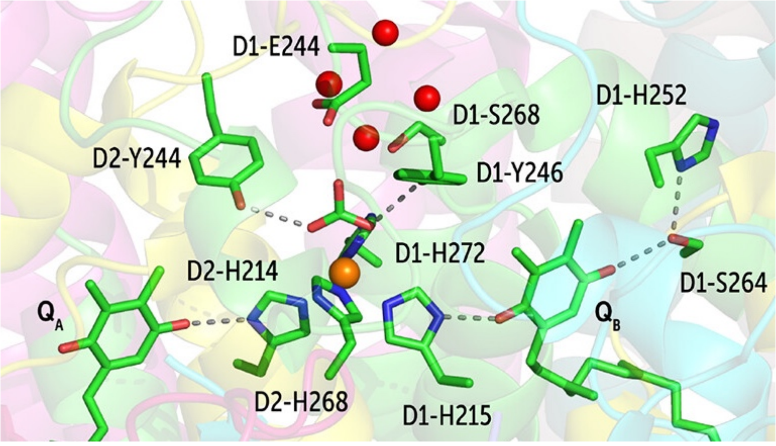


Fig. 1.7 Protein environment of the non-heme iron (**NHI)–quinone** complex on the electron acceptor side of cyanobacterial PSII (3WU2.pdb). Coordination to the NHI and hydrogen bonds are shown with dashed lines. The figure is reproduced in part with permission from *J. Am. Chem. Soc*. **2022**, *144*, 22035-22050.

The primary electron acceptor quinone (Q_A) is located between the acceptor $Pheo_{D1}$ and the NHI. It functions as an one-electron carrier ultimately delivering two electrons for the reduction of the substrate quinone (Q_B) through a series of proton-coupled electron transfer (PCET) steps.[115, 116, 151, 153, 154] The Q_BH_2 formed is released and enters the PQ/PQH_2 pool, from where it transfers electrons to the Cyt b_6f complex The ET from $Pheo_{D1}^-$ to Q_A is reported to occur in few hundred picoseconds (200–500 ps), based on kinetic measurements. This leads to the formation of a semiquinone anion radical ($Q_B^{\bullet-}$), which is stabilized by protonation of nearby amino acid side chains. The role of the NHI in type-II RC in general is still debated and the exact PT/ET mechanisms at the Q_B site are still unclear, though some studies claim that the (bi)carbonate regulates both ET and PT to Q_B, mediated by acceptor side water channels.[155-157] The modulation of the redox potential of Q_A and $Pheo_{D1}$ has photo-protective roles in oxygenic species,[158-163] as changes in the free energy gap between $S_2Q_A^{\bullet-}$ and $P_{D1}^+Pheo_{D1}^-$ can directly influence the kinetics and routes of charge recombination within the PSII-RC (see **Chapters 4** and **5** for further discussion on redox states of Q_A).

1.3.4 Water oxidation at the Oxygen Evolving Complex (OEC)

In PSII, the light-driven oxidation of water[63, 84, 164, 165] takes place in the OEC which constitutes an oxo-bridged tetramanganese–calcium cluster (Mn_4CaO_x).[67, 166-168] On the donor side of PSII, the electron hole at $P_{680}^{\bullet+}$ formed as a result of primary charge separation, is filled through oxidation of the redox-active D1-Tyr161 residue (Y_Z). The formation of the $Y_Z^{\bullet}$ radical is aided by a proton transfer to D1-His190. Tyr161(Y_Z) interacts closely with the Mn cluster of the OEC, and hence, $Y_Z^{\bullet}$ acts as its immediate oxidant.

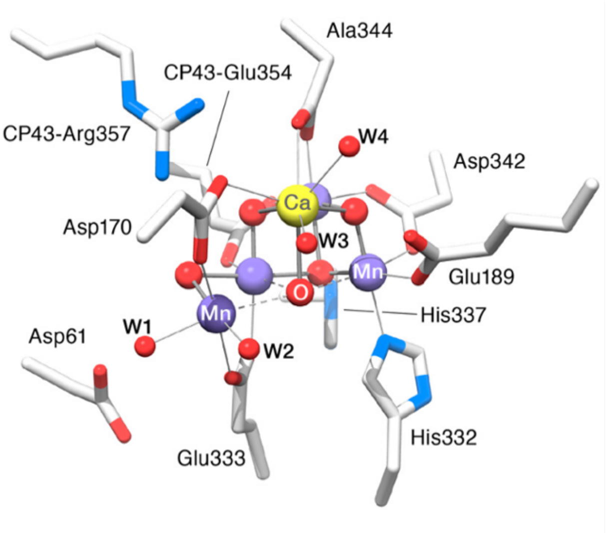


Fig. 1.8 X-ray crystallographic structure of the Mn_4CaO_x cluster (**OEC**) with protein derived ligands in dark-adapted PSII at 1.9 Å resolution. All residues belong to D1 (chain A) of the PSII dimer unless otherwise stated. The figure is reproduced in part with permission from *Acc. Chem. Res.* **2013**, *46*, 1588–1596.

The catalytic core of the OEC is an inorganic cluster composed of four manganese ions and one calcium ion interconnected by oxo or hydroxo ligands (Mn_4CaO_x). The cluster is coordinated by carboxylate residues (Asp and Glu), one His and four terminal H_2O/OH ligands (see **Fig. 1.8**).

Functionally, the OEC serves as both an accumulator of electron holes and an oxygen-evolving catalyst. The $Y_Z^{\bullet}$ radical oxidizes the Mn_4CaO_x cluster of the OEC in one-electron steps. Four such oxidations occur before dioxygen (O_2) is produced, indicating that four oxidizing equivalents are stored at the OEC before they are utilized in O–O bond formation. This allows the OEC to operate within a narrow range of redox potential. Two substrate water molecules are also bound at late stages of the catalytic cycle. The S-state progression is described by the Kok–Joliot cycle of S_i states (**Fig. 1.9a**),[169] where i=0–4: S_0 thus represents the most reduced state of the cycle and S_4 the most oxidized state, which then evolves O_2. S_1, known as the "dark-stable" state, is the state to which most OEC centers revert when PSII is in darkness. Some measurements showed that the reaction times for the different steps of the cycle lie in the micro- to millisecond range (40 µs –1.6 ms). Except for the S_4 state (and events following the formation of the $S_3Y_Z^{\bullet}$ intermediate and the reconstitution of S_0), the other S-states[170] are observable and have been extensively characterized by experimental[171] and computational means.[65, 172, 173] Also, the final step of the complete cycle leading to the O–O bond formation is still extensively discussed in literature.[174, 175]

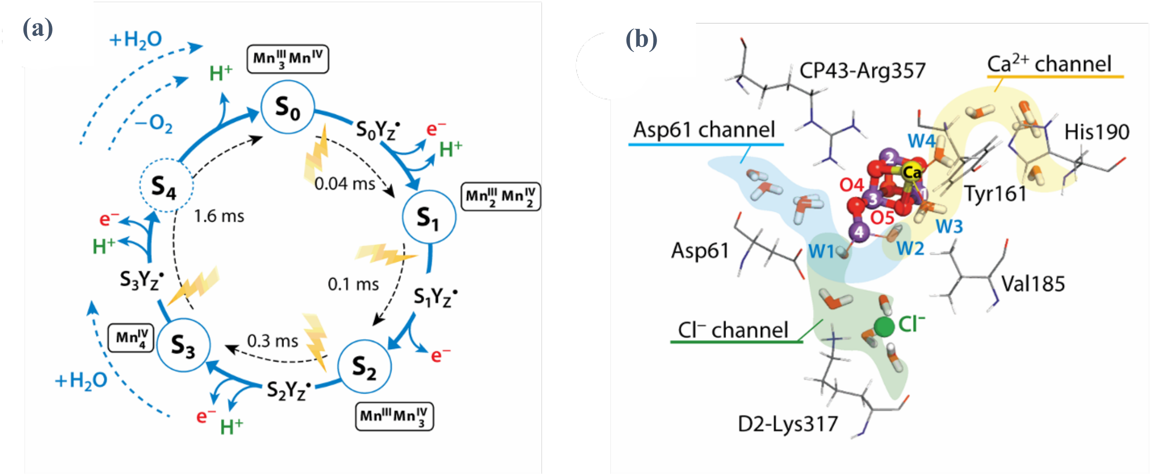


Fig. 1.9 **(a)** The S-state Kok-Joliot cycle showing the net oxidation state of the S_i (i=1–4) states and the release of electrons and protons. **(b)** The relevant channels for water and proton transfer (O1, O4 and Cl-1) are indicated as yellow, blue and green shaded areas respectively. The figure is reproduced with permission from *Annu. Rev. Biochem*. **2020**, *89*, 795-820.

In addition to the high oxidation potential generated at the RC, efficient catalytic activity at the OEC requires channels for substrate water delivery, proton release, and efficient diffusion of molecular O_2. Several structural[176] and theoretical studies (based on molecular dynamics simulations) demonstrated that PSII has three major water channel systems: Ca^{2+} or O1, Asp61 or O4 and the Cl^- or Cl-1 channels[157, 177] (**Fig. 1.9b**) that extend from the OEC onto the lumenal side of the thylakoid membrane. These channels have been proposed to play a key role in transporting protons and substrate waters during the S-state cycle of the OEC.

1.3.5 The CP43 and CP47 light harvesting proteins

In photosynthesis, light-harvesting pigment-protein complexes (PPC) consist of protein scaffolds that bind various light-absorbing pigments, such as Chl, BChl, and Car which are essential for efficiently utilizing solar energy.[44-48] PPCs perform critical functions including excitation energy transfer (EET), charge separation, photo-protection, and photo-acclimation.

The antenna complexes, responsible for harvesting solar energy and transferring excitation energy efficiently, are categorized into extrinsic and integral membrane antennas. Extrinsic antenna complexes, e.g. phycobilisomes in cyanobacteria or the Fenna-Matthews-Olson (FMO) complex in green bacteria, are located on the periphery of the core proteins and do not span the thylakoid membrane. In contrast CP43, CP47 in PSII and Light Harvesting Complex II (LHC II) in PS I, are intrinsic membrane-embedded antenna complexes to the RC.

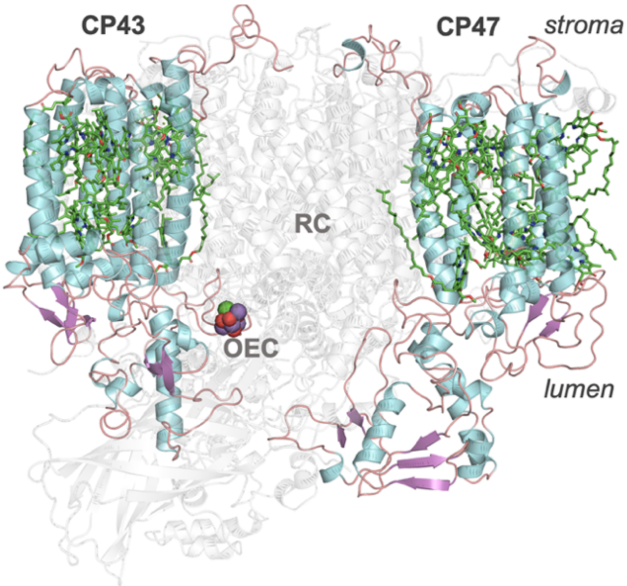


Fig. 1.10 The overall structure of the **CP43** (left) and **CP47** (right) intrinsic light harvesting antennae of the PSII monomeric complex along with the embedded chlorophylls (Chls); viewed perpendicular to the membrane normal. The approximate location of the RC and OEC are indicated.

The Chl a binding proteins CP47 (ca. 47 kDa) and CP43 (ca. 43 kDa), encoded by the psbB and psbC genes respectively serve as the proximal antenna to the PSII core complex. Each protein spans six transmembrane helices arranged in pairs, forming a trimer of dimers (**Fig. 1.10**). CP43 and CP47 harbors 13 and 16 Chls respectively, along with several beta-carotenoids (see **Fig. 1.5**). CP43 is located at the outer region of PSII close to D1(PsbA) and interacts with PsbO and PsbV extrinsic proteins. CP47 is positioned at the interface between the two monomers of a PSII dimer, forming close contacts with the D2 (PsbD) protein and other subunits such as PsbH, PsbL, and PsbM. They interact closely with D1/D2 (**Fig. 1.4**) to deliver excitation energy to the RC working either as light absorbers themselves or facilitating EET from peripheral LHCs. CP43 and CP47 also help maintain the overall structural integrity of the RC. CP43 additionally contributes to the stabilization of the OEC itself by providing a direct Mn-coordinating ligand (Glu354).

1.3.6 The extrinsic proteins of PSII

The PSII core complex (PSII-cc), comprising the D1, D2, CP43, CP47, and Cyt b559 (PsbE, PsbF) membrane subunits, is crucial for O_2 evolution and photoautotrophic growth cyanobacteria to higher plants. PSII complexes with only these intrinsic subunits can produce O_2 at low rates, however optimal O_2 evolution requires additional extrinsic subunits. These include PsbO, which is conserved in all photosynthetic organisms, and either PsbU, PsbV, CyanoQ, and CyanoP found in cyanobacteria, or PsbP, PsbQ, and PsbR, found in eukaryotes. **Fig. 1.11b** depicts the three extrinsic subunits in *T. vulcanus*: PsbO (33 kDa), PsbU (12 kDa) and PsbV (15 kDa) based on the high-resolution structure of PSII. The interactions of these extrinsic proteins, with the extensive lumenal domains of the intrinsic proteins CP43, CP47, D1 and D2, help maintain the proper environment for water oxidation.[90, 91, 93, 178-180]

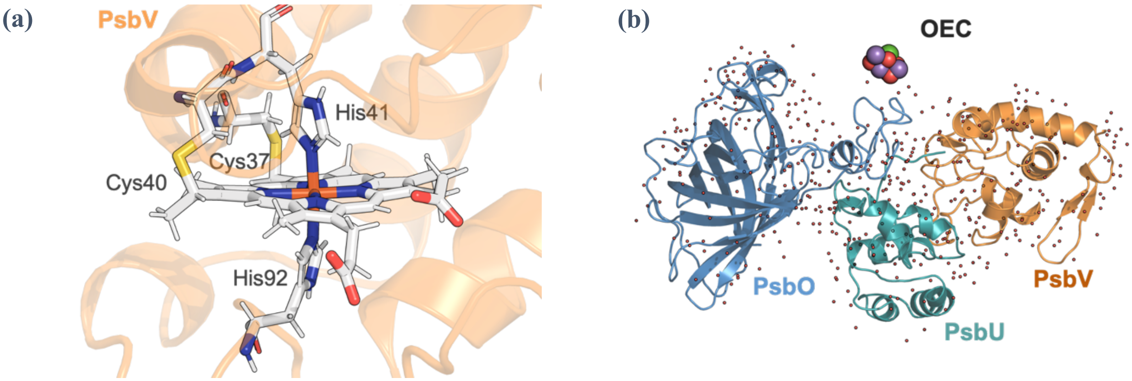


Fig. 1.11 (a) The three extrinsic proteins (**PsbO, PsbU**, **PsbV**) viewed with respect to the OEC (3WU2.pdb); red dots indicate position of crystal waters. **(b)** The heme active site of **Cytochrome c550** (PsbV) with the typical C−X−X−C−H binding motif. The porphyrin Fe center is hexa-coordinate with a bis-His ligation.

Across photosynthetic organisms, the PsbO protein stabilizes the Mn_4CaO_x cluster even at sub-optimal Cl^- concentrations and protects it from harmful exogenous reductants. Some studies suggest that PsbO and PsbU may aid in calcium binding at the active site, though the mechanisms are not fully understood. PsbV plays an accessory role in stabilizing the native conformation of the PSII-cc under physiological calcium and chloride levels. Additionally, PsbV has residues that are part of a longer hydrogen bond network linking the OEC to the PSII protein surface. Interestingly, the reported midpoint potential (E_m) of the heme active site (**Fig. 1.11a)** in free PsbV (−240 mV) differs significantly from that of PSII-bound PsbV (+200 mV), indicating that the physical association of PsbV with the PSII complex dramatically affects its E_m.[181-184] This may have potential implications towards the interaction of PsbV with its nearest redox neighbor, the OEC, though this hypothesis lacks direct evidence.

1.4 Spectroscopic Properties of Photosynthetic Pigments

Photosynthetic organisms employ a wide variety of porphyrin-type pigments for light harvesting, EET, charge separation, electron transfer (ET) and photo-protection. The main light absorbing pigments include chlorophylls (Chl a, Chl b, Chl c, Chl d, Chl f), bacteriochlorophylls (BChl a, BChl b, BChl c, BChl d, BChl e, BChl g), pheophytins (Pheo) and beta-carotenoids (β-Car). The Chls and BChls greatly vary in terms of their intrinsic photo-physical properties despite having structural similarities. BChl pigments are usually present in photosynthetic organisms living in the extreme and low-light conditions, while Chl a based photosynthesis is generally used by organisms surviving under normal or high-light conditions (e.g., higher plants). Interestingly, some long wavelength Chl variants[141-145,185] (Chl d and f) are synthesized and used in oxygenic photosynthesis by far-red light photo acclimated (FaRLiP) cyanobacteria.[79, 132, 139, 186-190]

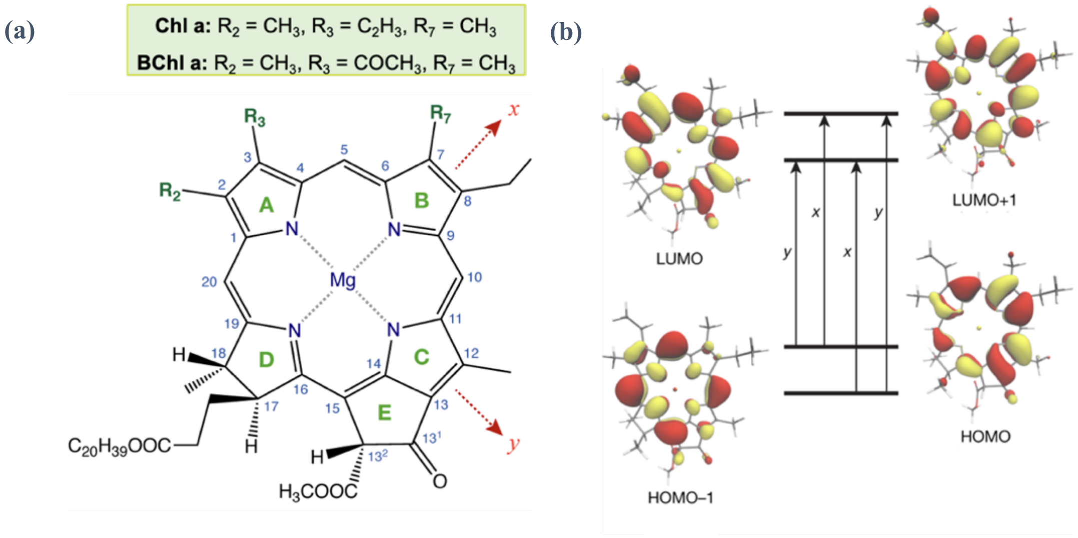


Fig. 1.12 (a) The structure of the chlorin macrocycle indicating the conventional atom numbering scheme and the molecular axes; the respective substituents for **Chl a** and **BChl a** are shown in the inset. In pheophytins (**Pheo**), the central Mg^{2+} ion is replaced by two protons on the N atoms of rings A and C. **(b)** Frontier molecular orbitals associated with the Q- and B- band transitions in Chl a according to the Gouterman model.[191] The arrows indicate orbital pairs with the highest contribution to the corresponding excitations; the labels indicate the approximate polarization of individual transitions. The figure is reproduced in part with permission from *J. Phys. Chem. B*, 2020, **124**, 8761-8771.[192]

The low-energy absorption spectrum of the Chl- or BChl-like pigments are described using the Gouterman model[191] based on the interactions between four frontier molecular orbitals: HOMO–1, HOMO, LUMO, and LUMO+1. Excitations among these orbitals produce the Q and B absorption bands (see **Fig. 1.12**). The Q-band appears after the absorption of red photons, while the B-band is a result of blue photon absorption. Each of these bands is further divided into two sub-bands, Q_y and Q_x, and B_y and B_x, based on the orientation of the transition dipole within the macrocyclic ring plane. The y-labeled features correspond to HOMO → LUMO and HOMO–1 → LUMO+1 transitions, whereas the x-labeled features correspond to HOMO–1 → LUMO and HOMO → LUMO+1 transitions. The Q_y excitation, being the lowest energy excitation, is most critical for light-harvesting and EET processes.[192]

Chlorophyll pigments play key roles such as EET and CS within the PSII-cc.[94, 95, 97, 98, 193, 194] The former mainly involves local excitations (LE), whereas the latter involves charge-transfer (CT) excitations,[103, 150] that act as precursors to primary CS and ET (e.g., $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ or $\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ states in the PSII-RC, see **Fig. 1.13** and **Chapters 4–6**). These pigments can also have characteristic CT states mixed with local excitons in some cases. Moreover, relative stabilities of the charge separated radical pairs determine pathways of charge recombination and secondary ET processes. Intrinsically, the spectroscopic and redox properties of the pigments can be tuned by the macrocyclic ring curvature, vinyl group rotation and various geometric and electronic factors imposed by the protein,[103, 146, 147, 149, 195] including axial ligation, hydrogen bonding, excitonic coupling, and local electrostatics.[94, 95, 136, 196-199] These combined intrinsic and extrinsic factors make each pigment photochemically unique within the protein scaffold,[200] optimizing their properties to maximize photosynthetic efficiency.

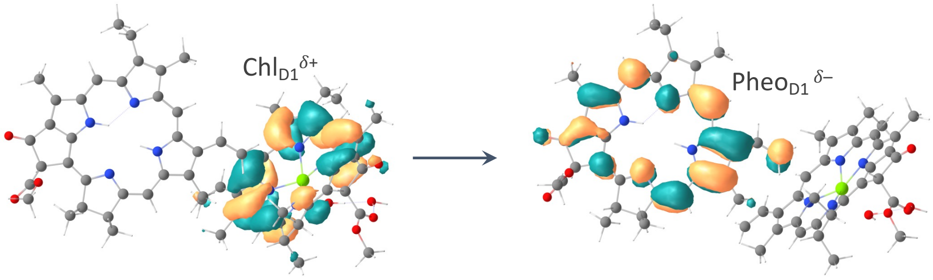


Fig. 1.13 Description of the lowest charge-transfer (**CT**) excited state within the PSII-RC with significant $\mathbf{Chl_{D1}}^{\delta+}\mathbf{Pheo_{D1}}^{\delta-}$ character in terms of Natural Transition Orbitals (NTOs) obtained from TD-DFT and QM/MM calculations.

This thesis attempts to address several key research questions in photosynthesis, with regards to each of the above aspects. We employ QM/MM based multiscale modeling approaches on the membrane-bound PSII protein, combining carefully benchmarked quantum chemical (QC) methods with classical MM/MD simulations to compute physiologically relevant spectroscopic and redox properties of photosynthetic cofactors. The work detailed in **Chapters 4–6** showcases the use of various multiscale computational approaches to achieve an electronic level description of the major light-activated processes in Photosystem II.

1.5 Scope of this Thesis

In this section, we briefly introduce the motivation and main research objectives in this thesis.

In **Chapter 2**, the applicability of the domain-based pair natural orbital implementation of coupled cluster theory (DLPNO-CC) is explored for calculating ionization energies and redox potentials of hydrated 3d transition metal ions. The study involves various solvation models, focusing on minimally explicitly hydrated QM clusters. The conductor-like polarizable continuum model (CPCM) is used to determine self-consistent redox potentials at the coupled cluster level. Additionally, multilayer (QM/QM) approaches combining higher-level DLPNO-CCSD(T) with a lower-level description of the second solvation layer are elucidated. The multilayer approach to DLPNO-CCSD(T) paves the way for employing chemically accurate yet computationally affordable local correlation methods in the investigation of more complex open-shell systems, both in the context of explicit solvation and in the case of redox-active cofactors embedded in protein matrices.

In **Chapter 3**, we provide a detailed description of the construction of the all-atom molecular-mechanics (MM) model of the membrane embedded PSII monomer and its variants. The MD and QM/MM protocol described in this Chapter follows the work described in **Chapters 4, 5** and **6**.

In **Chapter 4**, we investigate the nature and properties of triplet states within the PSII-RC using a multiscale QM/MM approach. The low-energy spectrum of excited singlet and triplet states, of both local and charge-transfer nature, is described using TD-DFT. We compute the EPR properties of relaxed triplet states and compare them with available experimental data. The electrostatic modulation of excited state energetics and redox properties of RC pigments by the reduced plastoquinone Q_A^- (closed RC) is described for the first time. This work provides a detailed electronic-level understanding of triplet states within the PSII-RC and form a refined basis for understanding primary and secondary ET, charge recombination and photo-protection mechanisms in PSII.

The D1 protein in cyanobacterial PSII is encoded by the *psbA* gene family that expresses distinct isoforms (PsbA1–3) depending on environmental conditions. Most differences in D1 isoforms are close to the active-branch RC pigments (P_{D1}, P_{D2}, Chl_{D1} and $Pheo_{D1}$). In **Chapter 5**, we combine MD simulations with multiscale TD-DFT/MM calculations on the membrane-bound PSII monomer of each PSII variant to compare the redox and excited state properties of RC pigments. We identify specific amino acid substitutions responsible for electrochromic shifts on distinct pigments and pigment groups. Our results indicate that the Pheo$_{D1}$ acceptor is the primary regulatory target. The redox properties of the Chl$_{D1}$–Pheo$_{D1}$ pair and Chl$_{D1}^{\delta+}$Pheo$_{D1}^{\delta-}$ CT states are distinctly modulated in the three isoforms. This work provides a microscopic description of how genetic variations modulate protein electrostatics and influence primary processes in photosynthetic RCs.

The CP43 protein interacts most closely with the active-branch (D1) of the RC. Understanding the function of CP43 requires detailed atomic-level insights into the properties of the embedded pigments. In **Chapter 6**, we employ a range of multiscale computational approaches to determine the site energies and excitonic profile of CP43 Chls, in the PSII monomer. In addition to TD-DFT/MM, we present a thorough analysis using the perturbed matrix method (PMM), which enabled us to utilize information from long-timescale MD simulations of native PSII-complexed CP43. Our study indicates a layered distribution of site energies and reveal specific groups of Chls that have shared contributions to low-energy excitations and this contribution changes as a function of conformational dynamics. We identified for the first time, a low-energy CT excited state within CP43 that involves two Chl a pigments. This work provides a refined basis for structure-based interpretation of spectroscopic observations and for further deciphering EET in oxygenic photosynthesis.

1.6 REFERENCES

1. P. Saura, M. Röpke, A. P. Gamiz-Hernandez and V. R. I. Kaila, in *Biomolecular Simulations: Methods and Protocols*, eds. M. Bonomi and C. Camilloni, Springer New York, New York, NY, 2019, DOI: 10.1007/978-1-4939-9608-7_4, pp. 75-104.
2. M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708-2728.
3. H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198-1229.
4. D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, . . . M. W. van der Kamp, *WIREs Computational Molecular Science*, 2019, **9**, e1393.
5. S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife and D. R. Salahub, *Int. J. Quantum Chem.*, 2018, **118**.
6. X. Sheng and F. Himo, *Acc. Chem. Res.*, 2023, **56**, 938-947.
7. M. R. A. Blomberg, T. Borowski, F. Himo, R.-Z. Liao and P. E. M. Siegbahn, *Chem. Rev.*, 2014, **114**, 3601-3658.
8. P. E. M. Siegbahn and F. Himo, *WIREs Computational Molecular Science*, 2011, **1**, 323-336.
9. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227-249.
10. M. G. Quesne, T. Borowski and S. P. de Visser, *Chem. Eur. J.*, 2016, **22**, 2562-2581.
11. E. Brunk and U. Rothlisberger, *Chem. Rev.*, 2015, **115**, 6217-6263.
12. S. C. L. Kamerlin and A. Warshel, *WIREs Computational Molecular Science*, 2011, **1**, 30-45.
13. W. Wang, O. Donini, C. M. Reyes and P. A. Kollman, *Annu. Rev. Biophys.*, 2001, **30**, 211-243.
14. S. A. Hollingsworth and R. O. Dror, *Neuron*, 2018, **99**, 1129-1143.
15. B. Mennucci, in *Green Chemistry and Computational Chemistry*, ed. L. Mammino, Elsevier, 2022, pp. 263-279.
16. R. E. Amaro and A. J. Mulholland, *Nat. Rev. Chem.*, 2018, **2**, 0148.
17. H. Klem, M. McCullagh and R. S. Paton, *Top. Catal.*, 2022, **65**, 165-186.
18. S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes and M. J. Ramos, *WIREs Computational Molecular Science*, 2017, **7**, e1281.
19. J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999-3094.
20. C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161-2200.
21. J. Chen, B. Chan, Y. Shao and J. Ho, *Phys. Chem. Chem. Phys.*, 2020, **22**, 3855-3866.
22. S. Bhattacharjee, M. Isegawa, M. Garcia-Ratés, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2022, **18**, 1619-1632.
23. A. Altun, F. Neese and G. Bistoni, *J. Chem. Theory Comput.*, 2020, **16**, 6142-6149.
24. M. Sparta, M. Retegan, P. Pinski, C. Riplinger, U. Becker and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 3198-3207.
25. Y. Guo, C. Riplinger, D. G. Liakos, U. Becker, M. Saitow and F. Neese, *J. Chem. Phys.*, 2020, **152**, 024116.
26. Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo and F. Neese, *J. Chem. Phys.*, 2018, **148**, 011101.
27. M. Saitow, U. Becker, C. Riplinger, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2017, **146**, 164105.
28. C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.
29. C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
30. C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652-1671.

31. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, . . . K. Morokuma, *Chem. Rev.*, 2015, **115**, 5678-5796.
32. B. M. Floser, Y. Guo, C. Riplinger, F. Tuczek and F. Neese, *J. Chem. Theory Comput.*, 2020, **16**, 2224-2235.
33. Q. M. Phung, M. Feldt, J. N. Harvey and K. Pierloot, *J. Chem. Theory Comput.*, 2018, **14**, 2446-2455.
34. M. Drosou, C. A. Mitsopoulou and D. A. Pantazis, *J. Chem. Theory Comput.*, 2022, **18**, 3538-3548.
35. M. Ansari, S. Bhattacharjee and D. A. Pantazis, *J. Am. Chem. Soc.*, 2024, **146**, 9640-9656.
36. S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.*, 1986, **7**, 230-252.
37. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Computational Molecular Science*, 2013, **3**, 198-210.
38. X. Zhu, P. E. M. Lopes and A. D. MacKerell Jr, *WIREs Computational Molecular Science*, 2012, **2**, 167-185.
39. B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, . . . M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1614.
40. M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, . . . W. F. van Gunsteren, *J. Comput. Chem.*, 2005, **26**, 1719-1751.
41. W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, . . . W. F. van Gunsteren, *J. Phys. Chem. A*, 1999, **103**, 3596-3607.
42. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236.
43. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657-1666.
44. Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046-4063.
45. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
46. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, . . . A. D. Mackerell Jr., *J. Comput. Chem.*, 2010, **31**, 671-690.
47. P. Reinholdt, J. Kongsted and F. Lipparini, *J. Chem. Theory Comput.*, 2022, **18**, 344-356.
48. M. Nottoli and F. Lipparini, *J. Chem. Phys.*, 2020, **153**, 224108.
49. C. M. Baker, *WIREs Computational Molecular Science*, 2015, **5**, 241-254.
50. I. Antes and W. Thiel, in *Combined Quantum Mechanical and Molecular Mechanical Methods*, American Chemical Society, 1998, vol. 712, ch. 4, pp. 50-65.
51. J. Blumberger, *Chem. Rev.*, 2015, **115**, 11191-11238.
52. K. M. Merz, Jr., *Acc. Chem. Res.*, 2014, **47**, 2804-2811.
53. D. Shevela, L. O. Björn and Govindjee, *Photosynthesis: Solar Energy for Life*, World Scientific, Singapore, 2017.
54. V. Krewald, M. Retegan and D. A. Pantazis, *Principles of natural photosynthesis, Solar energy for fuels*, 2016, 23-48.
55. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*, Wiley, Chichester, 2nd edn., 2014.
56. Y. Yoneda, E. A. Arsenault, S. Yang, Jr., K. Orcutt, M. Iwai and G. R. Fleming, *Nat. Commun.*, 2022, **13**, 2275.
57. T. Cardona, A. Sedoud, N. Cox and A. W. Rutherford, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 26-43.

58. E. Romero, I. H. M. Van Stokkum, V. I. Novoderezhkin, J. P. Dekker and R. Van Grondelle, *Biochemistry*, 2010, **49**, 4300-4307.
59. Y. Miloslavina, M. Szczepaniak, M. G. Müller, J. Sander, M. Nowaczyk, M. Rögner and A. R. Holzwarth, *Biochemistry*, 2006, **45**, 2436-2442.
60. K. Broess, G. Trinkunas, C. D. Van Der Weij-De Wit, J. P. Dekker, A. Van Hoek and H. Van Amerongen, *Biophys. J.*, 2006, **91**, 3776-3786.
61. G. H. Schatz, H. Brock and A. R. Holzwarth, *Biophys. J.*, 1988, **54**, 397-405.
62. D. A. Pantazis, in *Hydrogen Production and Energy Transition*, ed. M. Van de Voorde, De Gruyter, 2021, vol. 1, pp. 427-468.
63. N. Cox, D. A. Pantazis and W. Lubitz, *Annu. Rev. Biochem.*, 2020, **89**, 795-820.
64. J. P. McEvoy and G. W. Brudvig, *Chem. Rev.*, 2006, **106**, 4455-4483.
65. H. Li, Y. Nakajima, E. Nango, S. Owada, D. Yamada, K. Hashimoto, . . . J.-R. Shen, *Nature*, 2024, DOI: 10.1038/s41586-023-06987-5.
66. D. J. Vinyard, G. M. Ananyev and G. Charles Dismukes, *Annu. Rev. Biochem.*, 2013, **82**, 577-606.
67. P. E. M. Siegbahn, *Biochim. Biophys. Acta Bioenerg.*, 2013, **1827**, 1003-1019.
68. Y. Umena, K. Kawakami, J.-R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
69. N. Nelson and C. F. Yocum, *Annu. Rev. Plant Biol.*, 2006, **57**, 521-565.
70. B. Loll, J. Kern, W. Saenger, A. Zouni and J. Biesiadka, *Nature*, 2005, **438**, 1040-1044.
71. M. Sarewicz, M. Szwalec, S. Pintscher, P. Indyka, M. Rawski, R. Pietras, . . . A. Osyczka, *Science Advances*, 2023, **9**, eadd9688.
72. A. N. Tikhonov, *Photosynth. Res.*, 2023, DOI: 10.1007/s11120-023-01034-w.
73. L. A. Malone, M. S. Proctor, A. Hitchcock, C. N. Hunter and M. P. Johnson, *Biochim. Biophys. Acta Bioenerg.*, 2021, **1862**, 148380.
74. J. E. Johnson and J. A. Berry, *Photosynth. Res.*, 2021, **148**, 101-136.
75. A. N. Tikhonov, *Plant Physiology and Biochemistry*, 2014, **81**, 163-183.
76. K. Kato, R. Nagao, Y. Ueno, M. Yokono, T. Suzuki, T.-Y. Jiang, . . . J.-R. Shen, *Nat. Commun.*, 2022, **13**.
77. J. Li, N. Hamaoka, F. Makino, A. Kawamoto, Y. Lin, M. Rögner, . . . G. Kurisu, *Communications Biology*, 2022, **5**.
78. A. Naschberger, L. Mosebach, V. Tobiasson, S. Kuhlgert, M. Scholz, A. Perez-Boerema, . . . A. Amunts, *Nat. Plants*, 2022, DOI: 10.1038/s41477-022-01253-4.
79. E. Schlodder, F. Lendzian, J. Meyer, M. Çetin, M. Brecht, T. Renger and N. V. Karapetyan, *J. Am. Chem. Soc.*, 2014, **136**, 3904-3918.
80. S. Ohashi, T. Iemura, N. Okada, S. Itoh, H. Furukawa, M. Okuda, . . . M. Kobayashi, *Photosynth. Res.*, 2010, **104**, 305-319.
81. H. Yu, T. Hamaguchi, Y. Nakajima, K. Kato, K. Kawakami, F. Akita, . . . J.-R. Shen, *Biochim. Biophys. Acta Bioenerg.*, 2021, **1862**, 148471.
82. K. Kato, N. Miyazaki, T. Hamaguchi, Y. Nakajima, F. Akita, K. Yonekura and J.-R. Shen, *Communications Biology*, 2021, **4**.
83. M. Suga, F. Akita, M. Sugahara, M. Kubo, Y. Nakajima, T. Nakane, . . . J.-R. Shen, *Nature*, 2017, **543**, 131-135.
84. J.-R. Shen, *Annu. Rev. Plant Biol.*, 2015, **66**, 23-48.
85. T. Renger and F. Müh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3348-3371.
86. A. P. Casazza, M. Szczepaniak, M. G. Müller, G. Zucchelli and A. R. Holzwarth, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 1606-1616.
87. T. M. Bricker and L. K. Frankel, *Photosynth. Res.*, 2002, **72**, 131-146.
88. F. L. De Weerd, I. H. M. Van Stokkum, H. Van Amerongen, J. P. Dekker and R. Van Grondelle, *Biophys. J.*, 2002, **82**, 1586-1597.
89. L.-X. Shi and W. P. Schröder, *Biochim. Biophys. Acta Bioenerg.*, 2004, **1608**, 75-96.

90. T. M. Bricker, J. L. Roose, R. D. Fagerlund, L. K. Frankel and J. J. Eaton-Rye, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 121-142.
91. I. Enami, A. Okumura, R. Nagao, T. Suzuki, M. Iwai and J.-R. Shen, *Photosynth. Res.*, 2008, **98**, 349-363.
92. I. Enami, T. Suzuki, O. Tada, Y. Nakada, K. Nakamura, A. Tohri, . . . J.-R. Shen, *FEBS J.*, 2005, **272**, 5020-5030.
93. J. R. Shen and Y. Inoue, *Biochemistry*, 1993, **32**, 1825-1832.
94. R. Croce and H. v. Amerongen, *Science*, 2020, **369**, eaay2058.
95. F. Müh and A. Zouni, *Protein Sci.*, 2020, **29**, 1090-1119.
96. R. Croce, R. v. Grondelle, H. v. Amerongen and I. v. Stokkum, eds., *Light Harvesting in Photosynthesis*, Taylor & Francis/CRC Press, Boca Raton, 2018.
97. R. Croce and H. van Amerongen, *Nat. Chem. Biol.*, 2014, **10**, 492-501.
98. H. van Amerongen and R. Croce, *Photosynth. Res.*, 2013, **116**, 251-263.
99. M. Mamedov, Govindjee, V. Nadtochenko and A. Semenov, *Photosynth. Res.*, 2015, **125**, 51-63.
100. M. Capone, A. Sirohiwal, M. Aschi, D. A. Pantazis and I. Daidone, *Angew. Chem. Int. Ed.*, 2023, **62**, e202216276.
101. M. Gorka, A. Baldansuren, A. Malnati, E. Gruszecki, J. H. Golbeck and K. V. Lakshmi, *Frontiers in Microbiology*, 2021, **12**, 735666.
102. H. Tamura, K. Saito and H. Ishikita, *Chem. Sci.*, 2021, **12**, 8131-8140.
103. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Am. Chem. Soc.*, 2020, **142**, 18174-18190.
104. N. P. Pawlowicz, M. L. Groot, I. H. M. Van Stokkum, J. Breton and R. Van Grondelle, *Biophys. J.*, 2007, **93**, 2732-2742.
105. J. P. Dekker and R. Van Grondelle, *Photosynth. Res.*, 2000, **63**, 195-208.
106. Y. Sato and D. Sicotte, *Catalysts*, 2020, **10**, 718.
107. F. Müh, M. Plöckinger and T. Renger, *J. Phys. Chem. Lett.*, 2017, **8**, 850-858.
108. O. P. Kaminskaya and V. A. Shuvalov, *Dokl. Biochem. Biophys.*, 2016, **466**, 39-42.
109. H. A. Chu and Y. F. Chiu, *Front Plant Sci*, 2015, **6**, 1261.
110. J. A. Myers, K. L. M. Lewis, F. D. Fuller, P. F. Tekavec, C. F. Yocum and J. P. Ogilvie, *J. Phys. Chem. Lett.*, 2010, **1**, 2774-2780.
111. A. Kamlowski, L. Frankemöller, A. Van Der Est, D. Stehlik and A. R. Holzwart, *Ber. Bunsen-Ges. Phys. Chem.*, 1996, **100**, 2045-2051.
112. K. E. Shinopoulos and G. W. Brudvig, *Biochim. Biophys. Acta*, 2012, **1817**, 66-75.
113. D. Narzi, D. Bovi, P. De Gaetano and L. Guidoni, *J. Am. Chem. Soc.*, 2016, **138**, 257-264.
114. K. Saito, T. Ishida, M. Sugiura, K. Kawakami, Y. Umena, N. Kamiya, . . . H. Ishikita, *J. Am. Chem. Soc.*, 2011, **133**, 14379-14388.
115. S. D. Causmaecker, J. S. Douglass, A. Fantuzzi, W. Nitschke and A. W. Rutherford, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 19458-19463.
116. K. Saito, A. W. Rutherford and H. Ishikita, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 954-959.
117. F. Muh, C. Glockner, J. Hellmich and A. Zouni, *Biochim. Biophys. Acta*, 2012, **1817**, 44-65.
118. M. Judd, J. Morton, D. Nürnberg, A. Fantuzzi, A. W. Rutherford, R. Purchase, . . . E. Krausz, *Biochim. Biophys. Acta Bioenerg.*, 2020, **1861**, 148248.
119. Y. Takegawa, M. Nakamura, S. Nakamura, T. Noguchi, J. Selles, A. W. Rutherford, . . . M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2019, **1860**, 297-309.
120. W. Lubitz, F. Lendzian and R. Bittl, *Acc. Chem. Res.*, 2002, **35**, 313-320.
121. E. J. G. Peterman, H. van Amerongen, R. van Grondelle and J. P. Dekker, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6128-6133.

122. T. Renger and E. Schlodder, *ChemPhysChem*, 2010, **11**, 1141-1153.
123. G. Raszewski, W. Saenger and T. Renger, *Biophys. J.*, 2005, **88**, 986-998.
124. V. I. Novoderezhkin, J. P. Dekker and R. van Grondelle, *Biophys. J.*, 2007, **93**, 1293-1311.
125. T. Renger and E. Schlodder, *Journal of Photochemistry and Photobiology B: Biology*, 2011, **104**, 126-141.
126. M. L. Groot, N. P. Pawlowicz, L. J. G. W. Van Wilderen, J. Breton, I. H. M. Van Stokkum and R. Van Grondelle, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13087-13092.
127. G. Raszewski, B. A. Diner, E. Schlodder and T. Renger, *Biophys. J.*, 2008, **95**, 105-119.
128. I. V. Shelaev, F. E. Gostev, V. A. Nadtochenko, A. Y. Shkuropatov, A. A. Zabelin, M. D. Mamedov, . . . V. A. Shuvalov, *Photosynth. Res.*, 2008, **98**, 95-103.
129. A. Sirohiwal and D. A. Pantazis, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200356.
130. H. H. Nguyen, Y. Song, E. L. Maret, Y. Silori, R. Willow, C. F. Yocum and J. P. Ogilvie, *Science Advances*, 2023, **9**, eade7190.
131. Y. Silori, R. Willow, H. H. Nguyen, G. Shen, Y. Song, C. J. Gisriel, . . . J. P. Ogilvie, *J. Phys. Chem. Lett.*, 2023, **14**, 10300-10308.
132. D. J. Nürnberg, J. Morton, S. Santabarbara, A. Telfer, P. Joliot, L. A. Antonaru, . . . A. W. Rutherford, *Science*, 2018, **360**, 1210-1213.
133. F. Mokvist, J. Sjöholm, F. Mamedov and S. Styring, *Biochemistry*, 2014, **53**, 4228-4238.
134. J. R. Reimers, M. Biczysko, D. Bruce, D. F. Coker, T. J. Frankcombe, H. Hashimoto, . . . E. Krausz, *Biochim. Biophys. Acta Bioenerg.*, 2016, **1857**, 1627-1640.
135. A. Gelzinis, D. Abramavicius, J. P. Ogilvie and L. Valkunas, *J. Chem. Phys.*, 2017, **147**, 115102.
136. D. Narzi, E. Coccia, M. Manzoli and L. Guidoni, *Biophys. Chem.*, 2017, **229**, 93-98.
137. J. P. Allen and J. C. Williams, in *The Biophysics of Photosynthesis*, eds. J. Golbeck and A. van der Est, Springer New York, New York, NY, 2014, DOI: 10.1007/978-1-4939-1148-6_9, pp. 275-295.
138. S. I. Allakhverdiev, T. Tomo, Y. Shimada, H. Kindo, R. Nagao, V. V. Klimov and M. Mimuro, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 3924-3929.
139. T. Renger and E. Schlodder, *J. Phys. Chem. B*, 2008, **112**, 7351-7354.
140. H. Yamasaki, H. Nakamura and Y. Takano, *Chem. Phys. Lett.*, 2007, **447**, 324-329.
141. C. Friedl, D. G. Fedorov and T. Renger, *Phys. Chem. Chem. Phys.*, 2022, **24**, 5014-5038.
142. F. Müh, M. E.-A. Madjet and T. Renger, *Photosynth. Res.*, 2012, **111**, 87-101.
143. M. E. Madjet, A. Abdurahman and T. Renger, *J. Phys. Chem. B*, 2006, **110**, 17268-17281.
144. E. Cignoni, V. Slama, L. Cupellini and B. Mennucci, *J. Chem. Phys.*, 2022, **156**, 120901.
145. M. Nottoli, L. Cupellini, F. Lipparini, G. Granucci and B. Mennucci, *Annu. Rev. Phys. Chem.*, 2021, **72**, 489-513.
146. L. Cupellini, M. Corbella, B. Mennucci and C. Curutchet, *WIREs Computational Molecular Science*, 2019, **9**, e1392.
147. C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294-343.
148. O. Andreussi, S. Knecht, C. M. Marian, J. Kongsted and B. Mennucci, *J. Chem. Theory Comput.*, 2015, **11**, 655-666.
149. B. Mennucci, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6583-6594.
150. A. Sirohiwal and D. A. Pantazis, *Acc. Chem. Res.*, 2023, **56**, 2921-2932.

151. Y. Kato and T. Noguchi, *Photosynth. Res.*, 2022, DOI: 10.1007/s11120-021-00894-4.
152. E. Martin, R. I. Samoilova, K. V. Narasimhulu, T.-J. Lin, P. J. O'Malley, C. A. Wraight and S. A. Dikanov, *J. Am. Chem. Soc.*, 2011, **133**, 5525-5537.
153. R. J. Wei, Y. Zhang, J. Mao, D. Kaur, U. Khaniya and M. R. Gunner, *Photosynth. Res.*, 2022, DOI: 10.1007/s11120-022-00906-x.
154. N. Kulik, M. Kutý and D. Řeha, *Journal of Molecular Modeling*, 2020, **26**.
155. A. Fantuzzi, F. Allgöwer, H. Baker, G. McGuire, W. K. Teh, A. P. Gamiz-Hernandez, . . . A. W. Rutherford, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2116063119.
156. J. J. S. Van Rensen, C. Xu and Govindjee, *Physiol. Plant.*, 1999, **105**, 585-592.
157. A. Sirohiwal and D. A. Pantazis, *J. Am. Chem. Soc.*, 2022, **144**, 22035-22050.
158. A. Cuni, L. Xiong, R. Sayre, F. Rappaport and J. Lavergne, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4825-4831.
159. R. de Wijn and H. J. van Gorkom, *Biochim. Biophys. Acta Bioenerg.*, 2002, **1553**, 302-308.
160. F. Rappaport, M. Guergova-Kuras, P. J. Nixon, B. A. Diner and J. Lavergne, *Biochemistry*, 2002, **41**, 8518-8527.
161. F. Van Mieghem, K. Brettel, B. Hillman, A. Kamlowski, A. W. Rutherford and E. Schlodder, *Biochemistry*, 1995, **34**, 4798-4813.
162. B. Hillmann, K. Brettel, F. Van Mieghem, A. Kamlowski, A. W. Rutherford and E. Schlodder, *Biochemistry*, 1995, **34**, 4814-4827.
163. A. W. Rutherford, A. Osyczka and F. Rappaport, *FEBS Lett.*, 2012, **586**, 603-616.
164. K. Yamaguchi, S. Yamanaka, H. Isobe, M. Shoji, T. Kawakami and K. Miyagawa, in *Photosynthesis: Molecular Approaches to Solar Energy Conversion*, eds. J.-R. Shen, K. Satoh and S. I. Allakhverdiev, Springer International Publishing, Cham, 2021, DOI: 10.1007/978-3-030-67407-6_2, pp. 39-80.
165. D. A. Pantazis, *ACS Catal.*, 2018, **8**, 9477-9507.
166. D. A. Pantazis, W. Ames, N. Cox, W. Lubitz and F. Neese, *Angew. Chem. Int. Ed.*, 2012, **51**, 9935-9940.
167. K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber and S. Iwata, *Science*, 2004, **303**, 1831-1838.
168. J. Yano, J. Kern, K. Sauer, M. J. Latimer, Y. Pushkar, J. Biesiadka, . . . V. K. Yachandra, *Science*, 2006, **314**, 821-825.
169. J. Kern, R. Chatterjee, I. D. Young, F. D. Fuller, L. Lassalle, M. Ibrahim, . . . V. K. Yachandra, *Nature*, 2018, **563**, 421-425.
170. V. Krewald, M. Retegan, N. Cox, J. Messinger, W. Lubitz, S. Debeer, . . . D. A. Pantazis, *Chem. Sci.*, 2015, **6**, 1676-1695.
171. A. Bhowmick, R. Hussein, I. Bogacz, P. S. Simon, M. Ibrahim, R. Chatterjee, . . . V. K. Yachandra, *Nature*, 2023, **617**, 629-636.
172. M. Askerka, J. Wang, D. J. Vinyard, G. W. Brudvig and V. S. Batista, *Biochemistry*, 2016, **55**, 981-984.
173. M. Chrysina, M. Drosou, R. G. Castillo, M. Reus, F. Neese, V. Krewald, . . . S. DeBeer, *J. Am. Chem. Soc.*, 2023, **145**, 25579-25594.
174. Y. Guo, L. He, Y. Ding, L. Kloo, D. A. Pantazis, J. Messinger and L. Sun, *Nat. Commun.*, 2024, **15**, 5982.
175. P. Greife, M. Schönborn, M. Capone, R. Assunção, D. Narzi, L. Guidoni and H. Dau, *Nature*, 2023, **617**, 623-628.
176. R. Hussein, A. Graça, J. Forsman, A. O. Aydin, M. Hall, J. Gaetcke, . . . W. P. Schröder, *Science*, 2024, **384**, 1349-1355.
177. D. Kaur, K. Reiss, J. Wang, V. S. Batista, G. W. Brudvig and M. R. Gunner, *J. Phys. Chem. B*, 2024, **128**, 2236-2248.

178. K. Ifuku and T. Noguchi, *Frontiers in Plant Science*, 2016, **7**, 84.
179. C. Uno, R. Nagao, H. Suzuki, T. Tomo and T. Noguchi, *Biochemistry*, 2013, **52**, 5705-5707.
180. I. Enami, M. Iwai, A. Akiyama, T. Suzuki, A. Okumura, T. Katoh, . . . J.-R. Shen, *Plant Cell Physiol.*, 2003, **44**, 820-827.
181. P. Bernal-Bayard, C. Alvarez, P. Calvo, C. Castell, M. Roncel, M. Hervas and J. A. Navarro, *Physiol. Plant.*, 2019, **166**, 199-210.
182. M. Roncel, D. Kirilovsky, F. Guerrero, A. Serrano and J. M. Ortega, *Biochim. Biophys. Acta*, 2012, **1817**, 1152-1163.
183. F. Guerrero, A. Sedoud, D. Kirilovsky, A. W. Rutherford, J. M. Ortega and M. Roncel, *J. Biol. Chem.*, 2011, **286**, 5985-5994.
184. J. S. Vrettos, M. J. Reifler, O. Kievit, K. V. Lakshmi, J. C. de Paula and G. W. Brudvig, *J Biol Inorg Chem*, 2001, **6**, 708-716.
185. M. Chen, M. Schliep, R. D. Willows, Z.-L. Cai, B. A. Neilan and H. Scheer, *Science*, 2010, **329**, 1318-1319.
186. E. Elias, T. J. Oliver and R. Croce, *Annu. Rev. Phys. Chem.*, 2024, **75**, 231-256
187. C. J. Gisriel, G. Shen, D. A. Flesher, V. Kurashov, J. H. Golbeck, G. W. Brudvig, . . . D. A. Bryant, *J. Biol. Chem.*, 2023, **299**, 102815.
188. C. J. Gisriel, T. Cardona, D. A. Bryant and G. W. Brudvig, *Microorganisms*, 2022, **10**, 1270.
189. E. Schlodder, M. Çetin, H.-J. Eckert, F.-J. Schmitt, J. Barber and A. Telfer, *Biochim. Biophys. Acta Bioenerg.*, 2007, **1767**, 589-595.
190. S. Itoh, H. Mino, K. Itoh, T. Shigenaga, T. Uzumaki and M. Iwaki, *Biochemistry*, 2007, **46**, 12473-12481.
191. A. Ceulemans, W. Oldenhof, C. Gorller-Walrand and L. G. Vanquickenborne, *J. Am. Chem. Soc.*, 1986, **108**, 1155-1163.
192. A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsak and D. A. Pantazis, *J. Phys. Chem. B*, 2020, **124**, 8761-8771.
193. V. Mascoli, V. Novoderezhkin, N. Liguori, P. Xu and R. Croce, *Biochim. Biophys. Acta Bioenerg.*, 2020, **1861**, 148156.
194. V. I. Novoderezhkin, R. Croce, M. Wahadoszamen, I. Polukhina, E. Romero and R. Van Grondelle, *Phys. Chem. Chem. Phys.*, 2016, **18**, 19368-19377.
195. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2021, **17**, 1858-1873.
196. S. Sen, V. Mascoli, N. Liguori, R. Croce and L. Visscher, *J. Phys. Chem. A*, 2021, **125**, 4313-4322.
197. M. Aghtar, U. Kleinekathöfer, C. Curutchet and B. Mennucci, *J. Phys. Chem. B*, 2017, **121**, 1330-1339.
198. C. Ramanan, M. Ferretti, H. Van Roon, V. I. Novoderezhkin and R. Van Grondelle, *Phys. Chem. Chem. Phys.*, 2017, **19**, 22877-22886.
199. D. I. G. Bennett, K. Amarnath and G. R. Fleming, *J. Am. Chem. Soc.*, 2013, **135**, 9164-9173.
200. A. Sirohiwal and D. A. Pantazis, *Phys. Chem. Chem. Phys.*, 2021, **23**, 24677-24684.

2

Ionization Energies and Redox Potentials of Hydrated Transition Metal Ions: Evaluation of DLPNO Coupled Cluster Approaches[‡]

[‡]This chapter was reprinted with permission from “Ionization Energies and Redox Potentials of Hydrated Transition Metal Ions: Evaluation of Domain-Based Local Pair Natural Orbital Coupled Cluster Approaches” by Bhattacharjee, S.; Isegawa, M.; Garcia-Ratés, M.; Neese, F. and Pantazis, D. A., *J. Chem. Theory Comput.* **2022,** *18*, 1619-1632.
https://doi.org/10.1021/acs.jctc.1c01267

2.1 Introduction

Redox processes involving transition metal ions are important in a wide range of chemical and biological processes. For example, the variation of the redox level on transition metal sites[1, 2] plays an integral role in the function of synthetic catalysts[3-5] and is at the heart of fundamental enzymatic processes,[6-8] including the most critical energy converting transformations in biology.[6, 9-12] Obtaining accurate energetics[13-16] for spin-state and redox changes[17] in such systems is challenging and the treatment of the electronic structure problem places heavy demands both on the definition of the computational model in terms of the appropriate representation of the coordination environment,[18, 19] and on the electronic structure method.[20-26] The latter problem is particularly acute in view of the significant errors that can be encountered for larger transition metal systems.[13, 21, 27] Ionization energies[28-30] and redox potentials[14, 17, 20, 28-38] are in this respect crucial target properties that can be used to evaluate the capabilities and limitations of the different components that define the computational approach.

Density functional theory (DFT) methods[21, 33, 36-46] in conjunction with implicit solvation models[40, 47-50] are widely used for describing transition metal systems owing to the simplicity, low cost and often satisfactory performance of appropriately chosen functionals within sets of closely related chemical systems.[33, 51, 52] Nevertheless, DFT has limitations when dealing with complex electronic structure situations such as those encountered in open-shell transition metal complexes.[21] Numerous studies have highlighted the role of modern wave function based methods to address the challenge of spin-state or redox energetics in transition metal systems[53-58] and it is expected that such approaches may become a standard component of a future robust and generally applicable theoretical protocol.[59, 60] Radoń *et al.* applied multireference calculations (CASPT2[61] and NEVPT2[62, 63]) in studying ligand field transitions of aqua complexes of the first-row transition metal ions,[53, 64] pointing out that benchmark studies on transition metal clusters are prone to significant errors arising from the choice of solvation strategy. Noodleman and co-workers[37] applied a cluster model to the Mn^{2+}/Mn^{3+} and Fe^{2+}/Fe^{3+} pairs in aqueous solution and showed the importance of including explicit water molecules in the second solvation shell to increase the accuracy of predicted redox potentials. Uudsemaa *et al.*[36] also applied the cluster approach and pointed out discrepancies between experimental data and theoretical predictions for the spin state of Co and Ni ions in aqueous solution. Wang *et al.* calculated redox potentials with a QM/MM approach for 3d transition metals and highlighted the importance of solute–solvent hydrogen bonding.[35] In addition, previous studies employing reported the use of wave function theory in calculations of gas phase ionization energies and aqueous redox potentials using continuum solvation models for organic systems.[32] Studies of ionization energies and redox potentials have highlighted that the results can be significantly improved if coupled cluster theory[65-67] is used in place of DFT to calculate changes in electronic energies. In particular gas phase ionization energies were shown to improve considerably by using coupled cluster theory.[59, 65, 67, 68]

The most popular way to compute solvation energies are implicit solvation models.[47] An example is the polarizable continuum model (PCM)[69-71] and its different variants, like the conductor-like PCM (CPCM).[72, 73] Within the PCM, the solute-solvent interaction is represented by a collection of charges spread over the surface of a cavity that contains the solute. Although PCM describes electrostatic solvation effects, the non-electrostatic solvation component of the solvation process can be calculated by means of the Solvation Model based on Density[74] (SMD). A more complicated scheme is the conductor-like screening model for realistic solvents (COSMO-RS)[75] which combines quantum mechanics with statistical thermodynamics. Either the bare CPCM, the SMD or the COSMO-RS model have been used in combination with DFT to predict aqueous oxidation potentials or interaction energies of organic compounds.[48, 76] Studies on organic molecules showed that both COSMO[77] and SMD[74] perform similarly for the solvation energy of neutral species, but the accuracy is compromised with increasing charge, making the solvation energy the limiting factor in achieving the same level of accuracy for redox potentials as for ionization energies.[32]

For systems with specific solute–solvent intermolecular interactions, the nature of the solvent molecules is clearly different in the first solvation shell than in the bulk of the solvent. This aspect is not properly taken into account by implicit solvation models but can be addressed to some extent by the use of explicitly solvated cluster models, in which a number of solvent molecules that coordinate to the solute are treated at the same quantum chemical footing as the solute. At the same time, the treatment of the solvent as an unstructured continuum with a fixed dielectric constant can introduce severe errors particularly in cases of specific solute-solvent interactions such as hydrogen bonding in protic solvents.[35] In principle one can use an extensive multistep QM/MM approach to deal with the short-range interactions where more layers of solvent can be included[78] but even with a minimal inclusion of a single additional layer of solvent molecules, the improvement in the results can be impressive in cases of strong solvent-solute coupling.[34, 78-81] Such is the case for the transition metal cations in aqueous solution that form the subject of the present work.

Explicitly solvated systems are difficult to model and might even be impractical in combination with expensive electronic structure methods. The recent availability of a near linear-scaling local correlation method for open-shell systems, the domain-based local pair natural approach to coupled cluster theory with singles, doubles, and perturbative triples, DLPNO-CCSD(T),[82-85] paves the way for more affordable and at the same time accurate calculations of redox processes of even larger systems. As demonstrated in the study by Isegawa *et al.*, improved gas phase ionization energies do not necessarily translate into improved aqueous redox potentials, because the continuum solvation model may dominate as the main source of error for calculated redox potentials.[32] In the demanding case of aqueous transition metal complexes a major contributing factor is the change in solvation free energy accompanying the redox process.

The present work focuses on adiabatic ionization energies and redox potentials of explicitly hydrated 3d transition metal ions using the DLPNO-CCSD(T) method that was shown to recover most of the canonical correlation energy at a fraction of the computational cost compared to canonical coupled cluster.[68, 86-88] Since solvation-related errors tend to overshadow the improved electronic energies obtained from a reliable wave function method –in our case, DLPNO-CCSD(T)– here we adopt the explicit inclusion of a second solvation sphere, which allows a consistent treatment of close-range interactions at the same high level of electronic structure theory rather than by a continuum model. We place emphasis on the dependence of ionization energies on the explicit second solvation shell. We further use the recently implemented perturbation theory energy scheme with singles PTE(S)[89] as a protocol to compute solvation free energies for the oxidized and reduced species at the coupled-cluster level and thus derive standard redox potentials self-consistently. Correlations with experimental redox potentials are discussed. The conditions for obtaining systematically converged results are not self-evident, therefore we examine several methodological parameters that may affect the accuracy and reliability of the approach. Among others, we investigate the role of the different implementations of the perturbative triples corrections[82, 83, 90] and we also analyze the convergence of ionization energies with respect to the dimension of the PNO space.[91, 92] A major challenge involves establishing a suitable approach to deal with even bigger transition metal clusters, such as for example systems with more extended explicit solvent shells. Toward this goal, we demonstrate a multilayer DLPNO-CCSD(T) approach in which different PNO accuracy thresholds are employed for different regions/shells of a system,[93] and show that this approach holds great promise for the cost-effective treatment of large systems.

2.2 Theory and Methodology

2.2.1 Explicitly Hydrated Models

The adiabatic ionization energies (IEs) in aqueous solution for the first-row transition metals Ti–Cu are considered in the current study using (i) six-water coordinated models (**M-W6**), and (ii) 18-water coordinated models (**M-W18**) (**Fig. 2.1**). The first type of model involves six water molecules directly bonded octahedrally to the metal ion, whereas the latter further incorporates an explicit second shell of twelve additional water molecules. These twelve water molecules are hydrogen bonded to the first solvation shell leading to a $[M(H_2O)_6(H_2O)_{12}]^{n+}$ system. More than one configurations of the 18-water cluster have been considered previously.[36] Here we adopt the configuration that was reported by Radoń *et al.* to be the most stable conformation for such complexes in studies of spin-state energetics.[53, 64]

The exact coordination number of the metal ions in aqueous solution is not always obvious. For example, there have been similar conclusions from both theoretical calculations and experimental observations for four-, five- as well as six-waters coordinated to copper,[81, 94-98] suggesting they can potentially coexist owing to the very small energy differences involved. However, since hexa-coordination is the most common hydration pattern for the majority of the first-row transition metal ions, and to ensure consistency in the present approach, throughout this work we used models where the transition metal ions have six water molecules in their first coordination sphere.

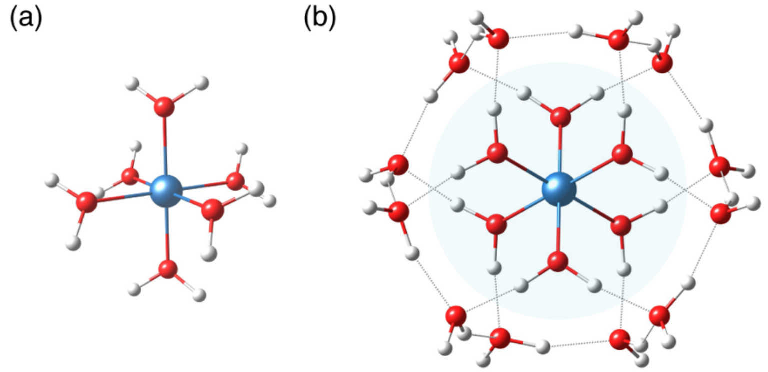


Fig. 2.1 (**a**) Structure corresponding to the cluster model of $[M(H_2O)_6]^{2+/3+}$ in which six water molecules coordinate to the central metal ion (**M-W6**). (**b**) Explicitly solvated cluster models of $[M(H_2O)_6.(H_2O)_{12}]^{2+/3+}$ (**M-W18**) [M= Ti, V, Cr, Mn, Fe, Co, Ni, Cu]. The most stable structure has been considered.

2.2.2 Electronic configurations

The set of the eight aqueous transition metal complexes were targeted among others because experimental redox potential values for most of them are known with reasonable accuracy and because most of them may undergo one-electron redox reactions in the chosen oxidation states without other associated chemical activity. In principle, both high (HS) and low spin (LS) states are possible for $[Cr(H_2O)_6]^{2+}$, $[Mn(H_2O)_6]^{2+}$, $[Mn(H_2O)_6]^{3+}$, $[Fe(H_2O)_6]^{2+}$, $[Fe(H_2O)_6]^{3+}$, $[Co(H_2O)_6]^{2+}$, $[Co(H_2O)_6]^{3+}$ and $[Ni(H_2O)_6]^{3+}$. However, experiments suggest that aqua complexes of Cr, Mn and Fe exist in the HS state whereas the LS state is predominant for Co(III) ions. In this study, we considered high spin states for all ions except $[Co(H_2O)_6]^{3+}$, and both spin states for $[Co(H_2O)_6]^{2+}$ and $[Ni(H_2O)_6]^{3+}$, although only the most stable one will be treated at all levels. The charge and corresponding spin multiplicities were kept consistent in the 18-water cluster models as well.

2.2.3 Geometry Optimizations

All geometry optimizations were carried out with a development version of ORCA 5.0.[99-101] All calculations were performed with the second-order Douglas-Kroll-Hess Hamiltonian (DKH2) to include scalar relativistic effects.[102-104] The different complexes were optimized with DFT using the hybrid TPSSh[105-107] functional with D3(BJ)[108-110] dispersion corrections and the DKH-def2-TZVP(-f)[111] basis set. Tight convergence and optimization criteria (TightSCF, TightOpt) and a fine grid (Grid6, Gridx6) was used. To speed up the calculations, the RIJCOSX[112-114] approximation was used in conjunction with the SARC/J fitting basis,[115-121] which is the decontracted version of the def2/J auxiliary basis sets for elements up to Kr.[122] The optimized coordinates for the M-W6 and M-W18 models are listed in the Supporting

Information. The effect of the bulk solvent (H_2O) on the M–O bond lengths was investigated through CPCM. In ORCA, the solvation charges on the surface of the solute cavity are treated as spherical Gaussians through the Gaussian charge scheme together with a switching function to accept or discard them.[123, 124] In particular, we adopt the GVDW scheme. More details on this scheme, that is, the type of solute cavity, number of charges per sphere, and radii for the spheres in the cavity can be found in the paper by Garcia-Ratés *et al.*[124] The CPCM scheme adopted in the DLPNO-CCSD(T) calculations is described in the corresponding subsection below.

2.2.4 Electronic Structure Calculations

For the DLPNO-CC calculations, Kohn-Sham determinants computed with the DFT-TPSSh functional were used as reference. This choice was found by experience to be associated with more well-behaved convergence of the CC calculations. It is noted that we give up on Brillouin's theorem due to this choice but the emerging off-diagonal Fock matrix elements are properly taken into account by the ORCA implementation. To avoid any misconception, it is stressed that despite the fact the reference determinant is a DFT determinant, the final DLPNO-CC energy does not contain any DFT component whatsoever. The second-order DKH2 Hamiltonian[102, 103] was employed in all calculations. For open-shell molecules, the energy was obtained on the basis of quasi-restricted orbitals (QROs).[86] The perturbative triples contributions were computed using the recently published iterative T_1 algorithm for both closed-shell[125] and open-shell systems.[82, 83, 90] All SCF calculations were performed in the absence of any approximations with a convergence criterion of 10^{-9} Hartree (VeryTightSCF). The 3s and 3p outer core orbitals were included in the correlation treatment, while the 1s and 2s inner core orbitals were kept frozen.[126] The large automatically generated "AutoAux" fitting basis set[127] was used where required in correlated wave function calculations. The three truncation parameters "T_{CutPNO}", "T_{CutPairs}" and "T_{CutMKN}", which define cutoffs for the occupation numbers in the pair natural orbitals, for the estimated pair correlation energies, and for the fitting domain selection, were chosen according to built-in settings, using the NormalPNO and TightPNO defaults. For each model (**M-W6**, **M-W18**), the correlation consistent triple zeta basis set cc-pwCVTZ-DK[28, 121] was used on the metal and cc-PVTZ-DK[128-131] for the rest of the molecule.

For a more detailed quantitative analysis of the DLPNO-CCSD(T) results, we used the open-shell variant of the local energy decomposition (LED) scheme[132-134] to obtain the inter-fragment energy terms for the individual layers of solvation. This approach quantifies the relative contributions of the metal, the first solvation sphere and the rest of the cluster respectively to the final energy difference for the redox pairs.

Recently, a systematic method to approach the complete PNO space limit in DLPNO-CCSD(T) calculations was proposed.[91] The correlation energies obtained by varying the T_{CutPNO} threshold parameters were extrapolated using a two-point extrapolation scheme, keeping all other parameters of the DLPNO calculations to the default TightPNO settings. The best fit for the dependence of the correlation energy on the parameter X (where $T_{\text{CutPNO}} = 10^{-X}$) is of the following functional form:

$$E^X = E + AX^{-\beta} \tag{2.1}$$

Here we tested this approach to investigate the dependence of the DLPNO-CCSD(T) ionization energies on the dimension of the PNO space ($T_{\text{CutPNO}} = 10^{-X}$, where $X = 5, 6, 7,$ and 8) using the Fe systems as a test case. The two-point extrapolated energy can be represented as

$$E\ (X/X+1) = E^X + F(E^{X+1} - E^X) \tag{2.2}$$

We use F=1.5 for the current work, as suggested originally.[91]

2.2.5 Calculation of Ionization Energies and Redox Potentials

Throughout this work the adiabatic ionization energy (IE) of the transition metal is defined as the difference of the total electronic energy between the $\mathbf{M}^{3+}$ and $\mathbf{M}^{2+}$ form in eV, computed at the DLPNO-CCSD(T) level of theory, without further thermodynamic corrections:

$$\text{IE} = E(\text{M}^{3+}) - E(\text{M}^{2+}) \tag{2.3}$$

The aqueous reduction potential (E^0) of the metal ion is defined as:

$$E^0 = \frac{\Delta G_{\text{ox}}}{nF} - \text{SHE} \tag{2.4}$$

$$\Delta G_{\text{ox}} = \text{IE} + \Delta G^0_{\text{solv.ox}} - \Delta G^0_{\text{solv.red}} \tag{2.5}$$

Reduction potentials are generally tabulated as standard half-cell potentials against a standard reference electrode. Considerable effort has been put towards establishing the absolute electrochemical half-cell SHE potential in different solvents, and different values in the range from 4.24 to 4.73 V have been suggested in the literature.[17, 135] Here we employ the value of 4.28 V (excluding surface potential) which is the most recommended value.[136] We obtain the above solvation free energy term (ΔG_{ox}) directly from the DLPNO-CPCM computations. An accurate estimation of the solvation free-energies[30] for the oxidized and reduced species will lead to an accurate prediction of the standard electrode reduction potentials for each redox pair and the energy obtained is assumed to contain intrinsically the correction to the solvation free energies for the oxidized and reduced species.[137] There exist different approaches to include the effect of the solvent in coupled cluster calculations, each of them with a different degree of complexity.[138, 139] The simplest of these schemes is the so-called "perturbation theory energy (PTE)" scheme, where the PCM contributions occur through the reference energy and the Fock matrix (solvated orbitals). A further level is the PTE(S) approach, where "S" stands for singles, which includes an extra solvation term in the correlation energy with respect to the PTE scheme. Neither the PTE scheme nor the PTE(S) scheme involve explicit corrections to the equations to compute the CC excitations ("T" amplitudes). In the present study, we use the PTE(S) scheme, which has been recently implemented in ORCA 5.0 for open-shell systems,[89, 140] to compute the solvation free energies both for the oxidized and reduced species. It is noted that the various approximate schemes show a high degree of consistency and hence that errors arising from the approximation of the solvation terms in the cluster equations must be very small, much smaller than the errors intrinsic in the implicit solvation schemes.

2.2.6 Multilevel QM/QM Scheme for Truncation Thresholds

The accuracy of DLPNO-CCSD(T) can also be controlled by fine tuning the T_{CutPNO} and T_{CutPairs} and T_{CutMKN} thresholds.[91, 92] When it comes to larger systems, the cost can still become limiting if high-accuracy settings are applied uniformly. In this work, we demonstrate that for the case of the **M-W18** systems one can effectively treat different parts of the molecule at different PNO settings instead of treating the entire molecule at a single level of accuracy.[93] This can be compared to a multilevel QM/QM approach where the different coordination spheres around the central metal ion are treated with different methods, but here the method is the same, albeit with different accuracy settings for each layer. In practice, we divided each M-W18 model into two hypothetical fragments or layers, based on the fact that the inner solvation shell is expected to be more critical in determining the absolute energies of the different oxidation states than the second shell. Therefore, we assigned the metal along with the six directly coordinated water molecules as the first layer and the rest of the waters as the second layer (**Fig. 2.1**). TightPNO settings ($T_{\mathrm{CutPairs}} = 10^{-5}$, $T_{\mathrm{CutPNO}} = 10^{-7}$, $T_{\mathrm{CutMKN}} = 10^{-4}$) were assigned to the first layer and NormalPNO ($T_{\mathrm{CutPairs}} = 10^{-4}$, $T_{\mathrm{CutPNO}} = 3.33 \times 10^{-7}$, $T_{\mathrm{CutMKN}} = 10^{-3}$) to the second layer. The inter-fragment interaction between the two layers was treated using TightPNO settings. An extension of this approach involved more approximate wave function methods for the low-level layer. Here we further elaborated on the multilevel scheme by treating the pair energies of the second layer at the second order Møller–Plesset (MP2)[141, 142] and at the Hartree–Fock (HF) level of theory. Global TightPNO settings and default FrozenCore settings for Orca 5 were used throughout in these calculations.

2.3 Results and Discussion

2.3.1 Geometries

All the geometry optimizations in this work have been carried out without any symmetry constraints and resulted in an approximately octahedral arrangement of the ligands around the central metal ion **(Fig. 2.1)**. The models under investigation could in principle possess molecular symmetry as high as S_6. However, the orbital degeneracies in the ground states for several of the aqua complexes lead to Jahn–Teller distortions. This is seen to result in the axial elongation/compression of the metal-oxygen (M-O) bond lengths in high-spin $[Cr(H_2O)_6]^{2+}$, $[Mn(H_2O)_6]^{3+}$, $[Fe(H_2O)_6]^{2+}$, low-spin $[Co(H_2O)_6]^{2+}$ and $[Ni(H_2O)_6]^{3+}$, and $[Cu(H_2O)_6]^{2+}$ respectively. The metal-ligand bond distances corresponding to the inner solvation sphere are listed in **Table 2.1**. In addition, the tetragonal distortion is strongest for the ^5E_g ground states in high-spin Cr^{2+} and Mn^{3+} (both d^4) as well as Cu^{2+} (d^9), arising from the lifting of the degeneracy for the single electron in the e_g level (M–L σ antibonding).

As expected, metal-ligand bond lengths are shorter for the divalent ions compared to the trivalent ones, which arises from a stronger metal-oxygen interaction in higher oxidation states. From the trends in M-O bond distances in **Table 2.1** one can also draw conclusions about the extent of short range (explicitly using 12-H_2O) and effective long range (implicitly using CPCM) solvation effects. The effect of solvation may be expected to be stronger for metal ions which show more variations in bond lengths. For almost all systems, the use of CPCM leads to contraction of the average metal-ligand bond lengths by ca. 0.04 Å with respect to the unsolvated **M-W6**. For the **M-W18** clusters, a further shortening by ca. 0.02 Å is observed. The slight contraction is usually accompanied by elongation in the O-H bond distances of the directly coordinated water ligands, implying that second-sphere solvation effects might result in first-sphere ligands to coordinate more strongly to the metal ion via the O atoms. The addition of explicit water molecules, however, does not systematically change the metal-ligand distances compared to the implicit case and the trend is more metal-dependent. The observation also highlights the importance of explicit solvation models for a few sensitive cases like Cu^{2+}, where the change in geometry is not captured well by the implicit model, which could be a consequence of the Jahn–Teller effect.

Table 2.1 Metal-ligand bond distances for the bare M-W6 complexes, for the M-W6 complexes with implicit CPCM(Water) solvation, for the M-W18 clusters, and for the M-W18 clusters with CPCM solvation, respectively. The range of values correspond to the Jahn-Teller distortion observed in ion complexes with a degenerate ground state. All optimizations were performed at the TPSSh-D3BJ/DKH-def2-TZVP(-f) level of theory.

TM Ion	Spin Multiplicity ($2S+1$)	M-W6	M-W6 + CPCM (H_2O)	M-W18
Ti^{2+}	3	2.196	2.173	2.165
Ti^{3+}	2	2.077	2.036	2.047
V^{2+}	4	2.138	2.128	2.127
V^{3+}	3	2.033	1.995	2.005
Cr^{2+}	5	2.073, 2.376	2.048, 2.359	2.051, 2.416
Cr^{3+}	4	1.994	1.967	1.975
Mn^{2+}	6	2.192	2.182	2.173
Mn^{3+}	5	1.956, 2.155	1.934, 2.112	1.933, 2.165
Fe^{2+}	5	2.113, 2.148	2.108, 2.130	2.112, 2.127
Fe^{3+}	6	2.040	1.996	2.016
Co^{2+}	4	2.098	2.090	2.082
Co^{3+}	1	1.918	1.888	1.901
Ni^{2+}	3	2.076	2.076	2.078
Ni^{3+} (LS)	2	1.880, 2.020	1.876, 2.036	1.886, 2.064
Ni^{3+} (HS)	4	2.000	1.990	1.970
Cu^{2+}	2	2.005, 2.281	2.001, 2.280	1.987, 2.331
Cu^{3+}	3	2.008	1.964	1.978

2.3.2 Ionization Energies for M-W6 Models

In recent work on the spin-state splitting of similar transition metal systems, an elaborate comparison has been made between the results from DLPNO-CCSD and canonical CCSD calculations.[88] The error associated with the DLPNO approximation in principle should yield absolute energy differences between the divalent and trivalent ions within the limits of chemical accuracy. However, non-negligible errors may arise from the treatment of perturbative triples. Until recently, the semi-canonical triples (denoted as T_0) had been the usual option as this method can be implemented quite efficiently for closed as well as open shell systems, leading to reasonably accurate relative energies with respect to the canonical results. However, it has been reported that particularly for certain open shell systems,[83, 85, 86] the T_0

results may deviate significantly from canonical triples and relative energy differences can be as high as 4 kcal/mol. Such errors may deteriorate the results for bigger systems like those investigated in our case. As an improvement, the iterative triples DLPNO-CCSD(T_1), recently implemented in ORCA for both closed and open shell species,[83, 85] yield in principle more accurate results on the triples correction.[88] The absolute energies from the SD, semi-canonical (T_0) as well as iterative triples correction (T_1) using the DLPNO approach, are provided in the SI. The behavior of the T_1 was almost consistent for the divalent and trivalent metal complexes studied here. If we look into the computed ionization energies, the DLPNO-CCSD(T_1) results have a mean difference of about 0.4 eV from those of the DLPNO-CCSD values (**Table 2.2**). Also, the differences in computed IEs between the semi-canonical and T_1 approach are rather small. On the other hand, the computational time for (T_1) was significantly higher compared to the semi-canonical approach. In addition, we note that the open shell DLPNO-CC calculations of the trivalent species were systematically more expensive than the divalent ones for all the metals, Co being the sole exception. This is due to the better convergence of the closed-shell algorithm.

Table 2.2 Comparison of $M^{3+/2+}$ Ionization Energies (in eV) at the DLPNO-CCSD(T_1), DLPNO-CCSD(T_0), and DLPNO-CCSD levels for the **M-W6** cluster. All values reported here were computed using the cc-pwCVTZ-DK and cc-PVTZ-DK basis sets on the metal and the water ligands respectively. Default TightPNO thresholds were used throughout.

Redox pair	**DLPNO-CCSD(T_1)**	**DLPNO-CCSD(T_0)**	**DLPNO-CCSD**
Ti^{2+} / Ti^{3+}	14.38	14.40	14.56
V^{2+} / V^{3+}	15.69	15.71	15.91
Cr^{2+} / Cr^{3+}	15.35	15.37	15.61
Mn^{2+} / Mn^{3+}	17.70	17.73	18.03
Fe^{2+} / Fe^{3+}	16.26	16.26	16.43
Co^{2+} / Co^{3+}	18.40	18.51	18.24
Ni^{2+} / Ni^{3+}	19.47	19.53	20.11
Cu^{2+} / Cu^{3+}	19.40	19.41	19.75

It is noteworthy that the ground state electronic configuration has a significant role to play for these species, Co and Ni being the only first-row transition metals frequently reported to exist in low spin states in their aqueous solutions. Further, trivalent Co is the only closed-shell species in our investigation. For Co, the extent of splitting of the d-orbitals for the divalent and trivalent ions, a coexistence of spin states, and/or dimerization in the solution phase or deprotonation of a water ligand might also be plausible explanations to theoretical predictions being different for the experimentally observed configuration. For most part of our analysis, we consider the spin state which is electronically more stable.

Table 2.3 Ionization energies (in eV) and computation times for DLPNO-CCSD(T1) calculations with TightPNO settings on the Fe clusters, varying the TCutPNO threshold. All calculations were performed on 8 cores with 24GB of memory available per core (%maxcore 24000, in ORCA nomenclature).

$[Fe(H_2O)_6]^{n+}$				**$[Fe(H_2O)_{18}]^{n+}$**		
	Fe(II)	**Fe(III)**		**Fe(II)**	**Fe(III)**	
T_{CutPNO}	**Time (hh:mm)**	**Time (hh:mm)**	**IE**	**Time (hh:mm)**	**Time (hh:mm)**	**IE**
10^{-5}	01:02	01:18	15.74	06:52	11:43	11.25
10^{-6}	01:44	02:12	16.15	21:35	37:01	11.67
10^{-7}	03:30	04:14	16.26	31:25	55:12	11.83
10^{-8}	06:57	08:06	16.31			

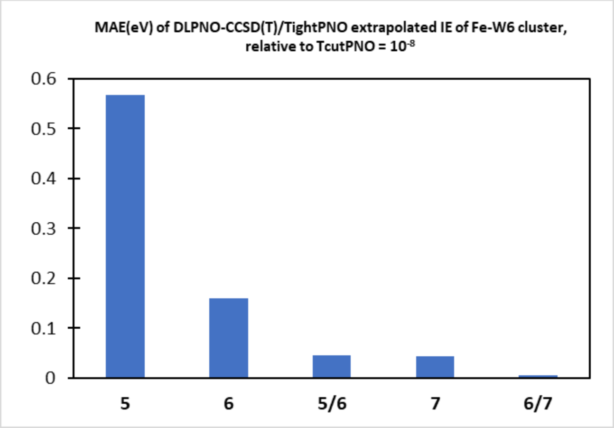


Fig. 2.2 Extrapolation to the PNO space limit. The numbers in the horizontal axis denote the exponent in $T_{CutPNO} = 10^{-x}$.

We also investigated the approach to the complete PNO space[91] limit of the DLPNO-CCSD(T) calculations, with respect to a particular basis set (cc-pwCVTZ-DK and cc-PVTZ-DK in our case) using the $[Fe(H_2O)_6]^{2+/3+}$ complex as a test case. The results are tabulated in **Table 2.3**. An asymptotic behavior is observed for the computed ionization energies (IE) on tightening the T_{CutPNO} threshold by a factor of 10. We further estimated the values for $[Fe(H_2O)_6]^{2+/3+}$ using a two-point extrapolation with the functional form described in **Equation 2.2**. The mean absolute errors of the extrapolated $T_{CutPNO} = 10^{-5}/10^{-6}$ and $T_{CutPNO} = 10^{-6}/10^{-7}$ energies with respect to $T_{CutPNO} = 10^{-8}$ are represented in **Fig. 2.2**. The accuracy obtained from the $E(5/6)$ extrapolation (16.33 eV) is close to that of $T_{CutPNO} = 10^{-7}$. In terms of computation time, there is a systematic scaling observed for both the models. For the 6-H_2O complex, the computational times are doubled with tightening of the PNO threshold from $T_{CutPNO} = 10^{-5}$ to 10^{-6} and again from 10^{-6} to 10^{-7}. For the larger 18-H_2O model that contains the second layer of water, the same tightening of the thresholds leads to steeper, approximately threefold increase of the run times for a change of T_{CutPNO} from 10^{-5} to 10^{-6} but a lower increase (ca. 1.5×) upon further tightening to 10^{-7}.

2.3.3 Effect of the Second Solvation Sphere

In the quest for an optimal protocol that incorporates a better consideration of the solvent without rendering the computational model inaccessible to correlated wave function methods, a modest step toward explicitly solvated cluster models is to treat an additional solvation sphere at the same quantum chemical level as the solute. In the following we demonstrate the applicability of the DLPNO-CCSD(T) approach to compute the open-shell systems with a second solvation sphere and show that the obtained results are chemically sound and at the same time computationally affordable. The methods investigated here should be contrasted with DFT based approaches that typically exhibit strong dependence both on the choice of a specific functional and on the nature of the metal under consideration.

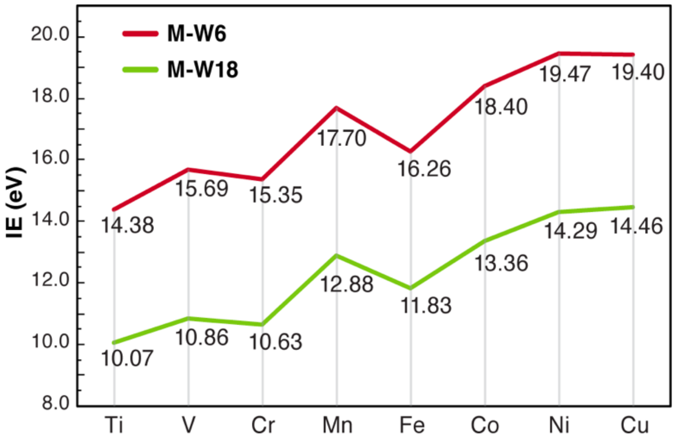


Fig. 2.3 Relative trends in DLPNO-CCSD(T)/TightPNO computed Ionization Energies (IEs) for each cluster model (M-W6 and M-W18).

The computed ionization energies for each of the metals using DLPNO-CCSD(T_1)/TightPNO are represented in **Fig. 2.3** for both models (**M-W6** and **M-W18**). The IEs show a general increasing trend across the period, as expected from the effective nuclear charges and ground state electronic configurations of the respective metals. In terms of numerical values, there is a consistent impressive difference of ca. 4–5 eV for the IEs corresponding to the M-W18 clusters with respect to the M-W6 values. There can be potentially two effects leading to this observation, namely the change in geometry and/or the electrostatic effects arising from the additional layer of water. However, when the ionization energies were computed using the geometries of the inner M[(H_2O)$_6$] system and excluding the additional layer of water, the effect was negligible. Therefore, we conclude that the significant difference of the IEs on adding the second layer derives predominantly from the electrostatic effect of the second layer that includes hydrogen-bonding, an aspect not considered by implicit solvation models.

To further probe the physical nature of the interaction between the first and second solvation layers, we performed an extensive local energy decomposition (LED) analysis for the specific example of the iron-water clusters. The LED analysis enables a rigorous decomposition of the total interaction energy into contributions arising from the reference (Hartree–Fock) component ($\Delta E_{\mathrm{int}}^{\mathrm{ref}}$) and the correlation energy, distinguished in the CCSD correlation energy $\Delta E_{\mathrm{int}}^{\mathrm{C-CCSD}}$, and the perturbative triples correlation energy contribution ($\Delta E_{\mathrm{int}}^{\mathrm{C-(T)}}$). **Fig. 2.4** depicts the distinct components that arise from the decomposition, and are defined as electronic preparation ($\Delta E_{\mathrm{el\text{-}prep}}$), electrostatic ($E_{\mathrm{elstat}}$), and exchange ($E_{\mathrm{ex}}$) in the case of ($\Delta E_{\mathrm{int}}^{\mathrm{ref}}$), and dispersive or non-dispersive terms for $\Delta E_{\mathrm{int}}^{\mathrm{C-CCSD}}$. Detailed results for the present test system are provided in **Table 2.4**, while here a summary of salient points will be given. Perhaps not surprisingly, the electrostatic interaction between the individual layers is quite dominant. Quantitatively, for both systems (with and without the second hydration shell), the electrostatic interactions within the first coordination sphere are stronger for the higher oxidation state by about three times compared to the reduced state (**Table 2.4**).

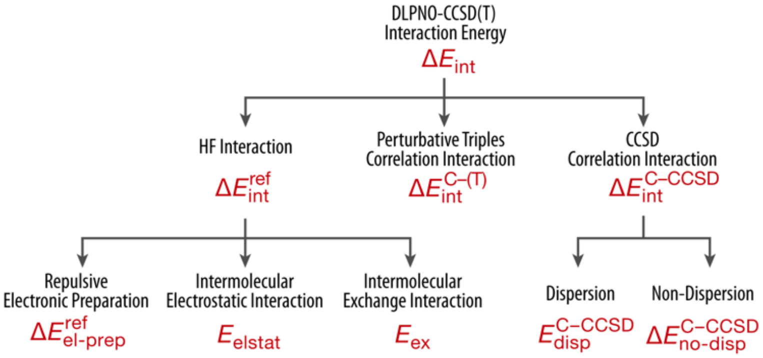


Fig. 2.4 Energy terms in the open-shell DLPNO-CCSD(T)/LED scheme.

It is noted that the combined electrostatic and exchange interactions for the metal (Fe) and the first solvation sphere (six directly coordinated H_2O molecules) is stronger in [Fe(H_2O)$_6$]$^{3+}$ compared to the [Fe(H_2O)$_{18}$]$^{3+}$ cluster, whereas for the lower oxidation state the interactions are slightly higher in the [Fe(H_2O)$_{18}$]$^{2+}$ cluster compared to the bare [Fe(H_2O)$_6$]$^{2+}$. Focusing on the 18-water cluster, the decomposition of the total interaction energy (ΔE_{int}) between the two solvation layers (**Table 2.4**) shows that the attractive interactions at the reference level are dominated by the electrostatic terms (E_{elstat}) for both oxidation states. The electronic preparation term ($\Delta E_{\text{el-prep}}$) however is very high and positive, which basically corresponds to the energy required to distort the electron densities of the individual layers from their ground state. It is important to note that the extent of these contributions depends on the oxidation state of the metal.

Table 2.4 Decomposition of DLPNO-CCSD(T_1) intra-fragment and inter-fragment interaction energies using the open-shell local energy decomposition (**LED**) approach, shown here for the $[Fe(H_2O)_6]$ and $[Fe(H_2O)_{18}]$ clusters. All values are reported in Hartrees.

	M-W18			**M-W6**				
	Fe^{3+}	Fe^{2+}	ΔE ($Fe^{3+} - Fe^{2+}$)	Fe^{3+}	Fe^{2+}	ΔE ($Fe^{3+} - Fe^{2+}$)	ΔE $[Fe(H_2O)_{18}]^{3+}$- $[Fe(H_2O)_6]^{3+}$	ΔE $[Fe(H_2O)_{18}]^{2+}$- $[Fe(H_2O)_6]^{2+}$
Intra-fragment terms								
6W	-447.6343	-453.3414	5.7071	-448.1687	-453.9766	5.8080	0.5343	0.6353
12W	-911.3893	-911.9813	0.5920					
Metal	-1253.2775	-1256.2922	3.0146	-1253.3711	-1256.3781	3.0070	0.0936	0.0860
Inter-fragment terms								
6W+12W								
Electrostatics (REF.)	-1.1828	-0.8019	-0.3809					
Exchange (REF.)	-0.2376	-0.1629	-0.0747					
Dispersion (strong pairs)	-0.0396	-0.0346	-0.0050					
Dispersion (weak pairs)	-0.0042	-0.0035	-0.0008					
M+6W								
Electrostatics (REF.)	-10.5449	-3.5861	-6.9589	-10.6421	-3.3617	-7.2804	0.0972	-0.2243
Exchange (REF.)	-1.3132	-0.4714	-0.8418	-1.3472	-0.4385	-0.9087	0.0340	-0.0329
Dispersion (strong pairs)	-0.0251	-0.0213	-0.0038	-0.0227	-0.0199	-0.0027	-0.0024	-0.0013
Dispersion (weak pairs)	-0.0109	-0.0062	-0.0047	-0.0110	-0.0061	-0.0049	0.0001	-0.0002
M+12W								
Electrostatics (REF.)	-0.8414	-0.2914	-0.5499					
Exchange (REF.)	-0.0495	-0.0068	-0.0427					
Dispersion (strong pairs)	-0.0005	0.0000	-0.0005					
Dispersion (weak pairs)	-0.0027	-0.0015	-0.0012					
Non dispersion (strong pairs)	-5.8454	-5.8345	-0.0108	-2.3242	-2.3174	-0.0069	-3.5211	-3.5171
Non dispersion (weak pairs)	-0.0107	-0.0201	0.0094	-0.0078	-0.0073	-0.0005	-0.0029	-0.0128

A similar trend is observed for the correlation energy contributions from CCSD, where the non-dispersive terms ($\Delta E_{\text{non-disp}}$), which represent the correction to HF-level electrostatics, are dominant. Hence, the larger stabilization of the Fe^{3+} complex is due to its larger electrostatic interaction compared to Fe^{2+}. The contribution from the perturbative triples ($\Delta E^{\text{C-(T)}}$) is comparatively negligible. The decomposition of the final difference in interaction energy between the two layers of solvation for the Fe$^{2+/3+}$ redox pair is provided in **Table 2.5**. We conclude that not only the total but also the individual interaction energy contributions between the co-ordination layers are dependent on the charge at the metal center. This is in line with the analysis by Wang *et al.*[35] that the heterogenous polarization of the solute electron density and the additional layer of water possibly leads to a decrease in the positive charge at the metal center, thereby lowering the energy difference between the redox pairs.

Table 2.5 Decomposition of Interaction energies between the two layers of solvation, for the [Fe(H_2O)$_{18}$] clusters using the DLPNO/CCSD(T) LED scheme. All values are in Eh.

	Reference energy			**Correlation energy**		
Ion	$\boldsymbol{E}_{\text{elstat}}$	$\boldsymbol{E}_{\text{ex}}$	$\Delta\boldsymbol{E}_{\text{elprep}}$	$\boldsymbol{E}_{\text{disp}}$	$\Delta\boldsymbol{E}_{\text{no-disp}}$	$\Delta\boldsymbol{E}^{\text{C-(T)}}$
Fe^{3+}	-12.5691	-1.6003	27.9052	-0.0830	-1.1259	-0.0152
Fe^{2+}	-4.6794	-0.6411	19.2412	-0.0558	-0.1666	-0.0126
ΔE_{int}	-7.8897	-0.9592	8.6639	-0.0272	-0.9593	-0.0026

2.3.4 Standard Reduction Potentials using a Cluster-Continuum Approach

Table 2.6 Solvation free energies and M$^{3+/2+}$ (M = Ti–Cu) standard reduction potentials computed using DLPNO-CCSD(T) in combination with PTE(S). The redox potentials are reported in V.

	M-W6		**M-W18**			
Redox pair	**Δ*G* solv. (CPCM)**	***E*⁰ vs SHE**	**Δ*G* solv. (CPCM)**	***E*⁰ vs SHE**	***E*⁰ ref.**	**Δ*E*⁰ (W6-W18)**
Ti^{2+} / Ti^{3+}	-10.40	-0.30	-6.93	-1.14	-0.90	0.84
V^{2+} / V^{3+}	-10.40	1.01	-6.90	-0.32	-0.26	1.33
Cr^{2+} / Cr^{3+}	-10.49	0.58	-7.02	-0.66	-0.41	1.24
Mn^{2+} / Mn^{3+}	-10.49	2.93	-7.02	1.58	1.54	1.35
Fe^{2+} / Fe^{3+}	-10.32	1.67	-6.92	0.63	0.77	1.04
Co^{2+} / Co^{3+}	-10.96	3.15	-7.13	1.95	1.92	1.20
Ni^{2+} / Ni^{3+} (LS)	-10.81	4.38	-7.05	2.96	[a]	1.42
Ni^{2+} / Ni^{3+} (HS)	-10.65	4.73	-6.98	3.15	[a]	1.58
Cu^{2+} / Cu^{3+}	-10.65	4.47	-7.08	3.10	[a]	1.37

[a] Experimental values do not exist for these pairs; estimates of 2.3V for Ni and 2.4V for Cu have been suggested.[143]

The **M-W18** clusters were used to combine DLPNO-CCSD(T) calculations and CPCM with the PTE(S)[89] scheme in the derivation of solvation energies that were used to estimate the standard reduction potentials of each M^{2+}/M^{3+} pair. The free energies of solvation and the calculated redox potentials with respect to the Standard Hydrogen Electrode (SHE) of 4.28 V, for both the **M-W6** and **M-W18** clusters are provided in **Table 2.6**. **Fig. 2.5** represents the correlation of the computed values with those reported from the experimental literature.[36, 143]

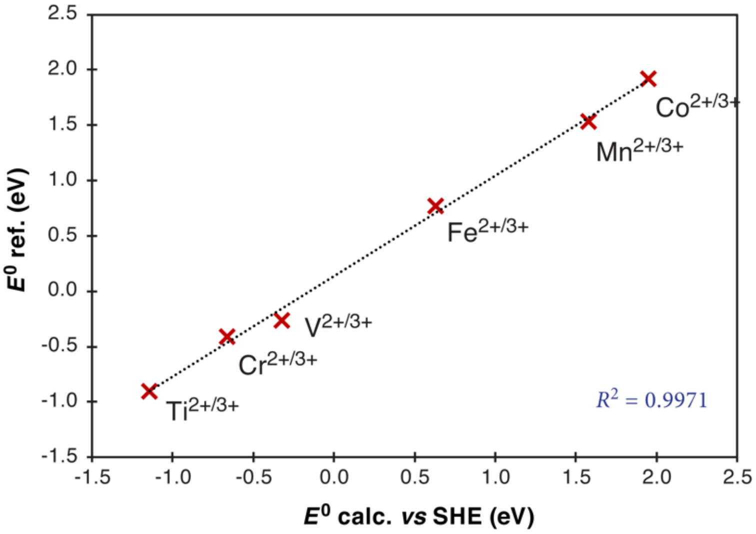


Fig. 2.5 Correlation plot of DLPNO/CCSD(T_1)/CPCM computed redox potentials with respect to experimental redox potentials for the M-W18 clusters (M = Ti–Co).

Most remarkably, the absolute solvation free energies are reduced by 3.4 to 3.8 eV compared to the bare M-W6 cluster on adding the extra layer of water. This leads to quantitative differences of more than 1 eV in the final redox potentials in all cases except Ti between the ΔE^0 computed with the **M-W6** and the **M-W18** models, but there is also an important qualitative distinction in terms of the change of sign for early transition metals. The experimentally observed M^{2+}/M^{3+} redox potentials are negative for Ti, V and Cr, and positive for Mn, Fe, and Co. This is only reproduced here with the **M-W18** models, while for V and Cr the hexa-aqua cluster predicts positive E^0 with respect to SHE. Overall, our estimated values for M-W6 have a mean absolute error of 1.07 V with respect to the reference values, which decreases to 0.13 V on the addition of a second layer of solvation. This result further stresses the importance of clearly identifying sources of error when explicit solvation is not considered to model redox processes in such systems.

It is noted that for Ni and Cu there are no reliably known experimental values, but only estimated suggestions (2.3 V for $Ni^{2+/3+}$ and 2.4 V for $Cu^{2+/3+}$). Therefore, we report the computed values here for these two pairs as reference DLPNO-CCSD(T) values without further analysis. Nevertheless, we note that the suggestions do not fit with the computed results, which nicely follow the trends for lighter elements up to Co and are fully consistent with the corresponding IE values.

At this point it is worth placing these results in the context of past studies that utilized DFT based cluster-continuum methods. In a well-known study, Noodleman and co-workers reported values of 1.59 V and 1.06 V for the Mn^{2+}/Mn^{3+} and Fe^{2+}/Fe^{3+} couple respectively.[37] Our method agrees quite well considering the respective experimental redox potentials reported to be 1.54 V and 0.77 V for the two metals. The values reported by Uudsemaa *et al.* using DFT computations on similar cluster models have a mean difference of 0.3 V from our estimations.[36] There also have been experimental reports on the spin states of Co^{2+} and Ni^{3+}, which stress the low-spin configurations as dominating in aqueous solution. Previous DFT based studies reported redox potential values for Co^{2+} and Ni^{3+} where the high spin state shows better agreement with experimental redox potentials. From our calculations, the trend is consistent for Co^{2+}. Also, here one should also keep in mind about the comparison of energetics between closed-shell and open shell species, which was not always accounted for in previous computational studies. Our results are consistent with the fact that the hydrated Co^{3+} complex should be considered in its low spin state for better agreement with experimental values. The values obtained with low-spin Ni^{3+} instead gives better agreement with previously estimated figures. Furthermore, the DLPNO-CCSD(T)/CPCM obtained energies for low spin Ni^{3+} complexes are consistently lower than the high spin counterpart by about 0.2–0.3 eV. As hypothesized, there can be several plausible explanations for this, like chemical transformations taking place in solution or other sources of error and uncertainty due to the co-existence of dimeric forms or of multiple spin states.

2.3.5 Evaluation of Multilayer DLPNO-based Approaches

In the preceding part of our study, we showed that accurate calculations of ionization energies and redox potentials at the full DLPNO-CCSD(T_1) level can be performed on the complete M-W18 systems. In this section we investigate if it is possible to obtain results of equivalent or comparable quality with lower computational cost by introducing approximations in the context of a multilayer approach. In the simplest example this corresponds to a two-level method for systems consisting of a clearly defined second coordination/solvation sphere, wherein a part of the system assumed to be chemically more important is computed with a higher-accuracy method than the rest of the molecule. The M-W18 systems are thus divided into an inner fragment (Layer 1), consisting of the metal ion surrounded by six waters, and an outer Layer 2 consisting of the second solvation sphere containing twelve waters. Layer 1 is always treated with the DLPNO-CCSD(T_1) method using TightPNO settings. For Layer 2 the following approximations have been considered:

(i) DLPNO-CCSD(T_1) with NormalPNO thresholds
(ii) Second-order Møller–Plesset (MP2) Perturbation Theory
(iii) Hartree-Fock (HF) theory

All of these methods are readily available in ORCA and are accessible through a suitable definition of fragments and uses of the existing multilayer DLPNO machinery. It is important to note that in all of the above two-layer approximations we chose to treat the more important inter-layer terms needed for the accurate description of the weak interactions with TightPNO thresholds. The resulting IEs are tabulated in **Table 2.7** and compared with those obtained from global TightPNO and NormalPNO settings on the entire system. **Fig. 2.6** compares the errors of the various approaches against the reference global TightPNO result. The results show that the two-layer-PNO approach where the second solvation shell is treated with NormalPNO settings is able to approximate the reference global-TightPNO calculations very well. The two-layer-PNO approach has a mean average error of -0.013 eV, with a fairly constant underestimation of the reference IE that is below 0.01 eV except for the titanium pair. Simultaneously, the fact that the intra-fragment terms of Layer 2 are set to the NormalPNO thresholds ensures savings in computational cost compared to a global TightPNO calculation. In the present case the savings are modest (about 20%), but they would be expected to increase with increasing size of the second ("low-level") solvation layer. The combined DLPNO+MP2 approach is less accurate than the two-layer-PNO and shows a larger spread of errors with both positive and negative signs. These errors display a rather regular trend with opposite maxima at the two ends of the series, i.e. underestimation of the IE for the Ti pair by 0.1 eV and overestimation for Cu by almost the same amount. As a result, the method has a mean average error of 0.021 eV but a mean absolute error of 0.054 eV. The cost of this approach for the present system is practically the same as the two-layer-PNO approach, but it is expected that the cost benefit will be more prominent with larger systems. Finally, when the second layer is treated at the HF level the gain in computational cost is more obvious, ca. 1.5 times faster than the global TightPNO reference. However, the errors are considerably higher (-0.27 eV on average, with a maximum error of -0.37 eV), which highlights the critical importance of treating electron correlation within the second solvation shell (Layer 2). The overall deviations in this case are larger than the deviations of the TightPNO DLPNO-CCSD(T) method relative to the reference values. Therefore, the approach cannot be recommended if the goal is to retain the accuracy of the reference method to the largest possible extent.

Table 2.7 DLPNO-CCCSD(T_1) computed ionization energies (in eV) for multilevel approaches and for NormalPNO, compared to the global TightPNO reference, ordered by increasing mean signed (MSE) and mean absolute (MAE) errors.

Redox pair	**Global TightPNO**	**Two-layer PNO**	**TightPNO + MP2**	**Global NormalPNO**	**TightPNO + HF**
Ti^{2+} / Ti^{3+}	10.07	10.02	9.97	9.97	9.70
V^{2+} / V^{3+}	10.86	10.85	10.83	10.83	10.54
Cr^{2+} / Cr^{3+}	10.63	10.62	10.69	10.64	10.41
Mn^{2+} / Mn^{3+}	12.88	12.87	12.87	12.86	12.57
Fe^{2+} / Fe^{3+}	11.83	11.82	11.88	11.73	11.53
Co^{2+} / Co^{3+}	13.36	13.35	13.43	13.37	13.19
Ni^{2+} / Ni^{3+}	14.29	14.28	14.33	14.19	14.07
Cu^{2+} / Cu^{3+}	14.46	14.45	14.55	14.30	14.17
MSE		-0.013	0.021	-0.061	-0.274
MAE		0.013	0.054	0.066	0.274

For comparison, **Table 2.7** includes the results of global NormalPNO DLPNO-CCSD(T_1) calculations. The global NormalPNO IEs are worse on average than those of the two-layer TightPNO+MP2 approach, with a mean average error of -0.061 eV. Nevertheless, the global NormalPNO results remain far superior to those of two-layer TightPNO+HF. Notably, global NormalPNO DLPNO-CCSD(T_1) calculations are ca. three times faster than TightPNO+HF. This suggests that judicious adjustment of PNO cutoffs is the optimal way of balancing both the accuracy and the cost of such calculations. The finer control over errors and convergence afforded by this approach makes it preferable to more conventional “QM/QM” approaches. In view of these results, we expect that multilayer DLPNO-based techniques will find increasing application in the future, not only in the context of explicit solvation but in any computational problem where similar chemically motivated partitions can be made, such as for example for metalloenzymes.

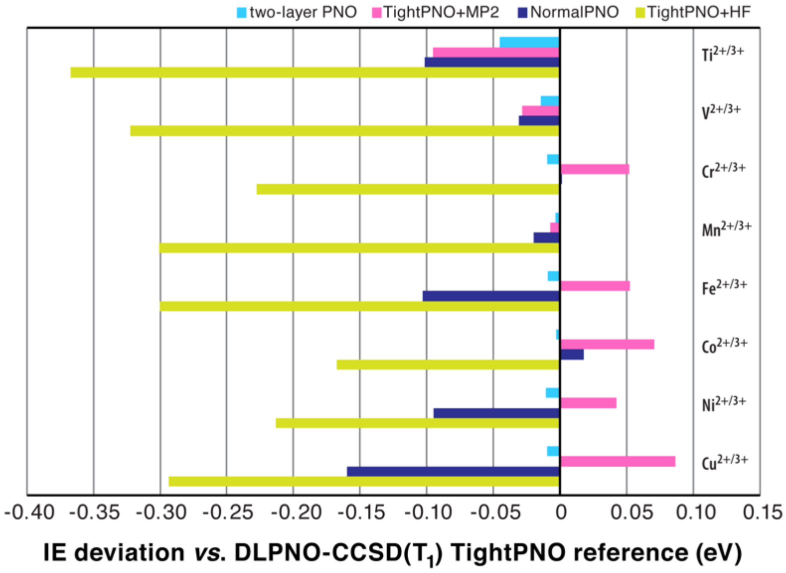


Fig. 2.6 Deviation of M^{2+}/M^{3+} ionization energies (in eV) for the **M-W18** clusters computed with the various two-layer approaches discussed in this work, as well as with global NormalPNO DLPNO-CCSD(T_1) calculations, compared to the reference global TightPNO results.

2.4 Conclusions

We used the domain-based pair natural orbital implementation of coupled cluster theory to estimate ionization energies and redox potentials of hydrated first-row (3d) transition metal ions in their 2+ and 3+ oxidation states. The systems were modeled with inclusion of an explicit second layer of water molecules, leading to 18-water clusters. Reference values were obtained with the DLPNO-CCSD(T_1) approach using global TightPNO settings. It was found that the perturbative triples excitations are necessary to obtain accurate ionization energies. The effect of the second hydration shell was quantified in terms of energetics and the interaction energies were analyzed using the local energy decomposition (LED) scheme for the case of the hydrated iron system. The recent implementation of the conductor-like polarizable continuum model (CPCM) with the PTE(S) scheme was used to determine self-consistent redox potentials at the coupled cluster level. Our results establish conditions for convergence of the DLPNO-CCSD(T_1) energetics and stress the necessity of explicit consideration of a second solvation sphere, whose effects cannot be simulated by a continuum solvation model. The minimal approach of adding a single layer of water is a major step in the right direction, even if it does not represent a conclusive and definitive treatment of the problem. A more refined computational protocol would have to consider the variability of the solvent shell, the dynamic nature of solvation, as well as the fact that changes in coordination number or geometry would necessarily be coupled to reorganization of the solvation layers. Nevertheless, the present

DLPNO-CCSD(T) approach that combines minimal explicit solvation in the form of a second layer of waters plus the PTE(S) model for CPCM, performs robustly and provides reliable estimates of reduction potentials that are within the accuracy of experimental values and largely consistent with previous DFT-based studies. The advantage of the present approach lies in the promise of delivering consistently reliable results for a variety of chemical systems without having to rely on the error cancellation that is a “feature” of DFT-based applications. An important new element of the present study is the multilayer approach to DLPNO-CCSD(T), which was evaluated for three distinct two-layer approaches that retain the high-level treatment of the central core consisting of the metal ion and the directly coordinated water molecules. It was found that an approach that relies on adjustment of PNO cutoffs for different layers as well as for their interaction terms, represents the most promising way of controlling the accuracy and cost of DLPNO-based calculations on large systems. Thus, the multilayer approach to DLPNO-CCSD(T) paves the way for employing chemically accurate yet computationally affordable local correlation methods in the investigation of more complex open-shell systems, both in the context of explicit solvation and in the case of redox-active molecules and metallocofactors embedded in biological matrices.

2.5 REFERENCES

1. A. Hemschemeier and T. Happe, *Nat. Rev. Chem.*, 2018, **2**, 231-243.
2. H. Li, S. Webb, J. Ivanic and J. Jensen, *J. Am. Chem. Soc.*, 2004, **126**, 8010-8019.
3. D. Wang and C. O. Bruner, *Inorg. Chem.*, 2017, **56**, 13638-13641.
4. Y. Yu, X. Liu and J. Wang, *Acc. Chem. Res.*, 2019, **52**, 557-565.
5. B. You, X. Liu, G. Hu, S. Gul, J. Yano, D.-e. Jiang and Y. Sun, *J. Am. Chem. Soc.*, 2017, **139**, 12283-12290.
6. H. Ogata, W. Lubitz and Y. Higuchi, *Dalton Trans.*, 2009, DOI: 10.1039/b903840j, 7577-7587.
7. J. Heimdal, M. Kaukonen, M. Srnec, L. Rulíšek and U. Ryde, *ChemPhysChem*, 2011, **12**, 3337-3347.
8. L. Rulíšek and U. Ryde, *Coord. Chem. Rev.*, 2013, **257**, 445-458.
9. J. Hirst and M. M. Roessler, *Biochim. Biophys. Acta Bioenerg.*, 2016, **1857**, 872-883.
10. R. H. Holm, P. Kennepohl and E. I. Solomon, *Chem. Rev.*, 1996, **96**, 2239-2314.
11. J. Liu, S. Chakraborty, P. Hosseinzadeh, Y. Yu, S. Tian, I. Petrik, . . . Y. Lu, *Chem. Rev.*, 2014, **114**, 4366-4469.
12. I. R. Vassiliev, M. L. Antonkine and J. H. Golbeck, *Biochim. Biophys. Acta Bioenerg.*, 2001, **1507**, 139-160.
13. Y. A. Aoto, A. P. de Lima Batista, A. Kohn and A. G. S. de Oliveira-Filho, *J. Chem. Theory Comput.*, 2017, **13**, 5291-5316.
14. D. Bím, L. Rulíšek and M. Srnec, *J. Phys. Chem. C*, 2018, **122**, 10773-10782.
15. K. P. Kepp, *Coord. Chem. Rev.*, 2013, **257**, 196-209.
16. Y. Minenkov, D. I. Sharapa and L. Cavallo, *J. Chem. Theory Comput.*, 2018, **14**, 3428-3439.
17. A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15068-16106.
18. T. V. Harris and R. K. Szilagyi, *J. Comput. Chem.*, 2016, **37**, 1681-1696.
19. P. J. Stephens, D. R. Jollie and A. Warshel, *Chem. Rev.*, 1996, **96**, 2491-2514.
20. M. Bruschi, R. Breglia, F. Arrigoni, P. Fantucci and L. De Gioia, *Int. J. Quantum Chem.*, 2016, **116**, 1695-1705.
21. J. N. Harvey, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, 2006, **102**, 203-226.
22. J. N. Harvey, *J. Biol. Inorg. Chem.*, 2011, **16**, 831-839.
23. L. Noodleman and W.-G. Han, *J. Biol. Inorg. Chem.*, 2006, **11**, 674-694.
24. J. Li, M. R. Nelson, C. Y. Peng, D. Bashford and L. Noodleman, *J. Phys. Chem. A*, 1998, **102**, 6311-6324.
25. M. H. M. Olsson, G. Hong and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 5025-5039.
26. R. A. Friesner and S. V. Jerome, *Coord. Chem. Rev.*, 2017, **344**, 205-213.
27. T. F. Hughes and R. A. Friesner, *J. Chem. Theory Comput.*, 2011, **7**, 19-32.
28. N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.*, 2006, **125**, 074110.
29. D. Ghosh, A. Roy, R. Seidel, B. Winter, S. Bradforth and A. I. Krylov, *J. Phys. Chem. B*, 2012, **116**, 7269-7280.
30. R. N. Tazhigulov and K. B. Bravaya, *J. Phys. Chem. Lett.*, 2016, **7**, 2490-2495.
31. J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi and M. Sprik, *Acc. Chem. Res.*, 2014, **47**, 3522-3529.
32. M. Isegawa, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2016, **12**, 2272-2284.
33. H. Neugebauer, F. Bohle, M. Bursch, A. Hansen and S. Grimme, *J. Phys. Chem. A*, 2020, **124**, 7166-7176.
34. B. Thapa and H. B. Schlegel, *J. Phys. Chem. A*, 2015, **119**, 5134-5144.

35. L. P. Wang and T. Van Voorhis, *J. Chem. Theory Comput.*, 2012, **8**, 610-617.
36. M. Uudsemaa and T. Tamm, *J. Phys. Chem. A*, 2003, **107**, 9997-10003.
37. J. Li, C. L. Fisher, J. L. Chen, D. Bashford and L. Noodleman, *Inorg. Chem.*, 1996, **35**, 4694-4702.
38. M. Toma, T. Kuvek and V. Vrček, *J. Phys. Chem. A*, 2020, **124**, 8029-8039.
39. T. Matsui, Y. Kitagawa, Y. Shigeta and M. Okumura, *J. Chem. Theory Comput.*, 2013, **9**, 2974-2980.
40. M.-H. Baik and R. A. Friesner, *J. Phys. Chem. A*, 2002, **106**, 7407-7412.
41. J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi and M. Sprik, *Acc. Chem. Res.*, 2014, **47**, 3522-3529.
42. N. J. Fowler, C. F. Blanford, J. Warwicker and S. P. de Visser, *Chem. Eur. J.*, 2017, **23**, 15436-15445.
43. T. Matsui and J.-W. Song, *Molecules*, 2019, **24**, 819.
44. R. A. Torres, T. Lovell, L. Noodleman and D. A. Case, *J. Am. Chem. Soc.*, 2003, **125**, 1923-1936.
45. Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **124**, 224105.
46. F. Arrigoni, R. Breglia, L. D. Gioia, M. Bruschi and P. Fantucci, *J. Phys. Chem. A*, 2019, **123**, 6948-6957.
47. B. Mennucci and R. Cammi, *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, John Wiley & Sons, Chichester U.K., 2007.
48. J. J. Guerard and J. S. Arey, *J. Chem. Theory Comput.*, 2013, **9**, 5046-5058.
49. C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161-2200.
50. L. Rulíšek, *J. Phys. Chem. C*, 2013, **117**, 16871-16877.
51. C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10757-10816.
52. T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4398-4401.
53. M. Radon, K. Gassowska, J. Szklarzewicz and E. Broclawik, *J. Chem. Theory Comput.*, 2016, **12**, 1592-1605.
54. P. Verma, Z. Varga, J. Klein, C. J. Cramer, L. Que and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2017, **19**, 13049-13069.
55. Q. M. Phung, M. Feldt, J. N. Harvey and K. Pierloot, *J. Chem. Theory Comput.*, 2018, **14**, 2446-2455.
56. F. Vlahovic, M. Peric, M. Gruden-Pavlovic and M. Zlatar, *J. Chem. Phys.*, 2015, **142**, 214111.
57. S. E. Neale, D. A. Pantazis and S. A. Macgregor, *Dalton Trans.*, 2020, **49**, 6478-6487.
58. D. Zhang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2020, **16**, 4416-4428.
59. M. Sparta and F. Neese, *Chem. Soc. Rev.*, 2014, **43**, 5032-5041.
60. F. Neese, A. Hansen, F. Wennmohs and S. Grimme, *Acc. Chem. Res.*, 2009, **42**, 641-648.
61. K. Andersson, P. Å. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218-1226.
62. C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger and J.-P. Malrieu, *J. Chem. Phys.*, 2001, **114**, 10252-10264.
63. C. Angeli, R. Cimiraglia and J.-P. Malrieu, *J. Chem. Phys.*, 2002, **117**, 9138-9153.
64. M. Radon and G. Drabik, *J. Chem. Theory Comput.*, 2018, **14**, 4010-4027.
65. R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.*, 2007, **79**, 291-352.
66. C. J. Stein, V. von Burg and M. Reiher, *J. Chem. Theory Comput.*, 2016, **12**, 3764-3773.
67. L. Veis, A. Antalik, J. Brabec, F. Neese, O. Legeza and J. Pittner, *J. Phys. Chem. Lett.*, 2016, **7**, 4072-4078.
68. D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. Martin and F. Neese, *J. Chem. Theory Comput.*, 2015, **11**, 1525-1539.

69. B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B*, 1997, **101**, 10506-10517.
70. R. Cammi and J. Tomasi, *J. Comput. Chem.*, 1995, **16**, 1449-1458.
71. S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117-129.
72. V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995-2001.
73. E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032-3041.
74. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378-6396.
75. A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224-2235.
76. J. Chen, B. Chan, Y. Shao and J. Ho, *Phys. Chem. Chem. Phys.*, 2020, **22**, 3855-3866.
77. A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, DOI: 10.1039/P29930000799, 799-805.
78. C. M. Sterling and R. Bjornsson, *J. Chem. Theory Comput.*, 2019, **15**, 52-67.
79. J. J. Guerard, P. R. Tentscher, M. Seijo and J. Samuel Arey, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14811-14826.
80. B. Thapa and H. B. Schlegel, *J. Phys. Chem. A*, 2017, **121**, 4698-4706.
81. V. S. Bryantsev, M. S. Diallo and W. A. Goddard, *J. Phys. Chem. A*, 2009, **113**, 9559-9567.
82. Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo and F. Neese, *J. Chem. Phys.*, 2018, **148**, 011101.
83. Y. Guo, C. Riplinger, D. G. Liakos, U. Becker, M. Saitow and F. Neese, *J. Chem. Phys.*, 2020, **152**, 024116.
84. C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.
85. M. Saitow, U. Becker, C. Riplinger, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2017, **146**, 164105.
86. A. Hansen, D. G. Liakos and F. Neese, *J. Chem. Phys.*, 2011, **135**, 214102.
87. D. G. Liakos and F. Neese, *J. Chem. Theory Comput.*, 2015, **11**, 4054-4063.
88. B. M. Floser, Y. Guo, C. Riplinger, F. Tuczek and F. Neese, *J. Chem. Theory Comput.*, 2020, **16**, 2224-2235.
89. M. Garcia-Ratés, U. Becker and F. Neese, *J. Comput. Chem.*, 2021, **42**, 1959-1973.
90. C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
91. A. Altun, F. Neese and G. Bistoni, *J. Chem. Theory Comput.*, 2020, **16**, 6142-6149.
92. D. G. Liakos and F. Neese, *J. Phys. Chem. A*, 2012, **116**, 4801-4816.
93. M. Sparta, M. Retegan, P. Pinski, C. Riplinger, U. Becker and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 3198-3207.
94. S. Gomez-Salces, F. Aguado, R. Valiente and F. Rodriguez, *Angew. Chem., Int. Ed.*, 2012, **51**, 9335-9338.
95. V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and W. A. Goddard Iii, *J. Phys. Chem. A*, 2008, **112**, 9104-9112.
96. K. J. de Almeida, N. A. Murugan, Z. Rinkevicius, H. W. Hugosson, O. Vahtras, H. Ågren and A. Cesar, *Phys. Chem. Chem. Phys.*, 2009, **11**, 508-519.
97. A. Pasquarello, I. Petri, P. S. Salmon, O. Parisel, R. Car, É. Tóth, . . . A. E. Merbach, *Science*, 2001, **291**, 856-859.
98. G. Chillemi, E. Pace, M. D’Abramo and M. Benfatto, *J. Phys. Chem. A*, 2016, **120**, 3958-3965.
99. F. Neese, *WIREs Computational Molecular Science*, 2011, **2**, 73-78.
100. F. Neese, *WIREs Computational Molecular Science*, 2017, **8**, e1327.
101. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
102. M. Reiher and A. Wolf, *J. Chem. Phys.*, 2004, **121**, 2037-2047.

103. M. Reiher and A. Wolf, *J. Chem. Phys.*, 2004, **121**, 10945-10956.
104. T. Nakajima and K. Hirao, *Chem. Rev.*, 2012, **112**, 385-402.
105. J. P. Perdew, J. Tao, V. N. Staroverov and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 6898-6911.
106. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
107. V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129-12137.
108. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
109. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
110. D. G. Smith, L. A. Burns, K. Patkowski and C. D. Sherrill, *J. Phys. Chem. Lett.*, 2016, **7**, 2197-2203.
111. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
112. K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283-290.
113. K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119-124.
114. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
115. D. A. Pantazis and F. Neese, *J. Chem. Theory Comput.*, 2011, **7**, 677-684.
116. J. D. Rolfes, F. Neese and D. A. Pantazis, *J. Comput. Chem.*, 2020, **41**, 1842-1849.
117. D. Aravena, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2016, **12**, 1148-1156.
118. D. A. Pantazis, X.-Y. Chen, C. R. Landis and F. Neese, *J. Chem. Theory Comput.*, 2008, **4**, 908-919.
119. D. A. Pantazis and F. Neese, *J. Chem. Theory Comput.*, 2009, **5**, 2229-2238.
120. D. A. Pantazis and F. Neese, *Theor. Chem. Acc.*, 2012, **131**.
121. N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.*, 2005, **123**, 064107.
122. F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057-1065.
123. D. M. York and M. Karplus, *J. Phys. Chem. A*, 1999, **103**, 11060-11079.
124. M. Garcia-Ratés and F. Neese, *J. Comput. Chem.*, 2020, **41**, 922-939.
125. C. Riplinger, P. Pinski, U. Becker, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2016, **144**, 024109.
126. G. Bistoni, C. Riplinger, Y. Minenkov, L. Cavallo, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 3220-3227.
127. G. L. Stoychev, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 554-562.
128. W. A. d. Jong, R. J. Harrison and D. A. Dixon, *J. Chem. Phys.*, 2001, **114**, 48-53.
129. T. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
130. K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548-10560.
131. D. E. Woon and T. H. D. Jr., *J. Chem. Phys.*, 1995, **103**, 4572-4585.
132. A. Altun, R. Izsák and G. Bistoni, *Int. J. Quantum Chem.*, 2021, **121**, e26339.
133. A. Altun, M. Saitow, F. Neese and G. Bistoni, *J. Chem. Theory Comput.*, 2019, **15**, 1616-1632.
134. W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2016, **12**, 4778-4792.
135. M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, . . . T. R. Tuttle, *J. Phys. Chem. A*, 1998, **102**, 7787-7794.
136. C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2006, **110**, 16066-16081.
137. J. Ho and M. Z. Ertem, *J. Phys. Chem. B*, 2016, **120**, 1319-1329.
138. R. Cammi, *J. Chem. Phys.*, 2009, **131**, 164104.

139. M. Caricato, *J. Chem. Phys.*, 2011, **135**, 074113.
140. A. Altun, M. Garcia-Ratés, F. Neese and G. Bistoni, *Chem. Sci.*, 2021, **12**, 12785-12793.
141. J. S. Binkley and J. A. Pople, *Int. J. Quantum Chem.*, 1975, **9**, 229-236.
142. W. Kutzelnigg, *Int. J. Quantum Chem.*, 2009, **109**, 3858-3884.
143. S. G. Bratsch, *J. Phys. Chem. Ref. Data*, 1989, **18**, 1-21.

3

Multiscale Modelling of Photosynthetic Systems

3.1 Introduction

The major part of this thesis employs two multiscale modelling methodologies: classical molecular dynamics (MD) and hybrid QM/MM simulations. The fundamental step towards constructing any MD or QM/MM model of a biomolecular system involves obtaining an accurate force-field based description of the complete system that includes the protein, solvent, cofactors and all related components. This chapter describes the complete protocol to build a classical molecular mechanics (MM) model of the membrane bound monomeric assembly of the Photosystem II (PSII) protein complex along with its genetic variants and mutated isoforms. The preliminary setup and MD simulations were performed in the AMBER20 package[1-3] while all the QM/MM simulations were performed using the multiscale module of the ORCA quantum chemistry package.[4-7] The methodology described here provides the basis of the computational work described in **Chapters 4, 5** and **6**.

3.2 System Preparation

3.2.1 Setup of the PSII monomer

For any QM/MM simulation, selecting an adequate starting structure and preparing it properly is important, and sufficient care and time is usually devoted to the process. In biomolecular simulations, the starting point is either an experimentally derived structure or a structural model of the protein of interest. In the case of PSII, a number of high-resolution structures are currently available due to advancement in protein crystallographic and cryo-EM procedures.[8-12] The 1.9 Å resolution X-ray crystal structure of *Thermosynechococcus vulcanus* (3WU2.pdb) reported by Umena et al. is used as the starting point throughout this work.[13] The crystal structure of PSII includes coordinates for 19 protein subunits, 54 cofactors and 20 special lipids. The PSII monomer was extracted from the dimeric system and first completed by adding the missing components as described in the work by Sirohiwal et al.[14] The standard protonation states of the amino acid residues and the co-factors were determined using the *reduce* module in AmberTools.[1] The unassigned special lipid molecules and co-factors were previously re-built based on other cyanobacterial PSII crystal structures (2AXT.pdb[15] and 4V62.pdb[16]). The missing atoms of the partially resolved cofactors were completed using PyMol visualization software[17] and all crystallographic waters were retained in our computational models. The orientation of the lipid bilayer was determined by aligning the z-axis of the PSII monomer along the membrane plane, using the OPM (Orientation of Protein in Membranes) webserver.[18]

The complete PSII monomer was then embedded inside a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane patch using the *Packmol-Memgen* module[19, 20] in AmberTools. Related studies on membrane-protein systems have reported that POPC is a reasonable choice because phosphatidylcholine is compatible with the functioning of PSII,[21] and also the thickness of POPC membrane is within ~1 Å of the cyanobacterial thylakoid membrane.[22, 23] A distance cutoff of 2 Å was applied while packing the membrane and solvent (water) molecules in order to avoid geometric clashes with the protein complex. A total of 784 lipid molecules were added to the upper and lower leaflets and the system was solvated with waters 17.5 Å above and below the protein (115,237 H_2O molecules). The system was neutralized with appropriate number of counterions[24] and a physiological salt concentration of 0.15 M was maintained by adding 296 Na^+ and 263 Cl^- atoms. The final PSII system with the hydrated lipid bilayer consisted of 503,328 atoms (see **Fig. 3.1**).

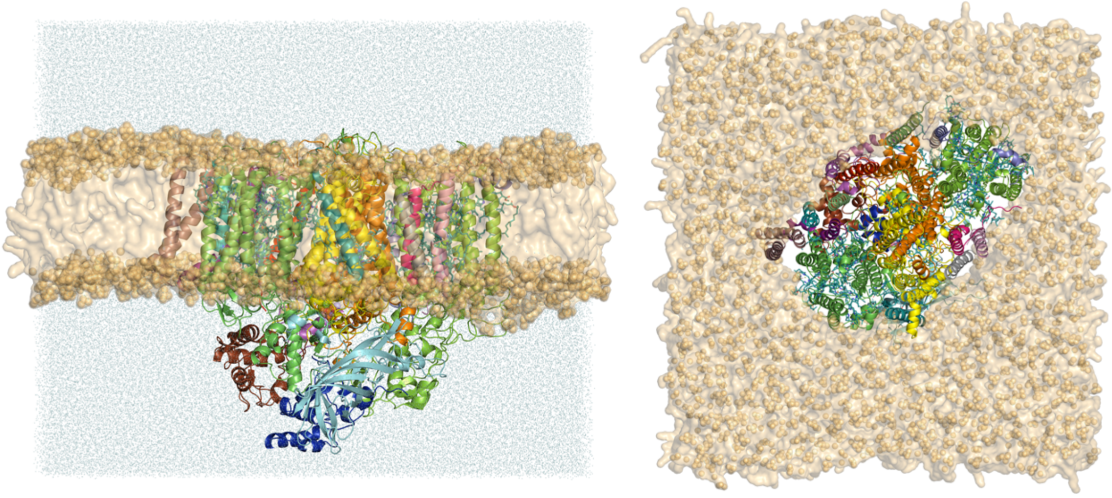


Fig. 3.1 Side-view and top-view of the molecular-mechanics (**MM**) setup of the membrane-bound PSII monomer. The PSII protein (**3WU2.pdb**) is embedded within a **POPC** bilayer and solvated with **TIP3P** waters; Na^+ and Cl^- ions are not shown for clarity. The final box dimension is 176 x176 x 160 $Å^3$. The figures are generated using PyMol.

3.2.2 Construction of the PsbA (D1) protein variants

The D1 protein subunit in cyanobacterial PSII is encoded by the psbA gene family that expresses three distinct isoforms (PsbA1–3) depending on environmental conditions.[25-32] In **Chapter 5**, we investigate how three naturally occurring genetic variants of this crucial PSII protein differentially adjust the optical and redox properties of the pigments in the RC. Towards this, individual MM models for each PsbA variant were constructed to evaluate the mutations in each D1 copy. The initial structure of the PSII monomer is based on the crystal structure of *T. vulcanus* (3WU2.pdb)[13] as described in the previous section, originally containing the PsbA1 variant. For the PsbA2 and A3 variants, the D1 protein (Chain A, 3WU2.pdb) was replaced with the corresponding D1 chains of PsbA2- and PsbA3-PSII (**Fig. 3.2**) based on the recent 1.9 Å crystal structures (7YQ2.pdb and 7YQ7.pdb).[27] It is important to note that the PsbA1 sequences of *T. vulcanus* and *T. elongatus* are essentially identical with the exception of a single difference at D1-286. Specifically, two protonation states for the substitution at D1-130 were considered for PsbA2 and PsbA3 variants, namely (a) D1-Glu for the deprotonated glutamate and (b) D1-Glu(H) for glutamic acid. In each of the five setups, the complete protein unit was embedded in a POPC lipid bilayer using *Packmol-Memgen* and solvated with TIP3P waters.[19, 33] The number of lipids and water molecules were kept same for each setup. All the models were neutralized with appropriate number of counterions and a physiological salt buffer.

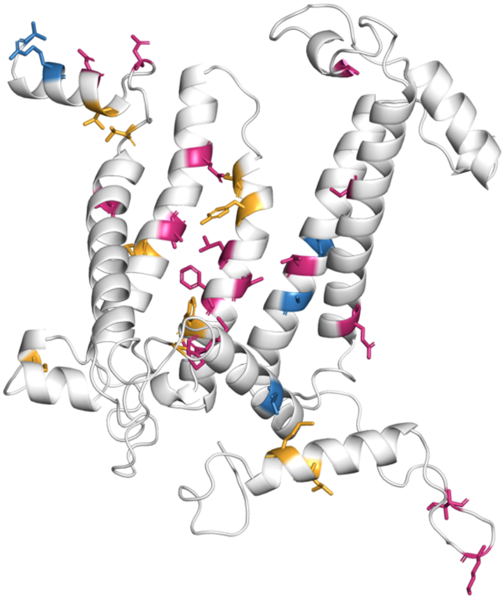


Fig. 3.2 The secondary structure of the **PsbA (D1)** protein in PSII depicting the location of the substitutions in the PsbA2 and PsbA3 variants with respect to PsbA1. The residues which differ in A1/A2 are shown in orange, A1/A3 in blue and residues different in all three A1/A2/A3 variants are shown in pink.

3.2.3 Models of the PSII “extrinsic” proteins[‡]

Cyanobacterial PSII contains three or four membrane extrinsic proteins in each monomer.[34, 35] Specifically, the 1.9 Å crystal structure of *T. vulcanus* revealed that three extrinsic proteins (PsbO, PsbU, and PsbV) are bound to the PSII core complex. Despite this structural information, the functional roles of these extrinsic proteins in PSII are still not well understood. All-atom MD simulations are a highly valuable computational technique for investigating the structural and dynamic changes at the interaction sites of PsbO, PsbU, and PsbV proteins with the PSII core complex.[22, 36, 37] For example, simulations have demonstrated that the removal of PsbO leads to perturbation of the OEC, reduced O_2 production, and altered H_2O dynamics.[38-40]

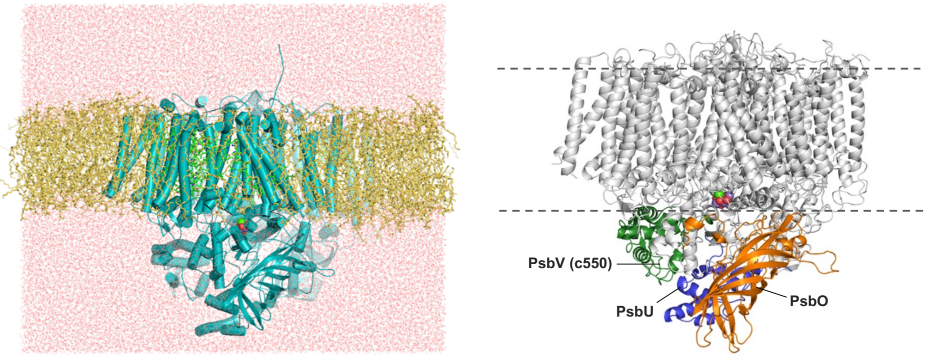


Fig. 3.3 The MM setup of the PSII monomer used for simulations of the protein “mutants”. The resulting MM models in each case consisted of the PSII monomer with various combinations of the extrinsic proteins.

We used classical MD simulations to investigate the response of PSII to perturbations, such as the selective removal or loss of extrinsic proteins. To this end, we constructed molecular models for three PSII “mutants” (ΔPsbO, ΔPsbU, and ΔPsbV), representing PSII in the absence of each extrinsic subunit. These "mutants" were prepared using the PSII monomeric system described above as the starting coordinates thus, the resulting MM models consisted of the PSII monomer with various combinations of the extrinsic proteins. For the ΔPsbO mutant, we removed chain O along with the associated Ca ion and 195 water molecules from the 3WU2.pdb crystal structure. Similarly, chain U and the associated crystal waters were removed to construct the ΔPsbU mutant, and chain V along with the heme cofactor was removed for the ΔPsbV mutant. The protonation states of the titratable residues were retained as in the original setup. Each of the resulting PDB files were renumbered using the *pdb4amber* tool[1] and the protein structures were then separately embedded into a hydrated POPC membrane patch using *Packmol-Memgen.*[19] All the models were neutralized with appropriate number of counterions and 0.15M physiological salt concentration. The total number of water molecules is found to be consistent with the volume of the cavity created as a result of removing the protein chains in each case. The final MM model used for the simulations of the PSII mutants is depicted in **Fig 3.3**.

3.2.4 Explicitly solvated model of Cytochrome c_{550} (PsbV)[‡]

The Cytochrome c_{550} or PsbV extrinsic protein of cyanobacterial PSII, is reported to have widely different chemical properties in its soluble and bound forms.[41-44] To investigate its possible role in photosynthetic ET, QM/MM models were constructed for each case i.e., PsbV with explicit solvation and the PSII bound-PsbV. For the starting coordinates of solvated PsbV, we extracted Chain V along with all its components (heme cofactor, crystal waters) from the 3WU2.pdb crystal structure of PSII. All titratable groups were considered in their standard protonation states i.e., all Asp/Glu groups were negatively charged, Arg/Lys were positively charged. The axial histidine ligands to the heme cofactor (His41, His92) were protonated at the N_δ position as "HID" while the cysteines (Cys37, Cys40) covalently linked to the heme porphyrin were modelled as "CYX". Further details on the definition of force field parameters for the heme-c cofactor and coordinated residues are provided in **Section 3.3**. The complete PsbV protein is then placed in a TIP3P water box with a solute-solvent radius of 15 Å, and neutralized with 6 Na^+ ions and a 0.15 M salt buffer to maintain the exact setup conditions as in the PSII-bound form. The final QM/MM setup for solvated PsbV consisted of 61,745 atoms and 19,794 water molecules (**Fig. 3.4a**).

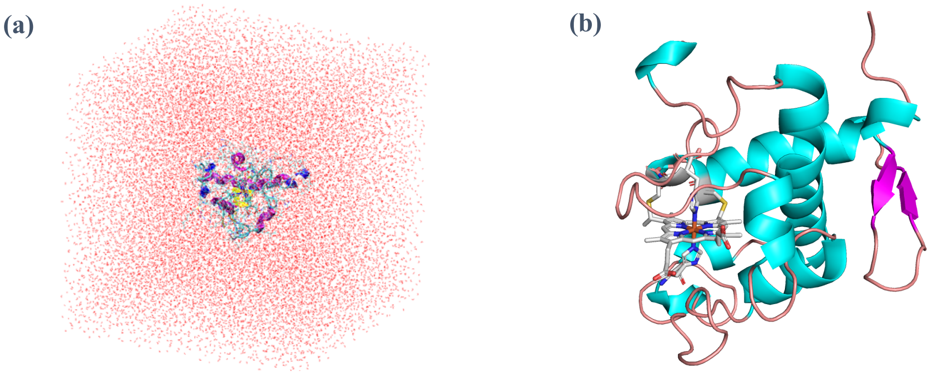


Fig. 3.4 (a) Structure of the isolated **Cytochrome c_{550}** (PsbV subunit) in PSII with the heme active site. **(b)** MM setup for the explicitly solvated Cyt c_{550} system.

[‡] *This work is still in progress so a separate chapter has not been included in this thesis.*

3.3 Definition of Force Field Parameters

The molecular mechanics model of any biomolecular system requires a complete and accurate description of the force-field for the protein and all its components. The amino acid residues and organic cofactors can be defined using the standard force fields and libraries. However, modelling metal active sites are significantly more challenging due to their variable coordination numbers, relatively labile chemical bonding, and diverse electronic states.[45, 46]

The description of the force-field in PSII can be categorized as (i) standard protein residues, POPC lipids, water, ions and (ii) non-standard residues, such as metallo-cofactors (OEC, NHI, Hemes, Chls), organic cofactors (pheophytins, β-carotenes, plastoquinones, special lipids). Consequently, the force-field description (partial charges, bonded and vdW terms) of the non-standard residues is case-specific and has to be manually determined. In our computational model, we used the Amberff14SB force-field[47, 48] for the standard amino acid residues, whereas the solvent (H_2O) was modelled as TIP3P.[33] The POPC lipid bilayer was defined using the LIPID17 force field.[49-51] For Na^+ and Cl^- ions, Joung-Cheatham parameters (*ionsjc_tip3p*)[52, 53] compatible with the TIP3P water models were used. The electrostatic charges for all the non-standard residues are derived based on the MK-RESP methodology.[54, 55] In case of the organic co-factors we imported the RESP charges derived from an earlier investigation[14, 37] (geometry optimization at the B3LYP/def2-SVP[56, 57] level followed by single-point calculations at the HF/6-31G* level of theory[47, 58] to obtain the partial charges). The atom-types were assigned using the *antechamber* module and the non-bonded terms were imported directly from the compatible GAFF2 library[59, 60] in AmberTools.

A bonded model was employed for determining the partial charges of the OEC and NHI clusters, consistent with past studies in our group.[14, 37, 61, 62] The QM cluster models used for deriving the charges included the metal sites along with the ligand residues or H_2O molecules in the first coordination sphere (**Fig. 3.5a**). The Mn_4CaO_x cluster was originally modeled[14] in its resting S_1 state, i.e., Mn_1(III)–Mn_2(IV)–Mn_3(IV)–Mn_4(III) and the protein ligands included Asp170, Glu354, Ala344, Asp342, Glu189, His332, Glu333 and four H_2O molecules. In case of the NHI site, the Fe was modelled as Fe(II) together with the bicarbonate (HCO_3^-), His214, His268, His215, and His272. For each of these cofactors the H-atoms were first optimized at B3LYP/def2-TZVP[56, 57] level of theory and single-point energy calculations were performed in ORCA using B3LYP/6-31G*. The MK-RESP charges were then derived in Multiwfn.[63, 64] During the fitting procedure, the charges of the backbone atoms of the ligands were constrained on the link atoms based on the original AMBER force field.[45, 47] For the work described in **Chapter 4**, the partial atomic charges of the primary quinone (Q_A) and the OEC were modified based on the same cluster model depicted in **Fig. 3.5a**. Specifically, the OEC was modelled in the S_2 state of the Kok–Joliot cycle,[65] i.e., with formal oxidation states Mn1(III)-Mn2(IV)-Mn3(IV)-Mn4(IV) whereas Q_A was modelled in its singly-reduced state (Q_A^-) using the same computational protocol as described above.

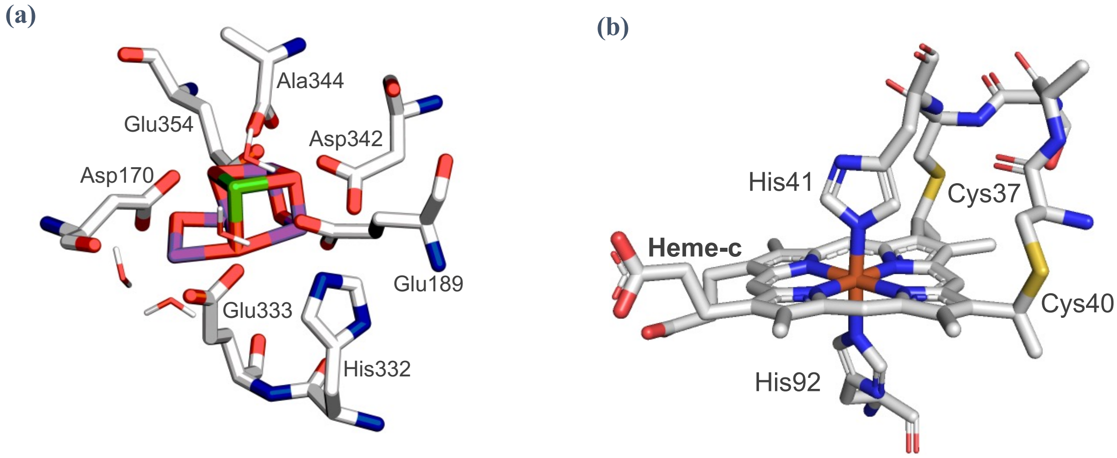


Fig. 3.5 Cluster models used for the derivation of the RESP charges for the **(a)** OEC and **(b)** Heme-c (**Cyt c_{550}**). The backbone atoms were not included in the RESP fitting, hydrogen atoms are not shown for clarity.

The RESP charges for the chlorophylls, pheophytins and heme-b (Cytb_{559}) are imported from earlier studies.[14] The RESP charges of the Heme-c active site in PsbV (Cytc_{550}) are manually derived in this work using the QM cluster shown in **Fig. 3.5b** and the protocol described above. The QM region for RESP fitting consisted of the Fe(II) heme porphyrin, along with the covalently linked cysteine residues Cys37, Cys40 ("CYX" in Amber nomenclature) and axially coordinated His41 and His92.

The protonation states of the axial His ligands to the Chl-Mg and Heme-Fe atoms were manually checked (i.e., "HID" and "HIE" in Amber nomenclature for protonation at N_δ and N_ε respectively) and the metal-ligand bonds were defined using the *tleap* module of AmberTools. For metal ions, the non-bonded parameters were based on oxidation states defined in the Ion-Oxygen Distance data library[45, 66, 67] compatible with the TIP3P model. The non-bonded parameters of the ligated protein residues (to the OEC and NHI) were directly imported from the AMBER library.[60] The bonded parameters for Chl a, Pheo a and heme porphyrin were obtained from literature[68] whereas customized in-house bonded parameters were used for the OEC site.[14] We used the previously assigned value of 150 kcal $\text{mol}^{-1}\text{Å}^{-2}$ for the bonds within the Mn_4CaO_x cluster, whereas the bonds between the Mn/Ca and coordinating ligand atoms had a force constant of 70 kcal mol^{-1}Å^{-2}. Similarly, the angles within the OEC were restrained with a value of 120 kcal $mol^{-1}rad^{-2}$ and angles between the OEC and coordinating residues with a value of 70 kcal $mol^{-1}rad^{-2}$. The bonded parameters for the NHI site were taken from literature.[23] The force-field parameters described here is henceforth applied to the lipid bilayer bound PSII monomer, and all the PSII–ΔPsbO, ΔPsbV and ΔPsbU "mutants" described in **Section 3.2**.

3.4 Molecular Dynamics of the PSII assembly

The timescales and molecular dynamics settings described in this section are specific to the PSII protein-bilayer assembly, however it is important to note that the individual steps can be generalized to most explicitly solvated membrane-protein complex with minor modifications.

3.4.1 Energy minimization

As a first step towards the thermal equilibration of our system, a stepwise minimization protocol is followed to remove the unfavourable geometric clashes within the protein and its various components. Firstly, all hydrogen atoms were optimized for a total of 2000 steps (1000 steps each for steepest descent and conjugate gradient) while maintaining positional restraints (force constant, 50 kcal mol^{-1}Å^{-2}) on all heavy atoms in the system. In the next step, the lipid bilayer, solvent (waters) and ions are relaxed for a total of 20,000 steps (50% each of steepest descent and conjugate gradient) with a restraint of 50 kcal mol^{-1}Å^{-2} on the protein residues and cofactors. Finally, the complete system was relaxed for a total of 15,000 steps while maintaining positional restraints (20 kcal mol^{-1}Å^{-2}) on the backbone C_α atoms of the protein. All the minimization steps were performed using the *sander.MPI* and *pmemd.MPI* engines in the AMBER20 program package.[1, 3]

3.4.2 Heating and Equilibration

The protein structural configuration obtained after the final step of the energy minimization is used for the equilibration molecular dynamics simulations. For the heating, the system is first slowly heated (5 ps) from 10–100K in the canonical (NVT) ensemble, followed by heating from 100K to the target temperature of 303K along 125 ps in the isothermal-isobaric (NPT) ensemble. The protein is then systematically relaxed in the NPT ensemble by gradually lowering the restraints on the C_α atoms by 2 kcal mol^{-1}Å^{-2} every 500 ps. Thereafter, the entire system is freely simulated for a total of 65 ns to ensure proper equilibration of the lipid bilayer. The temperature during the heating and equilibration is controlled using the *Langevin Dynamics* with a collision frequency of 5 ps^{-1}.[69] The pressure was regulated anisotropically using the *Berendsen* barostat[70] with a relaxation time of 2 ps and value of 1 bar. The PME approach[71] is used to treat all electrostatic interactions within 10 Å. The MD trajectory frames were saved every 2 ps. All the MD equilibration simulations were performed in the GPU version of the PMEMD engine (*pmemd.cuda*) in AMBER20.[72-74] The thermal equilibration of the system is monitored by the convergence of the total density, protein RMSD, area per lipid and electron density profile of the membrane (**Fig. 3.6**), using the *cpptraj* module.[75]

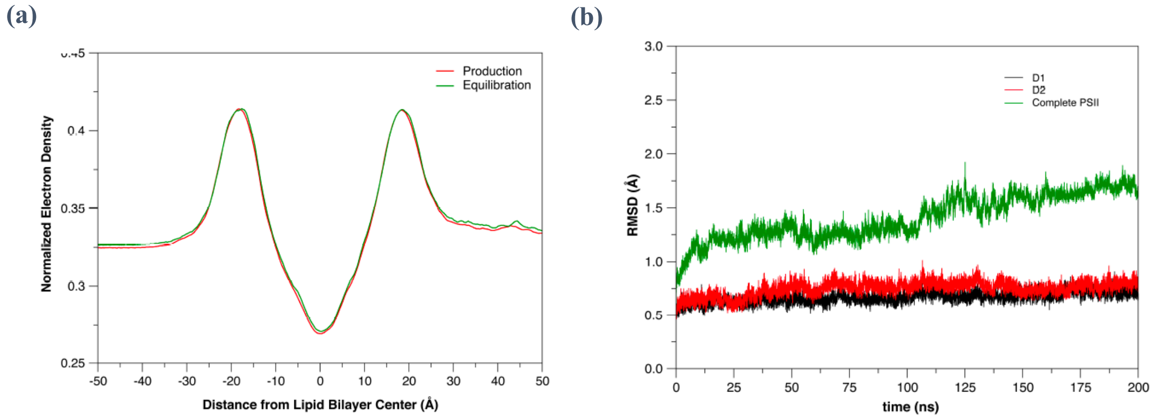


Fig. 3.6 (a) Electron density profile of the **POPC** lipid bilayer during the equilibration and production simulations. The membrane thickness (peak to peak distance) during the equilibration and production simulations agrees well with the experimental value (37 Å) of the pure POPC lipid bilayer. **(b)** Root mean squared displacement (**RMSD**) for C_α atoms along the MD trajectory[76] for the **D1** and **D2** polypeptides compared to the complete PSII monomer. The trajectory analysis is performed using the ***Cpptraj*** module.

3.4.3 Production simulations

The protein structural configuration obtained after the final step of the system equilibration is used for the production molecular dynamics simulations. For the PSII systems, the unbiased production run is generally performed in the NPT ensemble. The temperature during the production run is also controlled using the *Langevin Dynamics*[69] with a collision frequency of 1 ps^{-1}. The pressure was regulated anisotropically using the *Berendsen* barostat[70] with a relaxation time of 2 ps and value of 1 bar. The SHAKE algorithm[77] was used to constrain the bonds involving hydrogens, which allowed us to use an integration time-step of 2 fs. The PME approach[71] is used to treat all electrostatic interactions as described previously. The MD trajectory frames were saved every 2 ps. All the MD equilibration and production simulations were performed in the GPU version of the PMEMD engine (*pmemd.cuda*) in AMBER20.[72-74] The conformational analysis (e.g., distances, RMSD) was performed using *cpptraj.cuda*,[78] while VMD[79] and PyMol[17] were used for trajectory visualization.

3.5 MD simulations of PsbV

The settings described here are specific to the isolated PsbV protein (see **Section 3.2.4** for details on the setup), however it is important to note that the protocol is generally applicable to most explicitly solvated proteins.

3.5.1 System minimization

A two-step minimization protocol is followed to remove the geometric clashes within the PsbV protein. In the first step, hydrogens are optimized for 2000 steps (1000 steps each for steepest descent and conjugate gradient) while maintaining positional restraints (50 kcal $mol^{-1}Å^{-2}$) on all heavy atoms. Then, the solvent (waters) and ions are relaxed for 20,000 steps (50% each of steepest descent and conjugate gradient) with a restraint of 50 kcal $mol^{-1}Å^{-2}$ on the protein residues and the heme moiety. The optimization is performed using *sander* in AMBER20.

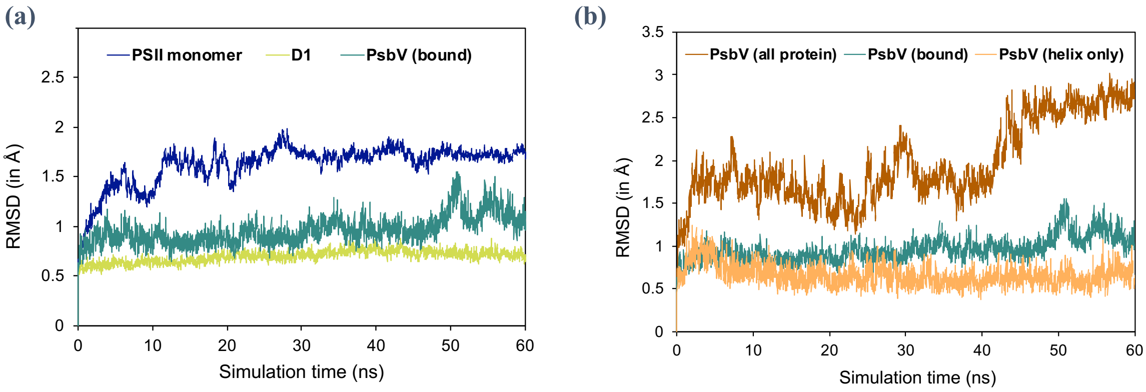


Fig. 3.7 Root mean squared displacement (**RMSD**) for C_α atoms along the MD trajectory for the **(a) D1** and **PsbV** polypeptides compared to the complete PSII monomer; the MM system is depicted in **Fig. 3.1. (b)** Solvated PsbV compared to the complexed protein. Trajectory analysis is performed using CPPTRAJ.

3.5.2 Equilibration Dynamics

The final configuration of the solvated PsbV protein from the energy minimization was used as a starting point of the heating and equilibration dynamics. The system was heated from 10 K to 300 K through 400 ps in the NVT ensemble and positional restraints (50 kcal $mol^{-1}Å^{-2}$) were maintained on the protein and heme moiety. The system was the relaxed in the NPT ensemble for 1000 ps by gradually decreasing the restraints. The thermal equilibration of the system is determined by the protein RMSD and box density. The protein structural configuration obtained after the final step of the system equilibration is used for the production molecular dynamics simulations. The production run is performed in the NPT ensemble for 60 ns without any restraints. The temperature during the dynamics was controlled using the *Langevin* thermostat with a collision frequency of 1 ps^{-1}. The pressure was regulated using the *Berendsen* barostat with a relaxation time of 2 ps and value of 1 bar. The SHAKE algorithm[77] was used to constrain the bonds involving hydrogens, which allowed us to use an integration time-step of 2 fs. The PME approach[71] is used to treat all electrostatic interactions as described previously. All the MD equilibration and production simulations were performed in the GPU version of the PMEMD engine[72-74] (*pmemd.cuda*) in AMBER20. The MD trajectory analysis (see **Fig. 3.7**) was performed using the *cpptraj* module, while VMD[79] and PyMol were used for visualization.

3.6 QM/MM Computations

The QM/MM based computations on PSII were mainly performed on the reaction center (RC), the CP43 core antenna complex and the Cyt c550 (PsbV) protein. All the QM/MM calculations were performed on the PSII monomer (described in **Section 3.2**). The results obtained from these calculations are discussed in detail in **Chapters 4, 5** and **6**.

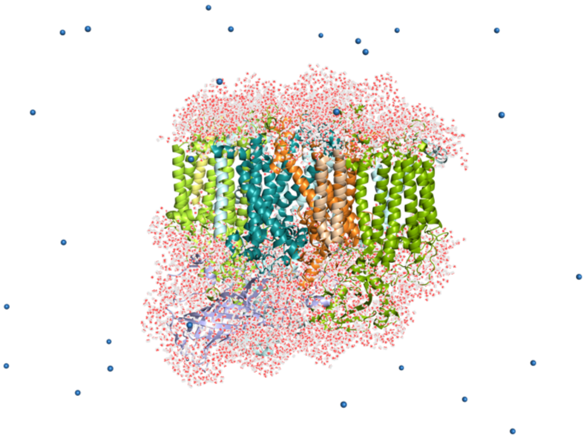


Fig. 3.8 The final setup of the PSII monomer used for all QM/MM calculations in this work. The figure is reproduced in part with permission from *Chem. Sci.* **2023,** *14*, 9503-9516.[80]

This section provides a detailed description about setting up QM/MM calculations in ORCA.[4, 5, 7] The QM/MM multiscale implementation in ORCA relies upon two files one containing the force field parameters (*.ORCAFF.prms*) and the other containing the structural coordinates of the protein configuration (MD snapshot in PDB format). The former is created by converting the AMBER topology file (*.prmtop*) generated by the *tleap* module during the system setup, using the *orca_mm* module. These files are directly read by the ORCA program to perform QM/MM calculations without the need for any external MM engine or QM/MM interface. All the QM/MM calculations reported in **Chapters 4, 5** and **6** are performed using the electrostatic embedding technique described in **Chapter 1** (**Section 1.2.3**). The QM/MM boundaries cutting through covalent bonds were capped by hydrogen link atoms and charge-shift method is used to avoid over-polarization of the QM region by the MM atoms in the vicinity.

3.6.1 Definition of the QM regions

The first most important step towards setting up any QM/MM calculation is the appropriate definition of the "QM region". In biomolecular systems the QM region likely comprises a small part of the protein where the main chemical activity takes place, for instance, the enzyme active sites or photoactive chromophores. The rest of the system, "MM region" is defined using point charges. In this section we describe in detail the choice of QM regions in the PSII complex.

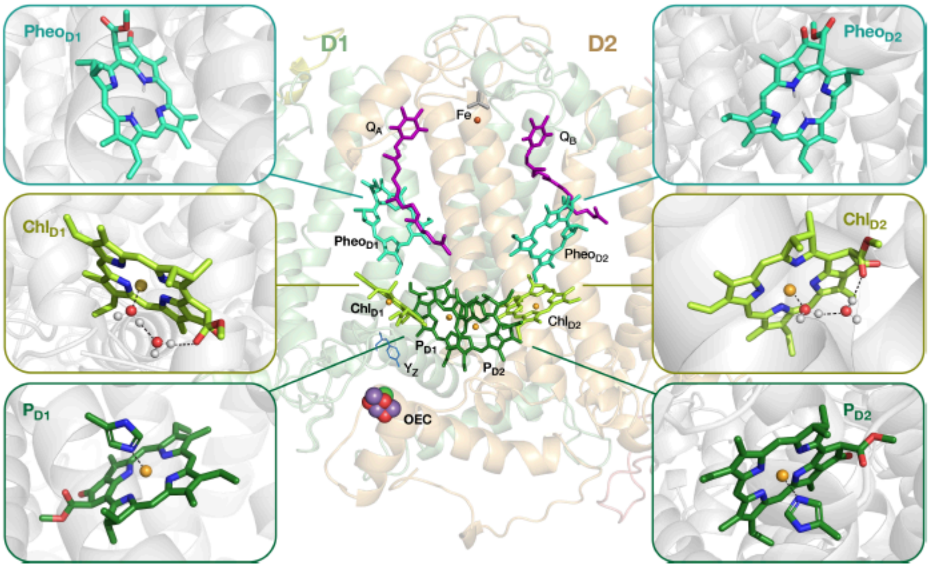


Fig. 3.9 Depiction of the arrangement of pigments in the PSII-RC within the D1 (green) and D2 (orange) proteins. The QM regions of the six chromophores along with their MM environment are shown in the insets. The QM regions shown here follows the work described in **Chapters 4** and **5**. The figure is reproduced with permission from *J. Chem. Theory Comput.* **2024**, https://doi.org/10.1021/acs.jctc.4c00578.

Reaction Center (RC): The work described in **Chapters 4** and **5** are based on the RC of PSII. The RC consists of six pigments arranged pseudo-symmetrically along the D1 and D2 subunits (**Fig 3.9**). The work performed on the RC mainly focuses on the excited states, redox and EPR properties of the pigments based on QM/MM calculations in order to account for the electronic properties and accurately describe the protein environment at the same time. The QM/MM results described in **Chapter 4** are based on the "crystal-like" protein snapshot derived from the early equilibration protocol. In the QM/MM set-up, the entire PSII monomer is considered with 8000 waters, typically including all the waters present inside the protein cavities and ~7 Å bulk-region around the protein. Appropriate number of Na^+ ions are retained to ensure a neutral system (**see Fig. 3.8**). The QM regions in the RC consisted of the individual pigments (P_{D1}, P_{D2}, Chl_{D1}, Chl_{D2}, $Pheo_{D1}$, $Pheo_{D2}$), pigment pairs (Chl_{D1}–$Pheo_{D1}$, Chl_{D2}–$Pheo_{D2}$, P_{D1}–P_{D2}) and pigment assemblies (Chl_{D1}–$Pheo_{D1}$–P_{D1}–P_{D2}, Chl_{D2}–$Pheo_{D2}$–P_{D1}–P_{D2} and Chl_{D1}–$Pheo_{D1}$–Chl_{D2}–$Pheo_{D2}$). For each pigment, the chlorin macrocycles and the phytyl chains were included in the QM region up to C^{17} (truncated as a methyl group) and the rest of the chain was treated in the MM region. The axial ligands to the Mg^{2+} (His198 for P_{D1}, His197 for P_{D2} and waters for Chl_{D1}, Chl_{D2}) are also included in the QM region.

CP43 antenna complex: The work described in **Chapter 6** is based on the CP43 core antenna protein of PSII. The CP43 protein contains 13 embedded Chl pigments, distributed in three layers, with four Chls present in the lumenal layer (C1-C4), eight in the stromal layer (C5-C9, C11-13), and one (C10) positioned in the center of the transmembrane region. The structure and function of the CP43 protein are discussed in detail in **Chapters 1** and **6**. The work on the CP43 antenna protein is focused on using the QM/MM approach to investigate low-energy excited states for all CP43 Chls, explicitly accounting for interactions of the pigments with the PSII protein. The QM/MM computations on the CP43 chlorophylls are performed on the equilibrated PSII model consisting of 76,035 atoms (**Fig. 3.8**). The detailed protocol for equilibration dynamics is discussed in **Section 3.4**. The chlorin macrocycles with the axial ligand to Mg^{2+} and the side chain of residues H-bonded to the keto group at the 13^1-carbon position (ring E) were included in the QM region. The phytyl chains were included in the QM region up to C^{17} and the rest of the chain was treated in the MM region similar to the RC pigments.

3.6.2 Geometry optimizations

For all QM/MM geometry optimizations, the total system is divided into two distinct regions or layers: Active and Static. The active region is allowed to move during the optimization procedure and is defined by the QM region along with a part of the MM region (around the QM region) whereas the static region remains fixed in the optimization procedure. The static part of the MM region can only influence the active region through the electrostatic effect of the point charges. The choice of the QM region during geometry optimizations cannot by generalized and needs to be carefully determined on a case-by-case basis. It is important to note that the active region should include the whole QM region and only a small part of the MM region (see **Fig. 3.10**).

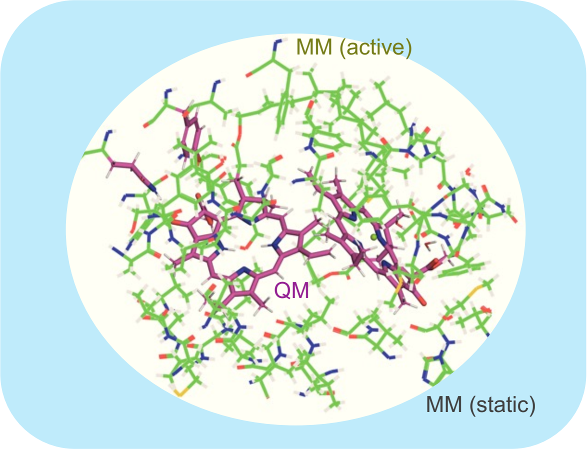


Fig. 3.10 Schematic representation of the QM and MM partitioning for RC pigments. The inner (purple) region indicates the **QM region** consisting of the pigments under investigation, while the outer region (green) represents the **active MM** atoms. The remaining MM system (blue) is treated as **static** point charges**.**

For a given structural configuration (snapshot) of the PSII protein, generally QM/MM geometry optimizations are performed on the individual pigments.[14, 81] For instance, Chl_{D1}, Chl_{D2}, Pheo_{D1}, and Pheo_{D2} are optimized individually while the $\text{P}_{\text{D1}}\text{P}_{\text{D2}}$ pair is treated as a single QM unit due to the dimeric nature and close proximity of the two Chls. However, in some cases one can also optimize the pigments pairwise i.e., Chl_{D1}–Pheo_{D1}, Chl_{D2}–Pheo_{D2}, P_{D1}–P_{D2} in order to account for the effects of pigment-pigment interactions within a single QM entity. It is noted that specific pigment pairs at the RC are structurally uncoupled and that the geometries obtained by directly optimizing a tetramer are essentially identical compared to the combination of pairwise-optimized structures.[82] For the work described in **Chapter 4**, the QM geometries derived from the pairwise optimization of the pigments is embedded within a single MM region and this QM/MM setup was used for computing the singlet and triplet charge-transfer excitations of the RC.[80] However, local properties like the triplet ground states energies and EPR parameters were computed on individually optimized geometries of each pigment in their triplet states. The QM/MM results obtained from this setup are discussed in detail in **Chapter 4**. Most importantly, the investigation of charge-transfer (CT) excitations in the PSII-RC requires special handling of the QM regions and it is recommended to optimize the pigments that maybe involved in CT excitations, as a pair. The work described in **Chapter 5**, follows the QM/MM optimization of the Chl_{D1}–Pheo_{D1} pigment pair, whereas the MM region is defined by the PSII protein containing a variant of the D1 protein. In situations, where the RC pigments are involved in H-bonding interactions with surrounding amino acid side chains or waters, the H-bonded components are also included in the QM region during geometry optimizations.

For the CP43 protein, a similar optimization protocol was employed as described above.[83] For each Chl, the chlorin macrocycles along with the axial ligand to Mg^{2+} and the side chain of residues H-bonded to the keto group at the 13^1-carbon position (ring E) are included in the QM region. For the Chl pairs, we performed a constrained optimizations where the QM/MM optimized geometries of individual pigments were combined and kept fixed while only the MM active region was relaxed. This setup follows the work described in **Chapter 6**.

In all QM/MM optimizations, the active region consisted of all protein residues and solvent molecules within 10 Å around the QM region, measured from the center of the chlorin macrocycle as illustrated in **Fig. 3.10**. All QM/MM geometry optimizations, employed the PBE[84] DFT functional and Def2-TZVP[85] basis set, along with D3(BJ) dispersion corrections.[86, 87] The resolution of identity (RI) approximation in combination with Weigend's universal def2/J[88, 89] auxiliary basis sets were used to speed up the calculations of Coulomb integrals. Tight convergence criteria and the default integration grid (DefGrid2 in ORCA) are used throughout.

3.6.3 Computation of Excitation Energies

For all QM based multiscale approaches, the choice of an appropriate QM method to describe the electronic structure and spectroscopic properties of the system is crucial. The excitation energies for all the photosynthetic pigments studied in this work, are computed with time TD-DFT methods, within the QM/MM framework. The TD-DFT calculations were performed using the range-separated ωB97X-D3(BJ) functional (modified version of ωB97X-V[90] with D3BJ correction) along with Def2-TZVP[85] basis sets without the Tamm–Dancoff approximation (TDA). The long-range-corrected functional has a fixed exact (Hartree–Fock) exchange of 16.7% (short-range) that increases to 100% at long range with a range-separation parameter of 0.30 bohr^{-1}. The performance of this functional towards the efficient treatment of excited states and electrochromic shifts of photosynthetic pigments has already been confirmed in the past via direct comparisons with similarity transformed equation of motion coupled cluster theory (STEOM-CCSD).[81, 91] In case of the RC triplet states described in **Chapter 4**, excitation energies were computed on the pair-optimized ground state singlet geometries (i.e., spin-restricted DFT reference), employing the TDA approximation, because the “triplet instability” problem of spuriously low-lying excitations for complex systems can be overcome to a large extent by using the TDA approach.[92, 93] The excited states for isolated pigments were computed using gas phase TD-DFT whereas in the case of the RC the electrostatic effects of the protein environment on the excited states were included through MM point charges of the entire PSII monomer. For the Chls in CP43, vertical excitation energies are computed on the optimized ground state geometries of each Chl, using full TD-DFT without TDA as described above. The first 8 excited states (roots) were computed, thus covering the entire Q-band range and further low-lying excited states for individual chlorophylls as well as for dimers.[61, 80, 94] Transition dipole moments (to be used in subsequent PMM calculations, see **Chapter 6** for more details) for the ground and first two excited states were additionally calculated for all the CP43 Chls.

The RIJCOSX approximation[88] and the corresponding auxiliary basis sets were used in all QM/MM calculations. VeryTightSCF convergence criteria is applied, along with dense integration grids (DefGrid2). The nature of the excited states is characterized based on Natural Transition Orbitals (NTOs)[95] and TD-DFT difference densities using the *orca_plot* module.

The QM/MM methods described here majorly follows the work in **Chapters 4, 5** and **6**. The other QM based multiscale methodologies employed to calculate spectroscopic and redox properties of specific PSII cofactors, are further described in great detail in each subsequent Chapter.

3.7 REFERENCES

1. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, . . . K. M. Merz, Jr., *J. Chem. Inf. Model.*, 2023, **63**, 6183-6191.
2. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Computational Molecular Science*, 2013, **3**, 198-210.
3. D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., . . . R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
4. F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.
5. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
6. F. Neese, *WIREs Computational Molecular Science*, 2017, **8**, e1327.
7. F. Neese, *WIREs Computational Molecular Science*, 2011, **2**, 73-78.
8. H. Li, Y. Nakajima, E. Nango, S. Owada, D. Yamada, K. Hashimoto, . . . J.-R. Shen, *Nature*, 2024, DOI: 10.1038/s41586-023-06987-5.
9. M. Suga, F. Akita, M. Sugahara, M. Kubo, Y. Nakajima, T. Nakane, . . . J.-R. Shen, *Nature*, 2017, **543**, 131-135.
10. M. Suga, F. Akita, K. Yamashita, Y. Nakajima, G. Ueno, H. Li, . . . J.-R. Shen, *Science*, 2019, **366**, 334-338.
11. K. Kato, N. Miyazaki, T. Hamaguchi, Y. Nakajima, F. Akita, K. Yonekura and J.-R. Shen, *Communications Biology*, 2021, **4**.
12. R. Hussein, A. Graça, J. Forsman, A. O. Aydin, M. Hall, J. Gaetcke, . . . W. P. Schröder, *Science*, 2024, **384**, 1349-1355.
13. Y. Umena, K. Kawakami, J.-R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
14. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Am. Chem. Soc.*, 2020, **142**, 18174-18190.
15. B. Loll, J. Kern, W. Saenger, A. Zouni and J. Biesiadka, *Nature*, 2005, **438**, 1040-1044.
16. A. Guskov, J. Kern, A. Gabdulkhakov, M. Broser, A. Zouni and W. Saenger, *Nature Structural & Molecular Biology*, 2009, **16**, 334-342.
17. W. L. DeLano, http://www.pymol.org/, 2002.
18. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg and A. L. Lomize, *Nucleic Acids Res.*, 2012, **40**, D370-376.
19. S. Schott-Verdugo and H. Gohlke, *J. Chem. Inf. Model.*, 2019, **59**, 2522-2528.
20. L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157-2164.
21. T. Noji, M. Kondo, K. Kawakami, J.-R. Shen, M. Nango and T. Dewa, *Res. Chem. Intermed.*, 2014, **40**, 3231-3241.
22. F. Guerra, M. Siemers, C. Mielack and A.-N. Bondar, *J. Phys. Chem. B*, 2018, **122**, 4625-4641.
23. N. Sakashita, H. C. Watanabe, T. Ikeda, K. Saito and H. Ishikita, *Biochemistry*, 2017, **56**, 3049-3057.
24. M. R. Machado and S. Pantano, *J. Chem. Theory Comput.*, 2020, **16**, 1367-1372.
25. P. Mulo, I. Sakurai and E.-M. Aro, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 247-257.
26. A. Boussac, J. Sellés and M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2023, **1864**, 148979.
27. Y. Nakajima, N. Ugai-Amo, N. Tone, A. Nakagawa, M. Iwai, M. Ikeuchi, . . . J.-R. Shen, *J. Biol. Chem.*, 2022, **298**, 102668.
28. M. Sugiura and A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2014, **1837**, 1427-1434.
29. M. Sugiura, C. Azami, K. Koyama, A. W. Rutherford, F. Rappaport and A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2014, **1837**, 139-148.

30. M. Sugiura, Y. Kato, R. Takahashi, H. Suzuki, T. Watanabe, T. Noguchi, . . . A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 1491-1499.
31. J. L. Hughes, N. Cox, A. W. Rutherford, E. Krausz, T.-L. Lai, A. Boussac and M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 11-19.
32. B. Loll, M. Broser, P. B. Kós, J. Kern, J. Biesiadka, I. Vass, . . . A. Zouni, 2008, **389**, 609-617.
33. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
34. T. M. Bricker, J. L. Roose, R. D. Fagerlund, L. K. Frankel and J. J. Eaton-Rye, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 121-142.
35. I. Enami, A. Okumura, R. Nagao, T. Suzuki, M. Iwai and J.-R. Shen, *Photosynth. Res.*, 2008, **98**, 349-363.
36. L. Kemmler, M. Ibrahim, H. Dobbek, A. Zouni and A.-N. Bondar, *Phys. Chem. Chem. Phys.*, 2019, **21**, 25449-25466.
37. A. Sirohiwal and D. A. Pantazis, *J. Am. Chem. Soc.*, 2022, **144**, 22035-22050.
38. S. Lorch, S. Capponi, F. Pieront and A.-N. Bondar, *J. Phys. Chem. B*, 2015, **119**, 12172-12181.
39. R. Nagao, T. Tomo and T. Noguchi, *Biochemistry*, 2015, **54**, 2022-2031.
40. L. Gerland, D. Friedrich, L. Hopf, E. J. Donovan, A. Wallmann, N. Erdmann, . . . H. Oschkinat, *ChemBioChem*, 2020, **21**, 1597-1604.
41. P. Bernal-Bayard, C. Alvarez, P. Calvo, C. Castell, M. Roncel, M. Hervas and J. A. Navarro, *Physiol. Plant.*, 2019, **166**, 199-210.
42. F. Guerrero, A. Sedoud, D. Kirilovsky, A. W. Rutherford, J. M. Ortega and M. Roncel, *J. Biol. Chem.*, 2011, **286**, 5985-5994.
43. M. Roncel, A. Boussac, J. L. Zurita, H. Bottin, M. Sugiura, D. Kirilovsky and J. M. Ortega, *J. Biol. Inorg. Chem.*, 2003, **8**, 206-216.
44. J. S. Vrettos, M. J. Reifler, O. Kievit, K. V. Lakshmi, J. C. de Paula and G. W. Brudvig, *J Biol Inorg Chem*, 2001, **6**, 708-716.
45. P. Li and K. M. Merz, *Chem. Rev.*, 2017, **117**, 1564-1686.
46. L. Hu and U. Ryde, *J. Chem. Theory Comput.*, 2011, **7**, 2452-2463.
47. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, . . . P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.
48. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
49. C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.*, 2014, **10**, 865-879.
50. C. J. Dickson, R. C. Walker and I. R. Gould, *J. Chem. Theory Comput.*, 2022, **18**, 1726-1736.
51. Å. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen, *J. Phys. Chem. B*, 2012, **116**, 11124-11136.
52. I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B*, 2009, **113**, 13279-13290.
53. I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B*, 2008, **112**, 9020-9041.
54. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
55. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, . . . P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
56. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372-1377.
57. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
58. E. Sigfridsson and U. Ryde, *J. Comput. Chem.*, 1998, **19**, 377-395.

59. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
60. X. He, V. H. Man, W. Yang, T.-S. Lee and J. Wang, *J. Chem. Phys.*, 2020, **153**, 114502.
61. A. Sirohiwal and D. A. Pantazis, *Acc. Chem. Res.*, 2023, **56**, 2921-2932.
62. A. Sirohiwal and D. A. Pantazis, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200356.
63. T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580-592.
64. J. Zhang and T. Lu, *Phys. Chem. Chem. Phys.*, 2021, **23**, 20323-20328.
65. D. A. Pantazis, W. Ames, N. Cox, W. Lubitz and F. Neese, *Angew. Chem. Int. Ed.*, 2012, **51**, 9935-9940.
66. P. Li, L. F. Song and K. M. Merz, Jr., *J. Phys. Chem. B*, 2015, **119**, 883-895.
67. P. Li, B. P. Roberts, D. K. Chakravorty and K. M. Merz, Jr., *J. Chem. Theory Comput.*, 2013, **9**, 2733-2748.
68. M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, 2003, **24**, 129-142.
69. R. J. Loncharich, B. R. Brooks and R. W. Pastor, *Biopolymers*, 1992, **32**, 523-535.
70. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
71. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
72. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878-3888.
73. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542-1555.
74. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374-380.
75. D. R. Roe and T. E. Cheatham, III, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
76. M. Capone, A. Sirohiwal, M. Aschi, D. A. Pantazis and I. Daidone, *Angew. Chem. Int. Ed.*, 2023, **62**, e202216276.
77. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
78. D. R. Roe and T. E. Cheatham Iii, *J. Comput. Chem.*, 2018, **39**, 2110-2117.
79. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33-38.
80. S. Bhattacharjee, F. Neese and D. A. Pantazis, *Chem. Sci.*, 2023, **14**, 9503-9516.
81. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2021, **17**, 1858-1873.
82. A. Sirohiwal and D. A. Pantazis, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200356.
83. S. Bhattacharjee, S. Arra, I. Daidone and D. A. Pantazis, *Chem. Sci.*, 2024, **15**, 7269-7284.
84. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
85. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
86. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
87. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
88. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
89. K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283-290.
90. N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904.
91. A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsak and D. A. Pantazis, *J. Phys. Chem. B*, 2020, **124**, 8761-8771.
92. D. Jacquemin, E. A. Perpète, I. Ciofini and C. Adamo, *J. Chem. Theory Comput.*, 2010, **6**, 1532-1537.

93. D. Hait, T. Zhu, D. P. McMahon and T. Van Voorhis, *J. Chem. Theory Comput.*, 2016, **12**, 3353-3359.
94. A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsák and D. A. Pantazis, *J. Phys. Chem. B*, 2020, **124**, 8761–8771.
95. R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775-4777.

4

Triplet States in the Reaction Center of Photosystem II[‡]

[‡] This chapter is reproduced in part with permission from "Triplet states in the reaction center of Photosystem II" by Bhattacharjee, S.; Neese, F.; Pantazis, D. A., *Chem. Sci.* **2023**, *14*, 9503-9516. https://doi.org/10.1039/D3SC02985A

4.1 Introduction

Oxygenic photosynthesis involves a series of light-dependent electron transfer reactions which are carried out by membrane-bound pigment-protein complexes.[1, 2] The reactions at these energy-converting enzymes generate a transmembrane electrochemical potential gradient to drive the synthesis of ATP. The first enzyme in the photosynthetic chain is Photosystem II (PSII), a dimeric multi-subunit protein–pigment complex responsible for the four-electron oxidation of water into molecular oxygen and two-electron reduction of a mobile plastoquinone acceptor (Q_B).[2-7] The light-driven charge separation and the initial electron transfer events occur at the reaction center (RC) of PSII. This comprises four chlorophyll molecules, namely the P_{D1} and P_{D2} central pair flanked by the "accessory" chlorophylls Chl_{D1} and Chl_{D2}, and two pheophytin molecules, $Pheo_{D1}$ and $Pheo_{D2}$. The RC pigments are arranged pseudo-symmetrically along the D1 and D2 heterodimeric subunits of PSII (**Fig. 4.1**) that are highly conserved across photosynthetic organisms.[8]

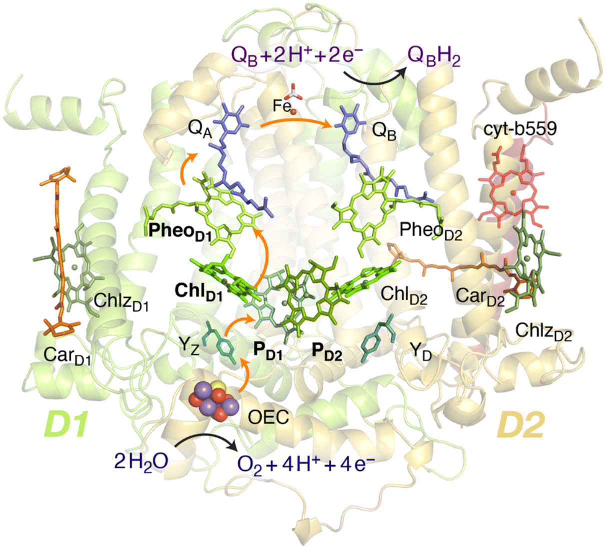


Fig. 4.1 Reaction center (RC) pigments and other important cofactors, with a schematic representation of electron flow along the D1 branch.

The excitation energy transfer from external light harvesting complexes and the internal antennae CP43 and CP47 initiates the electron transfer process along the D1 branch of the RC (Fig. 1). Charge-transfer excited states of mostly $\text{Chl}_{\text{D1}}{}^{\delta+}\text{Pheo}_{\text{D1}}{}^{\delta-}$ character are created, leading to formation of the primary charge separated radical pair $\text{Chl}_{\text{D1}}{}^{+}\text{Pheo}_{\text{D1}}{}^{-}$ [6, 9-18] and the cationic charge is then distributed over the $\text{P}_{\text{D1}}\text{P}_{\text{D2}}$ pair (often referred to as $\text{P}_{680}{}^{+}$).[10, 19-21] This highly oxidizing radical cation (estimated E_{m} of 1.1–1.3 V) is the strongest known oxidant in biology and drives water oxidation at the oxygen-evolving complex.[3, 22, 23] Under normal conditions, charge recombination of the initially formed radical pairs $[\text{Chl}_{\text{D1}}{}^{+}\text{Pheo}_{\text{D1}}{}^{-}]$[4, 11, 16] (or possibly $[\text{P}_{\text{D1}}{}^{+}\text{Pheo}_{\text{D1}}{}^{-}]$ in some reaction centers)[24] is prevented by forward electron transfer from Pheo_{D1} to the primary plastoquinone acceptor Q_{A} within a few hundred ps. This leads to formation of the "closed RC" state with a reduced Q_{A}.[5, 25-29] If the plastoquinone pool remain reduced, electron transfer from $\text{Q}_{\text{A}}{}^{-}$ to the mobile acceptor Q_{B} is inhibited, thus preventing further electron transfer from Pheo_{D1} to Q_{A}. This can facilitate charge recombination[30-35] within the RC and enable formation of chlorophyll triplet states prior to relaxation to the ground state.[14, 30-33, 36-46] Triplet states are detrimental as they can readily generate chemically active singlet oxygen ($^{1}\text{O}_2$) that reacts with the protein causing oxidative stress.[47, 48] The D1 protein embeds most crucial redox cofactors in PSII, including the oxygen-evolving complex (OEC), and thus photodamage can lead to a disruption of the entire photosynthetic machinery. Correlation has been reported between $^{1}\text{O}_2$ production and the extent of photodamage of the D1 protein on exposure to excess light.[49-51] All photosynthetic organisms therefore naturally adopt intrinsic strategies of photoprotection by efficiently quenching chlorophyll triplet states either by redox active cofactors (e.g. $\text{Q}_{\text{A}}{}^{-}$ in the RC)[29, 32, 52] or carotenoids[53-55] (e.g. in the bacterial RC or antenna complexes), but the exact molecular mechanisms of these phenomena are not fully understood. Therefore, it is useful to have a reliable description of the nature and localization of triplet states themselves, as an essential basis for understanding photoprotection mechanisms in PSII.

Chlorophyll triplet states, in addition to being highly reactive, serve as chemical probes to investigate primary electron transfer pathways and characterize the chemical environment of photosynthetic reaction center pigments.[44] Electron paramagnetic resonance (EPR) and electron–nucleus double resonance (ENDOR) spectroscopies[29, 37, 39, 44, 56-69] and other spectroscopic approaches including Fourier-transformed infrared (FT-IR) and optically detected magnetic resonance (ODMR)[36, 38, 39, 51, 54, 70-76] suggest that the "primary donor" triplet is located on an individual accessory chlorophyll (Chl_{D1} or Chl_{D2}) at cryogenic temperatures.[30, [56-58, 61, 77] It has also been suggested that the triplet is partially shared with other chlorophylls at the RC at higher temperatures, but this has not been well characterized.[33, 62] It is important to note that many studies report varying observations depending on the type of preparation and conditions used, as in the case of D_1D_2Cytb_{559} samples[60, 62, 78, 79] or samples with chemically reduced quinone (Q_A^{-}/Q_A^{2-}).[25, 29, 46, 52]

Various chemical and photo-physical properties of pigments such as site energies and redox potentials [10, 22, 43, 80, 81] are directly or indirectly controlled by the surrounding protein matrix,[82-84] as already established in the case of charge transfer states involving the RC pigments.[12, 24] From a methodological perspective, this establishes the need for multilayer approaches to provide an accurate quantitative description of how inter-pigment and pigment–protein interactions determine spectroscopic properties. Previous excited state calculations based on time-dependent density functional theory (TD-DFT) and quantum-mechanics/molecular-mechanics (QM/MM) simulations on pigment assemblies have shown that the lowest singlet excitations in the RC are characterized by a mixture of excitonic and [$\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$] or [$\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$] charge-transfer (CT) character.[12, 24, 85, 86] However, a coherent description of excited and ground triplet states is lacking. The excitation profiles of all RC pigments in their triplet states are important elements for establishing possible routes of triplet delocalization[87, 88] and triplet–triplet energy transfer (T-TET) onto other pigments within the PSII core complex.[89]

In this work, we use a membrane-bound model of an entire PSII monomer as the basis for multiscale QM/MM modelling to study singlet–triplet excitations as well as relaxed triplet states within the RC pigments. The quantum chemical descriptions of both local and charge-transfer excitations in oligomeric assemblies are obtained by range-separated TD-DFT. We employ our QM/MM approach to also compute EPR properties of all triplet states localized on each chromophore, and compare the results with available spectroscopic data.[37, 39, 65] Finally, we study how charge transfer pathways and triplet formation at the RC depend on the redox state of the primary quinone (Q_A) acceptor and of the OEC.[61, 90] Overall, the present work contributes to a more complete understanding of the nature of triplet states within the RC of PSII, of their electronic and spectroscopic properties, and of the electrostatic control exerted by the PSII protein matrix.

4.2 Methodology

4.2.1 QM/MM Setup

The classical membrane-embedded MM setup was built using the 1.9 Å resolution crystal structure of PSII from the thermophilic cyanobacterium *T. vulcanus* (PDB ID: 3WU2).[8] In the current study we chose a snapshot that resembles the X-ray structure configuration[8, 12] from an initial MD equilibration.[24] The details on the QM/MM setup are described in **Chapter 3** and shown in **Fig. 4.2**. The oxidation states of the Mn ions of the OEC were assumed to correspond to the dark-stable S_1 state of its catalytic cycle. In order to model the "closed" reaction center in the $S_2Q_A^-$ state, the AMBER parameter file was modified with the electrostatic charges for both cofactors (Q_A^- and OEC) based on the standard MK-RESP methodology.[91-93] For the semiquinone, geometry optimization was first performed at the B3LYP/def2-SVP level[94, 95] and then single-point calculations were performed at the HF/6-31G* level of theory in ORCA.[96] In order to compute the charges on the OEC (Mn_4CaO_5) a small cluster model was taken including the amino acid side chains directly coordinated to each metal site (see **Chapter 3** for details). The OEC was then modelled in the S_2 state of the Kok–Joliot cycle, i.e. with formal oxidation states Mn1(III)-Mn2(IV)-Mn3(IV)-Mn4(IV).[2, 3] The corresponding RESP charges are derived from B3LYP/6-31G*.[94, 95] The RESP fitting of the charges was performed using Multiwfn.[97] The charge on backbone atoms of the coordinated residues on the OEC is carefully restrained on the link atoms, according to the standard residue charges of the original AMBER force field.[93]

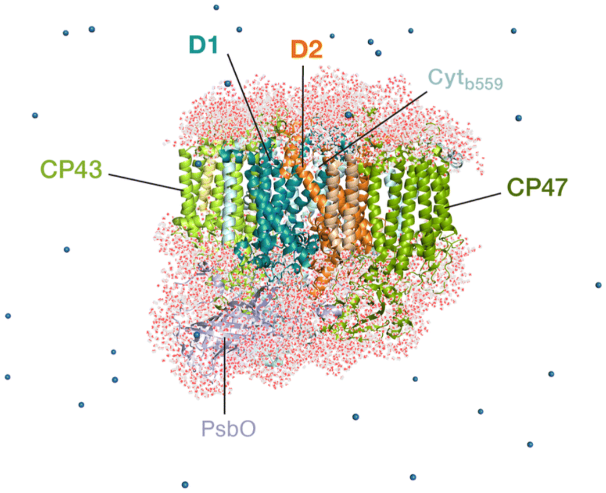


Fig. 4.2 The all-atom model of the PSII monomer used for the QM/MM computations, indicating selected major protein subunits. Appropriate number of sodium ions were retained to maintain charge neutrality of the model. The overall system contains 76, 056 atoms.

All QM/MM calculations were performed using the multiscale module of the ORCA 5.0 suite, which incorporates the electrostatic embedding technique.[96, 98, 99] The hydrogen link atom approach was employed to cut through C–C covalent bonds and the charge-shift (CS) scheme was used to avoid over polarization of the QM region. Along with the chlorin macrocycles, the axially coordinated ligands to Mg^{2+} were also treated at the QM level. For Chl_{D1} and Chl_{D2}, the water molecule hydrogen bonded to the axially ligated water and ester group attached to the 13^2-carbon position on ring E is also included in the QM region. Similarly, the axial histidines (His198 and His197) in case of P_{D1} and P_{D2} were also treated at the QM level. The phytyl chains were included in the QM region up to C^{17} (truncated as a methyl group) and the rest of the chain was treated in the MM region.

4.2.2 Geometry optimization

For geometry optimizations in the QM/MM framework, the complete system was further subdivided into two parts: active and static. The active region consists of atoms within the QM and MM regions, which are flexible during the optimization, whereas the remaining MM atoms are fixed and only contribute to the electrostatics. The original pair-optimized QM geometries (i.e., Chl_{D1}–$Pheo_{D1}$, Chl_{D2}–$Pheo_{D2}$ and P_{D1}–P_{D2}), for the ground state singlet states ($S = 0$) were used as starting structures.[24] The ground triplet states ($S = 1$) of all RC pigments were optimized individually except for the central pair ($P_{D1}P_{D2}$), which is considered as a single dimeric unit. For individual pigments, complete amino acid residues and waters within 10 Å around the QM region were included in the active region, whereas a larger active region was chosen around the $P_{D1}P_{D2}$ pair (~10 Å around each of P_{D1} and P_{D2}). The Perdew–Burke–Ernzerhof (PBE) functional[100] was used to optimize the QM regions using the def2-TZVP basis set,[101] along with D3(BJ) dispersion corrections.[102, 103] Dense DFT integration grids (DefGrid2 in ORCA convention) were used in all optimizations. The resolution of identity approximation (RI) was used to speed up the calculation of Coulomb integrals with the corresponding auxiliary basis set (def2/J).[104, 105] All QM/MM geometry optimizations were performed using the L-BFGS optimizer.[106]

4.2.3 Excitation energies

Vertical triplet excitation energies were computed on the pair-optimized ground state singlet geometries (i.e., spin-restricted DFT reference), employing the Tamm–Dancoff approximation (TDA) to TD-DFT. It has been shown that the “triplet instability” problem of spuriously low-lying excitations for complex systems can be overcome to a large extent by using the TDA approach.[107] In this work, we also considered the effect of this approximation on the singlet excitation energies of photosynthetic pigments, which remains a challenging problem for approximate TD-DFT.[108, 109] All TD-DFT calculations were performed using the range-separated ωB97X-D3(BJ) functional (modified version of ωB97X-V[110] with D3BJ correction) along with def2-TZVP basis sets. The long-range-corrected functional has a fixed exact HF exchange of 16.7% (short-range) that increases to 100% at long range with a range-separation parameter of 0.30 bohr^{-1}. The performance of this functional towards the efficient treatment of excited states and electrochromic shifts using TD-DFT has already been confirmed in the past via direct comparisons with similarity transformed equation of motion coupled cluster theory (STEOM-CCSD).[84, 111] The RIJCOSX approximation[104] was used to speed up the calculations and the corresponding auxiliary basis sets were used throughout. VeryTightSCF convergence criteria were applied throughout, along with dense integration grids (DefGrid2). The first 10 singlet (S = 0) and triplet (S = 1) excited states were computed for individual RC pigments as well as for oligomeric assemblies. This approach effectively describes the entire Q-band range and all low-lying excited states with local excitation (LE), charge-transfer (CT), and mixed LE/CT characters. The excited states for isolated pigments were computed using gas phase TD-DFT whereas in the case of the reaction center the electrostatic effects of the protein environment on the excited states were included through MM point charges of the entire PSII monomer. We further obtained the low-energy triplet (S = 1) excited states for pigment assemblies along the D1 [$P_{D1}P_{D2}Chl_{D1}Pheo_{D1}$] and D2 [$P_{D1}P_{D2}Chl_{D2}Pheo_{D2}$] branches (see **Fig. 4.1**). It is noted that specific pigment pairs at the RC are structurally uncoupled and that the geometries obtained by directly optimizing a tetramer are essentially identical compared to the combination of pairwise-optimized structures.[24] The inclusion of tetramers in QM optimizations do not obviously alter the excited state energetics and provide the same qualitative picture of low-lying CT states and local excitons as the pair-optimized structures.

4.2.4 EPR parameters

The isotropic hyperfine coupling constants of all the hydrogen atoms are computed on the localized triplet states of individual pigments (Chl_{D1}, P_{D1}, P_{D2}, Chl_{D2} and $Pheo_{D1}$) including the protein using the QM/MM approach. All EPR parameters were computed within the framework of a DFT-based coupled-perturbed self-consistent field approach (CP-SCF)[112] on the QM/MM optimized geometries of the triplet (S = 1) states with separate QM regions defined for each

RC chromophore. For the hyperfine coupling constants and *g*-tensors, we used the TPSSh functional[113] with Barone's EPR-II basis set[114] on hydrogen atoms and def2-TZVP on the remaining atoms in the QM region. The RIJCOSX approximation and VeryTightSCF convergence criteria were used along with the highest DefGrid3 integration grids.[104] The triplet *g*-tensors were computed in conjunction with the spin–orbit mean-field (SOMF) approximation for the spin–orbit coupling.[112, 115] The spin-spin contribution to the zero field splitting (ZFS) tensors[116] (*D* and *E* respectively) were computed using the restricted open shell Kohn-Sham (ROKS) framework, as this approach has been seen to yield better agreement with experimental results than unrestricted (UKS), for triplet states of several organic molecules involving π electrons.[117]

4.3 Results and Discussion

4.3.1 Singlet–triplet excitations in individual pigments

The electrostatic effects of the protein matrix are known to modulate the excited state properties of reaction center pigments.[84] Previous work identified that the protein matrix is exclusively responsible for creating transverse and lateral excitonic asymmetry among the pigments within the PSII-RC.[12, 82, 83] This asymmetry leads to trapping of the excitation energy and initiation of primary charge-separation in the D1 branch. In the presence of the protein matrix the pigment with the lowest site energy is computed to be $\mathrm{Chl_{D1}}$.[12] Detailed work on pigment assemblies additionally showed that the lowest singlet excited state is localized on the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair and that this is usually a mixture of excitonic and charge-transfer (CT) $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ character.[24] The corresponding CT state involving the [$\mathrm{Chl_{D2}}$–$\mathrm{Pheo_{D2}}$] pair on the inactive D2 is higher in energy, thus elucidating the excitonic asymmetry of the RC, where the protein matrix stabilizes excited CT states on the D1 branch. However, the explicit role of the protein electrostatics in controlling the excited state energetics of the triplet states has never been studied. It is also not clear if asymmetry exists at all in the case of triplet excitations. This information would be useful for understanding triplet-state formation and subsequently establish the role of protein matrix in photoprotection.

As a first step, we computed the singlet and triplet excitation energies of individual RC pigments using TD-DFT in the QM/MM framework. The Q and B bands of the absorption spectra of porphyrin-like macrocyclic compounds are described according to the Gouterman model,[118] which involves excitations within the four frontier molecular orbitals HOMO–1, HOMO, LUMO, and LUMO+1, delocalized over the chlorin ring.[111] For instance, the fundamental singlet excitation of the chlorophylls is the Q_y band (S_1), corresponding to HOMO→LUMO and secondarily to HOMO–1→LUMO+1 excitation. Based on the TD-DFT calculations, the lowest triplet excitations consist of two unpaired electrons, ferromagnetically coupled to each other in two singly occupied orbitals (SOMO 1, SOMO 2), also delocalized over the chlorophyll macrocycle.[119, 120] Our TDA-TDDFT results (see **Tables 4.5–4.9**) show that the two lowest energy triplet excited states (T_1,T_2) of RC chlorophylls are characterized by HOMO → LUMO (in the range of 1.22–1.30 eV) and HOMO–1→ LUMO (range of 1.73–1.78 eV) transitions. Furthermore, in all four central chlorophylls (i.e., Chl_{D1}, P_{D1}, P_{D2}, Chl_{D2}) the two lowest triplet excited states (T_1 and T_2) are energetically lower than the corresponding singlet excitations (S_1 and S_2). This observation suggests that the lowest triplet local excitations are likely to result from spin-orbit induced inter system crossing (ISC) from the corresponding first singlet excited state (S_1) of each chlorophyll.[84]

Table 4.1 Vertical excitation energies of the lowest singlet (**S_1**) and triplet (**T_1**) excited states along with the respective **S–T** gaps, in the presence and absence of the PSII protein matrix, calculated using TD-DFT (ωB97X-D3BJ/def2-TZVP). **ΔT1** represents the geometry relaxation of the first triplet state. The gas-phase excited state calculations were performed using the QM/MM optimized geometries. All values are reported in eV.

Method	**Δ SCF**		**TD-DFT (in protein)**			**TD-DFT (gas-phase)**		
RC Pigment	**T_1-S_0 (opt)**	**ΔT_1 (opt)**	**S_1**	**T_1**	**S_1-T_1**	**S_1**	**T_1**	**S_1-T_1**
Chl_{D1}	0.920	0.300	1.818	1.220	0.598	1.884	1.290	0.594
P_{D1}	0.994	0.311	1.859	1.305	0.554	1.898	1.323	0.575
P_{D2}	0.978	0.313	1.859	1.291	0.568	1.897	1.314	0.583
Chl_{D2}	0.970	0.318	1.878	1.288	0.590	1.900	1.319	0.581

The computation of singlet excitation energies without protein electrostatics demonstrates that both Chl_{D1} and Chl_{D2} pigments have similar site energy in the gas phase (1.88 eV and 1.90 eV, respectively, see **Table 4.1**). Moreover, the nature of excitations and participating orbitals for the chlorophyll triplet remains consistent even in the absence of the explicit PSII protein environment. On the other hand, calculations done with full inclusion of protein electrostatics red-shifts the first excited state for both pigments. This effect is more pronounced for $^1Chl_{D1}$ (1.82 eV) compared to $^1Chl_{D2}$ (1.88 eV). Interestingly, similar spectral shifts are obtained for the lowest triplet state (T_1), where we observed protein-induced red shifts highest for $^3Chl_{D1}$ (70 meV) followed by $^3Chl_{D2}$ (31 meV), $^3P_{D1}$ (18 meV) and $^3P_{D2}$ (23 meV). The excitation energy of $^3Pheo_{D1}$ was found to be 17 meV higher than $^3Chl_{D1}$, and about 10 meV higher than the T_1 states of P_{D1}, P_{D2} and Chl_{D2}. Clearly, the signature of transverse excitonic asymmetry within the RC is preserved for the lowest localized triplet excitations. Nevertheless, it will be interesting to see how the absolute S_1 and T_1 excitation energies and S_1–T_1 gap is modulated by the protein matrix as these states should be involved in S–T intersystem crossing. The vertical excitation energies of the lowest singlet and triplet state along with the respective S–T gaps, in the presence and absence of the protein, are listed in **Table 4.1**. It is important to note that the protein matrix induces an asymmetry in tuning the S_1–T_1 gap for the accessory chlorophylls Chl_{D1} and Chl_{D2}. In the case of Chl_{D1}, both S_1 and T_1 are red-shifted by ca. 70 meV in the presence of protein compared to the gas phase. In the case of Chl_{D2} the S_1–T_1 gap is 0.59 eV in the presence of the protein similar to gas phase (0.58 eV).

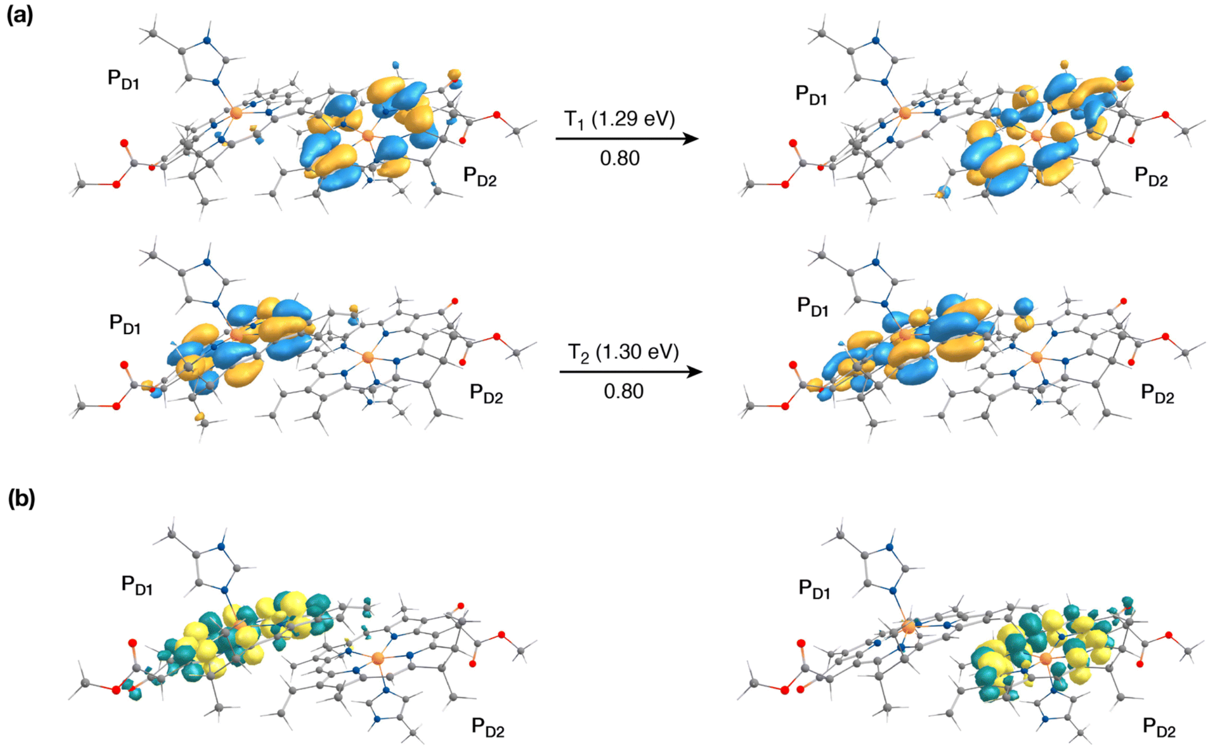


Fig. 4.3 The nature of the lowest triplet excitations shown for the $\mathbf{P_{D1}P_{D2}}$ pair: **(a)** donor and acceptor natural transition orbitals and relative contributions of the transition to the given excited state (here the NTOs coincide with the HOMO and LUMO orbitals of the individual pigments); **(b)** the corresponding difference densities for the lowest singlet-triplet excitations on P_{D1} and P_{D2}.

The singlet excited states on the central pair $P_{D1}P_{D2}$ in the presence of the protein point charges show that the lowest singlet excited states at 1.86 eV and 1.88 eV are a superposition of local excitons on P_{D1} and P_{D2}, respectively (**Table 4.9**). The lowest CT state involving the central pair ($P_{D1}^{\delta+}P_{D2}^{\delta-}$) is significantly higher (ca. 3.2 eV) than the S_0. On the other hand, in the case of triplet excitations, the two lowest triplet states are isoenergetic and correspond to triplet excitons localized on P_{D2} (T_1, 1.29 eV) and P_{D1} (T_2, 1.30 eV) respectively (see **Fig. 4.3**). Our results do not show any low-energy triplet state of the same character as the $^1[P_{D1}{}^{\delta+}P_{D2}{}^{\delta-}]$ CT state mentioned above. Moreover, each triplet excitation spanning a range of 1.40–1.50 eV is attributed to individual pigments (see **Table 4.9**), suggesting that the triplet excitons are entirely localized on either of the two chlorophyll molecules (P_{D1} or P_{D2}) and therefore there is no superposition, in contrast to the singlet excitons. The absence of a low-lying triplet state with CT character is also indicative of the fact that a radical-pair charge recombination may not be favorable to form $^3[P_{D1}P_{D2}]$ states in the RC. However, there can be mixed excitonic and triplet excitations at higher energies, similar to the singlet CT excitations. [12, 24]

4.3.2 Singlet–triplet excitations in pigment assemblies

In photosynthetic RCs the excitation profiles of individual pigments are far from complete, and a thorough understanding of the initial charge-separation and charge recombination events requires insights from excitation energetics of multiple pigments. For instance, $^1[P_{D1}{}^{\delta+}Pheo_{D1}{}^{\delta+}]$ and $^1[Chl_{D1}{}^{\delta+}Pheo_{D1}{}^{\delta-}]$ charge-transfer (CT) excitations were found significantly stabilized, lower than the local excitons, due to the differential effect of the protein matrix.[18, 24] Moreover, the lowest (Q_y) excitation of Chl_{D1} was found to be mixed significantly with the $^1[Chl_{D1}{}^{\delta+}Pheo_{D1}{}^{\delta-}]$ CT state.[18] A number of experimental studies suggest that based on the characteristic spin polarization pattern of the EPR spectra, the observable triplet state should be formed from a charge recombination of the primary radical pair.[37, 56, 58, 61, 64, 66] This further necessitates a quantitative description of the excitation profiles of groups of pigment assemblies, in order to establish a connection between the singlet-triplet CT excitations and the experimentally observable triplet state. Towards this objective, we first computed the low energy singlet and triplet excited states for the tetrameric pigment assemblies along the D1 $[P_{D1}P_{D2}Chl_{D1}Pheo_{D1}]$ and D2 $[P_{D1}P_{D2}Chl_{D2}Pheo_{D2}]$ branches.

The most common mechanism of triplet formation in organic chromophores involves a spin–orbit-induced intersystem crossing (ISC) but singlet fission, radical pair ISC, or spin–orbit charge-transfer ISC can result in triplet formation, particularly in systems with donor-acceptor pigment pairs.[54, 89, 121, 122] Similar studies on biomimetic assemblies have reported that low-lying CT states can promote triplet formation through a charge recombination of donor-acceptor radical pairs followed by ISC.[123, 124] Our TD-DFT results show that the lowest singlet excitations in the $[P_{D1}P_{D2}Chl_{D1}Pheo_{D1}]$ branch correspond to $^1[P_{D1}{}^{\delta+}Pheo_{D1}{}^{\delta-}]$ (1.548 eV) and

$^1[\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ (1.693 eV) CT states, respectively (**Table 4.2**). These results are further in line with recent QM/MM and TDDFT studies.[18, 24]

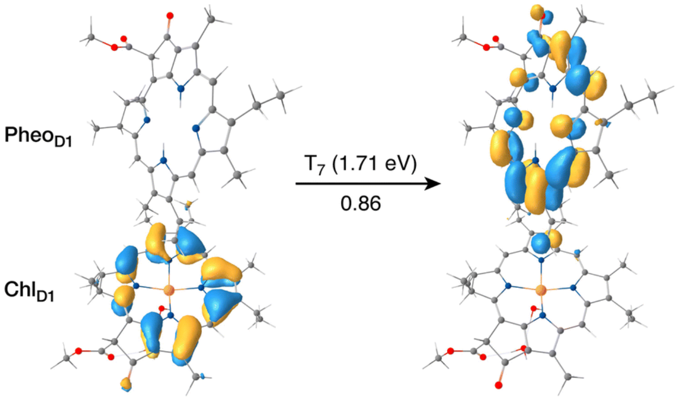


Fig. 4.4 The identity of the triplet excited state with significant **$[\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$** charge transfer character in terms of canonical molecular orbitals and their contribution to the given excitation (calculation performed on the **$[\mathrm{P_{D1}P_{D2}Chl_{D1}Pheo_{D1}}]$** tetramer, but only the implicated pigment pair is depicted).

Table 4.2 Excited state properties of the D1 tetramer (**$\mathrm{P_{D1}P_{D2}Chl_{D1}Pheo_{D1}}$**) computed using (TDA)-TDDFT with QM/MM at the ωB97X-D3BJ/def2-TZVP level of theory. The nature of excited states is labelled as local excitons (**LE**) or charge-transfer (**CT**), based on natural transition orbitals (NTOs) for singlet states and canonical molecular orbitals for the triplet states. E_S and E_T represent the singlet and triplet vertical excitation energies (VEE) in eV; f_{osc} are the corresponding oscillator strengths.

Roots	E_S	f_{osc}	**Transition**	E_T	**Transition**
1	1.548	0.00	CT ($\mathrm{P_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)	1.215	LE ($\mathrm{Chl_{D1}}$)
2	1.693	0.06	CT ($\mathrm{Chl_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)	1.291	LE ($\mathrm{P_{D2}}$)
3	1.801	0.32	LE ($\mathrm{Chl_{D1}}$)	1.303	LE ($\mathrm{P_{D1}}$)
4	1.807	0.02	CT ($\mathrm{P_{D2}} \rightarrow \mathrm{Pheo_{D1}}$)	1.386	LE ($\mathrm{Pheo_{D1}}$)
5	1.855	0.39	LE ($\mathrm{P_{D1}}$) + LE ($\mathrm{P_{D2}}$)	1.548	CT ($\mathrm{P_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)
6	1.882	0.05	LE ($\mathrm{P_{D1}}$) + LE ($\mathrm{P_{D2}}$)	1.681	LE ($\mathrm{Pheo_{D1}}$)
7	2.023	0.00	CT ($\mathrm{P_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)	1.708	CT ($\mathrm{Chl_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)
8	2.033	0.17	LE ($\mathrm{Pheo_{D1}}$)	1.731	LE ($\mathrm{Chl_{D1}}$)
9	2.251	0.00	CT ($\mathrm{Chl_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)	1.773	LE ($\mathrm{P_{D2}}$)
10	2.340	0.00	CT ($\mathrm{P_{D2}} \rightarrow \mathrm{Pheo_{D1}}$)	1.778	LE ($\mathrm{P_{D1}}$)
11	2.385	0.04	LE ($\mathrm{Chl_{D1}}$)	1.807	CT ($\mathrm{P_{D2}} \rightarrow \mathrm{Pheo_{D1}}$)
12	2.409	0.03	LE ($\mathrm{P_{D2}}$)	2.023	CT ($\mathrm{P_{D1}} \rightarrow \mathrm{Pheo_{D1}}$)

The results presented and analyzed in terms of natural transition orbital (NTO) compositions (see **Table 4.2** and **Fig. 4.4**) and (TDA)-TDDFT difference densities (**Fig. 4.5**) show that the lowest triplet excited state of the D1 tetramer (T_1 at 1.215 eV) is fully localized on $\mathrm{Chl_{D1}}$, which also exhibits the lowest site energy (S_1 at 1.801 eV) among all RC pigments. The second and third triplet states (T_2 at 1.291 eV and T_3 at 1.303 eV) are localized excitations on $\mathrm{P_{D2}}$ and $\mathrm{P_{D1}}$ respectively. These results are in line with those obtained for the pigment monomers and dimers. Most importantly, we identified the "spin-flipped" triplet states $^3[\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ (1.548 eV) and $^3[\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ (1.708 eV) that are isoenergetic with the lowest singlet CT states (see **Table 4.2**). The corresponding TD-DFT difference densities for the low-energy CT triplet excitations $^3[\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ and $^3[\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ are depicted in **Fig. 4.5**. It is noteworthy that all the RC pigments exhibit a triplet exciton lower than the above donor-acceptor CT states.

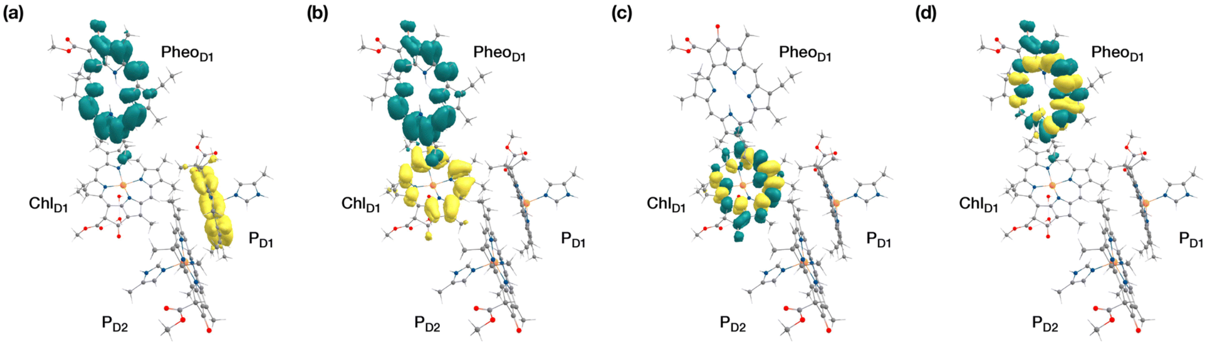


Fig. 4.5 Difference densities describing the lowest singlet-triplet excitations of the D1 branch in PSII: **(a)** the lowest triplet excitation with $^3[\mathbf{P_{D1}}^{\delta+}\mathbf{Pheo_{D1}}^{\delta-}]$ charge transfer character; **(b)** the lowest triplet excitation with $^3[\mathbf{Chl_{D1}}^{\delta+}\mathbf{Pheo_{D1}}^{\delta-}]$ charge transfer character; **(c)** local $^3\mathrm{Chl_{D1}}$ excitation; **(d)** local $^3\mathrm{Pheo_{D1}}$ excitation.

All the low-energy triplet states are dominated by local excitations on $\mathrm{Chl_{D1}}$, $\mathrm{P_{D1}}$, $\mathrm{P_{D2}}$ and $\mathrm{Pheo_{D1}}$, all lower in energy than the lowest triplet CT states. This is in contrast to singlet excitations wherein the low-energy profile is dominated by mixed local excitons and CT excitations or states with pure CT character. Also, most local excitons are blue-shifted compared to the donor–acceptor CT states. Overall, our results clearly demonstrate that low-energy singlet and triplet excited state manifolds differ significantly for primary donor–acceptor pairs in the RC. A detailed schematic representation comparing the complete low-energy spectrum (singlet and triplet excitations) of the RC is summarized in **Fig. 4.6**. Hence, on the basis of our calculations one would expect that the observable triplet state in the RC can be formed from recombination of either of these radical pairs that subsequently decays to the neutral ground-state chlorophyll triplet $^3\mathrm{Chl_{D1}}$. This mechanism is different from the formation of other triplet states (e.g. in light-harvesting antennae) where ^{3}Chl formation is mediated by triplet–triplet energy transfer (T-TET)[54, 55, 70] or direct intersystem crossing from a singlet excited state.[54] The singlet-triplet excitation spectra of the D2 tetramer $[\mathrm{P_{D1}P_{D2}Chl_{D2}Pheo_{D2}}]$

(see **Fig. 4.7, Table 4.10**) are also comprised of CT triplet excitations corresponding to $^3[\mathrm{P_{D2}}^{\delta+}\mathrm{Pheo_{D2}}^{\delta-}]$ (1.706 eV), $^3[\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D2}}^{\delta-}]$ (1.816 eV) and $^3[\mathrm{Chl_{D2}}^{\delta+}\mathrm{Pheo_{D2}}^{\delta-}]$ (2.032 eV) respectively. The lowest triplet exciton in the D2 side is localized on $\mathrm{Chl_{D2}}$ at 1.279 eV.

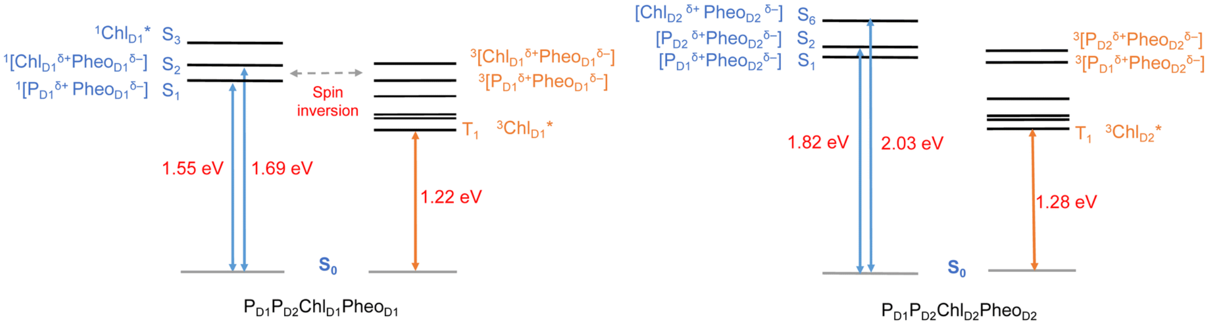


Fig. 4.6 Schematic representation of selected low-energy singlet and triplet excitations for the $\mathrm{P_{D1}P_{D2}Chl_{D1}Pheo_{D1}}$ and $\mathrm{P_{D1}P_{D2}Chl_{D2}Pheo_{D2}}$ tetrameric assemblies computed using (TDA)-TDDFT with QM/MM at the ωB97X-D3BJ/def2-TZVP level of theory.

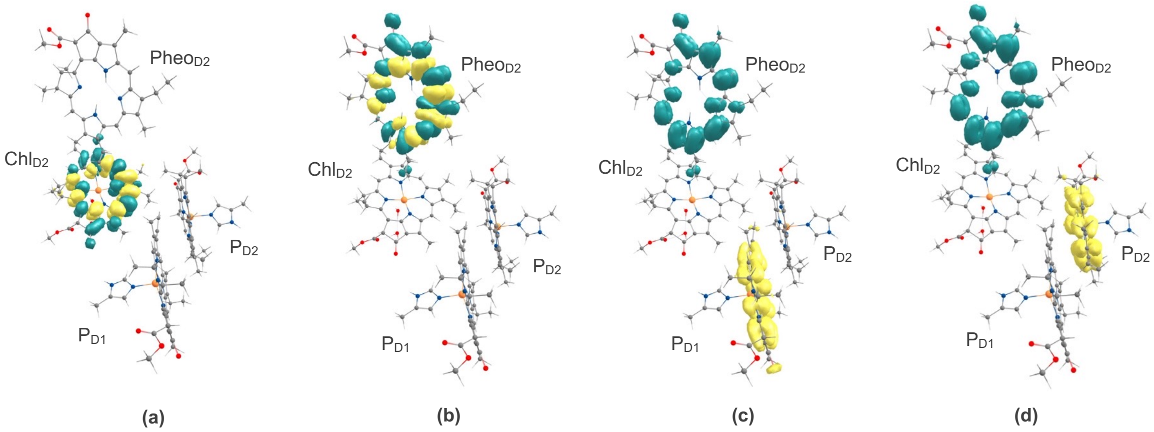


Fig. 4.7 Difference densities describing the lowest singlet-triplet excitations of the D2 branch in PSII: **(a)** local $^3\mathbf{Chl_{D2}}$ excitation; **(b)** local $^3\mathbf{Pheo_{D2}}$ excitation; **(c)** the lowest triplet excitation with $^3[\mathbf{P_{D1}}^{\delta+}\mathbf{Pheo_{D2}}^{\delta-}]$ charge transfer character; **(d)** the lowest triplet excitation with $^3[\mathbf{P_{D2}}^{\delta+}\mathbf{Pheo_{D2}}^{\delta-}]$ charge transfer character.

4.3.3 Relaxed triplet states

In the previous section we explored the influence of the protein matrix on the excitonic asymmetry for singlets and triplets, where the lowest energy excitons were found to be localized on $\mathrm{Chl_{D1}}$. Interestingly, while singlet excitation energy transfer (EET) within the RC seems unlikely due to rapid charge separation, the protein matrix tends to delocalize triplet states over the four chlorophyll pigments. Understanding this phenomenon of triplet delocalization among RC pigments is crucial for comprehending the mechanisms of photo-quenching and photoprotection in PSII.[51, 87] Moreover, obtaining accurate estimates of the triplet energy gaps among individual pigments is necessary to determine the actual rates of photo-quenching. To address this, we conducted further QM/MM geometry optimizations of the individual pigments (Chl_{D1}, Chl_{D2}, P_{D1}, and P_{D2}) in their singlet and triplet states, enabling us to estimate the adiabatic T_1–S_0 energy gaps for each chlorophyll.

Previous site-directed mutagenesis experiments on D1-H198Q, combined with low-temperature optical difference spectroscopy, conducted by Diner et al.,[9] reported shifts in the difference spectra of P_{D1}^{+}/P_{D1} and $Y_Z^{\bullet}/Y_Z$, as well as displacements in the midpoint potential of P_{D1}^{+}/P_{D1}. However, the mutation had no effect on the difference spectra or EPR properties corresponding to $^3P_{680}$. Schlodder et al.[125] performed similar studies on D1-T179H mutants, which involve the ligand H-bonded to the axially bound water of Chl_{D1}, and observed shifts in the Q*y* band and EPR signals upon triplet formation. The T-S absorption spectra of photosynthetic pigments in D_1D_2Cyt*b*559 complexes were also investigated by Renger et al.,[15, 42] and more recent phosphorescence measurements[73, 74] supported the notion that the triplet state is localized on an RC chlorophyll different from the one accommodating the stable positive charge. FT-IR measurements indicated that the triplet is localized on a chlorophyll distinct from the primary cation-stabilizing chlorophyll, based on the vibrational peak of the 13^1-keto C=O keto arising from differences in H-bonding interactions.[20] These experimental observations, combined with the latest experimental and theoretical descriptions of the primary events at the RC of PSII that identify Chl_{D1} as the primary donor, consistently support the idea that the accessory chlorophyll Chl_{D1} serves as the site of the most stable triplet state.

Here, we determined the TD-DFT vertical excitation energies for $^3[P_{D1}P_{D2}]$ and $^3Chl_{D2}$ to be 1.29 eV and 1.28 eV, respectively (see **Table 4.1**). Consequently, the lowest energy triplet excitation was found to be localized on Chl_{D1}, consistent with the above findings. Additionally, we observed that the QM/MM geometry relaxation had a similar effect of approximately 0.3 eV on the triplet state for each chlorophyll in the reaction center (**Table 4.1**). The EPR/ENDOR and FT-IR spectra obtained from temperature-dependent studies estimated energy differences between $^3Chl_{D1}$ and $^3P_{D1}$ of 8–13 meV from isolated RCs and 11 meV from core complexes.[9, 62, 74, 126] Our computational results align with these experimental observations, indicating that the triplet state on Chl_{D1} is also the lowest in energy among all pigments at the reaction center.[15, 42] However, given the close spacing of energy levels, it is expected that at higher temperatures, an equilibrium would exist among the triplet states of P_{D1}, P_{D2}, Chl_{D1}, and Chl_{D2}, resulting in the delocalization of the observable triplet state over more than one chlorophyll molecule. These conclusions are consistent with recent FT-IR studies conducted by Noguchi and co-workers.[87] Therefore, our findings support both the localization of the triplet on the specific chlorophyll center (Chl_{D1}) at low temperatures and the decrease in triplet intensities due to delocalization at ambient temperatures.

4.3.4 EPR parameters of triplet chlorophylls

Magnetic resonance studies coupled with photoexcitation, especially time-resolved EPR spectroscopy, have been widely applied to characterize the triplet states and organic radicals involving photosynthetic pigments.[36, 38, 40, 54, 55, 71, 75, 76, 127] The triplet states involving RC, antenna chlorophylls as well as carotenoids have been characterized using transient and pulse ENDOR spectroscopy,[37, 39, 66, 126, 128, 129] however a number of these studies led to varying observations depending on the type of preparation and conditions used, as in the case of $D_1D_2Cytb_{559}$ particles or PSII core complexes. DFT methods have also been used to quantify EPR properties of photosynthetic pigments but they have excluded so far the effect of protein electrostatics.[127, 128] Therefore, in order to reliable quantitative insights towards the influence of the local protein environment on the site of localization of the triplet states, here we compute for the first time the EPR properties of each RC pigment in their triplet ($S=1$) geometries using the present QM/MM setup.

The accurate determination of zero field splitting (ZFS) parameters D and E are important to characterize the spatial extent and specific location of the triplet-state spin densities. From a methodological perspective, the accuracy of the spin–spin contribution of the D-tensors (Dss) for organic radicals is significantly affected by the spin contamination, and ROKS approaches show better performance than UKS approaches for predicting the correct sign and the ZFS tensor orientation in organic triplets.[117] Based on our calculations (see **Table 4.3**) we observe good agreement despite a small systematic underestimation of the magnitude of the ZFS for the RC triplets, as also reported in the past for isolated Chl *a* triplets.[117] Our calculations nevertheless confirm that the lowest triplet state is localized on a monomeric chlorophyll at the RC, as can be concluded from the corresponding ZFS parameters and comparison with those of isolated Chl *a*. This appears to rule out the possibility that the observed triplet is delocalized at low temperatures. From the first series of EPR studies on chlorophyll triplets in photosynthetic RCs, Rutherford *et al.* [56, 61] and Van Mieghem *et al*. proposed that the observable triplet is localized on a pigment whose ring plane is tilted at an angle of 30° with respect to the membrane plane.[58] Following on the 1.9 Å crystal structure of PSII,[8] this was assumed to be either of the accessory chlorophylls, Chl_{D1} or Chl_{D2}.

Based on our QM/MM model and EPR calculations, we estimated an angle of about 37° between the chlorophyll *z*-axis and the approximate membrane plane, the *z*-axis of the ZFS tensor and the molecular *z*-axis (perpendicular to the porphyrin ring plane) being approximately collinear. However, one still cannot assign the triplet state of the RC to a specific pigment only based on the ZFS parameters.

Table 4.3 Computed and experimental ZFS parameters *D* and *E* and principal values of the *g*-tensors for RC pigments in their triplet states. All values are calculated using gauge independent atomic orbitals (GIAOs).

EPR/ENDOR		***D*** (cm^{-1})	***E/D***	g_{xx}	g_{yy}	g_{zz}	g_{iso}
^{3}P680	Niklas *et al.*, 2022[37]	0.0288	0.15	2.00310	2.00320	2.00220	2.00280
	Pashenko *et al.*, 2003[65]	0.0289	0.15	2.00324	2.00306	2.00231	2.00287
QM/MM		***D*** (cm^{-1})	***E/D***	***g*-shifts** (ppm)			
Pigment	**Method**			Δg_{xx}	Δg_{yy}	Δg_{zz}	g_{iso}
Chl$_{D1}$	TPSSh	0.0189	0.31	1077	1774	-383	822
	B3LYP (10% HF)			572	1331	-378	508
Chl$_{D2}$	TPSSh	0.0194	0.28	1054	1753	-446	787
	B3LYP (10% HF)			527	1292	-427	464
P$_{D1}$	TPSSh	0.0198	0.31	448	1250	-228	490
	B3LYP (10% HF)			524	1386	-288	541
P$_{D2}$	TPSSh	0.0199	0.28	990	1827	-476	780
	B3LYP (10% HF)			483	1425	-621	429

The calculated *g*-values g_{ii} ($i = x, y, z$) are given as *g*-shifts Δg_{ii} in parts per million (ppm) with $\Delta g_{ii} = 10^6 \times (g_i - g_e)$, where $g_e = 2.002319$ is the free electron *g*-value. The isotropic *g*-value is defined as one-third of the sum of the principal *g*-values.

A more sensitive tool that offer insights into the electronic nature of the triplet states is the electron-nuclear hyperfine coupling (HFC) for protons and heavier nuclei bound to or strongly interacting with it. We computed the ^{1}H HFC constants for each of the chlorophyll triplet states explicitly accounting for the protein electrostatics. From our calculations, we can assign the EPR coupling constants to each proton corresponding to the chlorophyll triplet state (**Table 4.4**). It has been argued based on experiments that ^{3}P$_{680}$ is localized on Chl$_{D1}$ or Chl$_{D2}$, based on the low number of contacts of the three methyl groups (2, 7 and 12). We also conclude that the peak corresponding to the highest positive HFC should be assigned to the freely rotating methyl group at position 12, followed by that of 2, and this is consistent for all the RC pigments. Our assignment of the hyperfine coupling constants is also consistent with DFT computed

Mulliken spin populations of the neighboring carbon atoms of the chlorin macrocycle (see **Fig. 4.9**). Overall, C^{12} has the highest spin population (0.293 in Chl$_{D1}$) in the chlorin ring, which consequently leads to a large proton hyperfine coupling in the C^{12} methyl protons. The spin population at C^2 and C^7 are comparatively lower. The assignment of the HFC constants at position 2 is also interesting, because the signal corresponding to these protons is not clearly assigned in ENDOR studies of isolated RC (D_1D_2Cyt*b*$_{559}$) samples.[39] Interestingly, the largest contribution for each chlorophyll is seen to arise for the methyl protons oriented towards the perpendicular z-axis of the molecule. The negative values of the HFCs are assigned to the methine (CH) protons on the plane of the chlorin macrocycle (5, 10 and 20) because their isotropic couplings arise from spin polarization effects. Among these methine (CH) protons the carbon with highest spin density leads to more a negative value of ^{1}H HFC due to a higher spin polarization and this trend is consistent among all the four RC pigments. In the recent work by Niklas and coworkers,[37] the hyperfine coupling constants for the protons at C^{17} and C^{18} were not clearly determined for $^3P_{680}$.

Table 4.4 Experimental and calculated hyperfine coupling constants (HFCs in MHz) for ^{3}P680, other ^{3}Chl a species, and triplet states of the pigments in PSII-RC, computed with the TPSSh functional, the EPR-II basis set on H atoms and the def2-TZVP basis set on remaining atoms.

	Triplet state	**10 (CH)**	**20 (CH)**	**5 (CH)**	**7 (CH_3)**	**12 (CH_3)**	**2 (CH_3)**	**18 (CH)**	**17 (CH)**	**3' (CH)**	**3'' (CH_2)**
ENDOR [37, 39, 127]	$^3P_{680}$	-10.03	-7.88	-4.79	0.62	10.35	4.80	2.99	n.d	0.91	-1.30
	^{3}Chl *a* (WSCP)	-10.20	-7.70	-5.70	1.10	10.70	4.70	2.60	n.d.		
	^{3}Chl *a* (MTHF)	-11.44	-7.20	-6.20	n.d.	7.40	n.d.				
DFT	^{3}Chl *a* (gas-phase)	-5.12	-5.21	-3.32	0.97	10.77	5.61	4.81	3.96	0.69	-1.64
	^{3}Chl *a* (MTHF)	-7.20	-7.32	-4.96	0.39	10.61	5.69	3.14	2.46	0.16	-1.65
	^{3}Chl$_{D1}$ (gas-phase)	-6.63	-6.77	-5.61	0.61	10.95	5.35	2.90	1.78	0.39	-2.04
QM/MM	^{3}Chl$_{D1}$	-6.98	-6.18	-5.64	1.25	12.27	5.68	2.59	1.28	0.52	-2.80
	^{3}Chl$_{D2}$	-7.41	-6.42	-5.63	0.71	12.41	5.31	3.06	1.58	0.64	-2.19
	$^3P_{D1}$	-6.02	-5.17	-5.59	1.05	10.86	4.78	2.47	1.19	0.58	-2.42
	$^3P_{D2}$	-1.82	-4.13	-1.08	1.07	11.67	5.93	4.18	4.69	3.59	-1.76

From our calculations, we observe that for all the chlorophylls the proton at position 18 has a higher isotropic ^{1}H HFC than position 17. Also, the corresponding spin population analysis of the macrocyclic carbon atoms indicate a higher spin density at C^{19} than C^{16}. This trend is also consistent among all the RC chlorophylls ($\mathrm{Chl}_{\mathrm{D1}}$, $\mathrm{Chl}_{\mathrm{D2}}$, P_{D1} and P_{D2}), and therefore our QM/MM calculations indicate the experimentally observed HFC of 2.99 in $^{3}\mathrm{P}_{680}$ likely arises from position 18. We have also identified contributions from the vinyl group (3′, 3″), the peaks of which were not clearly assigned in previous spectroscopic observations. The negative HFC at 3″ is likely due to spin polarization from $C^{3''}$, and the magnitude is consistent with the corresponding spin populations. However, the orientation of the vinyl group of P_{D2} is particularly noteworthy here. It is known that in P_{D2} the vinyl CH_2 is slightly out of plane from the chlorin macrocycle, and our results indicate that this significantly affects the spin density distribution of the vinyl carbons. This clearly explains why the ^{1}H HFC of the vinyl protons in P_{D2} differ significantly from the other RC chlorophylls. Our QM/MM methodology therefore not only reproduces the experimental EPR/ENDOR results obtained from intact PSII core samples but also accounts for local perturbations that might affect EPR signals from isolated RC samples. Overall, the triplet spin distribution of individual chlorophylls (**Fig. 4.8**) remain unchanged for isolated RC samples.[87]

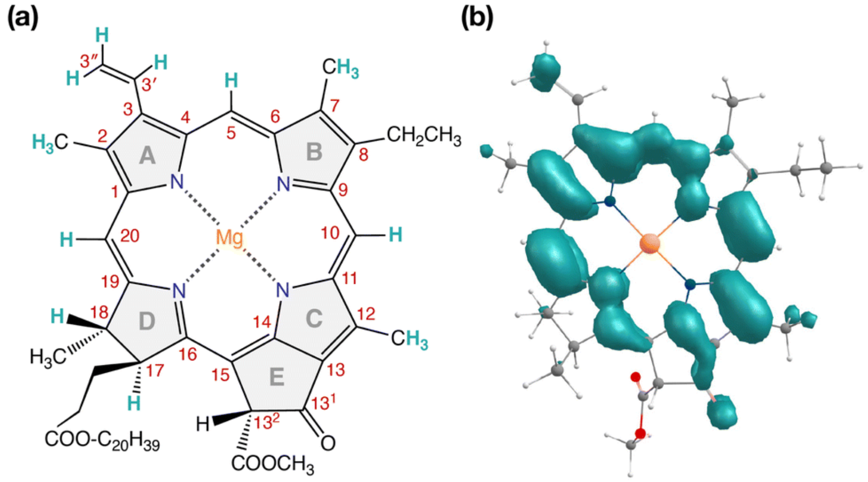


Fig. 4.8 (a) Structure of Chl a with carbon atom numbering and spectroscopically important hydrogen positions indicated. **(b)** Computed spin density distribution of triplet ($S = 1$) Chl a.

The EPR parameters however, are not sufficiently sensitive to the protein environment to enable confident differentiation between the chlorophylls of the RC and it is not possible to assign the spectroscopically observable triplet state to a single RC chlorophyll based on EPR parameters alone. Nevertheless, the lowest triplet excitations and the energetically most stable triplet state are found on $\mathrm{Chl}_{\mathrm{D1}}$ and, hence, the combined results of all our calculations show a clear preference to assign this state to a triplet state localized on the accessory chlorophyll $\mathrm{Chl}_{\mathrm{D1}}$.

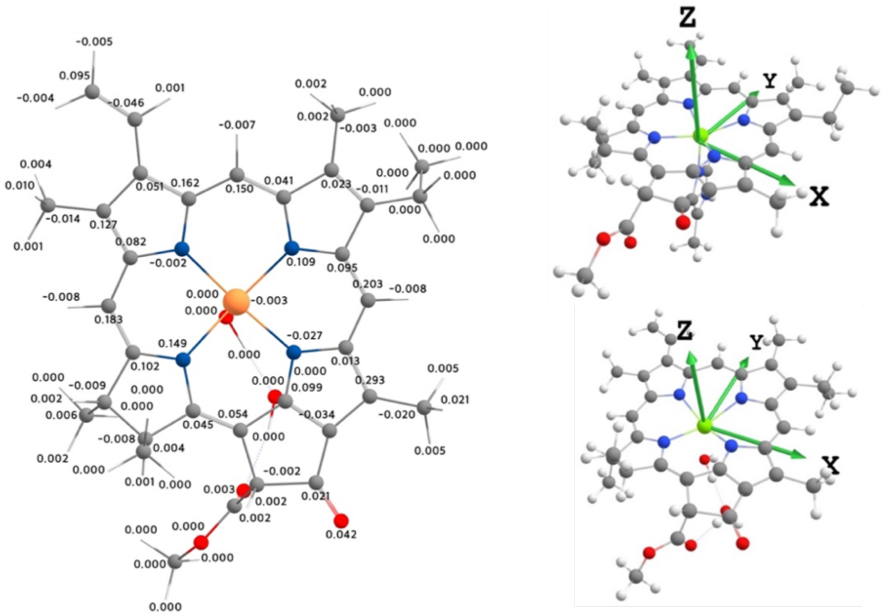


Fig. 4.9 (a) Mulliken spin population analysis of **Chl_{D1}** in protein calculated with DFT-TPSSh and QM/MM with EPR-II basis set on H-atoms and def2-TZVP on the remaining atoms. **(b)** The orientation of the ZFS D-tensor in **P_{D1}** and **Chl_{D1}** RC pigments, computed with DFT-TPSSh and QM/MM with EPR-II basis set on H-atoms and def2-TZVP on the remaining atoms.

4.3.5 Electrostatic effects by plastoquinone Q_A and the OEC

Until now, we discussed the optical properties of primary electron transfer processes in PSII, where the OEC is in its dark-stable state (S_1) and Q_A is a neutral plastoquinone. Under normal conditions, the charge recombination of the primary charge separated states $[Chl_{D1}^{+}Pheo_{D1}^{-}]$[4, 11, 16] (or $[P_{D1}^{+}Pheo_{D1}^{-}]$)[24] is prevented by forward electron transfer from $Pheo_{D1}$ to Q_A (within a few hundred ps). The oxidation of Y_Z (the redox-active tyrosine residue that interfaces the OEC with the RC) by P_{D1}^{+} occurs instead within 25 ns–50 µs.[5] Both processes contribute to formation and adjustment of an electrostatic gradient across the transmembrane region, which, coupled with intrinsic protein matrix effects, tunes the thermodynamics and kinetics of electron transfer pathways.

Based on reported timescales, the oxidation of OEC by the Y_Z (50 µs–4 ms), and electron transfer from Q_A to Q_B (0.2–0.8 ms) are the two main rate-limiting steps in PSII. These electron transfer processes thus eventually create the next stable intermediate of the RC with an oxidized OEC and reduced Q_A ($S_2Q_A^{-}$). On the other hand, in extreme conditions such as prolonged light exposure the plastoquinone (PQ) pool in thylakoid membranes can remain reduced, abolishing electron transfer from Q_A to Q_B and allowing Q_A^{-} to accumulate. This can further drive competing secondary electron transfer pathways leading to triplet formation in the RC.

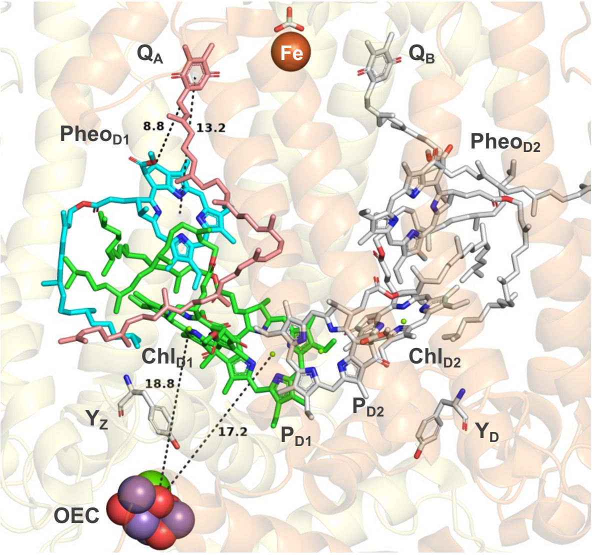
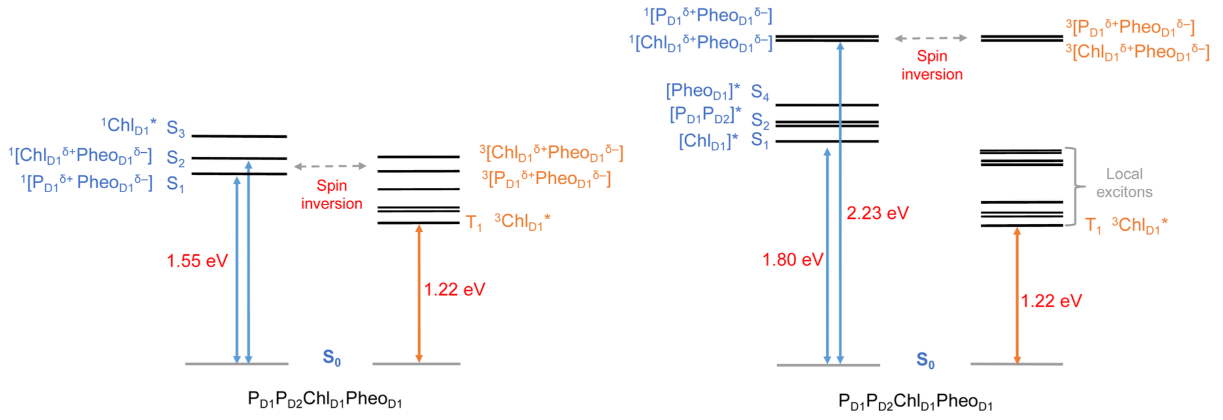


Fig. 4.10 Depiction of distances (in Å) of the redox active cofactors **Q_A** and **OEC** from the P_{D1}, Chl_{D1}, and $Pheo_{D1}$ pigments in the PSII-RC. Distances are measured from the center of each chlorin ring, unless otherwise specified.

In view of the above, as a next step we performed TD-DFT calculations on the "closed" RC, where the OEC is modelled in the S_2 state of the Kok–Joliot cycle and Q_A is reduced, i.e., the $S_2Q_A^-$ state. Our excited state calculations on the [$P_{D1}P_{D2}Chl_{D1}Pheo_{D1}$] assembly (**Table 4.11**) reveal interesting results. The low-energy spectrum (see **Fig. 4.11**) in the presence of the semiquinone Q_A^- is dominated by local excitations both for singlets and triplets, in stark contrast to the case when Q_A is neutral and available to accept electrons. The relative stability of site energies (Chl_{D1}, P_{D1}, P_{D2} and $Pheo_{D1}$) also explains the longer lifetime of chlorophyll excited states and high fluorescence yields observed in closed RCs.[26, 45] Moreover, the $^1[Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}]$ CT state is 2.23 eV higher than the ground state and thus considerably blue-shifted compared to the open RC (1.69 eV). This is in line with previous experimental hypotheses regarding reduced charge separation due to the electrostatic repulsion of Q_A^-.[26, 32, 79] Interestingly, we also find that the two low-energy CT states $^1[Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}]$ (2.231 eV) and $^1[P_{D1}^+Pheo_{D1}^-]$ (2.276 eV) are almost isoenergetic for the closed RC (**Table 4.11**). This is clearly an effect of a differential influence of oxidized OEC and Q_A^- on the primary donor-acceptor pairs with $Pheo_{D1}^-$ and P_{D1}^+ being more destabilized than Chl_{D1}^+ due to their spatial proximity to Q_A^- and/or the oxidized OEC respectively (**Fig. 4.10**; $Pheo_{D1}$ is the closest pigment to Q_A with an edge-to-edge distance of 8.8 Å, and a center-to-center distance of 13.2 Å, while P_{D1} is closest to OEC with a distance of about 17.2 Å).



(a) Open RC (S_1Q_A)

(b) Closed RC ($S_2Q_A^-$)

Fig. 4.11 Schematic representation of the singlet and triplet vertical excitation energies of the $P_{D1}P_{D2}Chl_{D1}Pheo_{D1}$ tetramer in: **(a)** the open RC **(S_1Q_A)**, and **(b)** the closed RC **(S_2Q_A–)**. All energies are reported relative to the ground state singlet (S_0).

Studies on charge recombination reactions have shown the formation of RC triplet states can be influenced not only by the presence of the semiquinone (Q_A^-) but also by the complete absence of Q_A (e.g., isolated D_1D_2Cytb_{559} samples) or the double reduction of Q_A.[45, 46] In some experiments conducted at cryogenic temperatures the spin-polarized triplet state was only detected when Q_A was doubly reduced (Q_A^{2-} or Q_AH_2) and not when it was singly reduced, which led to controversies about whether or not primary charge separation can occur in the presence of Q_A^-. Van Mieghem *et al.* monitored the light-induced triplet signals with different redox states of Q_A using EPR, and reported higher triplet yields but shorter life times ($t_{1/2} <$ 20μs) with Q_A^- (closed RC).[29, 32, 33] On the other hand, Feikema *et al.* based on time-resolved EPR measurements on PSII core samples reported that the yield of the triplet state with a singly reduced Q_A^- did not differ significantly from those with Q_AH_2.[29] In the case of Q_AH_2 however, the chlorophyll triplet was reported to have a much extended lifetime ($t_{1/2}$ ~1–2 ms) and this has been attributed to the absence of Q_A^- to quench chlorophyll triplet states in PSII. Moreover, flash-induced PSII activity measurements showed the extent of D1-photodamage due to $^1O_2$ to be most pronounced in the S_2 and S_3 states of the OEC, and this also has been correlated to other competing back reactions.[47, 50] Hence, the pathway of triplet formation and how the singlet–triplet excitations depend on the redox state of the Q_A and OEC remain crucial to understand both the control of primary processes by the transmembrane electrostatic gradient and the photoprotection mechanisms of PSII.

As seen from the excitation energy profiles (**Fig. 4.11**), the energetics of the singlet and triplet charge transfer (CT) excitations can be directly influenced by the redox state of surrounding cofactors, particularly Q_A. A more comprehensive overview of the singlet and triplet excitation energies, charge transfer pathways, charge recombination and triplet forming routes, are depicted in **Fig. 4.12**. Based on our results, it can be suggested that formation of triplet states at the RC should be preceded by charge recombination of the primary radical pair $[\mathrm{Chl_{D1}}^{+}\mathrm{Pheo_{D1}}^{-}]$ or $[\mathrm{P_{D1}}^{+}\mathrm{Pheo_{D1}}^{-}]$ formed from the corresponding CT states. Subsequently, a very important aspect when discussing molecular mechanisms of photoprotection involves the acceptor side of PSII. $\mathrm{Pheo_{D1}}$ is the site of the primary anion radical $\mathrm{Pheo_{D1}}^{-}$, following charge separation.[4, 11, 12, 15, 79] In normal physiological conditions the electron is rapidly transferred to Q_A ($\mathrm{Pheo_{D1}}^{-}Q_A \rightarrow \mathrm{Pheo_{D1}}Q_A^{-}$). The thermodynamic driving force for this step is governed by the relative midpoint potentials of $\mathrm{Pheo_{D1}}/\mathrm{Pheo_{D1}}^{-}$and Q_A/Q_A^{-} and is controlled by local pigment–protein interactions.[32] However, the reduction of Q_A to Q_A^{-} can lead to the following alternate possibilities: (a) direct charge recombination with P_{680}^{+} to ${}^{1}P_{680}^{*}$ and finally the ground state, (b) backward electron transfer onto $\mathrm{Pheo_{D1}}$ to form ${}^{1}[P_{680}^{+}\mathrm{Pheo_{D1}}^{-}]$ or (c) formation of the charge recombination triplet ${}^{3}[P_{680}^{+}\mathrm{Pheo_{D1}}^{-}]$ which finally localizes on $\mathrm{Chl_{D1}}$ i.e., the triplet route. Calculation of the $\mathrm{Pheo_{D1}}$ electron affinity suggests that $\mathrm{Pheo_{D1}}^{-}$ formation is disfavored by ca. 0.5-1 eV in the presence of a reduced Q_A^{-}. The electrostatic repulsion of Q_A^{-} destabilizes the primary radical pair $[P_{680}^{+}\mathrm{Pheo_{D1}}^{-}]$, but also inhibits forward electron transfer. This might cause spin inversion from ${}^{1}[P_{680}^{+}\mathrm{Pheo_{D1}}^{-}]$ to ${}^{3}[P_{680}^{+}\mathrm{Pheo_{D1}}^{-}]$, the excess excitation energy dissipated through the non-radiative triplet route (**Fig. 4.12**). Experiments suggest that the observable triplet in the closed RC has an extremely short lifetime ($t_{1/2} < 20\mu s$), and it has been proposed that this is because Q_A^{-} quenches RC triplet states through ${}^{3}\mathrm{Pheo_{D1}}$.[32] However, this mechanism of triplet quenching involving the semiquinone (Q_A^{-}) and ${}^{3}\mathrm{Pheo_{D1}}$ is not well understood. Based on our computed excitation profile of the closed RC (**Fig. 4.11**), we find numerous thermodynamically accessible triplet states that are localized on the individual pigments ($\mathrm{Chl_{D1}}$, $\mathrm{P_{D1}}$, $\mathrm{P_{D2}}$ and $\mathrm{Pheo_{D1}}$). All these local excitations are in fact lower in energy than the CT ${}^{3}[\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}]$ and ${}^{3}[\mathrm{P_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta+}]$ excitations, which is in contrast to the triplet energy profile of open RC $[S_1Q_A]$ (see **Fig. 4.11**).

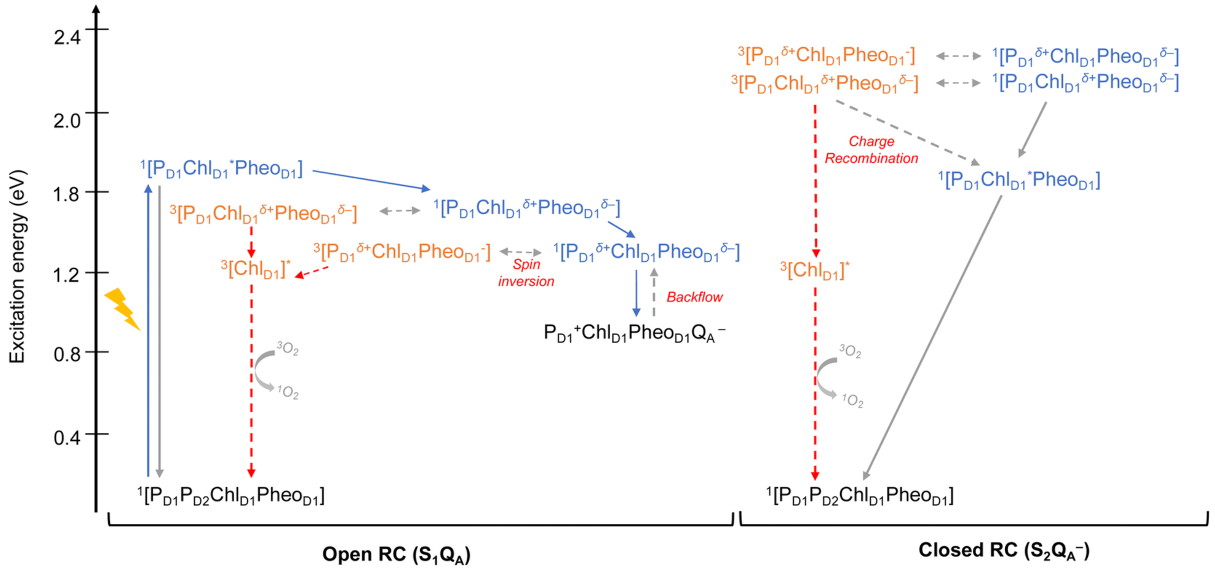


Fig. 4.12 Schematic representation of singlet and triplet excitation energies, charge transfer pathways, charge recombination and triplet forming routes in the PSII-RC (a) OEC in S1, Q_A is neutral (b) OEC in S2, Q_A is singly-reduced. All energies are computed in eV relative to the GS singlet (S_0). Singlet excited states are shown in blue, triplet states in orange, forward electron transfer with blue arrows, singlet/radiative decay routes with grey solid arrows, backflow/spin inversion/charge recombination with grey dotted arrows, triplet routes in red.

Specifically, all D1 pigments in the closed RC possess at least two triplet excitations (T_1 to T_8) energetically lower than the first CT state. Thus, non-radiative energy dissipation involving multiple RC pigments might be a possibility in the closed RC, in line with arguments regarding triplet delocalization pathways discussed in recent FT-IR studies.[87] When Q_A is doubly reduced as Q_AH_2, the $Pheo_{D1}{}^-$ anion is expected to be more stable in the absence of a negative charge in its vicinity.[25, 45, 46, 130] This can stabilize $^1[P_{680}{}^+Pheo_{D1}{}^-]$ and a subsequent spin inversion to $^3[P_{680}{}^+Pheo_{D1}{}^-]$ may again lead to more centers favoring the triplet route as opposed to a direct charge recombination to the singlet state.

It is known that formation of triplet states is detrimental to photosynthetic organisms as long-lived triplets in the RC can accelerate the formation of reactive oxygen species and subsequent photodamage.[43, 49, 51] In this respect, we provided a quantitative explanation of how the PSII protein matrix and redox active cofactors may work in tandem to tune the energetics of primary charge separation and triplet formation in photosynthetic reaction centers. Our results have implications for photoprotection mechanisms in both the open and the closed states of active PSII. The next line of photoprotection in the RC may involve the delocalization of triplet states away from Chl_{D1} onto other pigments at ambient temperatures to avoid the selective damage of the D1 protein. However, if this still leads to photoinactivation, the D1 protein is selectively degraded and regenerated, thereby allowing photosynthetic organisms to retain functionality even under extreme conditions.[48, 131, 132]

4.4 Conclusion

This work provides a detailed overview of the low-energy excitation spectrum of the PSII-RC, explaining the asymmetry of singlet-triplet excitations and charge transfer states along the D1 and D2 branches. The PSII protein environment explicitly controls the excitonic asymmetry of the RC, leading to low-energy charge-transfer excitations and triplet formation on the D1 side. Based on our calculations we may speculate that the observable triplet state at the RC seems unlikely to be formed by usual inter-system crossing (ISC), but through a radical-pair charge recombination preceding ISC. However, we stress that this cannot yet be concluded with confidence given that the kinetics of triplet formation and the factors/processes that determine them are not addressed in the present work. We identified all low-lying charge transfer excitations in the pigment assemblies, and find that the lowest triplet exciton is localized on $\mathrm{Chl_{D1}}$. With supporting insights from EPR calculations we confirm the localization of the most stable triplet state on $\mathrm{Chl_{D1}}$. Furthermore, we present the first theoretical description of the excited state properties of the “closed RC” ($\mathrm{S_2Q_A^-}$), explicitly quantifying the electrostatic effect of semiquinone ($\mathrm{Q_A^-}$) and how it influences excited state properties of RC pigments. Overall, this work provides a refined basis for the electronic-level understanding of primary and secondary electron transfer pathways, offering refined electronic structure information as a foundation for discussing possible photoprotection mechanisms in oxygenic photosynthesis.

4.5 Supplementary Tables

The singlet excitation energies are computed using full TDDFT without employing the Tamm–Dancoff approximation (TDA). Vertical triplet excitation energies were computed on the pair-optimized ground state singlet geometries, using the TDA approximation on TD-DFT.

Table 4.5 Excitation energies (ωB97X-D3BJ/def2-TZVP): **Chl_{D1} (gas-phase)**

Roots	**E_S (eV)**	**f_{osc}**	**Transition**	**E_T (eV)**	**Transition**
1	1.884	0.24	HOMO → LUMO (0.81) HOMO-1→LUMO+1 (0.13)	1.290	HOMO → LUMO (0.90) HOMO-1→ LUMO+1 (0.02)
2	2.387	0.04	HOMO-1 → LUMO (0.69) HOMO → LUMO+1 (0.26)	1.763	HOMO-1 → LUMO (0.89)
3	3.382	0.73	HOMO → LUMO+1 (0.59) HOMO-1 → LUMO (0.19)	2.212	HOMO → LUMO+1 (0.88)

Table 4.6 Excitation energies (ωB97X-D3BJ/def2-TZVP) and QM/MM: **Chl_{D1} (in protein)**

Roots	**E_S (eV)**	**f_{osc}**	**Transition**	**E_T (eV)**	**Transition**
1	1.818	0.29	HOMO → LUMO (0.84) HOMO-1→ LUMO+1 (0.11)	1.220	HOMO → LUMO (0.92)
2	2.376	0.05	HOMO-1 → LUMO (0.72) HOMO → LUMO+1 (0.23)	1.727	HOMO-1 → LUMO (0.90)
3	3.425	0.82	HOMO → LUMO+1 (0.70) HOMO-1 → LUMO (0.21)	2.280	HOMO → LUMO+1 (0.89)

Table 4.7 Excitation energies (ωB97X-D3BJ/def2-TZVP): **Chl_{D2} (gas-phase)**

Roots	**E_S (eV)**	**f_{osc}**	**Transition**	**E_T (eV)**	**Transition**
1	1.900	0.24	HOMO → LUMO (0.79) HOMO-1→ LUMO+1 (0.13)	1.319	HOMO → LUMO (0.88) HOMO-1→ LUMO+1 (0.03)
2	2.395	0.04	HOMO-1 → LUMO (0.66) HOMO → LUMO+1 (0.25)	1.745	HOMO-1 → LUMO (0.88)
3	3.388	0.73	HOMO → LUMO+1 (0.62) HOMO-1 → LUMO (0.20)	2.224	HOMO → LUMO+1 (0.87)

Table 4.8 Excitation energies (ωB97X-D3BJ/def2-TZVP) and QM/MM: **Chl_{D2} (in protein)**

Root	**E_S (eV)**	**f_{osc}**	**Transition**	**E_T (eV)**	**Transition**
1	1.878	0.25	HOMO → LUMO (0.81) HOMO-1→ LUMO+1 (0.12)	1.288	HOMO → LUMO (0.89)
2	2.406	0.05	HOMO-1 → LUMO (0.68) HOMO → LUMO+1 (0.24)	1.760	HOMO-1 → LUMO (0.88)
3	3.419	0.80	HOMO → LUMO+1 (0.65) HOMO-1 → LUMO (0.21)	2.250	HOMO → LUMO+1 (0.88)

Table 4.9 Singlet and triplet excitation energies (ωB97X-D3BJ/def2-TZVP): **P_{D1}–P_{D2}** pair (in protein). LE indicates local excitations. Arrows indicate charge transfer (CT) excitations.

Roots	E_S (eV)	f_{osc}	Transition	E_T (eV)	Transition
1	1.859	0.41	LE (P_{D1}) + LE (P_{D2})	1.291	LE (P_{D2})
2	1.885	0.07	LE (P_{D1}) + LE (P_{D2})	1.305	LE (P_{D1})
3	2.416	0.02	LE (P_{D1}) + LE (P_{D2})	1.773	LE (P_{D1})
4	2.427	0.08	LE (P_{D1}) + LE (P_{D2})	1.781	LE (P_{D2})
5	3.010	0.06	CT (P_{D1} → P_{D2})	2.270	LE (P_{D2})
6	3.188	0.33	CT (P_{D2} → P_{D1}) + LE (P_{D2})	2.278	LE (P_{D1})

Table 4.10 Singlet and triplet excitation energies (ωB97X-D3BJ/def2-TZVP): **P_{D1}–P_{D2}–Chl_{D2}–Pheo_{D2}** tetramer (open RC, S1QA).

Roots	E_S (eV)	f_{osc}	Transition	E_T (eV)	Transition
1	1.706	0.00	CT (P_{D1} → Pheo_{D2})	1.279	LE (Chl_{D2})
2	1.816	0.00	CT (P_{D2} → Pheo_{D2})	1.287	LE (P_{D2})
3	1.841	0.63	LE (Chl_{D2}) + LE (P_{D1})	1.300	LE (P_{D1})
4	1.854	0.10	LE (Chl_{D2}) + LE (P_{D2})	1.390	LE (Pheo_{D2})
5	1.885	0.06	LE (P_{D1}) + LE (P_{D2})	1.674	LE (Pheo_{D2})
6	2.032	0.01	CT (Chl_{D2} → Pheo_{D2})	1.706	CT (P_{D1} → Pheo_{D2})
7	2.047	0.16	LE (Pheo_{D2})	1.759	LE (P_{D1})
8	2.219	0.00	P_{D1} → Pheo_{D2} + Chl_{D2} → Pheo_{D2}	1.763	LE (Chl_{D2})
9	2.297	0.00	CT (P_{D2} →Pheo_{D2})	1.779	LE (P_{D2})
10	2.407	0.01	LE (Chl_{D2})	1.816	CT (P_{D2} → Pheo_{D2})
11	2.411	0.03	LE (P_{D2})	2.032	CT (Chl_{D2} → Pheo_{D2})
12	2.418	0.14	LE (P_{D2})	2.219	CT (P_{D1} → Pheo_{D2})

Table 4.11 Singlet and triplet excitation energies (ωB97X-D3BJ/def2-TZVP): **P_{D1}–P_{D2}–Chl_{D1}–Pheo_{D1}** tetramer **(closed RC, $S_2Q_A^-$)**

Roots	E_S (eV)	f_{osc}	Transition	E_T (eV)	Transition
1	1.796	0.40	LE (Chl_{D1})	1.223	LE (Chl_{D1})
2	1.862	0.37	LE (P_{D1}) + LE (P_{D2})	1.296	LE (P_{D2})
3	1.888	0.05	LE (P_{D1}) + LE (P_{D2})	1.306	LE (P_{D1})
4	1.990	0.18	LE (Pheo_{D1})	1.394	LE (Pheo_{D1})
5	2.231	0.00	CT (Chl_{D1} → Pheo_{D1})	1.670	LE (Pheo_{D1})
6	2.276	0.00	CT (P_{D1} → Pheo_{D1})	1.729	LE (Chl_{D1})
7	2.376	0.04	LE (Chl_{D1})	1.769	LE (P_{D1})
8	2.405	0.02	LE (P_{D1}) + LE (P_{D2})	1.770	LE (P_{D2})
9	2.421	0.21	LE (P_{D1}) + LE (P_{D2})	2.227	CT (Chl_{D1} →Pheo_{D1})
10	2.425	0.03	LE (Pheo_{D1})	2.256	LE (P_{D2})
11	2.436	0.00	CT (P_{D2} → Pheo_{D1}) (0.99)	2.276	CT (P_{D1} → Pheo_{D1})
12	2.732	0.00	CT (P_{D1} → Pheo_{D1}) (0.98)	2.277	CT (P_{D1} → Pheo_{D2}) + LE (P_{D1})

4.6 REFERENCES

1. R. E. Blankenship, *Molecular mechanisms of photosynthesis*, John Wiley & Sons, 2021.
2. J.-R. Shen, *Annu. Rev. Plant Biol.*, 2015, **66**, 23-48.
3. N. Cox, D. A. Pantazis and W. Lubitz, *Annu. Rev. Biochem.*, 2020, **89**, 795-820.
4. T. Cardona, A. Sedoud, N. Cox and A. W. Rutherford, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 26-43.
5. F. Muh, C. Glockner, J. Hellmich and A. Zouni, *Biochim. Biophys. Acta*, 2012, **1817**, 44-65.
6. Y. Takahashi, Ö. Hansson, P. Mathis and K. Satoh, *Biochim. Biophys. Acta Bioenerg.*, 1987, **893**, 49-59.
7. D. J. Vinyard, G. M. Ananyev and G. Charles Dismukes, *Annu. Rev. Biochem.*, 2013, **82**, 577-606.
8. Y. Umena, K. Kawakami, J.-R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
9. B. A. Diner, E. Schlodder, P. J. Nixon, W. J. Coleman, F. Rappaport, J. Lavergne, . . . D. A. Chisholm, *Biochemistry*, 2001, **40**, 9265-9281.
10. K. Saito, T. Ishida, M. Sugiura, K. Kawakami, Y. Umena, N. Kamiya, . . . H. Ishikita, *J. Am. Chem. Soc.*, 2011, **133**, 14379-14388.
11. Y. Yoneda, E. A. Arsenault, S. Yang, Jr., K. Orcutt, M. Iwai and G. R. Fleming, *Nat. Commun.*, 2022, **13**, 2275.
12. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Am. Chem. Soc.*, 2020, **142**, 18174-18190.
13. V. A. Nadtochenko, I. V. Shelaev, M. D. Mamedov, A. Y. Shkuropatov, A. Y. Semenov and V. A. Shuvalov, *Biochemistry (Moscow)*, 2014, **79**, 197-204.
14. W. Lubitz, F. Lendzian and R. Bittl, *Acc. Chem. Res.*, 2002, **35**, 313-320.
15. T. Renger and E. Schlodder, *ChemPhysChem*, 2010, **11**, 1141-1153.
16. E. Romero, I. H. M. Van Stokkum, V. I. Novoderezhkin, J. P. Dekker and R. Van Grondelle, *Biochemistry*, 2010, **49**, 4300-4307.
17. N. P. Pawlowicz, M. L. Groot, I. H. M. Van Stokkum, J. Breton and R. Van Grondelle, *Biophys. J.*, 2007, **93**, 2732-2742.
18. M. Capone, A. Sirohiwal, M. Aschi, D. A. Pantazis and I. Daidone, *Angew. Chem. Int. Ed.*, 2023, **62**, e202216276.
19. P. J. O'Malley, *J. Phys. Chem. B*, 2000, **104**, 2176-2182.
20. T. Noguchi, T. Tomo and Y. Inoue, *Biochemistry*, 1998, **37**, 13614-13625.
21. H. Tamura, K. Saito and H. Ishikita, *Chem. Sci.*, 2021, **12**, 8131-8140.
22. R. Takahashi, K. Hasegawa and T. Noguchi, *Biochemistry*, 2008, **47**, 6289-6291.
23. H. Tamura, K. Saito and H. Ishikita, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 16373-16382.
24. A. Sirohiwal and D. A. Pantazis, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200356.
25. I. Vass, G. Gatzen and A. R. Holzwarth, *Biochim. Biophys. Acta Bioenerg.*, 1993, **1183**, 388-396.
26. M. Szczepaniak, J. Sander, M. Nowaczyk, M. G. Müller, M. Rögner and A. R. Holzwarth, *Biophys. J.*, 2009, **96**, 621-631.
27. V. Martínez-Junza, M. Szczepaniak, S. E. Braslavsky, J. Sander, M. Nowaczyk, M. Rögner and A. R. Holzwarth, *Photochem. Photobiol. Sci.*, 2008, **7**, 1337.
28. G. Sipka, M. Magyar, A. Mezzetti, P. Akhtar, Q. Zhu, Y. Xiao, . . . G. Garab, *The Plant Cell*, 2021, **33**, 1286-1302.
29. W. O. Feikema, P. Gast, I. B. Klenina and I. I. Proskuryakov, *Biochim. Biophys. Acta Bioenerg.*, 2005, **1709**, 105-112.

30. Y. Takegawa, M. Nakamura, S. Nakamura, T. Noguchi, J. Selles, A. W. Rutherford, . . . M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2019, **1860**, 297-309.
31. S. Katagiri and Y. Kobori, *Applied Magnetic Resonance*, 2010, **37**, 177-189.
32. F. Van Mieghem, K. Brettel, B. Hillman, A. Kamlowski, A. W. Rutherford and E. Schlodder, *Biochemistry*, 1995, **34**, 4798-4813.
33. B. Hillmann, K. Brettel, F. Van Mieghem, A. Kamlowski, A. W. Rutherford and E. Schlodder, *Biochemistry*, 1995, **34**, 4814-4827.
34. D. Kleinfeld, M. Y. Okamura and G. Feher, *Biochim. Biophys. Acta Bioenerg.*, 1984, **766**, 126-140.
35. R. de Wijn and H. J. van Gorkom, *Biochim. Biophys. Acta Bioenerg.*, 2002, **1553**, 302-308.
36. D. Carbonera, M. Di Valentin, G. Giacometti and G. Agostini, *Biochim. Biophys. Acta Bioenerg.*, 1994, **1185**, 167-176.
37. J. Niklas, A. Agostini, D. Carbonera, M. Di Valentin and W. Lubitz, *Photosynth. Res.*, 2022, **152**, 213-234.
38. S. Santabarbara, G. Agostini, A. P. Casazza, C. D. Syme, P. Heathcote, F. Böhles, . . . D. Carbonera, *Biochim. Biophys. Acta Bioenerg.*, 2007, **1767**, 88-105.
39. F. Lendzian, R. Bittl, A. Telfer and W. Lubitz, *Biochim. Biophys. Acta Bioenerg.*, 2003, **1605**, 35-46.
40. S. Santabarbara, E. Bordignon, R. C. Jennings and D. Carbonera, *Biochemistry*, 2002, **41**, 8184-8194.
41. M. Volk, M. Gilbert, G. Rousseau, M. Richter, A. Ogrodnik and M.-E. Michel-Beyerle, *FEBS Lett.*, 1993, **336**, 357-362.
42. G. Raszewski, W. Saenger and T. Renger, *Biophys. J.*, 2005, **88**, 986-998.
43. A. W. Rutherford, A. Osyczka and F. Rappaport, *FEBS Lett.*, 2012, **586**, 603-616.
44. A. R. McIntosh and J. R. Bolton, *Nature*, 1976, **263**, 443-445.
45. I. Vass and S. Styring, *Biochemistry*, 1992, **31**, 5957-5963.
46. I. Vass, S. Styring, T. Hundal, A. Koivuniemi, E. Aro and B. Andersson, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 1408-1412.
47. H. Mattila, S. Mishra, T. Tyystjärvi and E. Tyystjärvi, *New Phytol.*, 2023, **237**, 113-125.
48. E.-M. Aro, I. Virgin and B. Andersson, *Biochim. Biophys. Acta Bioenerg.*, 1993, **1143**, 113-134.
49. I. Vass, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 209-217.
50. I. Vass, *Physiol. Plant.*, 2011, **142**, 6-16.
51. T. Noguchi, *Plant Cell Physiol.*, 2002, **43**, 1112-1116.
52. E. Schlodder, B. Hillmann, K. Brettel and F. Mallwitz, in *Photosynthesis: Mechanisms and Effects: Volume I–V: Proceedings of the XIth International Congress on Photosynthesis, Budapest, Hungary, August 17–22, 1998*, ed. G. Garab, Springer Netherlands, Dordrecht, 1998, pp. 1069-1072.
53. J. Feng, C.-W. Tseng, T. Chen, X. Leng, H. Yin, Y.-C. Cheng, . . . Y. Ma, *Nat. Commun.*, 2017, **8**, 71.
54. S. Santabarbara, A. Agostini, A. P. Casazza, G. Zucchelli and D. Carbonera, *Biochim. Biophys. Acta Bioenerg.*, 2015, **1847**, 262-275.
55. S. Santabarbara and D. Carbonera, *J. Phys. Chem. B*, 2005, **109**, 986-991.
56. A. W. Rutherford, D. R. Paterson and J. E. Mullet, *Biochim. Biophys. Acta Bioenerg.*, 1981, **635**, 205-214.
57. A. W. Rutherford and J. E. Mullet, *Biochim. Biophys. Acta Bioenerg.*, 1981, **635**, 225-235.

58. F. J. E. Van Mieghem, K. Satoh and A. W. Rutherford, *Biochim. Biophys. Acta Bioenerg.*, 1991, **1058**, 379-385.
59. G. Chen, Y. Allahverdiyeva, E.-M. Aro, S. Styring and F. Mamedov, *Biochim. Biophys. Acta Bioenerg.*, 2011, **1807**, 205-215.
60. F. Mamedov, R. Danielsson, R. Gadjieva, P.-Å. Albertsson and S. Styring, *Biochemistry*, 2008, **47**, 3883-3891.
61. A. W. Rutherford, *Biochim. Biophys. Acta Bioenerg.*, 1985, **807**, 189-201.
62. M. K. Bosch, I. I. Proskuryakov, P. Gast and A. J. Hoff, *J. Phys. Chem.*, 1996, **100**, 2384-2390.
63. A. Kawamori, T. A. Ono, A. Ishii, S. Nakazawa, H. Hara, T. Tomo, . . . S. A. Dzuba, *Photosynth. Res.*, 2005, **84**, 187-192.
64. M. Kammel, J. Kern, W. Lubitz and R. Bittl, *Biochim. Biophys. Acta Bioenerg.*, 2003, **1605**, 47-54.
65. S. V. Pashenko, I. I. Proskuryakov, M. Germano, H. J. Van Gorkom and P. Gast, *Chem. Phys.*, 2003, **294**, 439-449.
66. W. Lubitz, *Phys. Chem. Chem. Phys.*, 2002, **4**, 5539-5545.
67. A.-F. Miller and G. W. Brudvig, *Biochim. Biophys. Acta Bioenerg.*, 1991, **1056**, 1-18.
68. G. F. W. Searle, A. Telfer, J. Barber and T. J. Schaafsma, *Biochim. Biophys. Acta Bioenerg.*, 1990, **1016**, 235-243.
69. M. Okamura, K. Satoh, R. Isaacson and G. Feher, *Progress in Photosynthesis Research*, 1987, **1**, 379-381.
70. A. Agostini, L. Nicol, N. Da Roit, M. Bortolus, R. Croce and D. Carbonera, *Biochim. Biophys. Acta Bioenerg.*, 2021, **1862**, 148481.
71. D. Carbonera, *Photosynth. Res.*, 2009, **102**, 403-414.
72. T. Noguchi, T. Tomo and C. Kato, *Biochemistry*, 2001, **40**, 2176-2185.
73. A. A. Zabelin, K. V. Neverov, A. A. Krasnovsky, V. A. Shkuropatova, V. A. Shuvalov and A. Y. Shkuropatov, *Biochim. Biophys. Acta Bioenerg.*, 2016, **1857**, 782-788.
74. K. V. Neverov, A. A. Krasnovsky, A. A. Zabelin, V. A. Shuvalov and A. Y. Shkuropatov, *Photosynth. Res.*, 2015, **125**, 43-49.
75. A. Agostini, D. M. Palm, H. Paulsen and D. Carbonera, *J. Phys. Chem. B*, 2018, **122**, 6156-6163.
76. S. Santabarbara, R. C. Jennings and D. Carbonera, *Chem. Phys.*, 2003, **294**, 257-266.
77. D. E. Budil and M. C. Thurnauer, *Biochim. Biophys. Acta Bioenerg.*, 1991, **1057**, 1-41.
78. T. Noguchi, Y. Inoue and K. Satoh, *Biochemistry*, 1993, **32**, 7186-7195.
79. A. R. Holzwarth, M. G. Muller, M. Reus, M. Nowaczyk, J. Sander and M. Rogner, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6895-6900.
80. H. Ishikita, B. Loll, J. Biesiadka, W. Saenger and E.-W. Knapp, *Biochemistry*, 2005, **44**, 4118-4124.
81. F. Müh, F. Lendzian, M. Roy, J. C. Williams, J. P. Allen and W. Lubitz, *J. Phys. Chem. B*, 2002, **106**, 3226-3236.
82. L. Zhang, D.-A. Silva, H. Zhang, A. Yue, Y. Yan and X. Huang, *Nat. Commun.*, 2014, **5**.
83. F. Müh, M. Plöckinger and T. Renger, *J. Phys. Chem. Lett.*, 2017, **8**, 850-858.
84. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2021, **17**, 1858-1873.
85. V. I. Novoderezhkin, J. P. Dekker and R. van Grondelle, *Biophys. J.*, 2007, **93**, 1293-1311.
86. E. J. G. Peterman, H. van Amerongen, R. van Grondelle and J. P. Dekker, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6128-6133.

87. T. Hayase, Y. Shimada, T. Mitomi, R. Nagao and T. Noguchi, *J. Phys. Chem. B*, 2023, **127**, 1758-1770.
88. A. Kamlowski, L. Frankemöller, A. Van Der Est, D. Stehlik and A. R. Holzwart, *Ber. Bunsen-Ges. Phys. Chem.*, 1996, **100**, 2045-2051.
89. J. Ho, E. Kish, D. D. Méndez-Hernández, K. Wongcarter, S. Pillai, G. Kodis, . . . B. Robert, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E5513-E5521.
90. Y. Kato and T. Noguchi, *Photosynth. Res.*, 2022, **152**, 135-151.
91. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
92. E. Sigfridsson and U. Ryde, *J. Comput. Chem.*, 1998, **19**, 377-395.
93. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, . . . P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.
94. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372-1377.
95. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
96. F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.
97. T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580-592.
98. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
99. S. J. Fox, C. Pittock, T. Fox, C. S. Tautermann, N. Malcolm and C.-K. Skylaris, *J. Chem. Phys.*, 2011, **135**, 224107.
100. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
101. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
102. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
103. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
104. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
105. K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283-290.
106. D. C. Liu and J. Nocedal, *Mathematical Programming*, 1989, **45**, 503-528.
107. M. J. G. Peach, M. J. Williamson and D. J. Tozer, *J. Chem. Theory Comput.*, 2011, **7**, 3578-3585.
108. D. Jacquemin, E. A. Perpète, I. Ciofini and C. Adamo, *J. Chem. Theory Comput.*, 2010, **6**, 1532-1537.
109. D. Hait, T. Zhu, D. P. McMahon and T. Van Voorhis, *J. Chem. Theory Comput.*, 2016, **12**, 3353-3359.
110. N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904.
111. A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsak and D. A. Pantazis, *J. Phys. Chem. B*, 2020, **124**, 8761-8771.
112. F. Neese, *J. Chem. Phys.*, 2001, **115**, 11080-11096.
113. V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129-12137.
114. V. Barone, in *Recent Advances in Density Functional Methods*, WORLD SCIENTIFIC, 1995, vol. Volume 1, pp. 287-334.
115. F. Neese, *J. Chem. Phys.*, 2005, **122**, 034107.
116. F. Neese, *J. Chem. Phys.*, 2007, **127**, 164112.
117. S. Sinnecker and F. Neese, *J. Phys. Chem. A*, 2006, **110**, 12267-12275.
118. A. Ceulemans, W. Oldenhof, C. Gorller-Walrand and L. G. Vanquickenborne, *J. Am. Chem. Soc.*, 1986, **108**, 1155-1163.
119. D. A. Hartzler, D. M. Niedzwiedzki, D. A. Bryant, R. E. Blankenship, Y. Pushkar and S. Savikhin, *J. Phys. Chem. B*, 2014, **118**, 7221-7232.
120. D. M. Niedzwiedzki and R. E. Blankenship, *Photosynth. Res.*, 2010, **106**, 227-238.

121. P. D. Frischmann, K. Mahata and F. Würthner, *Chem. Soc. Rev.*, 2013, **42**, 1847-1870.
122. Y. Kobori, S. Yamauchi, K. Akiyama, S. Tero-Kubota, H. Imahori, S. Fukuzumi and J. R. Norris, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10017-10022.
123. M. L. Williams, I. Schlesinger, R. M. Jacobberger and M. R. Wasielewski, *J. Am. Chem. Soc.*, 2022, **144**, 18607-18618.
124. S. M. Harvey and M. R. Wasielewski, *J. Am. Chem. Soc.*, 2021, **143**, 15508-15529.
125. E. Schlodder, W. J. Coleman, P. J. Nixon, R. O. Cohen, T. Renger and B. A. Diner, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2008, **363**, 1197-1202.
126. M. Di Valentin, C. W. M. Kay, G. Giacometti and K. Möbius, *Chem. Phys. Lett.*, 1996, **248**, 434-441.
127. A. Agostini, D. M. Palm, F.-J. Schmitt, M. Albertini, M. D. Valentin, H. Paulsen and D. Carbonera, *Sci. Rep.*, 2017, **7**, 7504.
128. A. Agostini, M. G. Dal Farra, H. Paulsen, A. Polimeno, L. Orian, M. Di Valentin and D. Carbonera, *J. Phys. Chem. B*, 2019, **123**, 8232-8239.
129. A. Marchanka, W. Lubitz and M. Van Gastel, *J. Phys. Chem. B*, 2009, **113**, 6917-6927.
130. F. Mokvist, J. Sjöholm, F. Mamedov and S. Styring, *Biochemistry*, 2014, **53**, 4228-4238.
131. P. Mulo, I. Sakurai and E.-M. Aro, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 247-257.
132. A. K. Clarke, A. Soitamo, P. Gustafsson and G. Oquist, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 9973-9977.

5

Microscopic Basis of Reaction Center Modulation in PsbA Variants of Photosystem II[‡]

[‡] This work is adapted in part from "Microscopic Basis of Reaction Center Modulation in PsbA Variants of Photosystem II" by Bhattacharjee, S.; Gordiy, I.; Sirohiwal, A.; and Pantazis, D. A., **2024,** *submitted.*

5.1 Introduction

Photosystem II (PSII) of oxygenic photosynthesis is a dimeric multi-subunit protein-pigment complex (PPC) responsible for the four-electron oxidation of water into molecular oxygen and the two-electron reduction of a mobile plastoquinone.[1-6] Cyanobacterial PSII comprises 17 transmembrane subunits and 3 extrinsic proteins.[1, 7-12] The core complex (PSII-cc) consists of the six larger subunits D1 (PsbA), D2 (PsbD), CP43 (PsbC), CP47 (PsbB)[13] and Cyt$_{b559}$ (PsbE, PsbF).[14, 15] D1 and D2 bind most active components of the electron transfer chain including the reaction center (RC) pigments and the oxygen evolving complex (OEC).[16] The RC is the site of the light-driven charge separation and primary electron transfer. It consists of four chlorophylls, namely the P$_{D1}$P$_{D2}$ central pair flanked by Chl$_{D1}$ and Chl$_{D2}$, and two pheophytins Pheo$_{D1}$ and Pheo$_{D2}$, arranged pseudo-symmetrically along the D1 and D2 heterodimer **(Fig. 5.1).** Excitation energy transfer (EET) from external light harvesting complexes[17-19] and the core antennae [20] CP43 and CP47 triggers primary charge separation along the D1 branch of the RC.[17, 21, 22]

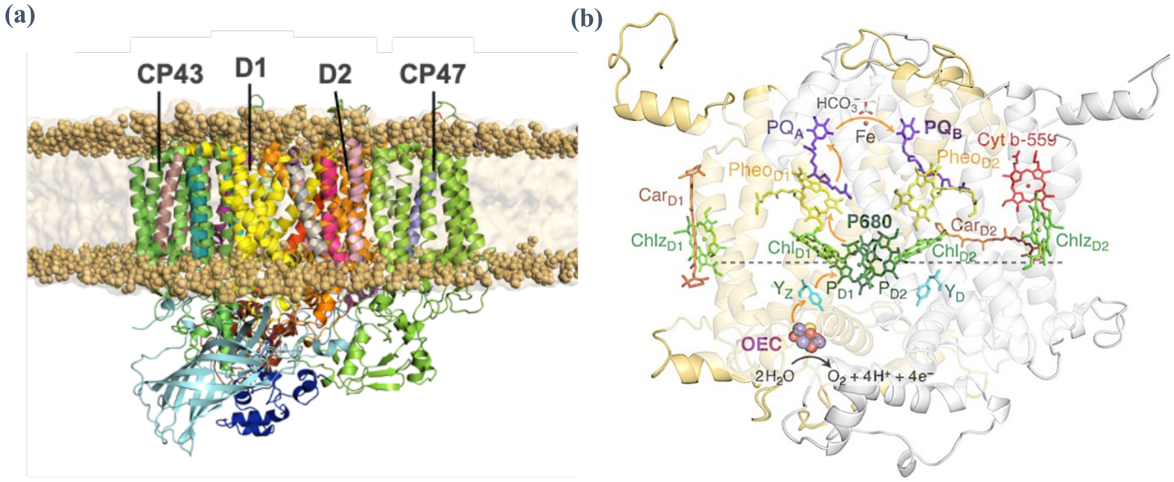


Fig. 5.1 (a) The overall structure of the membrane-bound PSII monomer with major subunits; **(b)** Reaction center (**RC**) cofactors oriented with respect to **D1** (yellow) and **D2** proteins (white).

The D1 protein, encoded by the *psbA* gene family,[23] undergoes light-induced turnover to protect PSII from photodamage.[24-26] In higher plants and green algae there is a single *psbA* gene[27] but cyanobacteria may possess multiple copies,[27-33] whose expression occurs in response to environmental factors such as high light, UV light, or varying temperature.[27, 29, 34, 35] In the mesophilic cyanobacterium *Synechocystis* PCC 6803, D1 is encoded by three *psbA* variants (*psbA*$_1$, *psbA*$_2$ and *psbA*$_3$), among which *psbA*$_2$ and *psbA*$_3$ encode the same D1 isoform expressed under various stress conditions, whereas *psbA*$_1$ encodes a distinct isoform. *Synechococcus* PCC 7942 also has three *psbA* genes that encode two different D1 isoforms, D1:1 by *psbA*$_1$ and D1:2 by *psbA*$_2$ and *psbA*$_3$. In the thermophilic cyanobacterium *T. elongatus*, three *psbA* genes have also been identified, but encoding three distinct D1 isoforms. Among these, *psbA*$_1$ is expressed under normal growth, *psbA*$_3$ is induced at high light,[30, 36, 37] and *psbA*$_2$ is partially activated under microaerobic conditions. Despite insights from studies that use site-directed mutagenesis combined with spectroscopy or with crystallographic analysis of available mutant strains,[28-30, 36-40] a microscopic understanding of how the different isoforms regulate the properties and function of the RC remains elusive.

Under physiological conditions, charge-transfer (CT) states of $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ character are created in the PSII-RC[41, 42] presumably leading to formation of a transient primary charge-separated radical pair $\mathrm{Chl_{D1}}^{+}\mathrm{Pheo_{D1}}^{-}$ before the electron hole rapidly delocalizes onto $\mathrm{P_{D1}P_{D2}}$ forming "$\mathrm{P_{680}}^{+}$",[43, 44] the highly oxidizing radical cation (E_m ~1.1–1.3 V) that drives water oxidation. Protein matrix electrostatics are the principal factor for differentiating the properties of otherwise chemically identical pigments, generating functional asymmetry in the RC and enabling formation of inter-pigment CT states prior to primary charge separation.[45] Given that each *psbA* variant has specific sequence differences, the RC pigments are embedded in a slightly different protein environment in each case. It is therefore essential to understand how these differences influence the excitation profile, redox properties, and primary CT states of RC pigments. Here we combine large-scale molecular dynamics (MD) simulations with multilevel quantum-mechanics/molecular-mechanics (QM/MM) calculations on membrane-bound PSII monomer models for the three PsbA-PSII variants. We compare the redox and optical properties of the critical $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair using long-range corrected time-dependent density functional theory (TD-DFT) and identify residues responsible for specific matrix-induced adjustments of CT states and redox properties in this crucial pigment pair.

5.2 Methodology

5.2.1 Preparation of MM models

Separate MM models for each PsbA variant were constructed to evaluate the mutations in each D1 copy. The initial structure of the PSII monomer is based on the high-resolution crystal structure of *T. vulcanus* (3WU2.pdb),**[10]** originally containing the PsbA1 variant. For the PsbA2 and A3 variants, the D1 protein (Chain A of 3WU2) was replaced with the corresponding sequences of PsbA2- and PsbA3-PSII based on the recent crystal structures (7YQ2.pdb and 7YQ7.pdb).**[28]** It is important to note that the PsbA1 sequences of *T. vulcanus* and *T. elongatus* are essentially identical with the exception of a single difference at D1-286. Two protonation states for the mutation at D1-130 were considered for PsbA2 and PsbA3 variants, namely (a) D1-Glu for the deprotonated glutamate and (b) D1-Glu(H) for glutamic acid. In each case, the complete protein unit was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer and solvated with TIP3P waters (see **Section 3.2** for details). The number of lipids and water molecules were kept same for each setup. All the models were neutralized with appropriate number of counterions and a 0.15M physiological Na^+Cl^- salt concentration. For the standard protein residues, waters and lipid bilayer we used the parameters from the AMBERff14SB, TIP3P and Lipid17 force fields respectively. The partial charges and forcefield parameters for the organic cofactors (plastoquinones, carotenoids, structural lipids) were obtained using GAFF2, parameters for the OEC were taken from earlier studies, while those of the remaining cofactors (non-heme Fe, chlorophylls, pheophytins and hemes) were obtained from literature.

5.2.2 Molecular Dynamics (MD) Simulations

A step-by-step minimization protocol is employed to remove energetically unfavorable geometric clashes in the system. In the equilibration phase, the systems are heated from 10 to 100 K in a succession of 5 ps in the NVT ensemble and further from 100 to 303 K in the NPT ensemble for 125 ps. The temperature during this step is maintained using the Langevin dynamics with a collision frequency of 5 ps^{-1}. During the equilibration phase, the C_α atoms of amino acids were restrained with a force constant of 20 kcal $mol^{-1}Å^{-2}$. Subsequently, the restraints on the C_α atoms of amino acids were systematically decreased (2 kcal $mol^{-1}Å^{-2}$/500 ps). The systems were further equilibrated for 65 ns to properly equilibrate the lipid bilayer. Thereafter, a series of production runs were initiated for 60 ns for each system (300 ns in total) without restraints and the temperature and pressure set at 303 K and 1 atm, respectively. During the entire procedure, the system pressure was controlled using the Berendsen barostat with anisotropic pressure scaling with a relaxation time of 2 ps. SHAKE algorithm was used to constrain the bonds involving hydrogens. The MD time step was 2 fs and frames were saved every 20 ps. The electrostatic interactions were treated using the Particle Mesh Ewald (PME) approach[46] with a 10 Å cut-off. The AMBER20 package was used to perform the energy minimizations and equilibration dynamics. The production MD simulations were performed in the GPU version of the pmemd module (*pmemd.cuda*) and trajectory analysis was performed using Cpptraj in AmberTools[47], VMD and PyMol.

5.2.3 Binding Free Energy Calculations

In order to compare the thermodynamic stability of $\mathrm{Pheo_{D1}}$ in each D1 isoform, binding free energies were calculated using the MM-PBSA module in AmberTools. $\mathrm{Pheo_{D1}}$ is considered as the ligand and the D1/D2 proteins as the receptor. In general, MM-PBSA combines molecular mechanics, implicit Generalized Born (GB)/Poisson-Boltzmann (PB) solvation schemes and solvent accessibility surface area calculations to estimate the binding free energies of a protein-ligand complex. A single trajectory approach is employed, where only the complex form is propagated, eliminating the need for separate molecular dynamics (MD) simulations for the ligand and protein. Solvation free energies are computed using both the GB and PB solvation schemes. Binding energy calculations for each MM system were carried out on 300 equidistant snapshots extracted from the initial 60 ns of production MD. The calculated binding energies were subsequently averaged over this ensemble. The solute dielectric constant is set to 2.0 in all cases. Our primary goal was only to compare the relative stabilities of $\mathrm{Pheo_{D1}}$ in each PsbA matrix, so the entropic contribution (ΔS) to the binding energy is not computed due to its high computational cost. The calculations are conducted for all the D1 isoforms described before. The binding free-energies were further decomposed into per-residue contributions to assess the influence of each residue towards the stability of $\mathrm{Pheo_{D1}}$ in each PSII variant. All calculations were performed using the parallelized *MMPBSA.py* module of the Amber20 package.

5.2.4 QM/MM Geometry optimizations

We chose 3x9 independent MD snapshots for QM/MM optimizations. The first snapshot resembles the 'crystal like' structure from the initial equilibration, while the remaining 8 were taken from the production MD of each PsbA1/A2/A3- PSII protein. For the QM/MM calculations the complete PSII monomer and all waters around the protein were retained (see

Section 3.6 for more details). Appropriate number of Na^+ ions were included to maintain overall neutrality of each system. QM/MM calculations are performed using electrostatic embedding scheme. The hydrogen link atom approach was employed to cut through C–C covalent bonds and the charge-shift (CS) method was used to avoid over-polarization of the QM atoms. The chlorin macrocycles and axially coordinated ligands to the Mg^{2+} were treated at the QM level. For Chl_{D1}, the water molecule hydrogen bonded to the axially ligated water and ester group attached to the 13^2-carbon position on ring E and all H-bonded sidechains in the vicinity of Chl_{D1}–$Pheo_{D1}$ pair were also included in the QM region. The phytyl chains were included up to C-17 (truncated as methyl) and the remaining chain was kept in the MM region. The complete system was further subdivided into two parts: active and static. The active region consisted of atoms within the QM and MM, which are flexible during optimization, whereas the remaining MM atoms were fixed and only contributed to the electrostatics. The Chl_{D1}–$Pheo_{D1}$ and P_{D1}–P_{D2} pair were optimized in their ground state ($S = 0$) geometry, for all D1 variants. Complete amino acid residues and waters within 10 Å from the center of each chlorin ring were considered in the active region. The PBE functional was used to optimize the QM regions using the def2-TZVP basis set, along with D3(BJ) dispersion corrections. Dense DFT integration grids (DefGrid2 in ORCA) and resolution of identity approximation (RI) was used with the corresponding auxiliary basis set (def2/J).

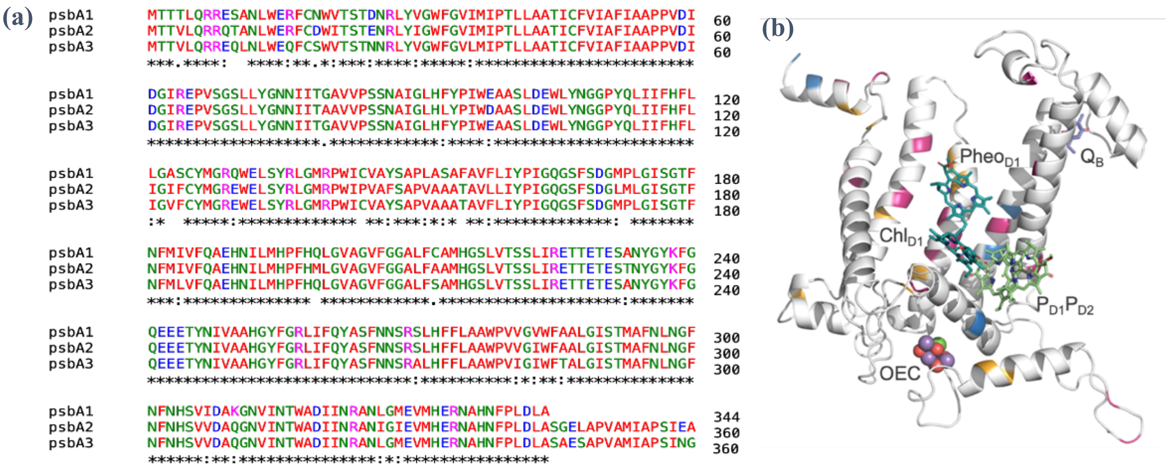
5.2.5 Excited State Calculations

The vertical excitation energies (8 roots) were computed on the pair-optimized ground state geometries using QM/MM and TD-DFT without TDA. All calculations are performed the using the range separated ωB97X-D3(BJ) functional (modified version of ωB97X-V with D3BJ correction) along with def2-TZVP basis sets. The RIJCOSX approximation[48] and the corresponding auxiliary basis sets were used throughout. VeryTightSCF convergence criteria were applied, along with dense integration grids (DefGrid2). To explicitly compare the effect of the D1 mutations and have a meaningful sampling of the protein conformation, further excited state calculations were performed on 60 snapshots obtained from unbiased production runs of the PSII-membrane complex, i.e., without any backbone restraints. The electrostatic effects of the protein environment on the excited states were included through explicit MM point charges in each case. All the QM/MM calculations in this work are performed using the multiscale module in ORCA 5.0.

5.3 Results

5.3.1 Structural analysis of PsbA variants.

The folded mature D1 protein in thermophilic cyanobacteria contains 344 residues within five transmembrane helices (TMH A–E) in each of the three PsbA variants.[28, 31] Sequence analysis shows 21 substitutions between PsbA1/PsbA3, 31 between PsbA1/PsbA2 and 27 between PsbA2/PsbA3, indicating that all three variants are highly conserved (**Fig. 5.2a**).



```
psbA1       MTTTLQRRESANLWERFCNWVTSTDNRLYVGWFGVIMIPTLLAATICFVIAFIAAPPVDI  60
psbA2       MTTVLQRRQTANLWERFCDWITSTENRLYIGWFGVIMIPTLLAATICFVIAFIAAPPVDI  60
psbA3       MTTVLQRREQLNLWEQFCSWVTSTNNRLYVGWFGVLMIPTLLAATICFVIAFIAAPPVDI  60
            ***.****:  ****:**.*:***:****:*****:************************

psbA1       DGIREPVSGSLLYGNNIITGAVVPSSNAIGLHFYPIWEAASLDEWLYNGGPYQLIIFHFL  120
psbA2       DGIREPVSGSLLYGNNIITAAVVPSSNAIGLHLYPIWDAASLDEWLYNGGPYQLIIFHFL  120
psbA3       DGIREPVSGSLLYGNNIITGAVVPSSNAIGLHFYPIWEAASLDEWLYNGGPYQLIIFHFL  120
            *******************.************:****:**********************

psbA1       LGASCYMGRQWELSYRLGMRPWICVAYSAPLASAFAVFLIYPIGQGSFSDGMPLGISGTF  180
psbA2       IGIFCYMGREWELSYRLGMRPWIPVAFSAPVAAATAVLLIYPIGQGSFSDGLMLGISGTF  180
psbA3       IGVFCYMGREWELSYRLGMRPWICVAYSAPVAAATAVFLIYPIGQGSFSDGMPLGISGTF  180
            :*  *****:************* **:***:*:* **:*************: *******

psbA1       NFMIVFQAEHNILMHPFHQLGVAGVFGGALFCAMHGSLVTSSLIRETTETESANYGYKFG  240
psbA2       NFMIVFQAEHNILMHPFHMLGVAGVFGGALFAAMHGSLVTSSLIRETTETESTNYGYKFG  240
psbA3       NFMLVFQAEHNILMHPFHQLGVAGVFGGALFSAMHGSLVTSSLIRETTETESANYGYKFG  240
            ***:************** ************.********************:*******

psbA1       QEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLAAWPVVGVWFAALGISTMAFNLNGF  300
psbA2       QEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLAAWPVVGIWFAALGISTMAFNLNGF  300
psbA3       QEEETYNIVAAHGYFGRLIFQYASFNNSRALHFFLAAWPVIGIWFTALGISTMAFNLNGF  300
            *****************************:**********:*:**:**************

psbA1       NFNHSVIDAKGNVINTWADIINRANLGMEVMHERNAHNFPLDLA                  344
psbA2       NFNHSVVDAQGNVINTWADIINRANIGIEVMHERNAHNFPLDLASGELAPVAMIAPSIEA  360
psbA3       NFNHSVVDAQGNVINTWADIINRANLGMEVMHERNAHNFPLDLASAESAPVAMIAPSING  360
            ******:**:***************:*:****************
```

Fig. 5.2 (a) Amino acid sequences of the three PsbA proteins (PsbA1: *T. vulcanus,* PsbA2 and PsbA3: *T. elongatus)*. The consensus sequence is marked as follows: identical by asterisks, similar by colon, less similar by dot and non-conserved by space. Acidic and hydrophilic residues (D, E) are colored blue, basic residues (K, R) pink, nonpolar/hydrophobic residues (A, F, I, L, M, V, W, P) are red, polar uncharged residues (C, G, Q, N, S, Y, T, H) green. The sequence comparison is performed using the Multiple Sequence Alignment (MSA) tool in Clustal Omega. **(b)** Location of the D1 protein substitutions with respect to PsbA1-PSII along with selected RC cofactors (Chl$_{D1}$, Pheo$_{D1}$, P$_{D1}$–P$_{D2}$, OEC and Q$_B$). Residues that differ only in the PsbA2 variant are marked in orange, PsbA3 variant in blue, and differences in both PsbA2 and A3 variants are in pink.

The segments with most significant variations are the N-terminus amino acids (1–30) with 10 variant residues followed by six variations in the TMH-C segment spanning residues (144–158) in the vicinity of the redox-active Y$_Z$. The nature and location of most PsbA substitutions **(Fig. 5.2b)** support the proposed influence on the spectroscopic and redox properties of the RC. Substitutions located near RC cofactors (**Table 5.1**) participating in the charge separation and electron transfer compose 64% (18 out of 28) of differences between PsbA1/PsbA2 and 57% (12 out of 21) between PsbA1/PsbA3.

Substitutions located more than 15 Å away from any RC cofactor are not expected to influence optical and redox properties, therefore, subsequent analysis is focused on substitutions that are closest to the active branch pigments. A general trend that can be observed for these substitutions is that PsbA1 residues with heteroatoms (sulfur or oxygen) are often substituted by aliphatic hydrophobic residues (e.g., M172L and M328I in PsbA2, and S153A, S270A in PsbA3). These heteroatoms possess easily polarizable free electron pairs, implying an increased importance of non-covalent interactions or an attenuation of directed electrostatic effects.

Table 5.1 Location and identity of all variant residues within the copies of PsbA1 (*T. vulcanus*), PsbA2 and PsbA3 of *T. elongatus* (residues within **15 Å** from any cofactor in RC). All residues belong to the D1 subunit of the PSII monomer. The values represent the distance between the closest sidechain atom from a heavy atom of the respective cofactor in the equilibrated MD snapshot for each variant.

Position	**A1**	**A2**	**A3**	**Nearest cofactor**	**A1**	**A2**	**A3**
123	Ala	Ile	Val	Chl$_{D1}$/Pheo$_{D1}$	8/9	6.1/7.7	6.9/7.8
124	Ser	Phe	Phe	Pheo$_{D1}$	13.1	12.8	12.8
144	Cys	Pro	Cys	Pheo$_{D1}$	7.8	6.3	7.8
80	Gly	Ala	Gly	Chl$_{D1}$	10.6	10.5	10.8
173	Pro	Met	Pro	Chl$_{D1}$/P$_{D1}$/OEC	8.1/9.9/9.4	9.1/9.9/7.1	8.0/9.8/9.3
199	Gln	Met	Gln	P$_{D1}$/Chl$_{D2}$	9.0/3.6	8.9/3.6	9.3/3.4
121	Leu	Ile	Ile	Chl$_{D1}$/Pheo$_{D1}$	13.7/14.9	13.7/14.3	13.6/14.4
130	Gln	Glu	Glu	Pheo$_{D1}$/Q$_A$	2.9/9.3	2.7/9.2	2.7/9.6
147	Tyr	Phe	Tyr	Pheo$_{D1}$	3.4	3.6	3.7
151	Leu	Val	Val	Pheo$_{D1}$/Chl$_{D1}$	8.6/8.8	7.8/9.2	7.3/9
153	Ser	Ala	Ala	P$_{D1}$/Chl$_{D1}$/Pheo$_{D1}$	3.7/5.4/7.2	4.0/5.3/8.0	3.8/5.8/8.1
158	Phe	Leu	Phe	Chl$_{D1}$/P$_{D1}$/Pheo$_{D1}$	4.1/9.1/8.9	5.4/9.9/10.7	3.8/9.3/9.0
172	Met	Leu	Met	P$_{D1}$/Chl$_{D1}$/OEC	5.7/3.8/11.0	6.8/4.2/11.3	5.8/3.8/11.0
184	Ile	Ile	Leu	P$_{D1}$/P$_{D2}$/OEC	4.2/6.4/10.4	4.2/6.6/10.7	4.0/5.9/10.3
212	Cys	Ala	Ser	P$_{D2}$/Pheo$_{D2}$/Pheo$_{D1}$/Q$_A$/Fe/Q$_B$	12.6/6.5/9.7/10.4/10.1/10.0	12.8/7.1/9.7/11.0/9.6/9.2	12.8/7.1/9.9/10.4/9.6/9.7
270	Ser	Ser	Ala	Fe/Q$_B$	12.1/8.5	12.1/8.9	11.0/8.4
281	Val	Val	Ile	P$_{D1}$/Pheo$_{D1}$	11.6/8.9	11.6/8.7	11.9/8.8
283	Val	Ile	Ile	P$_{D1}$/Chl$_{D1}$/Pheo$_{D1}$	5.1/8.6/3.8	4.1/8.4/3.8	4.0/8.9/3.8
286	Ala	Ala	Thr	P$_{D1}$/Chl$_{D1}$/Pheo$_{D1}$	3.4/8.3/8.1	3.4/8.2/8.0	3.6/7.9/7.8
326	Leu	Ile	Leu	OEC	12.7	12.8	12.2
328	Met	Ile	Met	P$_{D1}$/P$_{D2}$/OEC	5.4/7.9/9.6	6.8/9.7/8.4	5.3/8.8/9.4

Focusing on the donor side of PSII, the substitution at D1-172 is closest to the primary donor Chl$_{D1}$. In PsbA1 and PsbA3-PSII crystal structures[28] the M172 side chain is oriented towards the macrocyclic ring of Chl$_{D1}$ at a distance of 3.8 Å.[28] The PsbA2 variant has a leucine at this position slightly further at 4.2 Å. Only the D1-286 residue differs in the PsbA1 sequences of *T. elongatus* and *T. vulcanus*. In the former it is a threonine (T286) H-bonded to the acetyl moiety of P$_{D1}$, while the latter has a hydrophobic alanine residue. A286 is conserved in the PsbA2 of *T. elongatus* while T286 is found in the PsbA3 variant. In the sequences where T286 is replaced by A286, a water is H-bonded to the acetyl group of P_{D1} suggesting that this substitution may destabilize P_{D1}.[39] The D1-P173M substitution in PsbA2 is also shown to contribute towards structural differences compared to PsbA1/A3.[28] Particularly, recent crystallographic studies suggested that this substitution can lead to the loss of two water molecules due to the narrowing of the Cl-1 channel of the OEC,[28] but further investigations are required to understand the functional implications of this difference.

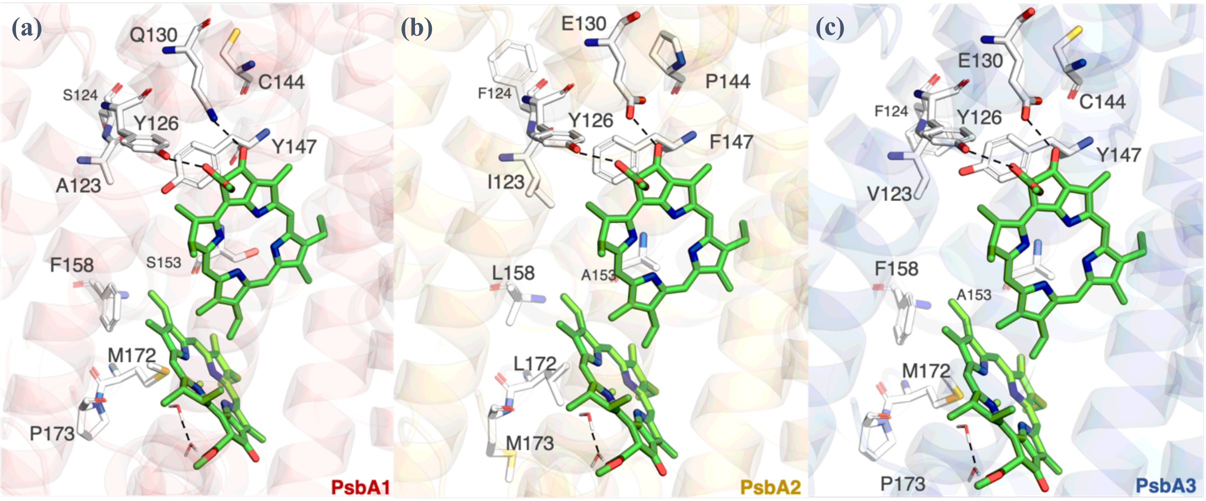


Fig. 5.3 Comparison of the D1 protein substitutions around the **Chl$_{D1}$–Pheo$_{D1}$** pair among the three PsbA (D1) variants based on QM/MM optimized structures in **(a)** PsbA1, **(b)** PsbA2 and **(c)** PsbA3. Selected hydrogen bonding interactions are indicated with dashed lines.

On the acceptor side, two important substitutions are found in the binding pocket of Pheo$_{D1}$ (**Fig. 5.3**). In PsbA1-PSII, the two closest substitution sites are D1-Q130, which is H-bonded to the 13^1-keto group of Pheo$_{D1}$, and D1-Y147 that acts as H-bonding partner to the C=O group of the phytyl chain. Based on recent crystal structures and previous sequence analysis[28, 31] Q130 of PsbA1 is replaced by a glutamate (E130) in PsbA2 and PsbA3, whereas Y147 is replaced by a non-polar F147 in PsbA2. Pheo$_{D1}$ has a third H-bonding partner (Y126) to the acetyl group, but this residue remains invariant across the three isoforms. The only amino acid change that occurs at the Q_B pocket is at D1-270, which is serine in PsbA1/A2 but alanine in PsbA3. This was proposed to result in loss of an H-bond between Ser270 and the sulfoquinovosyl-diacylglycerol (SQDG) molecule near Q_B, which affects the binding properties of Q_B and the exchange efficiency of Q_BH_2.[28, 49] Further important substitutions are located at position 212 where C212 in PsbA1 is substituted for alanine in PsbA2 and serine in PsbA3. This site is centrally located in between the D1/D2 branches close to Pheo$_{D2}$ but equidistant from non-heme Fe, Q_A, Q_B, Pheo$_{D1}$ and may have an electrostatic effect on secondary ET processes in the D2 side.

Overall, the differences between PsbA variants are expected to have the greatest influence on P_{D1}, Chl_{D1}, $Pheo_{D1}$ and the OEC, but proximity analysis suggests that the differences in each variant are most relevant for Chl_{D1} and $Pheo_{D1}$. The differences are farthest from the OEC (~10 Å), followed by $P_{D1}P_{D2}$. Therefore, these cofactors do not seem to be targeted for adjustment in the genetic variants and subsequent analysis in this work is focused on the properties of the Chl_{D1}–$Pheo_{D1}$ pair, the site of primary charge separation.

5.3.2 Functional Dynamics around $Pheo_{D1}$

Large-scale MD simulations enable us to obtain a realistic view of the H-bonding environment within the RC and to extensively sample the global conformational changes of the protein matrix. The whole system was systematically minimized and equilibrated using all-atom MD, and unbiased production runs were performed for 300 ns to obtain an ensemble of snapshots. Analysis of the MD trajectories indicates that the D1 polypeptides are relatively stable throughout the simulations compared to the dynamic evolution of the whole system (**Fig. 5.4**) and this observation is consistent for all PsbA variants.

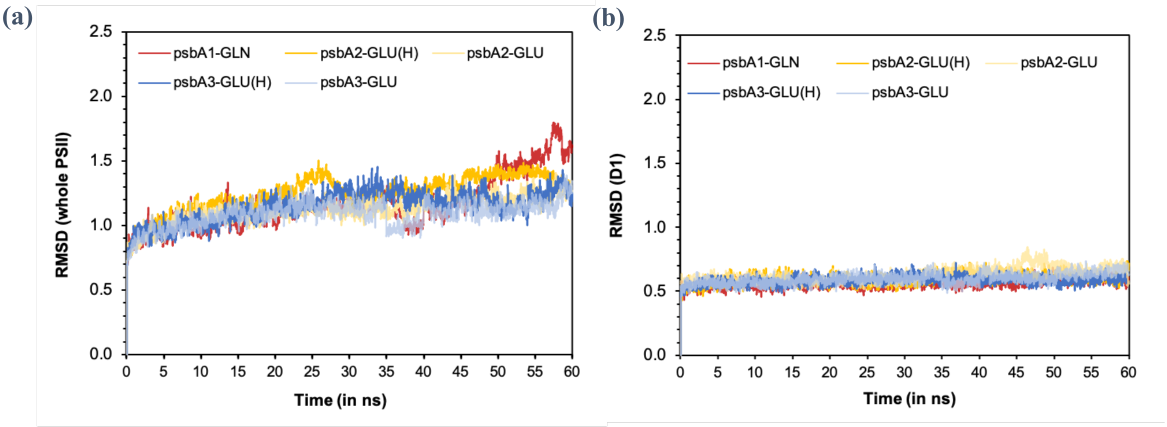
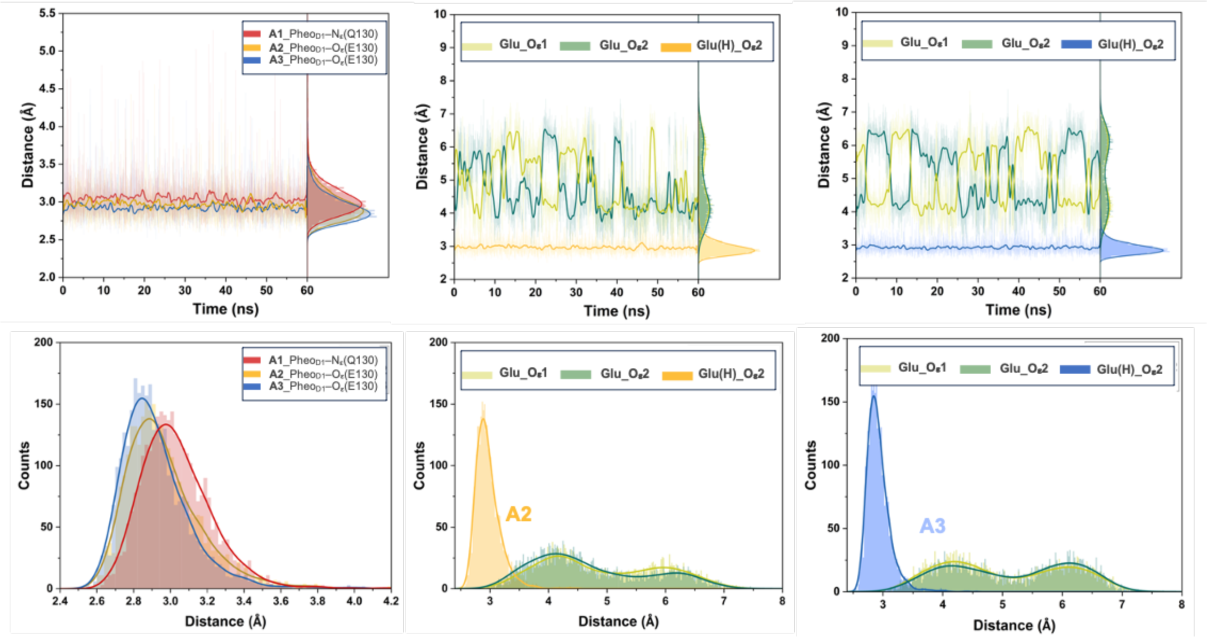


Fig. 5.4 (a) RMSD of the entire PSII monomer; **(b)** RMSD of the PsbA (D1) protein only for the A1–A2–A3 variants with different protonation states for **D1–E130** (The RMSD values are along **60 ns** of production MD). Disordered regions of the protein are not considered in the RMSD calculations.

Based on the relative RMSD (**Fig. 5.4**) and structural comparisons, we confirm that the three variants do not undergo any large-scale conformational changes within the MD timescales. The relative orientations of the six RC pigments also remained invariant. Recent crystallographic studies reported an RMSD of 0.27 Å for the C_α atoms between PsbA1/PsbA2-PSII, 0.25 Å for PsbA1/PsbA3-PSII, and 0.20 Å for PsbA2/PsbA3-PSII. In the current study, a comparative RMSD analysis of the whole PSII protein (**Fig. 5.4**) post-equilibration and throughout the production MD confirms that the multiple residue changes in the PsbA2 and PsbA3 variants do not have an effect on the overall protein structure.

Recent XRD structures[28] revealed that PsbA2-PSII has a longer H-bonding distance between Pheo$_{D1}$ and D1-Y126 and one H-bond less due to the D1-Y147F substitution, compared to that of PsbA1 and A3. In PsbA1/A3, Y147 acts as a direct H-bonding partner to the C=O group of the phytyl chain and stabilizes the Pheo$_{D1}$ moiety by 4.5 kcal mol^{-1} in PsbA1 and A3. Whereas in PsbA2, the hydrophobic F147 leads to a loss of this crucial H-bonding interaction consequently, making Pheo$_{D1}$ slightly unstable in PsbA2. The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) simulations reveal that with the exception of D1-130 and D1-147, all the residues that contribute to the stability of Pheo$_{D1}$ are conserved across the three variants. Based on the calculated binding energies, we find that the Pheo$_{D1}$ in PsbA2-PSII ($\Delta G_{\text{binding}}$ –78.37 kcal mol^{-1}) is less stable compared to PsbA1 ($\Delta G_{\text{binding}}$ –81.76 kcal mol^{-1}) and PsbA3 (–81.81 kcal mol^{-1}). The local energy decomposition and residue-wise decomposition of the total binding energy of Pheo$_{D1}$ for each model is depicted in **Table 5.4** and **Fig. 5.15**.



(a) (b) (c)

Fig. 5.5 (a) H-bonding distances between the Pheo$_{D1}$ (C=O) group and sidechain of D1-130 along the MD trajectory for PsbA1 (Pheo$_{D1}$–N$_{\varepsilon}$Q130), PsbA2 (Glu130, yellow) and PsbA3 (Glu130, blue) variants. **(b)** Distances between the Pheo$_{D1}$ C=O group and sidechain oxygen atoms (O$_{\varepsilon}$1, O$_{\varepsilon}$2) of D1-130 for deprotonated (green) Glu130, protonated Glu(H) in PsbA2-PSII (yellow) **(c)** same for PsbA3-PSII (blue).

5.3.3 Redox tuning of the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair

The physiochemical properties of the RC pigments are influenced by various "intrinsic" factors such as H-bonding, axial ligation, macrocyclic ring curvature, out-of-plane motion of π-conjugated groups, electronic couplings[50-57] and protein electrostatics. [41, 42, 45, 58, 59] However, the primary events in PSII are also affected by "extrinsic" factors like the wavelength and intensity of incident radiation. For instance, far-red light photoacclimated (FARLiP) cyanobacteria[29, 60-68] adapt by substituting different chlorophylls with red-shifted pigments (Chl d, f) to utilize longer wavelengths efficiently. On the other hand, the PsbA variants in thermophilic cyanobacteria illustrate how changes in protein electrostatics can modulate the physiological function of the RC. Under normal conditions, trapping of excitation energy and primary CS occur along the D1 branch to form the primary transient $\mathrm{Chl_{D1}}^{+}\mathrm{Pheo_{D1}}^{-}$ pair. Moreover, the nature of excitations and directionality of ET is not an intrinsic property of the pigments, but arises from protein electrostatics and pigment interactions.[41] Here we demonstrate how PsbA substitutions affect these critical RC cofactors.

Substitutions that cause variations in the H-bonding environment around the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair are assumed to directly tune specific chemical properties, in particular the redox potential ($\mathrm{E_m}$) of $\mathrm{Pheo_{D1}}/\mathrm{Pheo_{D1}}^{-}$.[30, 36, 38, 39, 69-72] Recent XRD structures[28] revealed that PsbA2-PSII has a longer H-bonding distance between $\mathrm{Pheo_{D1}}$ and D1-Y126 and one H-bond less due to the D1-Y147F substitution, compared to that of PsbA1 and A3. Based on our QM/MM computed energetics of the pigments, we see that the non-polar F147 residue in PsbA2 not only affects the H-bonding interaction but also makes the binding pocket of $\mathrm{Pheo_{D1}}$ more hydrophobic causing a destabilization of the $\mathrm{Pheo_{D1}}$ ground state in PsbA2-PSII.

A detailed analysis of the frontier molecular orbitals of the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair can help explain the electronic origin of the E_m shifts ($\mathrm{Pheo_{D1}}^{-}/\mathrm{Pheo_{D1}}$) for each D1 variant. We estimated the vertical electron affinity (EA) of $\mathrm{Pheo_{D1}}$ by the electronic energy difference of the $\mathrm{Pheo_{D1}}^{-}/\mathrm{Pheo_{D1}}$ redox couple. A comparison of the $\mathrm{Pheo_{D1}}/\mathrm{Pheo_{D1}}^{-}$ EAs across an ensemble of protein configurations (see **Fig. 5.6a)** suggests that $\mathrm{Pheo_{D1}}$ in PsbA3 has the highest average EA followed by PsbA1 and PsbA2. This trend is consistent across most configurations implying that the electron accepting tendency of $\mathrm{Pheo_{D1}}$ is highest in PsbA3, which causes its $\mathrm{E_m}$ to be more positive compared to that in PsbA1/PsbA2. This is opposite to what is reported for PsbA2 even though both PsbA2/A3 possess the same $\mathrm{Pheo_{D1}}$–E130 interaction. This demonstrates that the shift in the E_m ($Pheo_{D1}^{-}/Pheo_{D1}$) cannot be explained based on a single Q130E substitution, but is instead a cumulative effect of multiple substitutions.

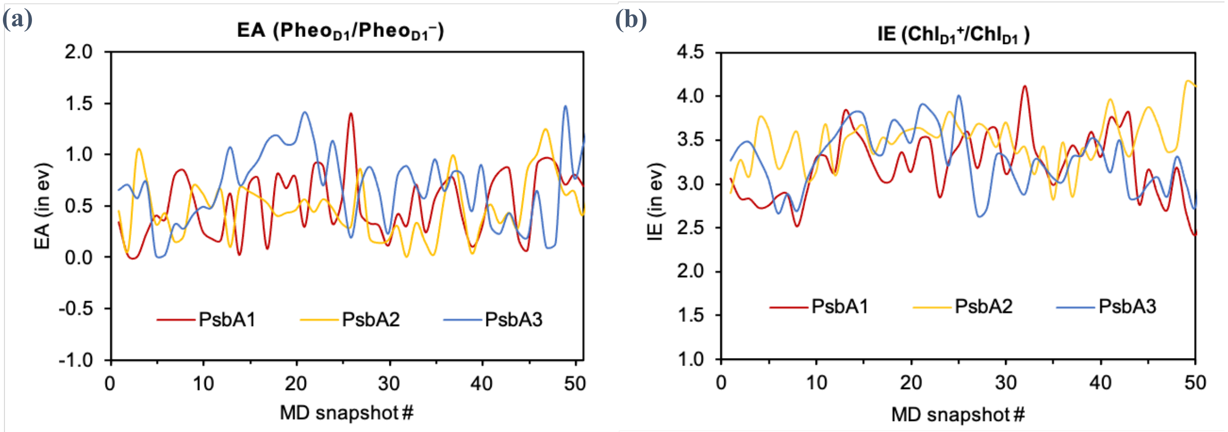


Fig. 5.6 Vertical **(a)** Electron Affinities (**EA**) for $Pheo_{D1}^{-}/Pheo_{D1}$ and **(b)** Ionization Energies (**IE**) for Chl_{D1}^{+}/Chl_{D1} in PsbA1–A2–A3 variants. The QM/MM single-point calculations are performed on 50 independent protein snapshots obtained from MD simulations.

We also computed the ionization energies (IE) of Chl_{D1} from the HOMO energies of the Chl_{D1}–$Pheo_{D1}$ pair applying Koopman's theorem and find that the IE also vary among the different variants (**Fig. 5.6b**), PsbA1: 3.15±0.41 eV, PsbA2: 3.54±0.31 eV, PsbA3: 3.32±0.35 eV. Chl_{D1} in PsbA1 is therefore the easiest to oxidize followed by PsbA3 and A2.

5.3.4 Global tuning of Chl_{D1}–$Pheo_{D1}$ excited states

In the previous section we determined how the interplay of protein electrostatics and structural modifications in the vicinity of the pigments can influence their redox properties. Here, we show how the D1 substitutions influence the optical properties of Chl_{D1}–$Pheo_{D1}$ and thereby the formation of the $Chl_{D1}^{+}Pheo_{D1}^{-}$ radical pair. First, the vertical excitation energies for the Chl_{D1}–$Pheo_{D1}$ pair for each variant were computed using QM/MM and TD-DFT. The QM region in each case consisted of the Chl_{D1}–$Pheo_{D1}$ pair along with axial waters and immediate H-bonded residues (**Fig. 5.3**). The nature of the excited states was determined using Natural Transition Orbitals (NTOs)[73] (**Fig. 5.7**). In PsbA1-PSII, the lowest excited state (S_1) is either localized on Chl_{D1}, or is a $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT state, possibly mixed with local excitation (LE) character. Our results on the "crystal like" snapshot (**Fig. 5.12**) show that Chl_{D1} is still the red-most pigment of the RC, with excitation energies of 1.806 eV (PsbA1), 1.816 (PsbA2) and 1.819 eV (PsbA3) respectively. Similarly, $Pheo_{D1}$ is blue shifted for all three variants: 2.048 eV (PsbA1), 2.029 eV (PsbA2) and 2.027 eV (PsbA3). The NTO coefficients indicate that the lowest excited state (S_1) also possesses variable amounts of $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT character. The

S_2 energy for Chl$_{D1}$–Pheo$_{D1}$ in PsbA3 is 2.427 eV, 2.392 eV in PsbA1 and 2.431 eV in PsbA2-PSII. The more important finding concerns the lowest Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ CT states for each variant which lies at 1.885 eV above the ground state for PsbA1-PSII, 1.932 eV for PsbA2 and 1.701 eV for PsbA3.

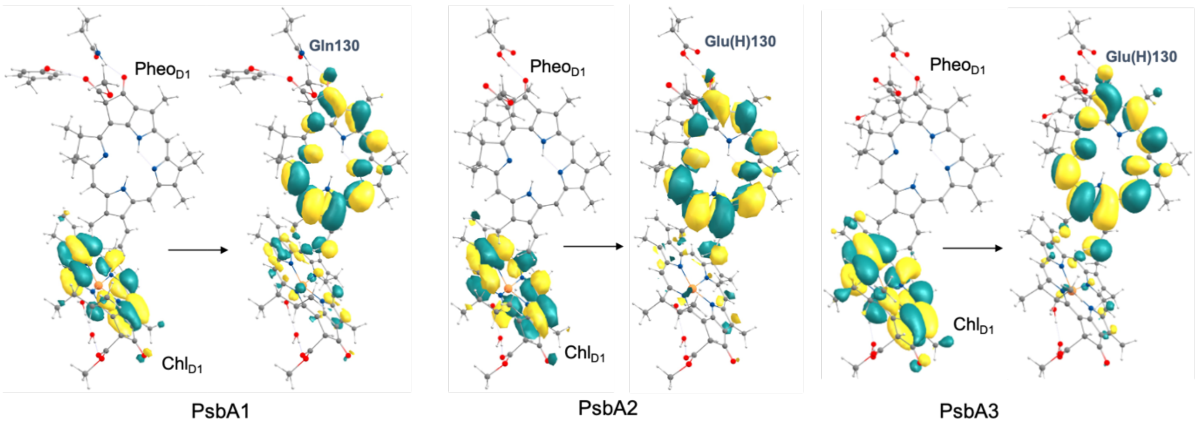
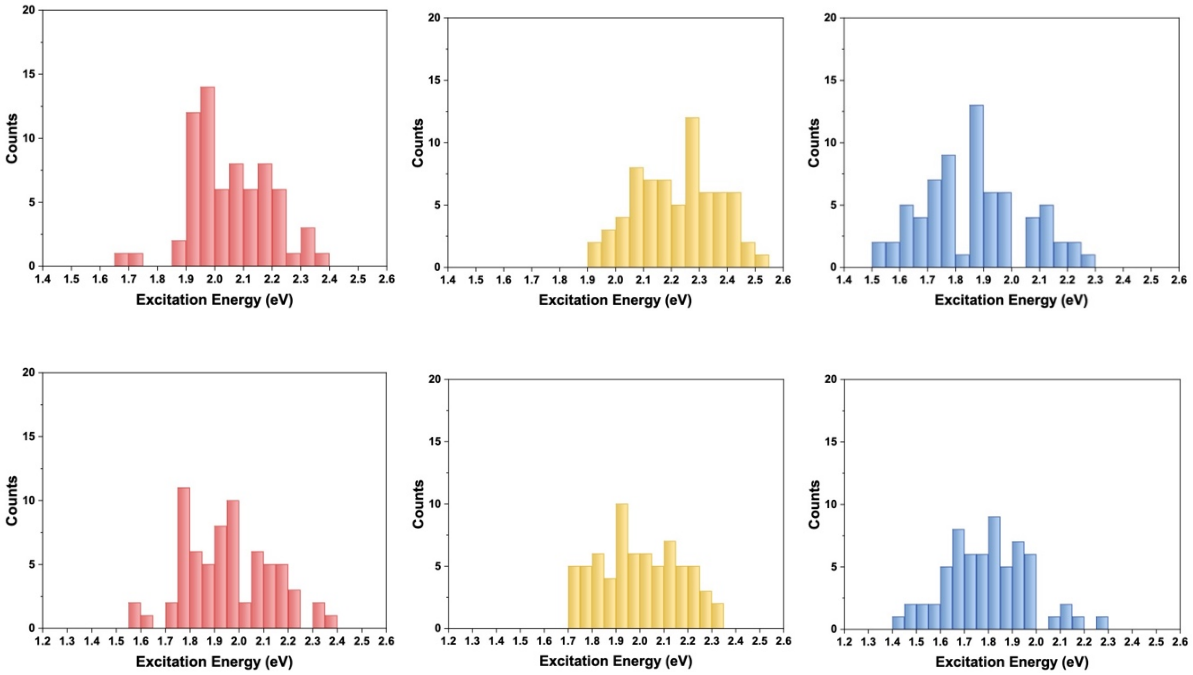


Fig. 5.7 The nature of the first root with significant CT character for the **Chl$_{D1}$–Pheo$_{D1}$** pairs in terms of Natural Transition Orbitals (NTOs) and relative contributions to a given excitation obtained from ωB97X-D3(BJ)/TD-DFT calculations, on the PSII monomer.

The excitation energies of the P$_{D1}$–P$_{D2}$ “special-pair” were also computed to determine if primary charge separation pathways differ among the D1 isoforms i.e., if P$_{D1}$ can act as the primary donor to yield P$_{D1}{}^{+}$P$_{D2}{}^{-}$ or P$_{D1}{}^{+}$Chl$_{D1}{}^{-}$ states. In the case of PsbA1-PSII, the Chl$_{D1}$ pathway is dominant with Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ being the lowest CT state. By contrast, the excitation profile of P$_{D1}$–P$_{D2}$ mainly consists of superpositions of individual local excitations while the lowest P$_{D1}{}^{\delta+}$P$_{D2}{}^{\delta-}$ CT state is as high as 3 eV. Current results are consistent with the findings on PsbA1 and provide a similar picture for the P$_{D1}$–P$_{D2}$ pair in PsbA2 and PsbA3-PSII (**Table 5.3**). The lowest P$_{D1}{}^{\delta+}$P$_{D2}{}^{\delta-}$ CT state lies at 3.03 eV for PsbA2 and 2.87 eV for PsbA3. The only difference is in the extent of mixing of LE character in the two lowest excited states. Unlike PsbA1, where the S_1 state has an almost equal contribution from P$_{D1}$ and P$_{D2}$, the S_1 state in PsbA2 and A3 shows a dominant contribution from P$_{D2}$ i.e., less mixing of the Q_y transitions. Nevertheless, our results rule out the possibility of a low-energy CT state within the P$_{D1}$–P$_{D2}$ pair and infer that the PsbA substitutions do not change the primary charge separation pathways within the PSII-RC.

The results remain robust when considering the dynamic evolution of the protein. TD-DFT calculations were performed on optimized QM/MM structures of the Chl$_{D1}$–Pheo$_{D1}$ pair on an ensemble of 65 distinct snapshots, chosen from the MD simulations of the PSII-membrane complex in each case. This provides the advantage that the excited states are computed on uncorrelated protein configurations which are not too close to the crystal structure minima but are properly hydrated and equilibrated with the protein and lipid bilayer at the same time. The relative trends in excited state energetics for each variant are summarized in **Fig. 5.8.**



(a) (b) (c)

Fig. 5.8 Relative distribution of the lowest excited state with dominant $\mathbf{Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}}$ CT character in the RC of **(a)** PsbA1, **(b)** PsbA2 and **(c)** PsbA3 variants. The QM-MM/TDDFT excited state calculations are performed on 65 independent protein configurations obtained from two sets of MD trajectories.

Focusing on the Chl_{D1}–$Pheo_{D1}$ dimer, we find that the lowest excited state (LE) of the RC remains on Chl_{D1} irrespective of protein dynamics. The overall trends indicate that the $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT states are slightly more stable i.e., red-shifted in PsbA3 (1.50–2.27 eV), compared to that in PsbA1 (1.66–2.35 eV) and PsbA2 (1.93–2.50 eV). Moreover, the probability of the $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT being dominant in the lowest excited state (S_1) of the dimer, was much higher in the case of PsbA3 (~45% snapshots) than for PsbA1 (~25%).

5.3.5 Electrostatic effects of D1 substitutions

The residue at D1-130 has been a major target for mutagenesis and the Q130E mutation in PsbA3 is reported to shift E_m ($Pheo_{D1}^{-}/Pheo_{D1}$) by 30–35 meV and partially modulate the energy of the $Chl_{D1}^{+}Pheo_{D1}^{-}$ radical pair.[74] Similar findings were observed in D1-Q130E mutants of *Synechocystis* PCC 6803,[75-77] and the corresponding D1-E130Q substitution in *Chlamydomonas reinhardtii*.[76, 78] On the other hand, thermoluminescence and fluorescence studies of *T. elongatus* with PsbA1 and PsbA3 PSII showed that multiple amino acid substitutions had significantly less impact on $S_2Q_A^-$ charge recombination, compared to a single mutation at D1-130.[77, 79, 80] This raises two important questions: (i) How does a single mutation at D1-130 affect the optical properties of the Chl_{D1}–$Pheo_{D1}$ pair? and (ii) Do the other D1 substitutions have compensatory effects and if so, which residues are responsible?

The first objective led us to construct independent QM/MM models for three "mutants" i.e., PsbA1-Q130E, PsbA2/PsbA3-E130Q, where only the D1-130 residue (Gln/Glu) was substituted keeping the rest of the PSII protein in the respective WT amino acid composition. Interestingly, our calculations on the D1-130 mutants depict that the effect of the point mutation differs for each D1 isoform (**Fig. 5.13**). In PsbA1-Q130E, the local excitations on Chl_{D1}–$Pheo_{D1}$ are unaffected, but the lowest $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT state is blue-shifted by 0.024 eV (194 cm^{-1}) compared to the WT. In PsbA2 the CT state remains invariant to the E130Q mutation, whereas in PsbA3-PSII the same mutation blue-shifts the CT states. Furthermore, identifying the other key D1 residues affecting electrochromic shifts in the optical absorption spectra of Chl_{D1}–$Pheo_{D1}$, especially spectral tuning of the $Chl_{D1}^{\delta+}Pheo_{D1}^{\delta-}$ CT states, is crucial. Here, we selected all naturally occurring mutations close to the Chl_{D1}–$Pheo_{D1}$ pair (10 Å) involving a substitution to a polar residue in PsbA2/A3 compared to PsbA1 (see **Table 5.1**). The substitutions investigated are: D1-130, D1-144, D1-147, D1-153, D1-158, D1-172 and D1-212.

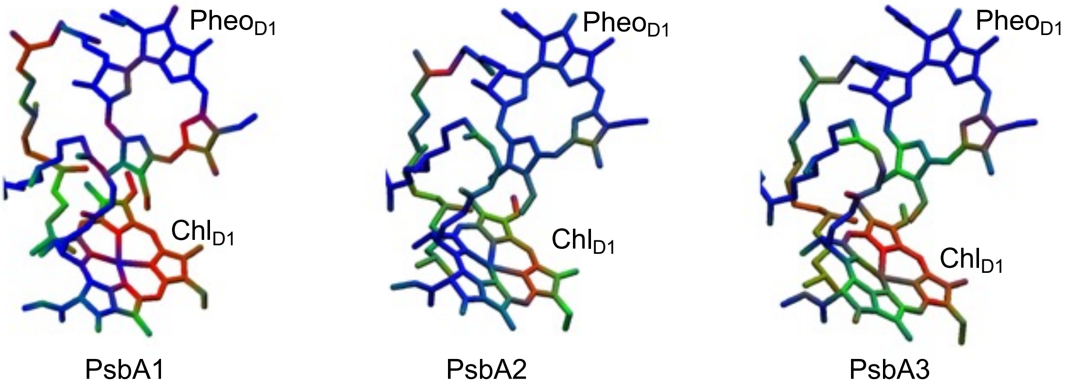


Fig. 5.9. Electrostatic potential (ESP) experienced by the **Chl_{D1}–$Pheo_{D1}$** pigment pair (in kT/e) inside the PSII protein for each variant. The calculations were performed using the APBS (Adaptive Poisson–Boltzmann Solver) in VMD on the equilibrated protein configuration in each case (shown for **snapshot 1**).

The contribution of each residue towards the overall electrochromic shifts for the Chl_{D1}–$Pheo_{D1}$ pair is illustrated in **Fig. 5.10**. In PsbA1-PSII the exciton on Chl_{D1} is most red-shifted by S153 (-97 cm^{-1}), M172 (-73 cm^{-1}) and F158 (-40 cm^{-1}) while $Pheo_{D1}$ is blue-shifted by Y147 (105 cm^{-1}) and Q130 (97 cm^{-1}). Interestingly, both Q130 and S153 stabilize the Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ states to a large extent, with minor contributions from Y147. D1-M172 and F158 are conserved in PsbA3 but substituted by L172 and L158 respectively in PsbA2 (**Table 5.1**). In PsbA2-PSII, A153 lowers the Chl$_{D1}$ excitation energy (-64 cm^{-1}), while E130 (210 cm^{-1}) and F147 (73 cm^{-1}) blue-shifts Pheo$_{D1}$. These findings demonstrate that even a subtle change in the H-bonding strength and hydrophobic environment in the D1-Q130E and D1-Y147F substitutions, directly influence the optical properties of Pheo$_{D1}$ in PsbA2 compared to PsbA1.

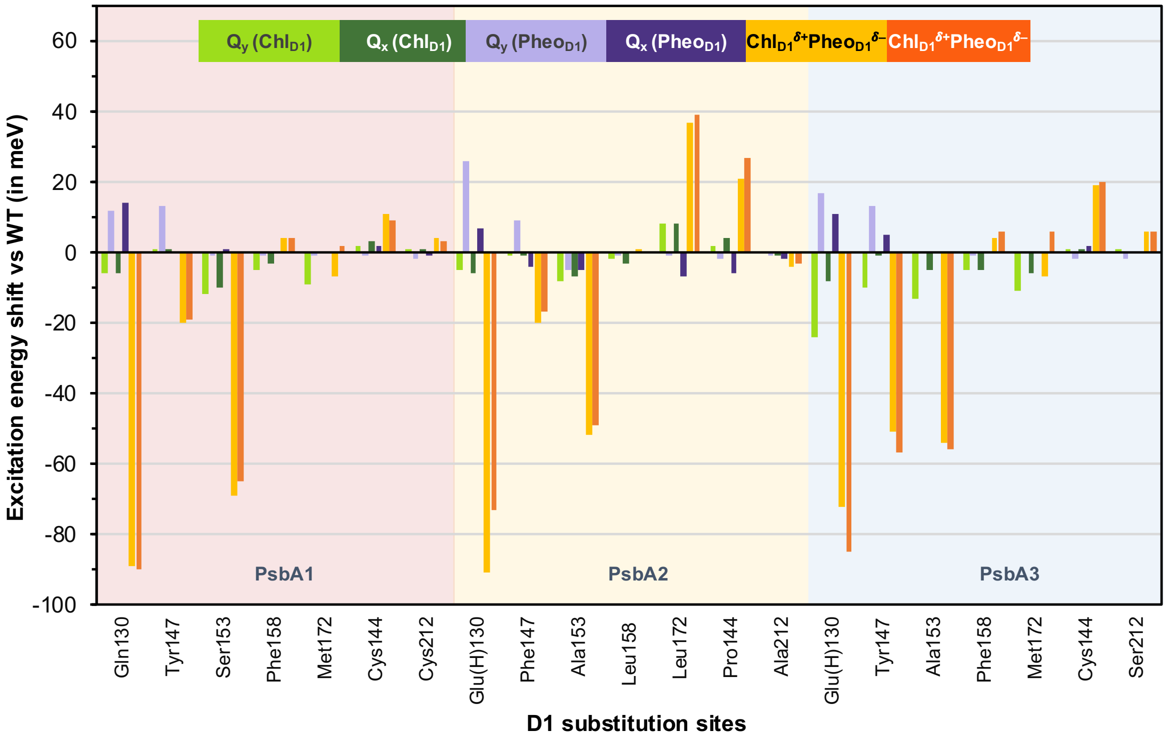


Fig. 5.10 Contributions of each D1 substitution in PsbA1–A2–A3 towards the overall excitation energy shifts for the **Chl$_{D1}$–Pheo$_{D1}$** pair in each variant. Negative values indicate a red-shift; positive values indicate a blue-shift in comparison to the WT protein. The shifts are computed using TD-DFT/MM on the equilibrated protein configuration (**snapshot 1**) in each case.

The more pronounced effect is seen in the Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ CT states, where all three residues E130, F147 and A153 red shift the excitation energy (**Fig. 5.10**). Consequently, one would expect that the Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ CT state in PsbA2 would be stabilized to a similar extent as that of PsbA1. However, our results indicate that the Chl$_{D1}{}^{\delta+}$Pheo$_{D1}{}^{\delta-}$ CT state in PsbA2 (1.932 eV) is instead blue-shifted compared to PsbA1 (1.885 eV) because of the additional D1-M172L and D1-C144P substitutions close to Chl$_{D1}$. The influence of the D1 substitutions is most pronounced in PsbA3-PSII where A153 and M172 is found to red-shift the energy of Chl$_{D1}$ whereas E130, Y147 both red-shifts Chl$_{D1}$ and blue shifts Pheo$_{D1}$. The energetics of CT states are also drastically affected by the PsbA1/A3 substitutions, with E130 contributing the most (581 cm^{-1}) followed by A153 (436 cm^{-1}) and Y147 (411 cm^{-1}).

The overall electrostatic potential (ESP) experienced by the Chl_{D1}–Pheo_{D1} pair in the D1 matrix are also distinct in each variant (**Fig. 5.9)**. Pheo_{D1} reside in a relatively more positive ESP pocket in PsbA2/A3 whereas the Chl_{D1} pocket in PsbA1/A3 is more electronegative. Therefore, the native protein matrix electrostatics in PsbA3 favor the formation of $\text{Chl}_{\text{D1}}{}^{\delta+}\text{Pheo}_{\text{D1}}{}^{\delta-}$ state more than PsbA1/A2.

5.4 Discussion

5.4.1 Structural changes around Chl_{D1}–Pheo_{D1}

The D1-Q130E substitution in cyanobacterial PSII-RC has been extensively discussed in literature. [30, 37, 39, 69] Site-directed mutagenesis and recent crystallographic studies showed that the D1-130 residue (Q130E) is 2.7–2.8 Å from the 13^1 C=O group of Pheo_{D1}, indicating a direct non-covalent interaction with Pheo_{D1}. This important structural difference is due to the change in the H-bonding environment of Pheo_{D1} caused by the substitution at D1-130. Despite varying H-bonding partners, the interaction between Pheo_{D1} and the PSII protein matrix at this position is conserved across all variants. However, our simulations raise a critical question regarding the protonation of E130 in PsbA2 and PsbA3-PSII. Based on the time evolution (**Fig. 5.5**) of H-bonding distances between D1-130 and the 13^1-keto group of Pheo_{D1}, we find that the H-bond between Pheo_{D1} and D1-130 is shorter in PsbA2- and PsbA3- PSII due to the Q130E change, requiring E130 to be protonated as glutamic acid.

Mutagenesis studies employing Raman,[81] EPR[82] and FT-IR spectroscopy[30, 69] proposed that the E130 side chain acts as a H-bond donor supporting the protonation of D1-E130. In our models where the E130 in PsbA2/A3 is deprotonated, the distance between D1-E130 and Pheo_{D1} fluctuates significantly within our MD timescales (**Fig. 5.11**) as the carboxylate group's flexibility does not facilitate a H-bond with Pheo_{D1}. A charged residue at this site is unexpected since lipid bilayer regions are typically hydrophobic. However, our MD analysis near Pheo_{D1} and deprotonated E130 did not reveal dynamic H-bonding networks or water channels, which makes an external protonation pathway to E130 unlikely. In addition, the number of water molecules and H-bonds within a 5 Å radius remains conserved, suggesting that E130 is unlikely to be protonated externally through other titratable amino acids.

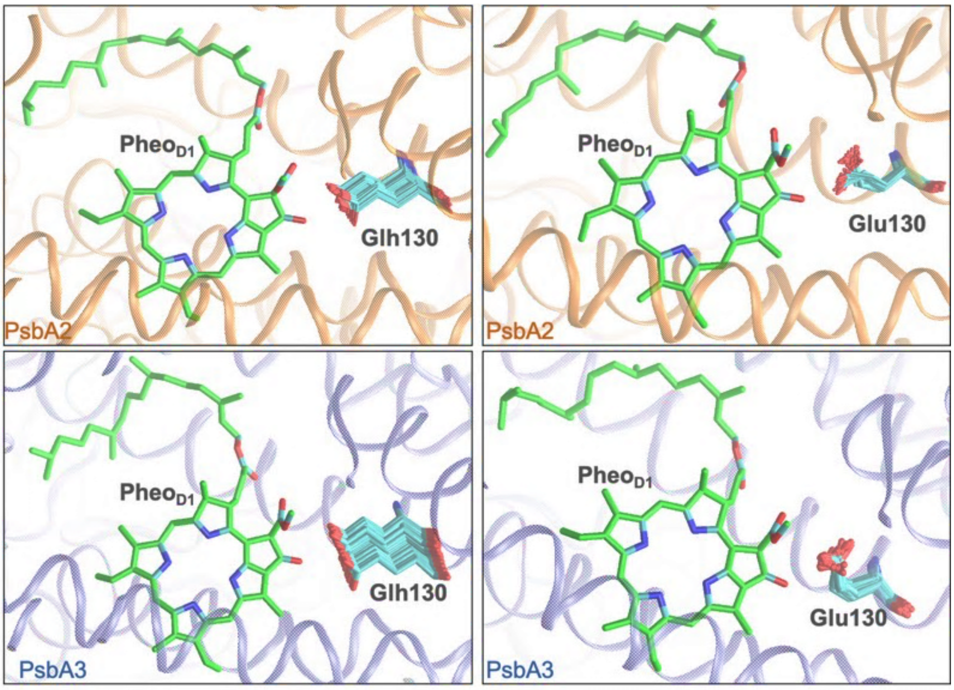


Fig. 5.11 The effect of protonation of D1-E130 on the orientation of the Glu/Glu(H) side chain during the production molecular dynamics simulations of PsbA2 and PsbA3-PSII. The conformations are averaged over 300 frames across 60 ns MD trajectories for each setup.

Although the major PsbA substitutions are in the vicinity of the RC pigments, there are a few notable differences closer to the OEC as well as to Q_B. The Cl-1 water channel leading to the Mn4 atom of the OEC begins at the D1-PsbO interface and includes several conserved D1 residues (Y73, E65, D59, R334, D61 and N181). Recent crystallographic models and kinetic measurements[38] suggest that the D1-P173M mutation led to a narrower Cl-1 channel in PsbA2 and disappearance of 2 waters near N181 possibly leading to slow proton egress during the S_1-S_2-S_3 transition in PsbA2-PSII. However, we did not observe any loss of crystal waters within the timescales of our simulations.

5.4.2 Influence on Redox Properties

Spectro-electrochemical measurements by Sugiura et al.[30] reported a more positive E_m of $Pheo_{D1}$ in PsbA3-PSII (−505 mV) compared to PsbA1 (−522 mV). This finding was partly attributed to the stronger H-bond between $Pheo_{D1}$–E130 in PsbA3 compared to $Pheo_{D1}$–Q130 in PsbA1 (**Fig. 5.5**). FTIR spectroscopy and DFT analysis also indicated shifts in E_m ($Pheo_{D1}^-$/$Pheo_{D1}$) upon modification of a H-bond donor in PsbA1 and A3.[69] However, the computational analysis relied upon implicitly solvated models of $Pheo_{D1}$ without explicitly accounting for the native protein environment. The E_m ($Pheo_{D1}^-$/$Pheo_{D1}$) of PsbA2-PSII remained unknown until a recent study by Boussac and co-workers reported it to be lower by ~30 mV in PsbA2 (−535 mV) when compared to PsbA3. This was estimated by the energy gap between $S_2Q_A^-$ and $P_{680}^+Pheo_{D1}^-$ [38] based on thermoluminescence and kinetics of proton release. All these studies suggested that $Pheo_{D1}$ has a higher reduction potential in PsbA3 than PsbA1/A2. Our trends in EA ($Pheo_{D1}^-/Pheo_{D1}$) align with the relative energy differences between PsbA1/A3,[30, 38, 39, 72] with $Pheo_{D1}$ in A3 having the highest EA but the trends in PsbA2 is not clearly distinguishable from A1. This indicates that the multiple residue substitutions affect EA ($Pheo_{D1}^-/Pheo_{D1}$) differently in PsbA3 and A2, despite both having the same D1-130E substitution. It is important to note that the exact EA differences cannot be directly compared with previously reported redox potential differences due to limitations in both experimental and theoretical methods. For instance, spectroscopic measurements of E_m ($Pheo_{D1}^-/Pheo_{D1}$) are often performed in dithionite treated PSII samples, possibly not reflecting the physiological state of the RC as Q_A is reduced to Q_A^- ("closed" RC). [30, 72] The reported values thus assume a consistent electrostatic effect of Q_A^- (c550 bandshift) across all PsbA variants but the shifts induced by Q_A^- may be significant towards the absolute differences of E_m ($Pheo_{D1}^-/Pheo_{D1}$). Additionally, protein matrix dynamics can induce redox potential differences similar to local electrostatic effects. The observed energy differences (0.017–0.035 eV) are also within the likely limitations of a QM approach, and extensive sampling would be needed to reduce the spread of values, which remains a significant challenge for computational studies of pigment assemblies. Overall, our results indicate that the PsbA substitutions adjust the redox properties of both pigments involved in primary charge separation to conserve the primary ET pathway.

5.4.3 Spectral tuning of Chl_{D1}–$Pheo_{D1}$ excited states

The electrostatic effects of the D1 variants are not only important to the intrinsic absorption properties of the pigments but also towards pushing the red limit of PSII. The specific D1-Q130E substitution exists in most cyanobacterial PsbA variants, especially those acclimated to high-light conditions (PsbA3). Interestingly, the exact Gln/Glu substitution is observed in FarLiP cyanobacteria[83] and some higher plants.[27] Consequently, specific variants of core PSII proteins available to different organisms are utilized to adjust photosynthetic efficiency in response to environmental conditions not only by presenting alternate localized electrostatic contributors to critical pigments like in the case of *psbA* variants, but also by adopting different conformations within the same variant. The present results demonstrate how the local electrostatic environment and conformational changes in the D1 protein matrix are responsible for enabling access to low-energy CT states in all D1 variants. Importantly, the shifts in $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ CT states and primary charge separation cannot be explained by a limited number of substitutions but result from global electrostatic optimization.

Mutagenesis and spectroscopic studies[40, 84] reported that $\mathrm{Q_A}$ reduction induces a red shift (~3.0 nm) in the $\mathrm{Pheo_{D1}}$ Q_x band (C550 bandshift) for PsbA3-PSII (547.3 nm), relative to PsbA1 (544.3 nm). This shift is attributed to the stronger H-bond to the keto of the $\mathrm{Pheo_{D1}}$ from the carboxylate group of D1-130 in PsbA3-PSII than that of PsbA1-PSII. The bandshift for PsbA2-PSII was proposed to be similar to that in PsbA3-PSII assuming that the same residue (D1-130) is responsible for the spectral shift in both proteins.[38, 39] Studies on *Synechococcus* PCC7942 and *Synechocystis* PCC6803 have reported a 25% increase in the quantum yield of primary CS with the high light D1 isoform (PsbA3) compared to the low light isoform (PsbA1), as well as with the Q130E mutation in PCC6803. (71-73) These findings align with previous studies indicating higher CS yields in WT-PsbA3 than E130Q mutants in *Chlamydomonas reinhardtii*.[76, 78] . Our results are consistent with the above experimental observations, but we attribute the differences to a combined effect of both redox and spectral tuning of the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair by D1 electrostatics. Firstly, we showed that $\mathrm{Pheo_{D1}}$ is more easily reduced (high EA) in PsbA3 than PsbA1/A2 (**Fig. 5.6a**), Secondly, the protein matrix in PsbA3 shifts $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ states towards longer wavelengths (**Fig. 5.8**). This suggests that the PsbA3 protein tunes the EA of $\mathrm{Pheo_{D1}}$ to enhance the forward ET and increase CS. These results are physiologically significant because not only PsbA3 exhibit increased photo-tolerance in cyanobacteria but also its PsbA homolog is most prevalent in eukaryotic PSII.[27] While the connection between photoprotection mechanisms and primary CT energetics cannot be directly inferred here, it is hypothesized that unlike PsbA1/A2, the more stable $[\mathrm{P_{680}}^{+}\mathrm{Pheo_{D1}}^{-}]$ in PsbA3 likely favors direct charge recombination over dissipation via the triplet route.[71]

In PsbA1/A2, our results do not show major differences in EA ($\mathrm{Pheo_{D1}}^{-}/\mathrm{Pheo_{D1}}$) but the IE ($\mathrm{Chl_{D1}}^{+}/\mathrm{Chl_{D1}}$) (**Fig. 5.6b**) and excitation profiles of the $\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$ pair vary. In PsbA2-PSII $\mathrm{Chl_{D1}}$ is a poor donor (higher IE) while $\mathrm{Pheo_{D1}}$ is a poorer acceptor (lower EA) **(Fig. 5.6a),** which blue-shifts the $\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$ CT states (**Fig. 5.8).** Interestingly, the D1 matrix in PsbA2 is found to *inherently disfavor* low-energy CT states at longer wavelengths, which could possibly lead to lower yields of primary CS within PsbA2 under normal light conditions compared to PsbA3 and PsbA1.This indicates that primary charge separation is more constrained in PsbA2 possibly leading to slower ET kinetics and $\mathrm{O_2}$ evolution. While there are no direct experimental observations on primary charge separation in PsbA2-PSII, some studies[38, 39] have reported a lower yield of O_2 production, linked to specific substitutions on the donor side that narrow the Cl-1 channel and slow Y_Z oxidation. Our findings offer a unique perspective indicating that the spectral-tuning of the primary CT state by the PsbA2 protein matrix could also limit the overall ET in PsbA2 variants. However, further spectroscopic and mutagenesis studies will be required to fully decipher the physiological significance of this finding.

5.5 Conclusion

We employ atomistic molecular dynamics combined with multiscale QM/MM calculations to investigate how PSII-PsbA genetic variants modulate property-function relationships in the reaction center pigments of thermophilic cyanobacteria. Our multiscale simulations and structural analysis reveal that the protonation of D1-E130 in the PsbA2 and A3 variants is key for determining the stability of the Pheo_{D1} acceptor, which in turn influences primary charge-separation. Specifically, the hydrogen bond between Pheo_{D1} and D1-130 fine-tunes the optical and redox properties of the Chl_{D1}–Pheo_{D1} pair across all three variants. Importantly, we demonstrate that the spectral tuning of the $\text{Chl}_{D1}{}^{\delta+}\text{Pheo}_{D1}{}^{\delta-}$ charge-transfer (CT) state is not due to a single substitution but rather a cumulative effect of all the multiple D1 substitutions. Our excited state calculations, on an ensemble of protein configurations, show that the PsbA3 variant red shifts while PsbA2 induces a blue shift in the lowest CT state, compared to PsbA1-PSII. We also identified the electrostatic effects of specific substitutions, suggesting potential targets for future mutagenesis experiments towards spectral tuning of the $\text{Chl}_{D1}{}^{\delta+}\text{Pheo}_{D1}{}^{\delta-}$ CT state. Overall, we conclude that the combined effect of both redox and spectral tuning of the Chl_{D1}–Pheo_{D1} pair by D1 electrostatics have functional implications for primary charge separation in each variant. This work illustrates the core principles of how protein electrostatics influence excited state properties, laying a foundation for engineered strategies aimed at precise tuning and functional optimization of photosynthetic systems.

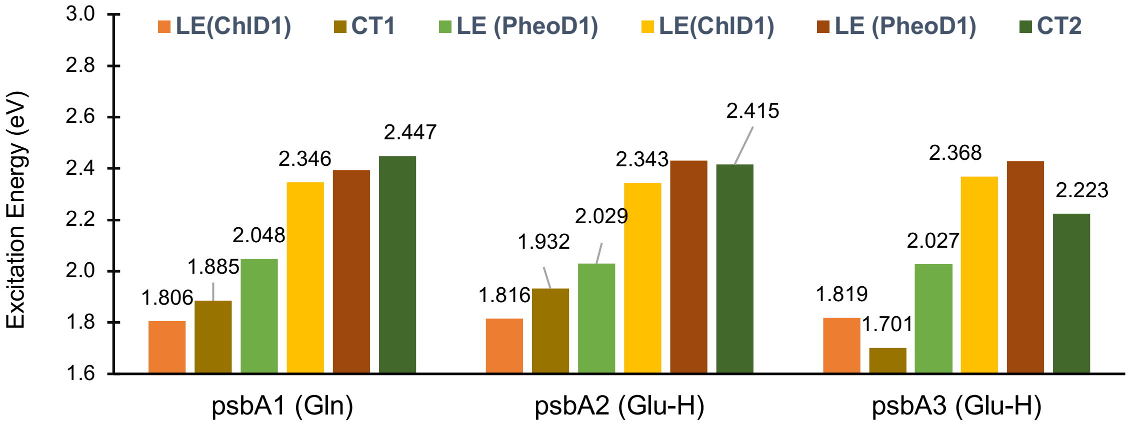
5.6 Supplementary Tables and Figures

Table 5.2 Vertical excitation energies (VEE) of **$\mathrm{Chl_{D1}}$–$\mathrm{Pheo_{D1}}$** in each PsbA variant, computed on the QM/MM optimized geometries of the pair embedded within the PSII matrix for "crystal-like" snapshot 1 in each case. E_S represents the absolute VEE while ΔE_S represent the shifts in excitation energy with respect to WT PsbA1 variant. All energy values are in eV.

NTO character	**PsbA1 (Gln)**	**PsbA2 (Glu)**		**PsbA3 (Glu)**		**PsbA2 (Glu-H)**		**PsbA3 (Glu-H)**	
	E_S	E_S	ΔE_S	E_S	ΔE_S	E_S	ΔE_S	E_S	ΔE_S
LE ($\mathrm{Chl_{D1}}$)	1.806	1.877	0.071	1.853	0.047	1.816	0.010	1.819	0.013
CT ($\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$)	1.885	2.735	0.850	2.701	0.816	1.932	0.047	1.701	-0.184
LE ($\mathrm{Pheo_{D1}}$)	2.048	1.904	-0.144	1.889	-0.159	2.029	-0.019	2.027	-0.021
LE ($\mathrm{Chl_{D1}}$)	2.346	2.360	0.014	2.343	-0.003	2.343	-0.003	2.368	0.022
LE ($\mathrm{Pheo_{D1}}$)	2.392	2.412	0.020	2.388	-0.004	2.431	0.039	2.427	0.035
CT ($\mathrm{Chl_{D1}}^{\delta+}\mathrm{Pheo_{D1}}^{\delta-}$)	2.447	3.155	0.708	3.140	0.693	2.415	-0.032	2.223	-0.224

Table 5.3 Vertical excitation energies (VEE) of **$\mathrm{P_{D1}}$–$\mathrm{P_{D2}}$** pair in each PsbA variant, computed on the QM/MM optimized geometries of the pair embedded within the PSII matrix for "crystal-like" snapshot 1 in each case. E_S represents the absolute VEE while ΔE_S represent the shifts in excitation energy with respect to WT PsbA1. All energy values are in eV.

State	**NTO character**	**PsbA1 (Gln)**	**PsbA2 (GluH)**		**PsbA3 (GluH)**	
		E_S	E_S	ΔE_S	E_S	ΔE_S
1	LE ($\mathrm{P_{D2}}$) + LE ($\mathrm{P_{D1}}$)	1.864	1.859	-0.005	1.868	0.004
2	LE ($\mathrm{P_{D1}}$) + LE ($\mathrm{P_{D2}}$)	1.884	1.891	0.007	1.897	0.013
3	LE ($\mathrm{P_{D2}}$)	2.403	2.404	0.001	2.426	0.023
4	LE ($\mathrm{P_{D1}}$)	2.422	2.434	0.012	2.442	0.020
5	CT ($\mathrm{P_{D1}}^{\delta+}\mathrm{P_{D2}}^{\delta-}$)	2.911	3.031	0.120	2.870	-0.041



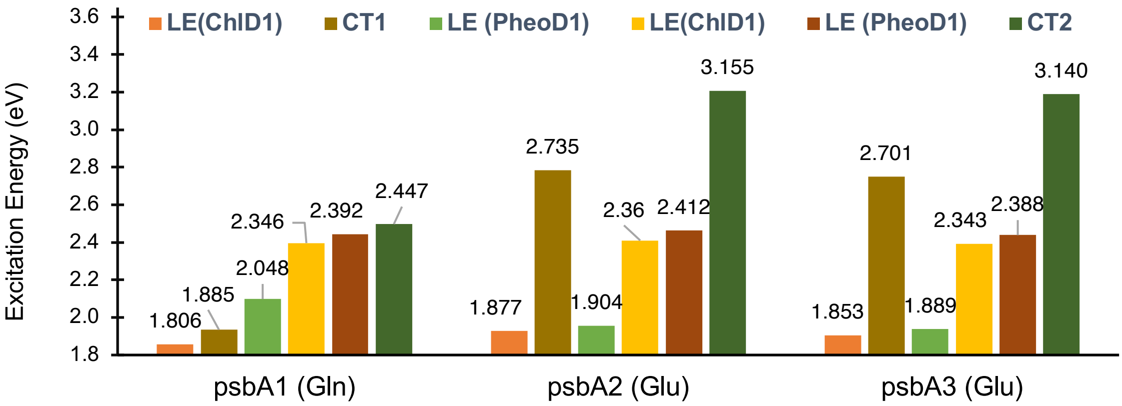


Fig. 5.12 Comparison of excited state energies on the 'crystal-like' (**snapshot 1**) with D1-Gln130 for PsbA1, **(a)** D1-GluH130 (uncharged/protonated) and **(b)** D1-Glu130 (charged/deprotonated) for the PsbA2 and PsbA3-PSII variants.

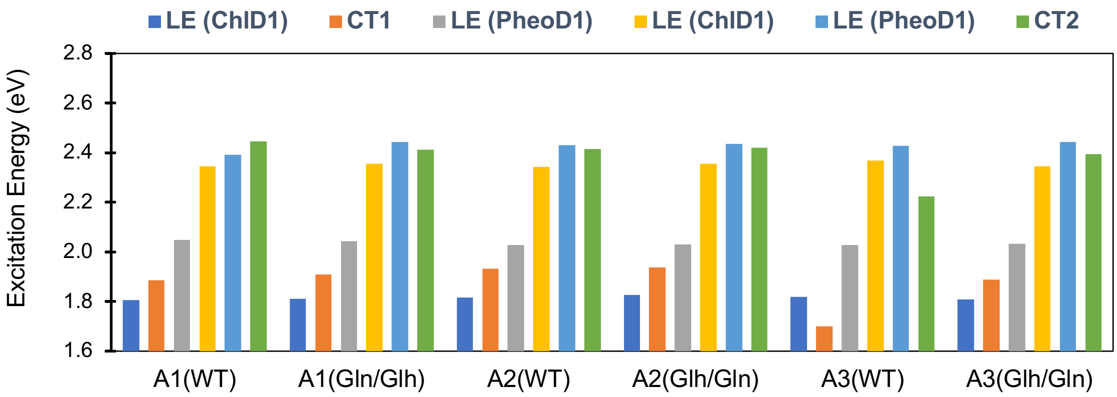


Fig. 5.13 Comparison of excited state energies on 'crystal-like' (**snapshot 1**) for D1-130 (Gln/Glu) mutants of PsbA1 and Glu/Gln mutants in PsbA2 and PsbA3-PSII, with respect to the wild-type (WT) D1 isoforms in each variant.

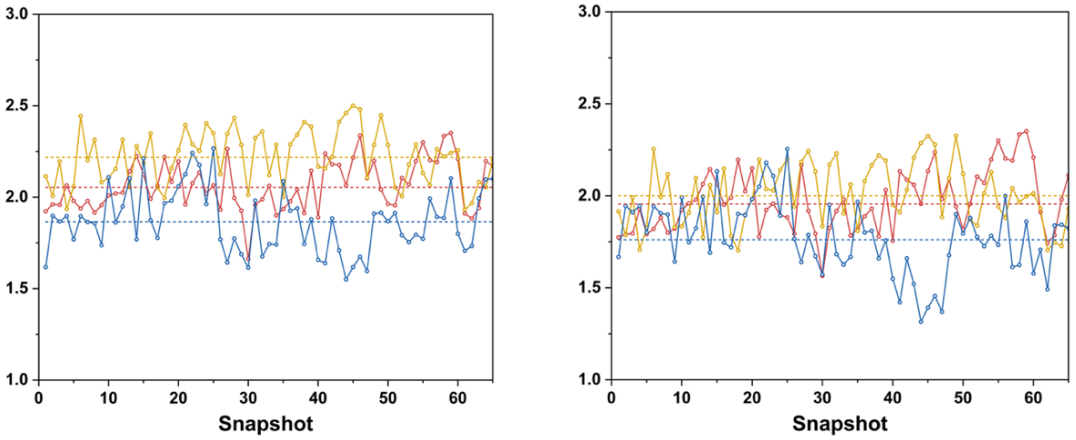


Fig. 5.14 Excitation energies (in eV) of the lowest root with dominant $\mathbf{Chl_{D1}}^{\delta+}\mathbf{Pheo_{D1}}^{\delta-}$ CT character in the RC of PsbA1 (in red), A2 (yellow) and A3 (blue) variants, based on QM-MM/TDDFT calculations are performed on 65 independent protein configurations obtained from two sets of MD trajectories.

Table 5.4 Local Energy Decomposition of the **Pheo$_{\mathbf{D1}}$** binding energy (in kcal mol^{-1}) in terms of vdW, electrostatic, non-polar, dispersion and solvation terms computed using Generalized Born (**MM-GBSA**) and Poisson Boltzmann (**MM-PBSA**) approaches respectively. The binding energies were computed using the first 60 ns of production MD over an ensemble of 300 equidistant snapshots.

GB Energy term	**PsbA1**	**PsbA2-Glu(H)**	**PsbA2-Glu**	**PsbA3-Glu(H)**	**PsbA3-Glu**
vdW	-94.9914	-95.2722	-95.8286	-95.7121	-94.7087
E_{EL}	-12.4012	-9.7478	-2.2946	-12.197	-4.1706
E_{PB}	25.0377	23.2989	23.7287	25.7132	23.8753
E_{NPOLAR}	-67.459	-66.9835	-66.989	-67.6079	-66.6251
E_{DISPER}	114.9312	114.3744	114.5556	114.9524	114.5588
ΔG_{gas}	-107.3926	-105.02	-98.1233	-107.9091	-98.8793
ΔG_{solv}	72.5099	70.6898	71.2954	73.0577	71.809
$\boldsymbol{\Delta G_{\mathrm{total}}}$	-34.8827	-34.3302	-26.8279	-34.8514	-27.0703
PB Energy term	**PsbA1**	**PsbA2-Glu(H)**	**PsbA2-Glu**	**PsbA3-Glu(H)**	**PsbA3-Glu**
vdW	-94.9914	-95.2722	-95.8286	-95.7121	-94.7087
E_{EL}	-24.8023	-19.4955	-4.5893	-24.394	-8.3411
E_{GB}	49.594	45.272	33.4546	49.8501	37.4964
E_{SURF}	-11.5603	-11.4077	-11.4026	-11.5551	-11.4187
ΔG_{gas}	-119.7937	-114.7677	-100.4179	-120.1061	-103.0499
ΔG_{solv}	38.0338	33.8642	22.0521	38.295	26.0777
$\boldsymbol{\Delta G_{\mathrm{total}}}$	-81.7599	-80.9035	-78.3658	-81.8111	-76.9721

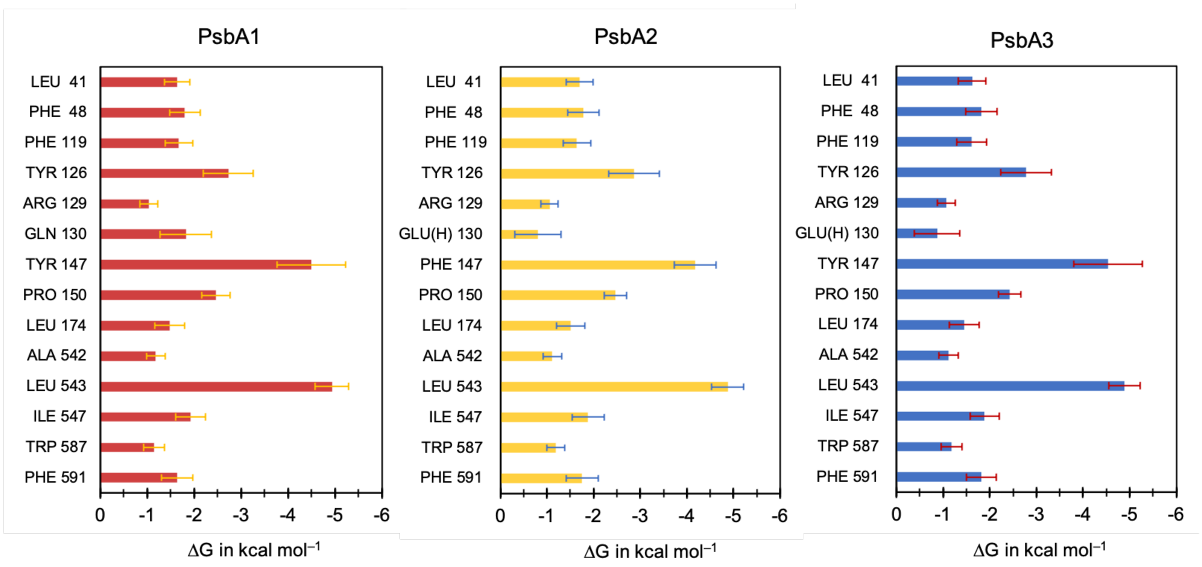


Fig. 5.15 The total contribution of selected residues to the Pheo$_{\mathrm{D1}}$ binding energy calculated with the MM-PBSA approach, for PsbA1–A2-A3. Residue numbers 1–344 belong to the PsbA (D1) protein, while the rest belong to PsbD (D2) in our setup.

5.7 REFERENCES

1. J.-R. Shen, *Annu. Rev. Plant Biol.*, 2015, **66**, 23-48.
2. V. Krewald, M. Retegan and D. A. Pantazis, Springer International Publishing, 2015, DOI: 10.1007/128_2015_645, pp. 23-48.
3. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*, Wiley, Chichester, 2nd edn., 2014.
4. F. Muh, C. Glockner, J. Hellmich and A. Zouni, *Biochim. Biophys. Acta*, 2012, **1817**, 44-65.
5. N. Nelson and C. F. Yocum, *Annu. Rev. Plant Biol.*, 2006, **57**, 521-565.
6. K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber and S. Iwata, *Science*, 2004, **303**, 1831-1838.
7. D. J. Vinyard, G. M. Ananyev and G. Charles Dismukes, *Annu. Rev. Biochem.*, 2013, **82**, 577-606.
8. C. Pagliano, G. Saracco and J. Barber, *Photosynth. Res.*, 2013, **116**, 167-188.
9. T. M. Bricker, J. L. Roose, R. D. Fagerlund, L. K. Frankel and J. J. Eaton-Rye, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 121-142.
10. Y. Umena, K. Kawakami, J.-R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
11. A. Guskov, J. Kern, A. Gabdulkhakov, M. Broser, A. Zouni and W. Saenger, *Nature Structural & Molecular Biology*, 2009, **16**, 334-342.
12. I. Enami, A. Okumura, R. Nagao, T. Suzuki, M. Iwai and J.-R. Shen, *Photosynth. Res.*, 2008, **98**, 349-363.
13. T. M. Bricker and L. K. Frankel, *Photosynth. Res.*, 2002, **72**, 131-146.
14. H. A. Chu and Y. F. Chiu, *Front Plant Sci*, 2015, **6**, 1261.
15. L. K. Thompson, A. F. Miller, C. A. Buser, J. C. De Paula and G. W. Brudvig, *Biochemistry*, 1989, **28**, 8048-8056.
16. N. Cox, D. A. Pantazis and W. Lubitz, *Annu. Rev. Biochem.*, 2020, **89**, 795-820.
17. Y. Yoneda, E. A. Arsenault, S. Yang, Jr., K. Orcutt, M. Iwai and G. R. Fleming, *Nat. Commun.*, 2022, **13**, 2275.
18. S. Yang, Jr., E. A. Arsenault, K. Orcutt, M. Iwai, Y. Yoneda and G. R. Fleming, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**.
19. D. I. G. Bennett, K. Amarnath and G. R. Fleming, *J. Am. Chem. Soc.*, 2013, **135**, 9164-9173.
20. Y. Shibata, S. Nishi, K. Kawakami, J.-R. Shen and T. Renger, *J. Am. Chem. Soc.*, 2013, **135**, 6903-6914.
21. M. Gorka, A. Baldansuren, A. Malnati, E. Gruszecki, J. H. Golbeck and K. V. Lakshmi, *Frontiers in Microbiology*, 2021, **12**, 735666.
22. V. I. Novoderezhkin, E. G. Andrizhiyevskaya, J. P. Dekker and R. Van Grondelle, *Biophys. J.*, 2005, **89**, 1464-1481.
23. P. Mulo, C. Sicora and E.-M. Aro, *Cell. Mol. Life Sci.*, 2009, **66**, 3697-3710.
24. T. Noguchi, *Plant Cell Physiol.*, 2002, **43**, 1112-1116.
25. I. Vass, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 209-217.
26. E.-M. Aro, I. Virgin and B. Andersson, *Biochim. Biophys. Acta Bioenerg.*, 1993, **1143**, 113-134.
27. P. Mulo, I. Sakurai and E.-M. Aro, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 247-257.
28. Y. Nakajima, N. Ugai-Amo, N. Tone, A. Nakagawa, M. Iwai, M. Ikeuchi, . . . J.-R. Shen, *J. Biol. Chem.*, 2022, **298**, 102668.
29. É. Kiss, P. B. Kós, M. Chen and I. Vass, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 1083-1094.

30. M. Sugiura, Y. Kato, R. Takahashi, H. Suzuki, T. Watanabe, T. Noguchi, . . . A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 1491-1499.
31. B. Loll, M. Broser, P. B. Kós, J. Kern, J. Biesiadka, I. Vass, . . . A. Zouni, 2008, **389**, 609-617.
32. A. K. Clarke, A. Soitamo, P. Gustafsson and G. Oquist, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 9973-9977.
33. L. Garczarek, A. Dufresne, N. Blot, A. M. Cockshutt, A. Peyrat, D. A. Campbell, . . . C. Six, *The ISME Journal*, 2008, **2**, 937-953.
34. C. I. Sicora, S. E. Appleton, C. M. Brown, J. Chung, J. Chandler, A. M. Cockshutt, . . . D. A. Campbell, *Biochim. Biophys. Acta Bioenerg.*, 2006, **1757**, 47-56.
35. M. Tichý, L. Lupínková, C. Sicora, I. Vass, S. Kuviková, O. Prášil and J. Komenda, *Biochim. Biophys. Acta Bioenerg.*, 2003, **1605**, 55-66.
36. P. B. Kós, Z. Deák, O. Cheregi and I. Vass, *Biochim. Biophys. Acta Bioenerg.*, 2008, **1777**, 74-83.
37. J. Sander, M. Nowaczyk, J. Buchta, H. Dau, I. Vass, Z. Deák, . . . M. Rögner, *J. Biol. Chem.*, 2010, **285**, 29851-29856.
38. A. Boussac, J. Sellés and M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2023, **1864**, 148979.
39. M. Sugiura and A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2014, **1837**, 1427-1434.
40. J. L. Hughes, N. Cox, A. W. Rutherford, E. Krausz, T.-L. Lai, A. Boussac and M. Sugiura, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 11-19.
41. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Am. Chem. Soc.*, 2020, **142**, 18174-18190.
42. M. Capone, A. Sirohiwal, M. Aschi, D. A. Pantazis and I. Daidone, *Angew. Chem. Int. Ed.*, 2023, **62**, e202216276.
43. R. Takahashi, K. Hasegawa and T. Noguchi, *Biochemistry*, 2008, **47**, 6289-6291.
44. T. Noguchi, T. Tomo and Y. Inoue, *Biochemistry*, 1998, **37**, 13614-13625.
45. A. Sirohiwal and D. A. Pantazis, *Acc. Chem. Res.*, 2023, **56**, 2921-2932.
46. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
47. D. R. Roe and T. E. Cheatham, III, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
48. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
49. Y. Sugo, K. Saito and H. Ishikita, *Biochemistry*, 2022, **61**, 1836-1843.
50. K. Saito, K. Mitsuhashi, H. Tamura and H. Ishikita, *Biophys. J.*, 2023, **122**, 470-483.
51. C. Friedl, D. G. Fedorov and T. Renger, *Phys. Chem. Chem. Phys.*, 2022, **24**, 5014-5038.
52. Y. Kobori, N. Ponomarenko and J. R. Norris, *J. Phys. Chem. C*, 2015, **119**, 8078-8088.
53. M. E.-A. Madjet, F. Müh and T. Renger, *J. Phys. Chem. B*, 2009, **113**, 12603-12614.
54. V. I. Novoderezhkin, J. P. Dekker and R. van Grondelle, *Biophys. J.*, 2007, **93**, 1293-1311.
55. M. E. Madjet, A. Abdurahman and T. Renger, *J. Phys. Chem. B*, 2006, **110**, 17268-17281.
56. E. T. Johnson, F. Müh, E. Nabedryk, J. C. Williams, J. P. Allen, W. Lubitz, . . . W. W. Parson, *J. Phys. Chem. B*, 2002, **106**, 11859-11869.
57. S. Bhattacharjee, S. Arra, I. Daidone and D. A. Pantazis, *Chem. Sci.*, 2024, **15**, 7269-7284.
58. A. Sirohiwal, F. Neese and D. A. Pantazis, *Chem. Sci.*, 2021, **12**, 4463-4476.
59. S. Bhattacharjee, F. Neese and D. A. Pantazis, *Chem. Sci.*, 2023, **14**, 9503-9516.
60. E. Elias, K. Brache, J. Schäfers and R. Croce, *J. Am. Chem. Soc.*, 2024, **146**, 3508–3520.
61. E. Elias, T. J. Oliver and R. Croce, *Annu. Rev. Phys. Chem.*, 2024, **75**, 231-256

62. Y. Silori, R. Willow, H. H. Nguyen, G. Shen, Y. Song, C. J. Gisriel, . . . J. P. Ogilvie, *J. Phys. Chem. Lett.*, 2023, **14**, 10300-10308.
63. C. J. Gisriel, G. Shen, D. A. Flesher, V. Kurashov, J. H. Golbeck, G. W. Brudvig, . . . D. A. Bryant, *J. Biol. Chem.*, 2023, **299**, 102815.
64. C. J. Gisriel, G. Shen, M.-Y. Ho, V. Kurashov, D. A. Flesher, J. Wang, . . . D. A. Bryant, *J. Biol. Chem.*, 2022, **298**, 101424.
65. C. J. Gisriel, T. Cardona, D. A. Bryant and G. W. Brudvig, *Microorganisms*, 2022, **10**, 1270.
66. C. Gisriel, G. Shen, V. Kurashov, M.-Y. Ho, S. Zhang, D. Williams, . . . D. A. Bryant, *Science Advances*, 2020, **6**, eaay6415.
67. D. J. Nürnberg, J. Morton, S. Santabarbara, A. Telfer, P. Joliot, L. A. Antonaru, . . . A. W. Rutherford, *Science*, 2018, **360**, 1210-1213.
68. F. Mokvist, J. Sjöholm, F. Mamedov and S. Styring, *Biochemistry*, 2014, **53**, 4228-4238.
69. Y. Shibuya, R. Takahashi, T. Okubo, H. Suzuki, M. Sugiura and T. Noguchi, *Biochemistry*, 2010, **49**, 493-501.
70. A. R. Holzwarth, M. G. Muller, M. Reus, M. Nowaczyk, J. Sander and M. Rogner, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6895-6900.
71. A. W. Rutherford, A. Osyczka and F. Rappaport, *FEBS Lett.*, 2012, **586**, 603-616.
72. Y. Kato, M. Sugiura, A. Oda and T. Watanabe, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17365-17370.
73. R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775-4777.
74. Y. Kato, T. Shibamoto, S. Yamamoto, T. Watanabe, N. Ishida, M. Sugiura, . . . A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 1998-2004.
75. S. A. P. Merry, P. J. Nixon, L. M. C. Barter, M. Schilstra, G. Porter, J. Barber, . . . D. R. Klug, *Biochemistry*, 1998, **37**, 17439-17447.
76. F. Rappaport, M. Guergova-Kuras, P. J. Nixon, B. A. Diner and J. Lavergne, *Biochemistry*, 2002, **41**, 8518-8527.
77. K. Cser and I. Vass, *Biochim. Biophys. Acta Bioenerg.*, 2007, **1767**, 233-243.
78. A. Cuni, L. Xiong, R. Sayre, F. Rappaport and J. Lavergne, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4825-4831.
79. L. B. Giorgi, P. J. Nixon, S. A. P. Merry, D. M. Joseph, J. R. Durrant, J. D. L. Rivas, . . . D. R. Klug, *J. Biol. Chem.*, 1996, **271**, 2093-2101.
80. M. Sugiura, C. Azami, K. Koyama, A. W. Rutherford, F. Rappaport and A. Boussac, *Biochim. Biophys. Acta Bioenerg.*, 2014, **1837**, 139-148.
81. P. Moenne-Loccoz, B. Robert and M. Lutz, *Biochemistry*, 1989, **28**, 3641-3645.
82. P. Dorlet, L. Xiong, R. T. Sayre and S. Un, *J. Biol. Chem.*, 2001, **276**, 22313-22316.
83. C. J. Gisriel, *Biochim. Biophys. Acta Bioenerg.*, 2024, **1865**, 149032.
84. N. Cox, J. L. Hughes, R. Steffen, P. J. Smith, A. W. Rutherford, R. J. Pace and E. Krausz, *J. Phys. Chem. B*, 2009, **113**, 12364-12374.

6

Excitation landscape of the CP43 photosynthetic antenna complex from multiscale simulations[‡]

[‡]This chapter was reprinted with permission from “Excitation landscape of the CP43 photosynthetic antenna complex from multiscale simulations” by Bhattacharjee, S.; Arra, S.; Daidone, I.; Pantazis, D. A., *Chem. Sci.* **2024,** *15*, 7269-7284.
https://doi.org/10.1039/D3SC06714A.
[‡]Original data from this work are provided as an open-access data set hosted by the Open Research Data Repository of the Max Planck Society at https://doi.org/10.17617/3.S2DPUE.

6.1 Introduction

Oxygenic photosynthesis is a process of fundamental biological and geochemical significance, encompassing light harvesting, charge separation, and water oxidation, centered on Photosystem II (PSII).[1-3] The cyanobacterial PSII is a dimeric membrane-bound pigment–protein complex comprising 20 protein subunits (17 membrane-intrinsic and 3 extrinsic), along with nearly 100 cofactors in each monomer.[4-6] The D1 (PsbA) and D2 (PsbD) proteins harbor the reaction center (RC) of PSII, the set of 4 Chls and 2 pheophytins responsible for charge separation that eventually drives H_2O oxidation to molecular O_2 by the OEC,[4, 7-11] coupled to plastoquinone reduction.[12] Two transmembrane Chl-binding proteins CP43 (PsbC) and CP47 (PsbB), with approximate molecular weights of 43 kDa and 47 kDa respectively, are essential intrinsic core antenna proteins of PSII.[13, 14] They interact closely with the D1 and D2 proteins (**Fig. 6.1**) to deliver excitation energy to the RC,[15, 16] working either as light absorbers themselves or facilitating excitation energy transfer (EET) from peripheral light-harvesting complexes.[3, 17-25]

CP43 is closest to the D1 protein, which hosts the branch of RC pigments that are active in charge separation[26] and also accommodates the Mn_4CaO_x cluster of the oxygen-evolving complex (OEC), the site of water oxidation. In addition to its role in EET, CP43 plays a pivotal role in maintaining the overall structural integrity of the RC and contributes to the stabilization of the OEC itself by providing a direct manganese-coordinating ligand (Glu354) as well as crucial second-sphere functionality (Arg357). As an essential core antenna complex, CP43 has been the subject of numerous studies that attempted to elucidate its spectroscopic properties and excitonic structure.[13, 27-31]

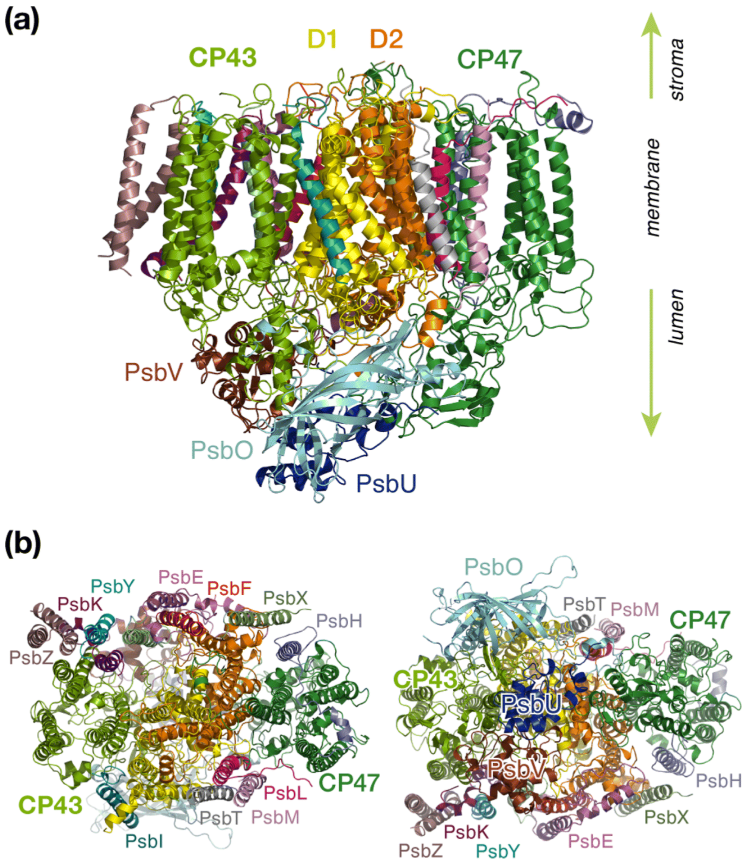


Fig. 6.1 PSII monomer with labeled subunits: **(a)** side view perpendicular to the membrane normal, **(b)** stromal and lumenal views.

CP43 contains 13 embedded Chl *a* pigments and 4 β-carotenes. **Fig. 6.2** depicts the spatial arrangement of the CP43 Chls and indicates the labeling used in the present work, which follows the numbering recommended by Müh and Zouni.[18] **Table 6.1** describes distinctive characteristics of each Chl in terms of their axial ligation and the hydrogen-bonding interaction at the 13^1-keto group. Additionally, to facilitate comparisons with previous studies that follow structure-specific numbering of CP43 Chls, **Table 6.1** lists the corresponding numbering of the CP43 Chls in selected PSII crystallographic models. As shown in **Fig. 6.2**, the Chls are distributed in three layers, with four Chl molecules present in the lumenal layer (C1-C4), eight in the stromal layer (C5-C9, C11-13), and one (C10) positioned in the center of the lipid bilayer. This arrangement is reminiscent of the distribution of Chls in the CP47 antenna.[13, 32]

Table 6.1 Nomenclature for the CP43 Chls. His$_\delta$ and His$_\varepsilon$ denote the binding mode of histidine residues with respect to the N$_\delta$ and N$_\varepsilon$ site, respectively. Axial ligation on the opposite and the same side of the phytyl chain are denoted as α and β type ligation. The Chls in the lumenal and stromal domain are denoted as L and S, respectively. All amino acid residues are from CP43 (PsbC subunit) unless otherwise indicated.

Site	Axial	13^1-Keto H-bond	Location	2AXT[33]	3BZ1[34]	3WU2-A[35]	3WU2-B	4IL6[36]
C1	His$_\varepsilon$237(α)	H$_2$O	L	33	474	501	902	501
C2	His$_\varepsilon$430(α)	Tyr297	L	34	475	502	903	502
C3	His$_\varepsilon$118(α)	-	L	35	476	503	904	503
C4	H$_2$O(α)	H$_2$O, LMG-519	L	37	477	504	905	504
C5	His$_\varepsilon$441(α)	H$_2$O, Arg449	S (close to D1)	41	478	505	906	505
C6	His$_\varepsilon$251(α)	-	S	42	479	506	907	506
C7	H$_2$O(α)	His$_\delta$164	S	43	480	507	908	507
C8	His$_\varepsilon$444(α)	-	S (close to D1)	44	481	508	909	508
C9	His$_\varepsilon$53(β)	Ser275	S	45	482	509	910	509
C10	His$_\varepsilon$56(α)	-	Center (close to PsbK)	46	483	510	911	510
C11	Asn39(α)	H$_2$O, Arg41	S (close to PsbK)	47	484	511	912	511
C12	His$_\varepsilon$164(α)	H$_2$O	S	48	485	512	913	512
C13	His$_\varepsilon$132(α)	Tyr131, LMT 102	S	49	486	513	914	513

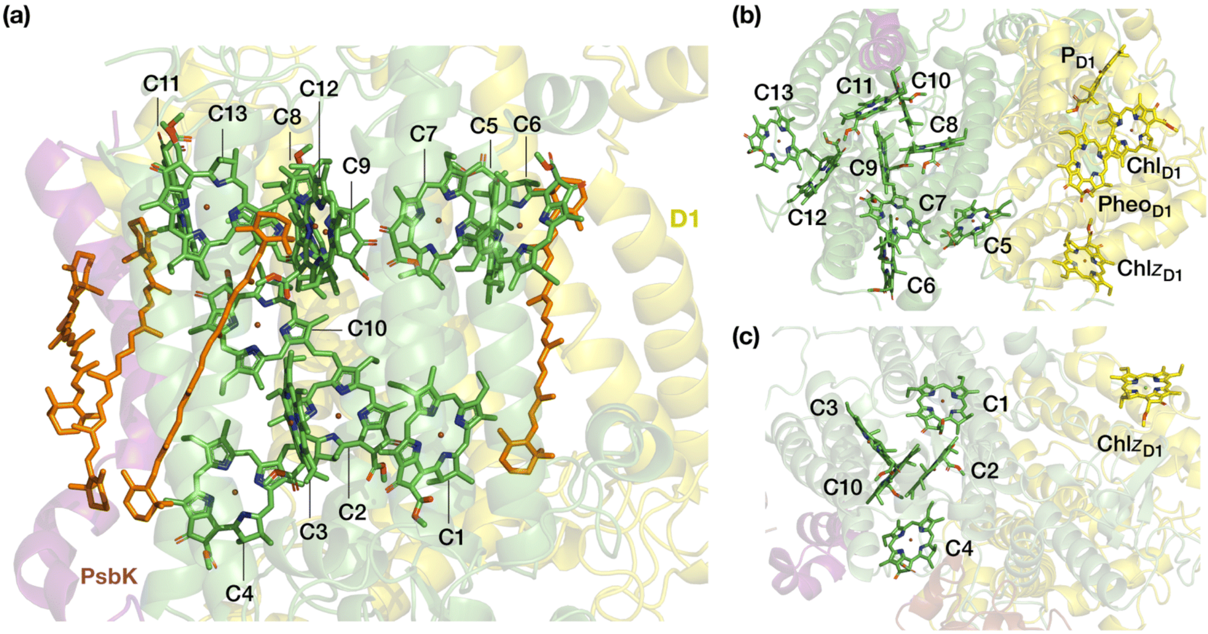


Fig. 6.2 The pigments of the CP43 complex with their labels: **(a)** side view, indicating also positions of carotenoids; **(b)** stromal ("top") view, depicting also proximal pigments of the reaction center that belong to the D1 (PsbA) chain; **(c)** lumenal ("bottom") view. Chl C10 is shown in both panels on the right to aid orientation. C5 is the closest CP43 Chl to D1 pigments Pheo$_{D1}$ and Chlz$_{D1}$ (*ca.* 21–22 Å, see **Table 6.7** for details).

Spectroscopic characterizations of intrinsic antenna complexes CP43 and CP47 are often performed on samples extracted from PSII. In the case of isolated CP43 samples thus only 13 Chls contribute to the spectra, facilitating analysis and fitting of the spectra to some extent. However, isolated samples depart from the native structure and possibly lack the structural integrity of the PSII-complexed system, potentially introducing inconsistencies in the resulting data sets. Additionally, it is possible that isolated preparations or certain treatments may result in deformation or even loss of one or more Chls, as suggested for CP47.[32] The CP43 core antenna in PSII is anticipated to possess two quasi-degenerate low-energy "trap" states,[31, 37] inferred from hole-burning (HB) studies,[38] other spectroscopic[39, 40] and structure-based calculations.[27-29] However, the exact assignment of these low energy excitonic states remains a topic of active debate in literature. Shibata et al. reported the presence of two red-shifted pigment domains within the PSII core complex emitting at 685 nm and 695 nm,[16] the former assigned to CP43. Previously, Hughes et al. claimed that both states are localized on one Chl, but one is excitonically coupled to other states.[37] Conversely, Raszewski, Renger and Müh have associated one CP43 trap state with a localized exciton in the lumenal layer whereas the other trap state was concluded to be on a delocalized domain in the stromal layer.[29, 41]

The kinetics of EET are also debated, with different groups arriving at significantly different values for the transfer times from the CP43/CP47 antennae to the RC compared to the rate of primary charge separation (CS) at the RC.[15, 16, 42, 43] The conclusions depend on the distinct assumptions and theoretical models employed. For example, Holzwarth and co-workers reported timescales of a few ps for EET to the RC based on transient absorption and fluorescence kinetics,[44, 45] whereas Renger arrived at estimates approximately an order of magnitude slower for the forward and half of that for the backward EET process.[41, 46] This also led to alternate EET mechanisms being proposed, namely the "exciton-radical pair equilibrium" (ERPE) model[30, 44] and the "transfer-to-trap limited" (TTTL) model.[45, 47] It has also been suggested that CS and EET may occur on the same timescale.[14] More recently, Yang et al. investigated the EET dynamics of the PSII core complex using two-dimensional electronic-vibrational (2DEV) spectroscopy[19] and suggested that C5 (current labeling) in CP43 and the peripheral D1 Chl known as Chlz_{D1} likely form the pathway for energy transfer from CP43 to the RC. The results were consistent with the TTTL model in the sense that EET from CP43 to Chlz_{D1} was found to be faster than subsequent EET to other D1 pigments, a step which thus constitutes the kinetic bottleneck.

Spectroscopic studies towards determining possible EET pathways in photosynthetic light harvesting complexes (LHCs) still face two main challenges.[17, 48, 49] First, the closely spaced pigments exhibit significant excitonic couplings rendering it impossible to make a direct correlation between the absorption bands and individual pigments they originate from. Second, the highly congested excitonic manifold makes it non-trivial to assign site energies to specific Chl molecules. That is why theoretical approaches that complement experimental data (e.g. spectral densities, absorption and fluorescence spectra) by simulating the structures and estimating the site energies of pigment-protein complexes have long played an important role in the study of antenna complexes.[25, 28, 29, 50-58] Past investigations employed quantum mechanics/molecular mechanics (QM/MM) and molecular dynamics (MD) simulations based on predetermined potential energy surfaces (PES) derived from DFT,[59] or "on-the-fly" PES obtained from the semi-empirical DFTB approach.[60] However, it remains challenging even with approximate QM methods to perform state of the art simulations on groups of Chl pigments in systems such as CP43 and CP47 while simultaneously fully accounting for the short- and long-range effects of the protein matrix. Furthermore, even in existing theoretical studies there have been discrepancies regarding the relative ordering of site energies and identities of the low energy trap states.[38, 39] In two investigations, Müh, Renger and co-workers identified three red Chls in the isolated CP43 protein.[28, 29] The excitonic couplings and local transition energies of Chls were computed using Poisson-Boltzmann (PB) electrostatics in both cases, albeit the latter study was based on the high-resolution crystal structure of PSII.[28, 29] Saito et al. evaluated EET coupling between Chls in the PSII monomer based on the QM/MM diabatization scheme.[61] In a more recent computational study, Sarngadharan et al. determined the site energies and excitonic couplings of CP43 Chls[27] using time-dependent long-range corrected density functional tight binding approach (TD-LC-DFTB) and QM/MM/MD trajectories. All the above theoretical studies seemed to have reached a consensus that the red-most Chls in CP43 are delocalized and likely belong to two separate "domains" of the transmembrane region. It is also important to note that all or most of the theoretical studies are based on the isolated CP43 protein, without the remaining PSII matrix. Moreover, axial ligation to the central Mg atom, pigment-protein interactions (e.g., H-bonding to the keto group), and pigment-pigment interactions (e.g. in Chl dimers) are known to directly influence excited state properties of Chls,[32, 62, 63] it is therefore crucial to treat these interactions at the QM level along with the chromophore, an aspect which has been overlooked in several past studies.

Here we utilize a comprehensive large-scale QM/MM approach employing range-separated time-dependent density functional theory (TD-DFT) to investigate the low-energy excited states for all CP43 Chls. The calculations explicitly account for interactions of the pigments with the complete membrane-embedded all-atom PSII monomer. Beyond individual pigments, we apply the same methods on pigment dimers in order to study coupled pairs and to investigate the presence of possible charge-transfer (CT) states.[62, 64, 65] Crucially, our work incorporates the perturbed matrix method (PMM)[66-69] that enables us to leverage the full information from long-timescale molecular dynamics simulations for the extraction of site energies and excitonic couplings. In addition to a new set of site energies and excitonic couplings, the results identify a previously unknown low-lying state with significant charge transfer character among CP43 Chls, which may have important implications for the functional role of the protein. Combined with a refined analysis of the contributions of the different monomeric and dimeric pigment groups to the first exciton, the present results provide a detailed map of static and dynamic properties of the CP43 pigments and contributes to the improved understanding of this essential photosynthetic antenna.

6.2 Methodology

6.2.1 Molecular Dynamics (MD) simulations

The initial structure of the Photosystem II (PSII) monomer used in this work is based on the high-resolution crystal structure of *T. vulcanus* (PDB ID: 3WU2).[35] The complete protein unit was embedded in a POPC lipid bilayer of dimension 176 × 176 Å^2 using Packmol-Memgen,[70] thoroughly solvated with a TIP3P water box and neutralized with appropriate number of counterions (see **Section 3.2** for more details**)**. For the protein residues, waters and POPC bilayer we used the standard parameters from the AMBERff14SB,[71] TIP3P[72] and LIPID17[73] force fields, respectively. The partial charges and force field parameters for the organic cofactors (plastoquinones, carotenoids, structural lipids) were obtained using GAFF2,[74] bonded parameters for non-heme Fe and OEC were taken from earlier studies,[26, 75, 76] while those of the remaining cofactors (Chl a, Pheo a, heme) were obtained from literature.[77, 78] The nonbonded parameters for the metal ions were based on their respective oxidation states using data sets available for the TIP3P water model.[79] For Na$^+$ and Cl$^-$ ions, we used the Joung–Cheatham parameters compatible with the TIP3P water model.[80, 81]

A step-by-step minimization protocol was employed to remove energetically unfavorable geometric clashes in the system **(see Section 3.4).** In the heating phase, the system is first heated from 10 to 100 K in a succession of 5 ps in the NvT ensemble, followed by further heating from 100 to 303 K in the NpT ensemble for a total of 125 ps. The temperature during this step is maintained using the Langevin dynamics[82] with a collision frequency of 5 ps^{-1}. During the equilibration phase, the C_α atoms of amino acids were restrained with a force constant of 20 kcal $mol^{-1}Å^{-2}$. Subsequently, the restraints on the C_α atoms were systematically reduced (2 kcal $mol^{-1}Å^{-2}$/500 ps) and the system was further equilibrated for 65 ns to equilibrate the POPC bilayer. Thereafter, production run was initiated without any restraints and the temperature and pressure set at 303 K and 1 atm, respectively. During the entire procedure, the temperature was controlled using Langevin dynamics with a collision frequency of 1 $ps^{-1}$and the system pressure was controlled using the Berendsen barostat[83] with anisotropic pressure scaling with a relaxation time of 2 ps. We employed the SHAKE algorithm[84] to constrain the bonds involving hydrogens, therefore a time step of 2 fs could be used. The electrostatic interactions were treated using the Particle Mesh Ewald (PME) approach[85] with a 10 Å cut-off. The AMBER20 package[86] was used to perform the energy minimizations and equilibration dynamics while the production MD simulations were performed in the GPU version of the pmemd module (*pmemd.cuda*).[87]

6.2.2 QM/MM geometries

The QM/MM computations on the CP43 Chls are performed based on an equilibrated PSII-lipid bilayer model of the 3WU2 crystal structure (**see Section 3.2**).[35, 62, 63] The detailed protocols for equilibration and production dynamics are discussed in **Section 3.4**. We consider the entire PSII monomer and a water layer (7 Å around the protein), including all the waters present in the protein cavity and various channels. The final system used in the QM/MM calculations consisted of 76,035 atoms. All QM/MM calculations were performed using the multiscale module of the ORCA 5.0 suite,[88, 89] employing the electrostatic embedding technique. The hydrogen link atom approach was employed to cut through C–C covalent bonds and the charge-shift (CS) scheme was used to avoid overpolarization of the QM region. For each Chl, the chlorin macrocycles along with the axial ligand to Mg^{2+} and the side chain of residues H-bonded to the keto group at the 13^1-carbon position (ring E) are included in the QM region. The phytyl chains were included in the QM region up to C^{17} (truncated as a methyl group) and the rest of the chain was treated in the MM region. For geometry optimizations, the complete system was further subdivided into two parts: active and static. The active region consists of atoms within the QM and MM regions that remain free to move during optimization, whereas the remaining MM atoms are fixed and only contribute to the electrostatics. Complete amino acid residues and waters within 10 Å from the center of each chlorin ring (taken as the Mg^{2+}) were considered in the active region. For the Chl pairs, we performed a constrained optimization where the QM/MM optimized geometries of individual pigments were combined and kept fixed while only the MM active region was relaxed. The Perdew–Burke–Ernzerhof (PBE) functional[90, 91] was used to optimize the QM/MM geometries using the def2-TZVP basis set[92] and D3(BJ) dispersion corrections[93, 94] throughout. The resolution of identity approximation (RI)[95] was used to speed up the calculation of Coulomb integrals with the corresponding auxiliary basis set (def2/J).[96]

6.2.3 Calculation of excited states

Vertical excitation energies are computed on the optimized ground state geometries using full TD-DFT without TDA. All calculations were performed using the range separated ωB97X-D3(BJ) functional (modified version of ωB97X-V[97] with D3BJ correction) along with def2-TZVP basis sets. This long-range-corrected functional has a fixed exact HF exchange of 16.7% (short-range) that increases to 100% at long range with a range-separation parameter of 0.30 bohr^{-1}. The first 8 excited states (roots) were computed, thus covering the entire Q-band range and further low-lying excited states for individual Chls as well as for dimers.[26, 63, 98] The electrostatic effects of the protein environment were included through MM point charges. The RIJCOSX approximation[99] and the corresponding auxiliary basis sets were used throughout. VeryTightSCF convergence criteria were applied, along with dense integration grids (DefGrid2). The vertical excitation energies (VEE) and transition dipole moments (to be used in subsequent PMM calculations) for the ground and first two excited states were calculated for all the QM regions specified above. The nature of the excited states was characterized based on Natural Transition Orbitals (NTOs)[100] using the orca_plot module.

6.2.4 Perturbed Matrix Method (PMM) Calculations

The Perturbed Matrix Method (PMM) shares conceptual similarities with many QM/MM approaches.[101-104] In this methodology, the system is partitioned into two distinct entities: the Quantum Center (QC), encompassing the specific portion treated at the quantum level, and the remaining component, referred to as "the environment", described as a semiclassical perturbation acting upon the QC. Differently from conventional QM/MM schemes, PMM employs long timescale molecular dynamics (MD) simulations for the entire system, encompassing both the QC and the environment, all governed by the same Hamiltonian—that is, the same classical force field. In the MD-PMM framework, the unperturbed quantum properties of the isolated QC are first calculated at the QM level on the QC geometry optimized at the QM/MM level (see below). Following this, for each configuration of the simulated system (i.e., for each frame in the classical molecular dynamics simulation), the electrostatic effect of the instantaneous atomistic configuration of the environment is included as a perturbing term within the QC Hamiltonian operator. The electronic Hamiltonian operator $\widehat{H}$ of the QC embedded in the perturbing environment can be thus expressed as follows:

$$\widehat{H} = \widehat{H}^0 + \widehat{V} \tag{6.1}$$

where $\widehat{H}^0$ is the QC unperturbed electronic Hamiltonian (i.e., as obtained considering the isolated QC) and $\widehat{V}$ is the perturbation operator. The perturbation operator $\widehat{V}$ can be derived using a multipolar expansion centered on the QC's center of mass:

$$\widehat{V} = q_T\mathcal{V} - \mathbf{E} \cdot \widehat{\boldsymbol{\mu}} \tag{6.2}$$

with q_T the total QC charge, $\mathcal{V}$ the electrostatic potential exerted by the perturbing environment, $\mathbf{E}$ the perturbing electric field and $\widehat{\boldsymbol{\mu}}$ the dipole operator. Finally, at each MD frame, the diagonalization of the Hamiltonian matrix provides a set of eigenvectors and eigenvalues representing the perturbed eigenstates and energies of the QC. In the present study we are mainly interested in the site energy, namely the difference between the perturbed energies of the first-excited and ground states of each QC. This physical quantity is hence calculated at each frame of the MD simulation, and the average evaluated in the MD ensemble. The unperturbed properties for each QC are calculated using the corresponding QM/MM optimized geometry, which, thus, remains fixed in the MD-PMM calculations.

In case of the interacting chromophores, i.e., a set of interacting QCs, we may consider the possible excitation coupling occurring among the QCs.[105] In the present case, the interactions between the electronic excitations localized on each Chl are considered, with the exception of C2–C10 and C7–C9 that were considered as dimers due to their electronic coupling (see above). To this aim, the perturbed Hamiltonian operator for the 11 QCs (9 single Chls and 2 dimers) is considered and, in matrix notation, is expressed as follows:

$$\widetilde{H} = E_0\widetilde{I} + \Delta\widetilde{H} \tag{6.3}$$

where E_0 is the electronic ground-state energy, $\widetilde{I}$ is the identity matrix and $\Delta\widetilde{H}$ is the excitation matrix, the diagonal elements of which are given by the single chromophores perturbed excitation energies. The non-diagonal elements of the excitation matrix i.e., the excitonic couplings, are obtained by truncating the expansion of the interaction operator of the chromophores at the dipolar term i.e., using the point dipole approximation (PDA). Thus, the electronic coupling between two QCs is treated as a dipole-dipole interaction and the k,k' interaction operator is given by:

$$\hat{V}_{k,k'} = \frac{q_{T,k}q_{T,k'}}{R_{k,k'}} + \frac{q_{T,k}\hat{\boldsymbol{\mu}}_{k'}\cdot\mathbf{R}_{k,k'}}{R_{k,k'}^3} - \frac{q_{T,k}\hat{\boldsymbol{\mu}}_{k}\cdot\mathbf{R}_{k,k'}}{R_{k,k'}^3} + \frac{\hat{\boldsymbol{\mu}}_{k}\cdot\hat{\boldsymbol{\mu}}_{k'}}{R_{k,k'}^3} - 3\,\frac{\hat{\boldsymbol{\mu}}_{k'}\cdot\mathbf{R}_{k,k'}\hat{\boldsymbol{\mu}}_{k}\cdot\mathbf{R}_{k,k'}}{R_{k,k'}^5} \tag{6.4}$$

where $\hat{\boldsymbol{\mu}}_k$ is the k^{th} chromophore dipole operator and $\mathbf{R}_{k,k'}$ is the k' to k chromophore displacement vector defined by the corresponding chromophore centers of mass. It should be noted that this is the general expression for charged QCs. In the present case, all QCs are neutral and thus the first three terms of the equation are zero. The transition dipoles in **Equation 6.4** the perturbed transition dipoles as obtained by considering each single chromophore embedded in the field produced by the rest of the environment, including the other QCs. Despite the fact that transition charges from the electrostatic potential (TrESP) method generally show a superior performance with respect to the PDA for excitonic couplings,[106] recent studies on the water-soluble Chl-binding protein (WSCP)[107] and on the Fenna–Matthews–Olson (FMO) complex[59] showed that the dipole approximation can give an accuracy comparable to TrESP. In the case of the FMO complex, the agreement between PDA and TrESP couplings is quantitative; in the case of the WSCP, PDA couplings were found to be only a few percent larger. In the present study we opted for PDA due to its computational efficiency and because it is already implemented within the PMM framework, facilitating a smoother workflow. Except for the computation of the excitonic states, which were performed using an in-house Fortran code, the basic PMM code is available as an open-source program (PyMM).[108]

Finally, by diagonalizing the excitation matrix, the perturbed excitation energies and eigenstates (i.e., the exciton states) are obtained. In the present case we utilized 3000 frames, corresponding to the final 30 ns of a previously performed 200 ns-long MD trajectory,[64,109] saved at intervals of 10 ps. This specific time frame was previously employed for the redox potential calculations[65] and was selected to ensure a reasonably converged root mean square deviation (RMSD) of the entire system with respect to the starting, crystallographic structure. In principle, there is no computational constraint on the number of frames that can be utilized within the MD-PMM framework, as it solely depends on the length of the MD trajectory.

The PMM approach has previously been employed in the study of Chls and pheophytins of the PSII-RC.[65] This application demonstrated very good agreement with experimentally-derived thermodynamic parameters such as reduction potentials, and with experimental kinetics properties such as rate constants of primary electron transfer within the RC. Here, we further validated the MD-PMM approach by computing the absorption line shape of CP43 at room temperature (see **Fig. 6.3**). Notably, there is a close match with the experimental line shape. Specifically, the experimental exciton splitting at 293 K (that manifests as a slight shoulder at low energy) is accurately reproduced (12 nm in the computed spectrum versus 13 nm in the experimental one). The total bandwidth is slightly overestimated (34 nm versus the experimental 24 nm), most probably due to an overestimation of the inhomogeneous broadening provided by the perturbation calculated along the MD trajectory.

6.2.5 Calculation of the absorption spectrum with MD-PMM

For computing the absorption spectrum, after calculating the perturbed frequencies (ν) and transition dipoles ($\boldsymbol{\mu}_{j,i}$) for each exciton state over the frames of the MD trajectory, we evaluate the excitation energy distribution using a suitable number of frequency bins. Subsequently, we utilize the excitation energy distribution and the corresponding transition dipoles to determine the molar extinction coefficient $\varepsilon_{0,i}$ for the transitions from the ground to the ith excited state, thus providing the absorption spectrum according to:

$$\epsilon_{0,i} = \sum_{\nu_{\text{ref}}} \frac{\Gamma_A(\nu_{\text{ref}}) h\nu}{N} \frac{e^{-(\nu-\nu_{\text{ref}})^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \tag{6.5}$$

$$\Gamma_A(\nu_{\text{ref}}) = \frac{|\mu_{0,j}|^2_{\nu_{ref}}}{6\epsilon_0 c\hbar^2} \tag{6.6}$$

Here, ν_{ref}, $n(\nu_{\text{ref}})$ and $|\boldsymbol{\mu}_{0,j}|^2_{\nu_{ref}}$ are the frequency at the center of each bin, the corresponding number of MD frames and mean transition dipole square norm within the bin, respectively, $\hbar = h/2\pi$ with h the Planck constant, ϵ_0 is the vacuum dielectric constant, c is the speed of light and σ^2 is the variance associated to the semiclassical intra-QC vibrations neglected in the evaluation of the unperturbed properties. In the present case, the value of σ^2 ($\sigma = 10^{-4}$ Hartree) has been estimated on the basis of the full width at half maximum (FWHM) of the experimental spectrum of Chl a in toluene,[110] approximating the vacuum condition. The calculated spectral line shape in the Q_y region is reported in **Fig. 6.3**. Note that the width of the calculated bands (inhomogeneous broadening) is due to the explicit effect of the semiclassical fluctuations of the QCs and their environment along the MD trajectory and does not include any adjustable parameter.

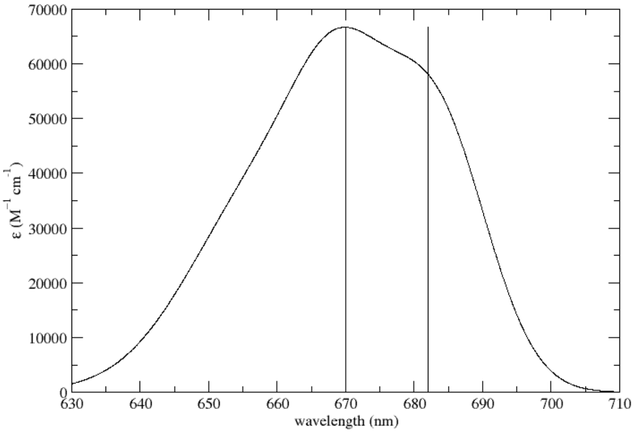


Fig. 6.3 Room-temperature absorption spectrum of the CP43 complex calculated with the **MD-PMM** approach. The spectrum has been shifted by 0.000457 Hartree to lower energies in order to match the position of the main experimental peak measured at 293 K.

6.3 Results and discussion

6.3.1 Chlorophyll site energies

The determination of the low-energy excitation energy profile for CP43 and CP47 core antenna pigments is key in identifying the characteristic pathways of EET within PSII. Theoretical parameters derived from calculations are frequently employed to estimate excitation energies by fitting them to experimental spectra. It is well established that the PSII matrix largely influences the photochemical properties of Chls through electrostatics, structural distortions, axial ligation and secondary sphere interactions like H-bonding networks. Most importantly, local protein electrostatics are known to be predominantly responsible for fine-tuning the excitation energies of photosynthetic pigments.[62, 63, 111] Here we consider all the above factors by employing multiscale modeling techniques within the QM/MM framework to determine the site energies of all CP43 Chls embedded in the "native" PSII system.

The geometries of all individual Chls were optimized using DFT and QM/MM as described in the Methodology. Following complete optimizations, we computed the vertical excitation energies of each CP43 Chl *in vacuo* using TD-DFT on the QM/MM optimized geometries. **Table 6.2** provides a comprehensive overview of the vertical excitation energies of the lowest excited state (S_1 or Q_y), and the protein electrochromic shifts of each CP43 Chl. We find that the *in vacuo* first excited state energies lie within the range of 1.868−1.966 eV (see **Table 6.2**). This demonstrates that intrinsic structural features of the pigment itself, like the macrocyclic ring curvature and the nature of axial ligation, play already major roles in differentiating the site energies of each Chl. We subsequently performed excited state calculations on the CP43 protein *excluding* the rest of the PSII monomer, so that we can compare the effect of having an "isolated" CP43 chain as opposed to a complete PSII monomer (we emphasize that this is not equivalent to simulating an experimentally extracted CP43 antenna because we are not simulating here the conformational state of the latter, a problem that presents a distinct challenge). The results show that the local protein electrostatics already shift the Chl site energies toward the red or blue regime relative to their gas-phase values (**Table 6.2**).

Overall, the site energies for the isolated CP43 lie in the range of 1.850−1.998 eV. The six Chls C1, C2, C3, C5, C8 and C9 are red-shifted, whereas the remaining C4, C6, C7, C10, C11, C12 and C13 Chls are blue-shifted with respect to the gas-phase excitation energies (**Fig. 6.4**). Interestingly, 3 out of the 6 red-shifted Chls are on the lumenal side of the protein (**Fig. 6.2c**). These findings are consistent with the stromal versus lumenal (membrane-transverse) trends obtained for the pigments in the RC and the CP47 antenna of PSII.[32, 62] Chls C5 and C1 are the pigments that shift most to the red due to protein matrix electrostatics.

Table 6.2 TD-DFT (ωB97X-D3(BJ)/def2-TZVP) site energies *E* (in eV) of all CP43 Chls and oscillator strengths of the **S_1 (Q_y)** transition computed *in vacuo*, with QM/MM on an isolated CP43 protein, and with **QM/MM** in the complete PSII monomer, compared with the mean site energy obtained by **PMM**. The geometries in all cases are derived from QM/MM optimizations within the PSII monomer. Shifts are reported in meV with respect to the gas-phase values.

	TD-DFT Gas-phase		TD-DFT | QM/MM Isolated CP43			TD-DFT | QM/MM Full PSII monomer			PMM	
Site	***E***	***f*osc**	***E***	***f*osc**	**Shift**	***E***	***f*osc**	**Shift**	***E*mean**	**Std.Dev.**
C1	1.868	0.22	1.850	0.20	-18	1.846	0.19	-22	1.873	0.005
C2	1.879	0.22	1.870	0.20	-9	1.867	0.20	-12	1.894	0.010
C3	1.868	0.22	1.866	0.20	-2	1.861	0.20	-7	1.867	0.005
C4	1.913	0.23	1.927	0.19	14	1.928	0.17	15	1.945	0.014
C5	1.966	0.23	1.944	0.22	-22	1.938	0.25	-28	1.915	0.011
C6	1.895	0.22	1.954	0.21	59	1.970	0.21	75	1.963	0.026
C7	1.886	0.23	1.897	0.20	11	1.885	0.21	-1	1.890	0.008
C8	1.893	0.22	1.883	0.23	-10	1.915	0.23	22	1.895	0.005
C9	1.897	0.23	1.881	0.23	-16	1.881	0.22	-16	1.912	0.008
C10	1.901	0.22	1.939	0.23	38	1.955	0.23	54	1.942	0.013
C11	1.916	0.22	1.998	0.22	82	1.983	0.24	67	1.964	0.020
C12	1.922	0.22	1.938	0.21	16	1.926	0.21	4	1.925	0.010
C13	1.917	0.23	1.930	0.22	13	1.961	0.22	44	1.937	0.018

The identity of the red-shifted Chls remains consistent when we inspect the full results for the complete PSII monomer, but the effect of the protein matrix is now significantly more pronounced. The global PSII protein electrostatics further stabilize the site energies of Chls C1, C2, C3 and C5, with C5 still being the most red-shifted pigment. This result is significant for two reasons. First, C5 (C505 in 3WU2) is located on the periphery of CP43 and is in direct contact with the D1 (PsbA) subunit of PSII. Second, recent two-dimensional electronic-vibrational (2DEV) spectroscopy measurements on the PSII core complex[19] suggested that Chl C5 along with the peripheral Chl Chlz_{D1} likely mediate EET from CP43 to the PSII-RC. The electrochromic shift on Chl C9 (16 meV) remains invariant in the isolated CP43 and in the PSII monomer. This is expected as well because it is deeply buried within the CP43 protein matrix and is the pigment least affected by structural deformation of CP43 upon isolation. Another interesting result is seen for C8, the only pigment to show opposite shifts in the intact PSII monomer compared to isolated CP43. C8 is also located close to the D1 (PsbA) subunit, making it susceptible to conformational changes during extraction from the native PSII. The D1 electrostatics induce an additional blue-shift in the excitation energy of C8. The remaining blue-shifted Chls (C6, C11, C13) are mostly located in the stromal layer of CP43, except for C10 which is in the middle of the transmembrane region. C6 and C13 lie on the periphery of CP43 while C10 and C11 are located close to the PsbK subunit (**Fig. 6.2b**). It is suggested that the salt bridge between D2-R233 and CP43-E29 may affect the interaction between C11 and the protein, thus lowering its site energy slightly in the intact PSII compared to extracted samples.[28]

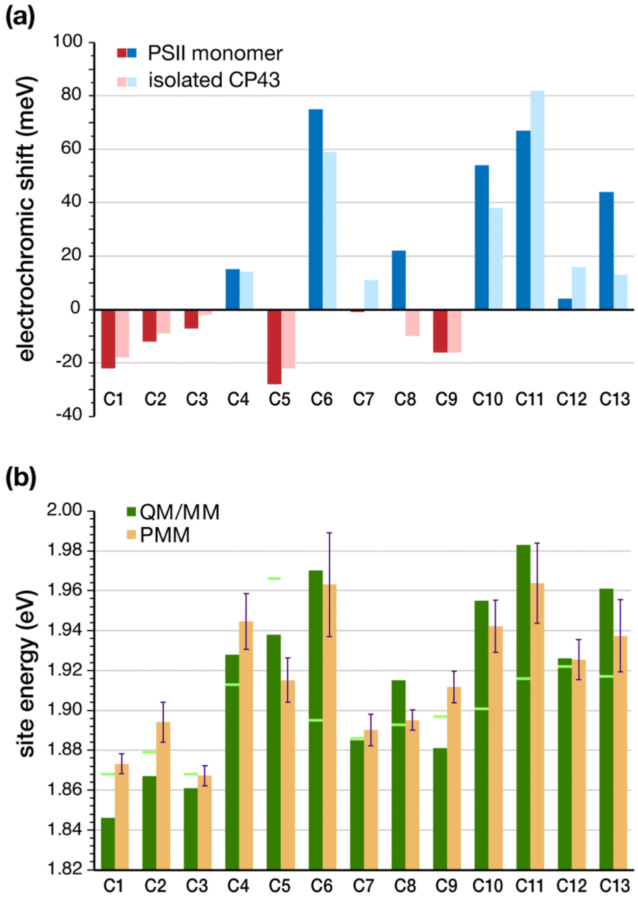


Fig. 6.4 (a) TD-DFT computed electrochromic shifts in site energies (meV) for CP43 Chls compared to gas-phase calculations. The electrochromic shift is defined as **S_1 (embedded) – S_1 (vacuum)** and is shown for both the whole PSII monomer and for the CP43 protein only. **(b)** Chl site energies computed with TD-DFT QM/MM on the reference snapshot (gas-phase values indicated as light green lines), compared to the mean site energies obtained from MD-PMM calculations averaged over 3000 frames from the equilibrated portion of the MD simulation.

Based on the results on CP43 Chls in the intact PSII monomer, the computed site energies lie in the range of 1.846–1.983 eV while the second excited states (S_2/Q_x) range from 2.260–2.468 eV (see **Table 6.4**). Moreover, the energy gap between the first two excited states is also seen to vary among the different domains, with the lowest being for C1 and C2 in the lumenal domain, and the highest being for site C8 in the stromal layer close to PsbA (D1). The Q_x–Q_y

energy gap is modulated by both intrinsic (macrocyclic ring curvature) and extrinsic components (axial ligation, H-bonding, protein electrostatics), as also seen for the case of CP47 and RC Chls.[32, 111]

It is noted that the differences in site energies of the CP43 Chls do not exclusively arise from protein electrostatics, as significant shifts are observed for the pigments in their QM/MM optimized geometries when the excited state calculations are performed *in vacuo*. We analyzed the contribution of these other effects through excited state calculations on the pigments *in vacuo*, but without the influence of the axial ligation and H-bonding interactions. Differences in the Q_y energy with that of optimized Chl a give the contribution of the QM/MM geometry optimization on the site energies. The results are summarized in **Fig. 6.5.**

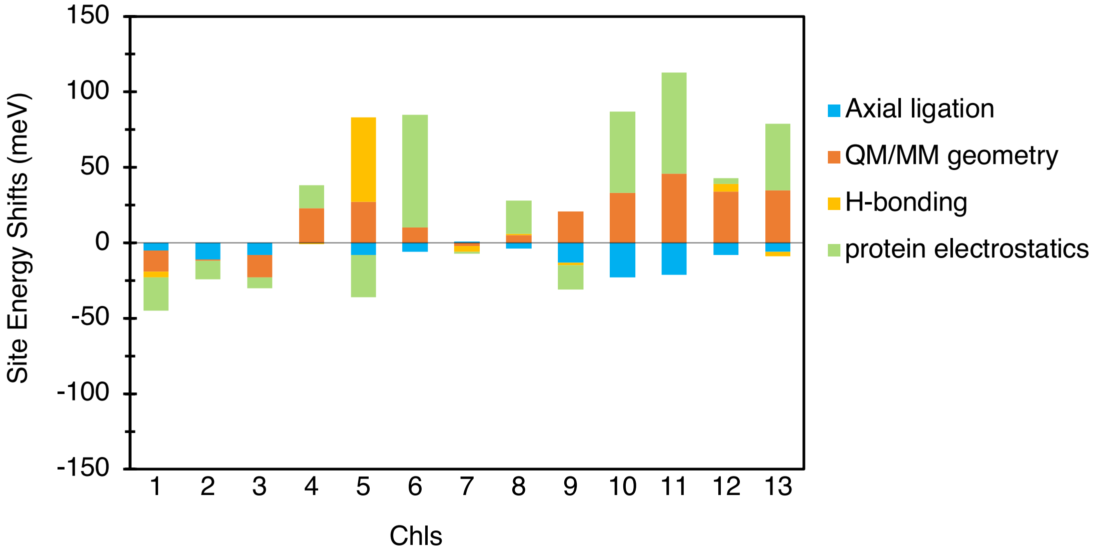


Fig. 6.5 Decomposition of the electrostatic and non-electrostatic contributions (axial ligation, H-bonding, QM/MM geometry optimization) to the site energy shifts of the CP43 Chls in PSII. All the energy differences are calculated with TD-DFT and QMMM.

Our results suggest that protein electrostatics still have the dominant contribution to site energy shifts, but there are significant contributions from the QM/MM geometry optimizations (i.e., intrinsic structural differentiation), especially contributing to the blue-shift of Chls C4 and C9–C13. The axial ligands do not affect the results significantly, but only slightly red-shift the S_1 state in each case. One important finding is that CP43-R449, which is H-bonded to a water bridging the keto group of C5, individually blue-shifts its site energy to a great extent but overall, the protein matrix overcompensates and tunes it in the opposite direction. These findings resemble past observations on Chls of CP47, which also had a similarly wide distribution of site energies even in the absence of protein electrostatics.[32] Both antennae are in this respect different from results on PSII RC pigments, where all 4 Chls (Chl$_{D1}$, Chl$_{D2}$, P$_{D1}$ and P$_{D2}$) had essentially the same site energies in the absence of the protein electrostatics despite structural differences.[62] Given that the influence of the axial ligands is not a decisive factor, it seems that the most plausible explanation for the difference is that the protein matrix allows greater flexibility for the CP43/CP47 Chls, whereas the RC pigments need to be more rigid to allow for efficient electron transfer.

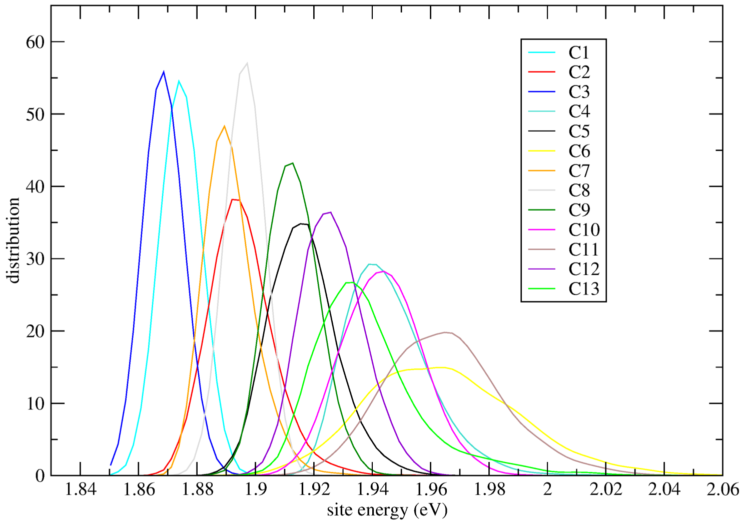


Fig. 6.6 Distributions of the site energy for all Chl-a molecules (which include their respective axial ligands) in the CP43 complex performing the **MD-PMM** calculations over the last 30 ns of the MD trajectory of the PSII complex (3000 frames in total).

An effective approach for assessing the spectral density of a chromophore within a photosynthetic pigment–protein complex (PPC) involves simulating the fluctuations in site energy resulting from environmental effects.[50, 53, 112] However, current limitations arise from the substantial size and the intricacy of these complexes and from the impossibility of simulating a membrane embedded PPC immersed in a water box while applying continuously QM methods for meaningful sampling.[113-115] Therefore, to account for the effect of conformational changes of the PSII protein environment on the individual site energies, the site energies of individual Chls were also calculated by means of the MD-PMM procedure. The energies are obtained at each time frame of the MD simulation over a specific interval. The corresponding site energy distributions are reported in **Fig. 6.6**. It can be observed that the distributions of the individual pigment molecules differ in their peak positions, widths, and skewness, although all distributions are roughly Gaussian in shape. The corresponding mean site energies are also reported in **Table 6.2** and compared with the results obtained by means of the TD-DFT calculations. **Fig. 6.4b** shows that the two sets of site energies have the same overall trend and that also the quantitative agreement is quite high, with maximum absolute deviations (MAD) of 20-30 meV. The maximum deviations are seen for pigments C1, C2, C5, C9 and C13.

In one line of investigation, Raszewski et al., have associated the "red" C685 localized on C4 (or C504 in 3WU2) in the lumenal layer of CP43.[41] In subsequent studies relying on high-resolution PSII structures,[28] the other "trap" state was suggested to be a delocalized excited state involving multiple pigments (C7, C9, and C11) located in the stromal layer. These assignments of trap states were derived largely from theoretical simulations of optical spectra that employed computed excitonic couplings and adapted site energies. A somewhat different perspective was presented by Shibata et al.,[16] who reported the presence of two red-shifted Chl domains within PSII-cc, emitting at 685 nm and 695 nm respectively, based on time-resolved fluorescence spectroscopy. In their study, only the 685 nm band was attributed to either C9 or C7 sites within CP43. A detailed comparison of the Chl site energies obtained from different literature sources is provided in **Table 6.6**. It is important to note that the results presented in the current work are not directly comparable to site energies obtained by approaches that involve fitting to various experimental data sets and different theoretical methods. Nevertheless, we can compare the relative trends in site energies with respect to the Chl with the lowest site energy.

Based on the site energies (obtained here using QM/MM and the PMM methodologies, see **Table 6.2** and **Fig. 6.4b**), the identity of the red Chls in CP43 differs from the previous assignments in literature (see **Table 6.6**). We identify the red-most Chl to be in the lumenal layer of CP43, with major contributions from sites C1 (C501 in 3WU2) followed by C3 and C2. The recent computational studies by Sarngadharan et al.[27] reflect the same trend of the red most Chls being on C1 and C3 in the lumenal layer, consistent with the current results (**Table 6.2**). The second lowest site energies are on Chls C9 and C7 from the stromal layer. This is in line with the work by Jankowiak et al. on non-photochemical HB studies who first reported that the two "trap" states[38] are localized in different layers of the thylakoid membrane. Another important finding by Hughes et al. identified that at least one of the red Chls has the 13^1-keto group hydrogen bonded to a protein residue.[37] Based on the 1.9 Å crystal structure, both Chls C9 and C7 are H-bonded to C-Ser275 and C-His164 respectively while C1 is H-bonded to a H_2O via the 13^1-keto group. By contrast C11 (C511 in 3WU2), a proposed site for the red most Chl in some assignments, is seen to possess a significantly high first excitation energy based on our calculations. This Chl has a unique axial ligation to the sidechain of PsbC-N39 and is in close contact with PsbK (see **Fig. 6.2**) and the N-terminal loop region of CP43 exposed to bulk water. Based on the PMM results in **Table 6.2**, the site energy of C11 has a significantly high standard deviation, which denotes that changes in the local protein environment induce large fluctuations in the computed site energies. There have also been discrepancies regarding the site energy of Chl C4 (C504 in 3WU2). Some studies based on refinement fits of optical spectra identified C4 to have a low excitation energy,[28] but our findings as well as previous structure-based simulations on the CP43 subunit do not agree with this assignment. However, it is also important to note that the inclusion of inter-pigment interactions, particularly electronic and excitonic couplings between specific Chls or groups of Chls may change the relative trends in site energies.

At this point it is important to consider the above results in relation to the D1 protein. The peripheral Chls Chlz_{D1} and Chlz_{D2} are discussed in the context of EET from the CP43 and CP47 core antennae, respectively, to the RC. Even though these two pigments are bound to the D1 and D2 proteins, they are thought to be functionally associated to the CP43/CP47 core antennae.[18] Notably, Chlz_{D1} is in van-der-Waals contact (4.2 Å distance) with a β-carotenoid (Car$_{D1}$) to which it can be excitonically coupled, and is close to the D1 active branch RC pigments (**Fig. 6.2**, ca. 21 Å), and hence assumed to play a direct role in EET between CP43 and the RC. The recent study by Yang et al. reported a faster EET from CP43 to Chlz_{D1} compared to EET from Chlz_{D1} to other D1 pigments,[19] concluding that the latter process is the actual rate limiting step in the overall EET pathway, consistent with previous studies by Renger and co-workers.[42] Nguyen et al. performed similar studies on D1D2Cyt$_{b559}$ complexes and identified a distinct excitonic state prior to primary charge separation (Trap*) likely belonging to one of the Chl*z* pigments.[116] The above findings are not in agreement with the theoretical study of Hsieh et al. who reported that the Chl*z* sites do not mediate EET into the RC based on molecular dynamics simulations of the PSII core complex.[117] Here, we explicitly compare the excited state energetics of the peripheral D1 Chl Chlz_{D1} with the CP43 Chls, including the full effect of PSII protein electrostatics (QM/MM) and conformational changes (PMM). Based on our results, Chlz_{D1} exhibits a lower Q_y (S_0–S_1) excitation energy (1.839 eV) than all CP43 Chls. Even with the inclusion of the dynamics based on PMM calculations, we find that Chlz_{D1} has a consistently lower mean site energy (1.855 eV) than all other CP43 Chls. Based on our findings, we cannot explicitly conclude if Chlz_{D1} can mediate EET from CP43 to the RC. Also, we do not locate the lowest site energy on the nearest CP43 pigment, C5. Nevertheless, one possible implication could be that the low-lying excited state on the peripheral Chlz_{D1} may act as a protective "trap" state to quench excess excitation energy from the RC, and avoid photodamage of the D1 protein.

6.3.2 Excitonic couplings from PMM

In LHCs, the site energies of individual chromophores can be effectively calculated by means of TD-DFT methods as shown in the previous section. However, in most cases treating groups of chromophores at one QM level may be challenging. Consequently, most computational studies of LHCs employ a combination of high-level methods to compute site energies of the individual Chls and use more simplified theoretical models to approximate the coupling interactions between them.[51] The coupling terms gives an estimate of the interaction between a pair of electronic excitations localized on different Chls and together with site energies is necessary to build a complete excitonic model of light harvesting proteins.[18, 118, 119] In this work, the excitonic couplings between Q_y transitions of Chls in CP43 were calculated using the PMM approach on the QM/MM optimized geometries and a modified version of the PDA method described earlier (see **Section 6.2.4**).

The calculated excitonic coupling values are listed in **Table 6.3**. The largest couplings mainly involve the following pigment pairs: C2−C4, C2−C10, C5−C7, C6−C7, C7−C9, C8−C9, C9−C11, and C10−C11. Except for the C5−C9 dimer, all other couples exhibit relatively small center-to-center distances (see **Table 6.7**). The Chls C7, C9 and C10 are the most strongly coupled, i.e., they strongly interact with at least three other Chl pairs with significantly high coupling values. The largest excitonic couplings (i.e., with absolute values greater than 100 cm^{-1}) are found for the C2−C4, C8−C9, C9−C11 and C7−C9 pairs.

The only exception in our estimated excitonic coupling values compared to those in previous literature is seen for the C2−C10 Chl pair which yields a high excitonic coupling value of 73 cm^{-1}. Moreover, the location of the C2−C10 is such that it connects the two layers of pigments in CP43 (**Fig. 6.2c**). A high excitonic coupling constant therefore implies that this dimer may play a role in EET between the stromal and lumenal layers of CP43. Similar arguments were made by Saito et al. in recent computational studies of EET coupling in the PSII-cc.[61] It is to be noted that excitonic coupling values in closely spaced pigments are highly sensitive to the approximations used to estimate them.[120] For instance, here the point dipole approximation is seen to predict a large positive coupling for the C2−C10 pair, but the TrEsp method in past investigations reported small coupling values for the same pair.[27, 29]

Table 6.3 Excitonic coupling constants between CP43 Chls computed using the **MD-PMM** approach. All values are reported in cm^{-1}. Absolute values greater than 50 cm^{-1} are shown in bold.

	C2	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**
C1	10.89	21.90	3.09	-3.07	-12.22	21.05	7.15	10.89	2.70	-2.52	0.63	-0.11
C2		8.14	**154.07**	-8.67	-16.15	-1.01	16.94	23.26	**73.09**	5.05	12.33	-10.27
C3			-36.43	4.48	-0.45	7.88	-4.39	-11.26	23.04	8.19	25.02	5.71
C4				-3.40	4.57	-2.08	-1.53	-0.97	-34.68	-4.46	-4.72	5.09
C5					25.46	**-98.32**	-5.90	-31.38	-1.18	-0.81	-1.55	5.03
C6						**-77.69**	-1.14	-32.26	-12.14	-4.85	-4.15	10.23
C7							17.84	**-112.59**	-4.61	-9.06	-34.46	26.34
C8								**130.37**	**66.94**	1.78	2.11	-5.16
C9									16.20	**-117.64**	36.43	-6.15
C10										**-97.23**	19.40	-11.76
C11											25.24	-13.91
C12												-48.50

It is noted that the couplings reported here are slightly higher in magnitude compared to those reported in previous theoretical studies,[27, 29] but are of the same magnitude as the values estimated based on PDA by Ishikita and co-workers on CP43[61] and those by Grondelle and co-workers on the PSI-Lhca4 complex.[121] This trend may be attributed to several effects, namely the overestimation of gas-phase transition dipoles by TD-DFT,[41] or the fact that in the current implementation of the PMM procedure the perturbation does not include polarization effects of the environment,[50, 122] which may be responsible for a screening of the couplings. Consequently, in order to account for delocalized excited states within strongly coupled group of pigments, one needs to employ more accurate quantum chemical methodologies.

6.3.3 Excited states of Chl dimers

In photosynthetic light harvesting complexes and RC proteins, closely spaced pigments possess excited states that are electronically or excitonically coupled to each other. Although individual site energies and excitonic couplings already reveal a lot about the excitonic manifold of light harvesting pigments, understanding the mechanism of EET within the core antennae as well to and from the RC additionally requires direct insight into the excitation profiles of multiple coupled chromophores. In the case of CP43, Müh et al. reported two degenerate low-energy exciton transitions that represent the lowest excited states of the two "domains" in the lumenal (containing C2 and C4) and stromal layers (containing C5 and C7−C11).[28] Earlier studies based on Stark and triplet-minus-singlet (T−S) spectra[31] provided evidence about partially delocalized excited states in the CP43 antenna as well, but no investigations exist so far to directly describe short-range effects on the excited states of pigment pairs in the CP43 or CP47 core antennae. Toward this objective, we computed the low energy excited states for specific Chl pairs within the "intact" CP43 in PSII-cc. The pigment pairs in CP43 were selected based on the following criteria: (a) center-to-center (Mg−Mg) distances less than 10 Å (b) the excitonic couplings calculated here (see **Table 6.3**). The nature of the excited states for each Chl dimer (C2−C4, C2−C10, C5−C7, C6−C7, C7−C9, C8−C9, C8−C10, C9−C10, C9−C11, C10−C11 and C12−C13) is characterized based on analysis of the natural transition orbital (NTO) coefficients for each transition. The TD-DFT results are summarized in **Table 6.8**.

The four Chl pigments C1−C2−C3−C4 constitute the lumenal layer of CP43 with an average Mg−Mg distance of 12.3 Å (see **Fig. 6.2c**). As discussed in Section 3.1 based on our TD-DFT and PMM results on individual pigments, we identified C1 and C3 to possess the lowest site energies in the "intact" CP43 antenna. C2 and C4 are the most closely arranged pigments in this layer and our findings suggest that this pigment pair has a strong positive coupling constant (154 cm^{-1}), consistent with previous studies.[27, 29, 61] However, based on our calculations, their individual site energies are not isoenergetic and our calculations (**Table 6.8**) on the C2−C4 dimer do not reveal any delocalized exciton. The energy corresponding to the lowest excited state on C2 remains unaffected, but the exciton localized on C4 is only slightly red-shifted (29 meV).

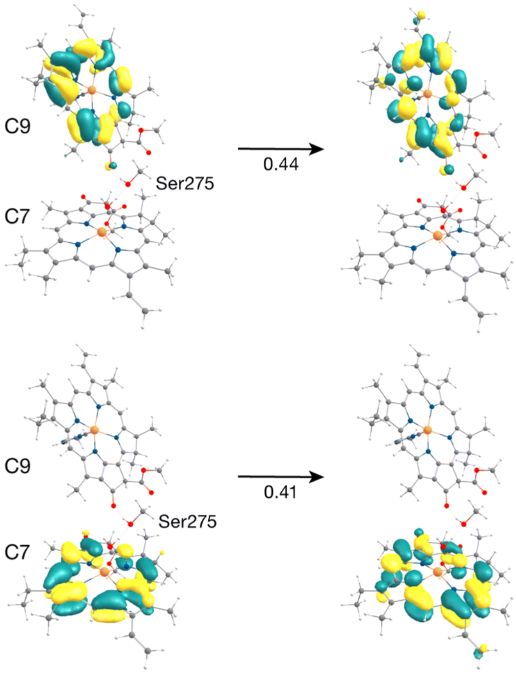


Fig. 6.7 Donor–acceptor NTOs for the S_1 state (1.868 eV) of the **C7−C9** dimer in the "crystal-like" snapshot, demonstrating that the state has mixed local excitation character, with approximately the same coefficients for the two excitations.

The C7−C9 pair is in the stromal layer of CP43 with an Mg−Mg distance of 11.4 Å. Interestingly, both lowest excited states S_1 (1.868 eV) and S_2 (1.891 eV) are superpositions of the Q_y excitation energies from C9 and C7 (analogous to the $P_{D1}P_{D2}$ special pair in the RC).[62] The NTO pairs for the corresponding S_1 and S_2 transitions are shown in **Fig. 6.7**. The delocalized lowest excited state (S_1) for this dimer is seen to possess an even lower excitation energy while the S_2 state has a slightly higher energy compared to the similar Q_y energy of the individual pigments C9 and C7 (1.881 and 1.885 eV respectively). The further red-shift of the lowest excited state in the dimer allows us to conclude that C9 and/or C7 may contribute majorly and almost equally to the red trap state in the stromal layer of CP43. This is in line with previous assignments of the lowest excitonic states made by Renger and coworkers.[28, 29] Based on the $S_0 \rightarrow S_1$ excitonic couplings reported by Müh et al.,[28] Saito et al.[61] and Sarngadharan et al.,[27] and those obtained in this work using the MD-PMM approach, both C7 and C9 are part of a strongly coupled group of Chls in the stromal layer. The low energy excitation profiles of the remaining Chl pairs in the cytoplasmic layer (C5−C7, C6−C7, C8−C10, C9−C10, C9−C11, C10−C11, C12−C13) majorly constitute localized excitations corresponding to the Q_y and Q_x transitions on individual pigments.

One motivation for the study of pigment dimers is the identification of possible charge transfer states.[123-130] In photosynthetic pigment-protein complexes CT states may participate in either spectral-tuning[124] or photoprotection through the rapid quenching of excess excitation energy.[125] It has been suggested that the latter process likely involves an excitation energy transfer from an excited Chl monomer to a strongly coupled Chl leading to a rapid non-radiative decay process via short-lived intermediate CT states.[126] Fleming and co-workers reported CT quenching in the minor antenna complexes of PSII (CP29, CP26 and CP24).[127, 128] Ramanan et al. have reported the presence of mixed excitonic-CT states in Chl heterodimers in the low-energy manifold of the LHCII complex.[129] More recently, Ostroumov et al. reported the presence of far-red emitting Chl–Chl CT states as intermediates in the excited state quenching of LHCII.[130] Stark spectroscopy and more recent computational studies on another light harvesting protein, the PSI-Lhca4 complex demonstrated that a distinctive red-shifted emission originates from the mixing of the lowest exciton state with a CT state of an excitonically coupled dimer.[131] Intermolecular charge-transfer (CT) states among photosynthetic pigments ($\mathrm{Chl_{D1}}^+\mathrm{Pheo_{D1}}^-$ or $\mathrm{P_{D1}}^+\mathrm{Pheo_{D1}}^-$) directly control primary charge separation and charge recombination processes in the PSII-RC,[26, 63, 65] but the presence of CT states in CP43 or CP47 core antenna proteins of PSII has not been reported. The present TD-DFT calculations on Chl dimers enable us to approach this question for the case of CP43.

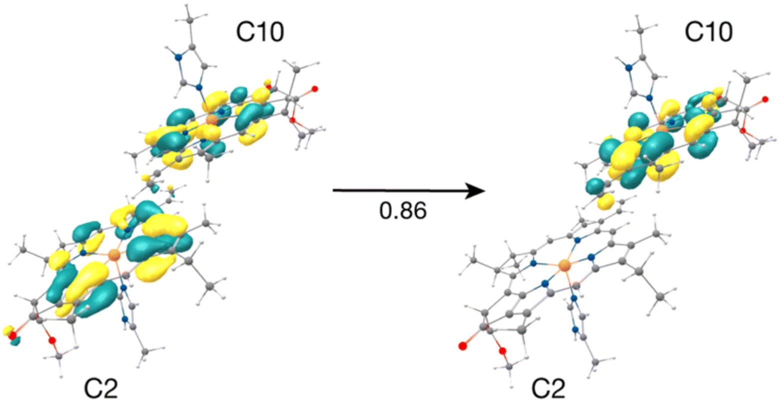


Fig. 6.8 Donor–acceptor NTOs for the S2 state (1.943 eV) of the **C2–C10** dimer in the 'crystal-like' snapshot, which shows considerable **C2** → **C10** charge-transfer (CT) character mixed with the local excitation on **C10**.

The result that stands out concerns the C2−C10 pair. These Chls (**Fig. 6.8,** Mg–Mg distance 10.2 Å) constitute a stacked dimer with parallel ring planes located approximately in the center of the CP43 transmembrane region connecting the stromal and lumenal layers (see **Fig. 6.2c**). Our results on the "crystal-like" configuration (snapshot 1) indicate that the first excited state for this dimer is localized on C2 (S_1, 1.843 eV), a likely candidate for the lowest excitation energy in CP43. Most importantly, we find the second S_2 (1.943 eV), and third (S_3, 2.001 eV) excited states of the C2−C10 dimer to have C2→C10 CT character mixed with the lowest excited state of C10. A detailed analysis of the NTOs involved in the transition i.e., the precise decomposition of the transition based on the NTO coefficients, shows that the second (S_2) and third (S_3) excitations are represented by a "delocalized" donor orbital within the C2−C10 dimer, with the acceptor NTO localized on C10, resulting in substantial CT character. The NTOs for the S_2 state are shown in **Fig. 6.8**.

Until now we looked into the excited state properties of the Chl dimers of CP43 based on a single "crystal-like" structural configuration of PSII (snapshot 1). Now we investigate how the dynamics of PSII influence the CT excitation character in the C2–C10 pair. For this purpose, we performed the same type of excited state calculations, each with individually optimized QM/MM geometries, on nine additional structurally independent snapshots (snapshots 2–10) obtained from unbiased production MD with consecutive intervals of 5 ns. The QM/MM-TDDFT results on the C2–C10 pair are provided in **Table 6.9**. **Figs. 6.10** and **6.11** depict the NTOs and difference densities respectively for the lowest CT state of the pair, which shows that the effect of the protein matrix is similar both in the crystal-like conformation of the protein and in the selected MD snapshots. The relative order of site energies remains the same, i.e., Chl C2 has a lower site energy than C10, but the C2→C10 CT character is distributed differently among the S_1, S_2 and S_3 states depending on the protein configuration. Specifically, in the crystal-like snapshot both S_2 and S_3 states have CT character mixed with local excitation on C10 while in half of the selected MD snapshots the S_3 state has dominant C2→C10 CT character (see **Table 6.9**). Interestingly we also find that the S_1 state has some CT character mixed with the local excitation on C2 for two of the examined protein configurations. Our findings thereby demonstrate that the conformational dynamics of PSII tune the extent of LE–CT mixing of the C2–C10 pair in its lowest excited states and thus allows the C2→C10 CT states to span an energy range that can bring considerable CT character as low as 1.81 eV. Such dynamic evolution of CT character has also been demonstrated in the low-energy excitation profile of RC pigments.[26, 62, 64, 65] The C2−C10 dimer is not the only Chl pair in CP43 to possess a CT state; our results locate higher-energy excited states (above 3 eV) with pure CT character for the C9−C11, C9−C10, C8−C10, C10−C11 pairs (**Table 6.8**), however the C2−C10 pair is unique in having a very low-lying CT state, essentially interleaved with the lowest locally excited states of the whole system.

Understanding the molecular mechanism of formation of low-lying CT states in LHCs has crucial implications not only for mechanisms of light harvesting and EET but also for photoprotection. Recently, Sláma et al. employed multiscale modelling approaches to show that the low-lying red states and red-shifted fluorescence bands in PSI-Lhca4 originate from the interplay of exciton and CT states within a Chl pair.[131] Such energetically low-lying mixed excitonic-CT states have also been assigned to Chl heterodimers in the excitonic manifold of the major plant light harvesting complex LHCII based on 2DES studies.[129] Mixed exciton-CT states may also act as an energy sink and thus determine pathways of EET in normal light conditions, or in some cases, may participate as intermediate trap states for excitation energy quenching in excess light.[130] Raszewski and Renger[41] in their seminal work on the PSII core complex antennae proposed that CP43/CP47 may switch from a "light-harvesting" mode for open RCs to a "photoprotective" mode for closed RCs. Therefore, our identification of a C2→C10 CT state coupled to the excitonic manifold of C10 may have the following functional implications: (a) facilitate EET within CP43 from the lumenal to the stromal layer, (b) spectral tuning of the "low-lying" red Chls in CP43, or (c) act as intermediate "trap" states for quenching of excess excitation energy away from the RC. The latter is particularly relevant for CP43 owing to its close proximity to the RC pigments and the D1 protein.

6.3.4 Global CP43 excitonic states from the MD-PMM approach

In the previous section we examined excited states of selected Chl dimers using a static TD-DFT approach. Here we report the global excitonic states considering all 13 Chls in CP43 by means of the MD-PMM procedure using as basis the single Chls for C1, C3, C4, C5, C6, C8, C11, C12 and C13 and the two dimers, C7−C9 and C2−10. The C7−C9 and C2−C10 dimers, instead of the single C2, C7, C9 and C10 Chls, were used because these pairs of pigments showed a non-negligible electronic coupling and distinctive nature of the excited states (see **Sections 6.3.2 and 6.3.3**). A schematic representation of the spatial arrangement of the pigments is shown in **Fig. 6.9a**. The distribution along the MD simulation of the vertical transition energies of the first excitonic state is shown in **Fig. 6.9b**.

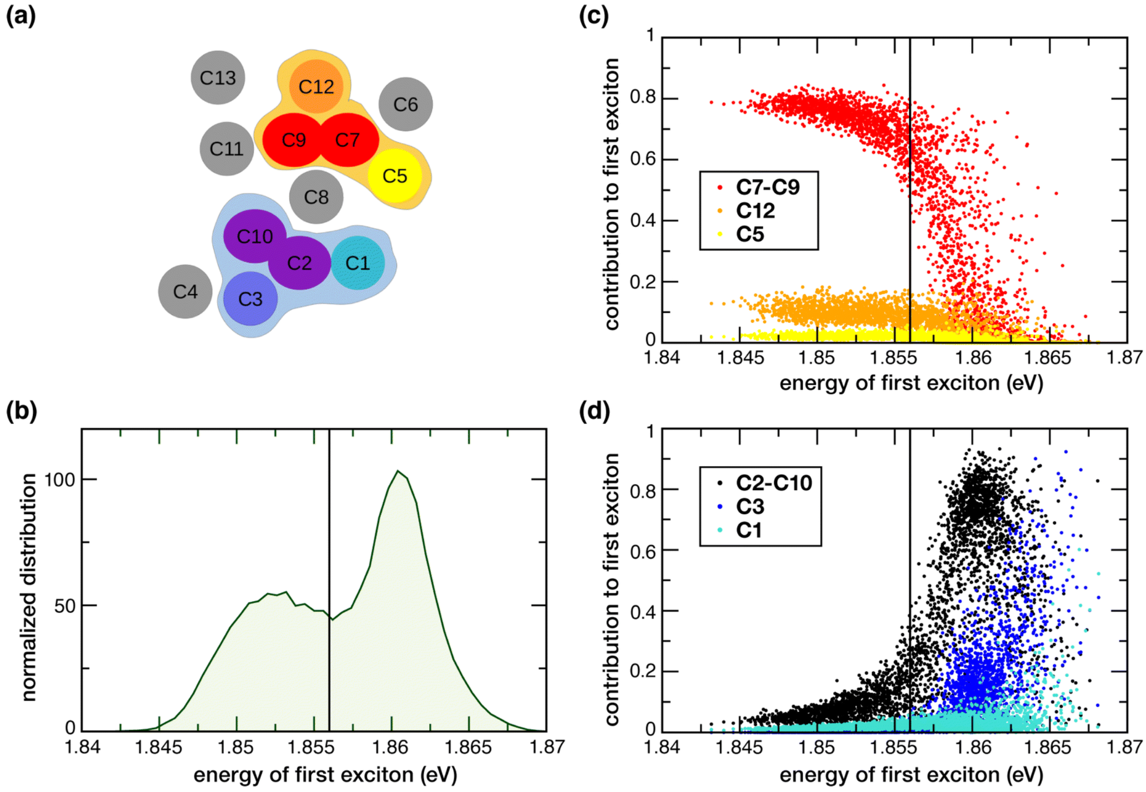


Fig. 6.9 (a) Chl network in CP43. Two groups of pigments are identified according to their contribution to the first excitonic state. The pigments highlighted in light orange contribute to the lowest-energy subpopulation, while the ones in light blue to the second lowest-energy subpopulation. **(b)** Energy distribution along the MD simulation of the first excitation energy of the first excitonic state (first exciton). **(c)** Contribution of **C5** (yellow), **C12** (orange) and the **C7−C9** dimer (red) to the first exciton as a function of the first excitation energy for each MD configuration. **(d)** Contribution of **C1** (cyan), **C3** (blue) and the **C2−C10** dimer (black) to the first exciton as a function of the excitation energy for each MD configuration.

It is important to note that the low-energy exciton is not always localized on the same Chl sites. We found two red-shifted trap-states that change depending on conformational fluctuations: one at lower energy (1.852 eV on average) with contribution from C7−C9 dimer, C12, C5, the other at slightly higher energy (1.860 eV on average) with contribution from C2−10 dimer, C3 and C1.

The lowest exciton shows two subpopulations that depend on fluctuations in the protein conformation, which implies two major conformational basins with slightly different excitonic energies. The analysis of the contribution of the pigments and dimers used as basis for the excitonic coupling calculations shows that the C7−C9 and C2−10 dimers contribute the most. Minor contributions arise from Chls C1, C3, C5, C8 and C12. Most importantly, two rather distinct groups of coupled pigments participating to the two subpopulations mentioned above, can be identified: the one at lower excitonic energy (peaked at around 1.852 eV) characterized by the contributions from the C7−C9 dimer, C12 and, as minor contribution, C5 and the other (peaked at around 1.860 eV) characterized by contributions from the C2−C10 dimer, C3 and, as minor contribution, C1 (the C8 contribution is not mentioned because it provides a similar contribution to the two groups). The weight of the contributions of the different monomeric and dimeric pigments to the first exciton is highlighted in **Fig. 6.9**. The C7−C9 dimer provides the major contribution (with a weight of around 0.8) to the most red-shifted trap, along with C12 (with a weight of around 0.12−0.15) and C5 (with a weight of around 0.08−0.05) (see **Fig. 6.9c**). Instead, The C2−C10 dimer provides the major contribution (with a weight in the range of 0.6-0.8) to the second trap, along with C3 (with a weight in the range of 0.10−0.30) and C1 (with a weight of around 0.05−0.10) (see **Fig. 6.9d**). In one investigation, Müh et al.[28] reported that two degenerate low-energy exciton transitions represent the lowest excited states of the two domains in the lumenal (containing C2 and C4) and stromal layers (containing C5 and C7−C11). Our results here support a similar scenario, in the sense of two excitonic domains in the lumenal and stromal layer, but with different contributions from the different pigments. It would nevertheless be interesting to have an estimate of the EET kinetics from the second trap (C2−C10, C3, C1) to the first (mostly red-shifted) trap (C7−C9, C12, C5), but this remains beyond the scope of the current work.

Based on our results, 3 out of the 4 Chls with the lowest site energies i.e., C3, C7, C9 (**Fig. 6.4b**) are located in the center of each layer. Interestingly, these pigments show strong excitonic coupling with the other group of Chls (**Table 6.3**). For instance, C9 is strongly coupled to C7 (-112.59 cm^{-1}) which is excitonically coupled to C5 (-98.32 cm^{-1}). We found that this pair possesses delocalized excited states (**Fig. 6.7**) in the low-energy regime, and contributes almost equally and majorly to the first excitonic state of CP43 (**Fig. 6.9c**). This identity of the low-energy trap state agrees with earlier assignments made by Müh et al.[28] and Shibata et al.[16] suggesting that Chls C9 and C7 belong to the same excitonic domain, as also seen in this work. Although some studies claim that the low energy sinks in the CP43/CP47 core antenna should be in close proximity to the RC pigments to facilitate efficient EET, this argument remained controversial as some simulations established that EET from the stromal layer to the RC to be equally efficient as EET from the lumenal layer.[41] Towards this, our assignment of a coherent excitonic domain comprising C7, C9 and C5 in the stromal layer has important functional implications towards EET from CP43 to the RC because C5 is indeed the closest CP43 pigment to peripheral $\mathrm{Chl}z_{\mathrm{D1}}$ as well as the RC pigments ($\mathrm{Chl}_{\mathrm{D1}}$ and $\mathrm{Pheo}_{\mathrm{D1}}$). This finding can be considered to be in agreement with recent studies by Yang et al.[19] because we identify C5 as part of the lowest excitonic domain in C43 and therefore as a pigment that can mediate energy transfer from CP43 to the RC.

Our calculation of the global excitonic states reveals another distinct excitonic domain comprising the C2−C10 dimer, C3 and C1 (**Fig. 6.9d**). Our TD-DFT results on C2−C10 indicated a low-lying CT state mixed with the excited state of C10 (**Fig. 6.8**). While recent computational studies have investigated the role of mixed excitonic-CT states towards the spectral tuning of “red” fluorescence states in PSI-Lhca4,[131] the involvement of a CT state in spectral tuning of the “red” states in PSII core antenna complexes have never been reported. Based on our current results, the C2−C10 interaction red-shifts the Q_y energy of C2 by ca. 24 meV, and most importantly this pair also forms a coherent excitonic domain along with the other “red” pigments C3 and C1 (**Fig. 6.4b**). Both results imply that the C2→C10 CT state likely plays a role in the spectral tuning of the “red” trap state within the lumenal layer of CP43.

6.4 Conclusion

In this work, we investigated the low-energy excitation manifold of the CP43 core antenna in PSII, utilizing a multiscale TD-DFT QM/MM approach combined with large-scale MD-PMM calculations. The site energies computed with both TD-DFT and PMM provide a qualitative agreement for the red-most pigments being distributed in two groups: C7 and C9 in the stromal, C1 and C3 in the lumenal layer. The excitonic couplings derived using PMM facilitated the identification of specific Chl pairs that exhibit strong inter-pigment interactions. TD-DFT calculations on selected Chl pairs revealed a delocalized excited state on the C7−C9 dimer.

Notably, calculations on the C2−C10 dimer identified a low-lying charge-transfer state mixed with the local excitation on C10. Further studies will be required to understand the physiological significance of this finding.

Finally, we reported global excitonic states involving all 13 Chls in CP43 by means of the MD-PMM procedure. Our findings led us to conclude that the lowest excited state is not localized on the same Chl site at each system configuration but its nature depends on the dynamics of the protein matrix. Specifically, we find two red-shifted excitonic domains, one in each layer of the thylakoid membrane. The lower energy excitonic state has contributions from C7−C9, C12 and C5. Although unique, this identity of the low-energy trap state aligns in several ways with previous literature assignments.[16, 19, 28] The concept of domains has been previously discussed in the literature, but our results present a new perspective. We demonstrate that the identity of these domains is influenced by conformational motions, and their relative energy can be thus modulated through conformational dynamics. This is particularly significant at room temperature, where conformational dynamics play a crucial role.

The coherent excitonic domain in the stromal layer (comprising C7, C9 and C5) carries functional implications for EET from CP43 to the RC because C5 is the closest CP43 pigment to peripheral Chlz_{D1} as well as the RC pigments.[19] The other excitonic domain involves the C2−C10 dimer along with C3 and C1 in the lumenal layer. Most importantly, the mixed excitonic-CT state on C2−C10 may play a role in the spectral tuning of the "red" pigments in the lumenal layer of CP43 or act as intermediate for quenching of excess excitation energy in PSII (photoprotection). Overall, this study establishes a refined basis for future kinetic modelling of EET pathways as well as for structure-based interpretation of spectroscopic properties of CP43, and contributes to an improved understanding of light harvesting and excitation energy transfer in oxygenic photosynthesis.

6.5 Supplementary Tables and Figures

Table 6.4 TD-DFT (ωB97X-D3(BJ)/def2-TZVP) excitation energies E (in eV) of the S_2/Q_x states and corresponding oscillator strengths for the CP43 Chls computed *in vacuo*, with QM/MM on an isolated CP43 protein, and with QM/MM in the complete PSII monomer. The geometries in all cases are derived from QM/MM optimizations within the PSII monomer. Shifts are reported in meV with respect to the gas-phase values.

	TD-DFT Gas-phase		**TD-DFT | QM/MM Isolated CP43**			**TD-DFT | QM/MM Full PSII monomer**		
Site	E	f_{osc}	E	f_{osc}	**Shift**	E	f_{osc}	**Shift**
C1	2.354	0.04	2.296	0.03	-58	2.260	0.02	-94
C2	2.372	0.04	2.307	0.03	-65	2.280	0.02	-92
C3	2.370	0.02	2.303	0.02	-67	2.283	0.01	-87
C4	2.427	0.03	2.359	0.01	-68	2.375	0.02	-52
C5	2.475	0.16	2.417	0.12	-58	2.440	0.15	-35
C6	2.383	0.06	2.457	0.09	74	2.464	0.11	81
C7	2.421	0.05	2.392	0.04	-29	2.382	0.05	-39
C8	2.394	0.05	2.412	0.03	18	2.452	0.07	58
C9	2.404	0.06	2.354	0.06	-50	2.342	0.06	-62
C10	2.362	0.07	2.424	0.11	62	2.450	0.12	88
C11	2.371	0.07	2.473	0.18	102	2.468	0.17	97
C12	2.418	0.08	2.433	0.09	15	2.409	0.09	-9
C13	2.423	0.07	2.413	0.08	-10	2.449	0.12	26

Table 6.5 TD-DFT Transition dipole moments of the CP43 Chls *in vacuo*.

	S_0–S_1					**S_0–S_2**				
Chl	**T (debye)**	**T^2 (au^2)**	**T_x (au)**	**T_y (au)**	**T_z (au)**	**T (debye)**	**T^2 (au^2)**	**T_x (au)**	**T_y (au)**	**T_z (au)**
C1	5.4223	4.5507	1.6598	0.9657	0.9292	1.7877	0.4947	-0.0116	-0.3280	-0.6220
C2	5.5418	4.7535	0.7559	1.9560	-0.5969	1.9422	0.5839	0.0872	-0.4145	0.6360
C3	5.6286	4.9036	1.6110	-1.4899	-0.2976	1.6609	0.4270	-0.2455	0.2258	0.5619
C4	5.5886	4.8341	-1.4121	-1.5491	-0.6636	1.8808	0.5475	0.1257	0.3485	0.6405
C5	5.5772	4.8146	1.3595	-1.2380	-1.1974	2.8896	1.2924	-0.1356	0.1439	1.1195
C6	5.4903	4.6656	0.5175	2.0624	0.3799	2.5398	0.9984	-0.4580	-0.5103	-0.7268
C7	5.6640	4.9655	2.0657	0.4489	-0.7050	2.0711	0.6639	-0.6712	-0.3362	-0.3168
C8	5.5771	4.8142	-1.5829	1.2619	-0.8464	2.3161	0.8303	0.2265	-0.3135	0.8250
C9	5.5291	4.7318	0.7763	2.0238	-0.1832	2.1829	0.7375	0.0540	0.5222	-0.6796
C10	5.4721	4.6347	0.7634	1.8828	-0.7121	2.8377	1.2464	0.2472	-0.5897	0.9152
C11	5.4245	4.5545	-0.8384	0.8208	-1.7827	2.7056	1.1331	0.1478	-0.0549	-1.0527
C12	5.4414	4.5830	1.0225	-1.8759	0.1356	2.6893	1.1194	-0.2525	0.7593	0.6922
C13	5.5280	4.7298	-1.7554	-0.9651	0.8468	2.8140	1.2257	0.5441	-0.0466	-0.9631

Table 6.6 Comparison of CP43 Chl site energies from the literature. All values are in eV.

Site	Muh et al.[28] [a]	Muh et al.[28] [b]	Shibata et al.[16] [c]	Shibata et al.[16] [d]	Sarngadharan et al.[27]	Saito et al.[61]	QM/MM (this work)	PMM (this work)
C1	1.839	1.843	1.853	1.842	1.818	2.127	1.846	1.873
C2	1.851	1.857	1.860	1.863	1.831	2.154	1.867	1.894
C3	1.843	1.840	1.853	1.862	1.817	2.114	1.861	1.867
C4	1.845	1.843	1.831	1.829	1.866	2.142	1.928	1.945
C5	1.846	1.847	1.853	1.834	1.842	2.152	1.938	1.915
C6	1.861	1.859	1.853	1.855	1.851	2.129	1.970	1.963
C7	1.839	1.838	1.833	1.841	1.842	2.122	1.885	1.890
C8	1.834	1.832	1.846	1.856	1.841	2.122	1.915	1.895
C9	1.828	1.834	1.825	1.818	1.848	2.096	1.881	1.912
C10	1.836	1.835	1.884	1.902	1.845	2.115	1.955	1.942
C11	1.823	1.823	1.862	1.830	1.831	2.136	1.986	1.964
C12	1.833	1.839	1.849	1.842	1.839	2.110	1.926	1.925
C13	1.861	1.858	1.873	1.870	1.865	2.142	1.961	1.937

[a] Based on 3ARC[35]

[b] Based on 4IL6[36]

[c] Fit of linear optical spectra

[d] Fit of absorption and LD spectra

Table 6.7 Relative center-to-center distances (in Å) of the 13 CP43 Chls, Chlz$_{D1}$ and D1-RC pigments based on the 3WU2 crystal structure.[35] Distances less than 12 Å are shown in red, 12–15 Å are shown in pink. In CP43, C5 is closest the D1 pigments (Chlz$_{D1}$, Pheo$_{D1}$ and Chl$_{D1}$) while C4 is closer to P$_{D1}$ (distances in green).

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**Chlz$_{D1}$**
C2	**13.1**													
C3	**12.7**	**12.1**												
C4	22.0	**11.2**	16.3											
C5	23.9	20.9	29.3	30.5										
C6	16.9	24.1	25.3	35.1	16.9									
C7	18.8	18.9	22.1	29.4	**12.1**	**8.7**								
C8	26.5	17.5	24.9	23.6	**14.6**	24.1	15.5							
C9	22.5	16.4	19.1	23.6	17.9	19.9	**11.7**	**9.1**						
C10	20.8	**10.2**	15.1	**14.3**	21.6	25.8	18.3	**11.4**	**9.6**					
C11	31.2	21.9	24.5	24.4	25.9	30.5	22.2	**11.9**	**10.7**	**11.8**				
C12	20.0	20.7	16.4	28.6	24.7	16.7	**13.3**	20.8	**12.0**	17.5	20.0			
C13	28.0	26.0	19.9	30.6	32.8	26.8	22.5	24.7	16.1	19.9	18.3	**10.4**		
Chlz$_{D1}$	33.7	34.3	43.7	42.4	**21.9**	31.2	31.4	34.9	38.9	39.8	46.6	43.8	53.0	
Chl$_{D1}$	41.0	33.3	45.4	35.4	**28.5**	44.0	39.2	31.9	39.6	36.3	41.8	49.1	55.1	24.7
Pheo$_{D1}$	39.0	30.8	42.6	34.7	**21.5**	38.3	32.7	24.4	32.7	31.4	34.7	43.0	48.7	24.3
P$_{D1}$	38.9	29.1	40.8	**28.1**	30.6	45.1	39.3	29.7	37.0	31.6	37.4	46.8	51.3	32.0

Table 6.8 QM/MM and TD-DFT (ωB97X-D3(BJ)/def2-TZVP) computed excitation energies (in eV) and corresponding oscillator strengths and NTO analysis for selected Chl pairs in CP43. The geometries in all cases are derived from QM/MM optimizations of the pigments embedded inside the complete PSII monomer for **snapshot 1**. (LE: local excitations, CT: Charge Transfer states).

Pair	Mg-Mg (Å)	State	Es (eV)	f_{osc}	NTO character	weight
C12-C13	10.1	1	1.920	0.21	LE (C12)	0.84
					LE (C12)	0.19
		2	1.967	0.24	LE (C13)	0.83
					LE (C13)	0.19
		3	2.410	0.06	LE (C12)	0.77
					LE (C12)	0.24
		4	2.46	0.15	LE (C13)	0.81
					LE (C13)	0.19
		5	3.222	0.12	delocalized exciton + CT	0.88
C2-C10	10.4	1	1.843	0.13	LE(C2)	0.84
					LE(C2)	0.20
		2	1.943	0.26	**LE (C10) + CT (C2 → C10)**	**0.86**
					LE (C10)	0.10
		3	2.001	0.07	**LE (C10) + CT (C2 → C10)**	**0.89**
					LE (C10)	0.10
		4	2.255	0.03	LE(C2)	0.69
					LE(C2)	0.34
		5	2.465	0.17	LE (C10)	0.83
					LE (C10)	0.19
C6-C7	9.0	1	1.885	0.13	LE(C7)	0.84
					LE(C7)	0.20
		2	1.963	0.26	LE (C6)	0.84
					LE (C6)	0.19
		3	2.384	0.07	LE(C7)	0.74
					LE(C7)	0.28
		4	2.453	0.03	LE (C6)	0.79
					LE (C6)	0.21
		5	3.218	0.17	delocalized exciton + CT	0.80
C8-C9	9.5	1	1.882	0.23	LE(C9)	0.82
					LE(C9)	0.20
		2	1.913	0.24	LE (C8)	0.83
					LE (C8)	0.18
		3	2.352	0.04	LE(C9)	0.77
					LE(C9)	0.25
		4	2.441	0.06	LE (C8)	0.75
					LE (C8)	0.26
		5	3.251	0.33	LE(C9)	0.67
					LE(C9)	0.27
C9-C10	9.3	1	1.888	0.17	LE(C9)	0.83
					LE(C9)	0.20
		2	1.942	0.25	LE(C10)	0.84
					LE(C10)	0.18
		3	2.365	0.04	LE(C9)	0.76
					LE(C9)	0.26
		4	2.433	0.11	LE(C10)	0.81
					LE(C10)	0.21
		5	3.094	0.00	CT (C10 → C9)	1.00
C7-C9	11.4	1	1.868	0.39	LE (C9)	0.44
					LE (C7)	0.41
		2	1.891	0.09	LE (C7)	0.43

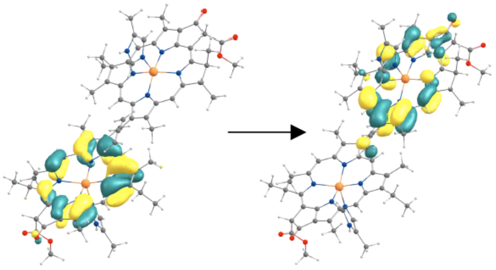
					LE (C9)	0.40
		3	2.350	0.07	LE (C9)	0.73
					LE (C9)	0.24
		4	2.37	0.04	LE (C7)	0.69
					LE (C7)	0.28
		5	3.278	0.30	LE (C9)	0.44
					LE (C7)	0.28
C9-C11	11.0	1	1.864	0.30	LE (C9)	0.83
					LE (C9)	0.20
		2	1.981	0.20	LE (C11)	0.83
					LE (C11)	0.18
		3	2.329	0.07	LE (C9)	0.76
					LE (C9)	0.27
		4	2.466	0.14	LE (C11)	0.83
					LE (C11)	0.19
		5	3.071	0.00	CT (C9 → C11)	1.00
C2-C4	11.8	1	1.868	0.32	LE(C2)	0.75
					LE(C2)	0.20
		2	1.899	0.098	LE(C4)	0.73
					LE(C4)	0.22
		3	2.291	0.02	LE(C2)	0.69
					LE(C2)	0.34
		4	2.320	0.00	LE(C4)	0.63
					LE(C4)	0.40
		5	3.166	0.64	LE(C4)	0.61
					LE(C4)	0.42
C8-C10	11.2	1	1.899	0.23	LE(C8)	0.80
					LE(C8)	0.17
		2	1.950	0.24	LE(C10)	0.80
					LE(C10)	0.17
		3	2.439	0.16	LE(C8)	0.62
					LE(C8)	0.22
		4	2.445	0.04	LE(C10)	0.67
					LE(C10)	0.18
		5	3.127	0.00	CT (C10 → C8)	1.00
C10-C11	11.9	1	1.958	0.34	LE(C10)	0.80
					LE(C10)	0.17
		2	1.994	0.18	LE(C11)	0.79
					LE(C11)	0.17
		3	2.451	0.26	LE(C10)	0.75
					LE(C10)	0.17
		4	2.484	0.11	LE(C11)	0.76
					LE(C11)	0.16
		5	2.546	0.00	CT (C10 → C11)	1.00
C5-C7	12.1	1	1.882	0.35	LE(C7)	0.78
					LE(C7)	0.18
		2	1.933	0.17	LE(C5)	0.79
					LE(C5)	0.16
		3	2.392	0.05	LE(C7)	0.72
					LE(C7)	0.30
		4	2.440	0.13	LE(C5)	0.83
					LE(C5)	0.19
		5	3.331	0.50	LE(C7)	0.64
					LE(C7)	0.34

Table 6.9 QM/MM and TD-DFT (ωB97X-D3(BJ)/def2-TZVP) computed excitation energies (in eV) and corresponding oscillator strengths and NTO analysis for the **C2-C10** pair in CP43. The geometries in all cases are derived from QM/MM optimizations of the pigments embedded inside the complete PSII monomer in **snapshots 2-10** (LE: local excitations, CT: Charge Transfer states).

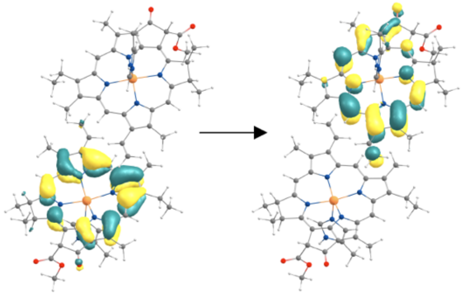
Snapshot	Mg-Mg (Å)	State	Es (eV)	f_{osc}	NTO character	weights
2	10.6	1	1.875	0.20	LE(C2)	0.83
					LE(C2)	0.22
		2	1.967	0.26	LE (C10)	0.85
					LE (C10)	0.18
		3	2.230	0.00	**CT (C2 → C10)**	**0.98**
		4	2.320	0.04	LE(C2)	0.69
					LE(C2)	0.34
		5	2.478	0.15	LE (C10)	0.80
					LE (C10)	0.21
3	10.3	1	1.838	0.12	**LE (C2) + CT (C2 → C10)**	**0.87**
					LE(C2)	0.17
		2	1.901	0.10	**LE (C2) + CT (C2 → C10)**	**0.94**
					LE(C2)	0.06
		3	1.998	0.23	LE (C10)	0.85
					LE (C10)	0.18
		4	2.273	0.02	LE(C2)	0.68
					LE(C2)	0.35
		5	2.447	0.01	**CT (C2 → C10)**	**0.98**
4	10.3	1	1.868	0.18	LE(C2)	0.83
					LE(C2)	0.22
		2	1.979	0.27	LE (C10)	0.85
					LE (C10)	0.17
		3	2.099	0.00	**CT (C2 → C10)**	**0.99**
		4	2.298	0.04	LE(C2)	0.69
					LE(C2)	0.34
		5	2.491	0.18	LE (C10)	0.81
					LE (C10)	0.20
5	10.1	1	1.869	0.18	LE(C2)	0.83
					LE(C2)	0.22
		2	1.953	0.28	LE (C10)	0.85
					LE (C10)	0.18
		3	2.237	0.00	**CT (C2 → C10)**	**0.98**
		4	2.279	0.04	LE(C2)	0.69
					LE(C2)	0.34
		5	2.461	0.14	LE (C10)	0.80
					LE (C10)	0.22
6	10.1	1	1.855	0.19	LE(C2)	0.83
					LE(C2)	0.22
		2	1.981	0.29	LE (C10)	0.85
					LE (C10)	0.17
		3	2.237	0.00	**CT (C2 → C10)**	**0.90**
		4	2.288	0.03	LE(C2)	0.58
					LE(C2)	0.44
		5	2.509	0.17	LE (C10)	0.81
					LE (C10)	0.21
7	10.2	1	1.842	0.18	LE(C2)	0.84
					LE(C2)	0.21
		2	1.966	0.15	**LE (C10) + CT (C2 → C10)**	**0.92**
					LE (C10)	0.10
		3	2.018	0.14	**LE (C10) + CT (C2 → C10)**	**0.88**

					LE (C10)	0.12
		4	2.271	0.04	LE(C2)	0.69
					LE(C2)	0.34
		5	2.509	0.14	LE (C10)	0.87
					LE (C10)	0.14
8	10.4	1	1.859	0.17	LE(C2)	0.83
					LE(C2)	0.21
		2	1.939	0.31	LE (C10)	0.85
					LE (C10)	0.17
		3	2.102	0.00	**CT (C2 → C10)**	**0.99**
		4	2.282	0.04	LE(C2)	0.70
					LE(C2)	0.33
		5	2.461	0.15	LE (C10)	0.81
					LE (C10)	0.20
9	9.9	1	1.803	0.11	**LE (C2) + CT (C2 → C10)**	**0.88**
					LE(C2)	0.15
		2	1.895	0.13	**LE (C2) + CT (C2 → C10)**	**0.91**
					LE(C2)	0.10
		3	1.955	0.23	LE (C10)	0.86
					LE (C10)	0.17
		4	2.240	0.03	LE(C2)	0.69
					LE(C2)	0.34
		5	2.440	0.05	**LE (C10) + CT (C2 → C10)**	**0.94**
10	10.5	1	1.856	0.18	LE(C2)	0.83
					LE(C2)	0.23
		2	1.923	0.29	LE (C10)	0.85
					LE (C10)	0.18
		3	2.257	0.00	**LE (C2) + CT (C2 → C10)**	**0.59**
					LE (C10)	0.42
		4	2.309	0.04	**LE (C2) + CT (C2 → C10)**	**0.76**
					LE(C2)	0.24
		5	2.436	0.13	LE (C10)	0.80
					LE (C10)	0.22

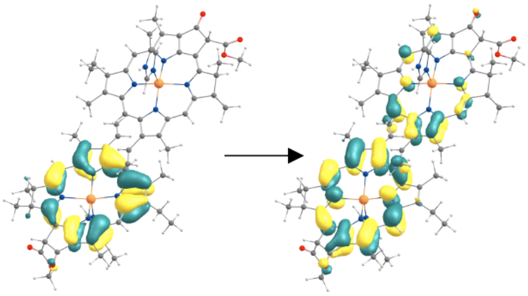
Snapshot 2 (5 ns), S3

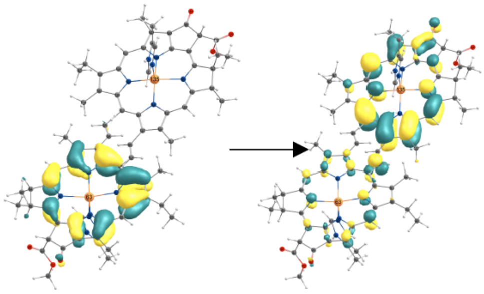


Snapshot 8 (35 ns), S3

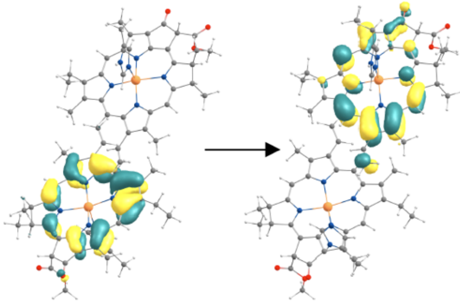


Snapshot 3 (10 ns), S1 and S2

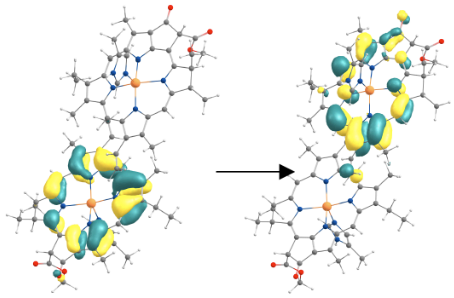




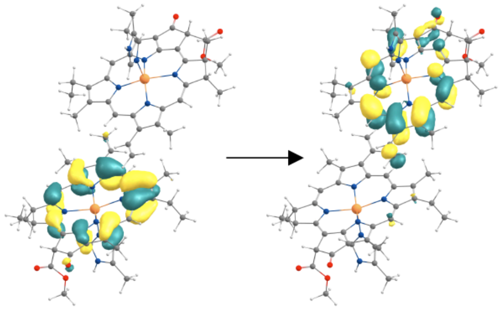
Snapshot 4 (15 ns), S3



Snapshot 5 (20 ns), S3



Snapshot 6 (25 ns), S3



Snapshot 7 (30 ns), S2

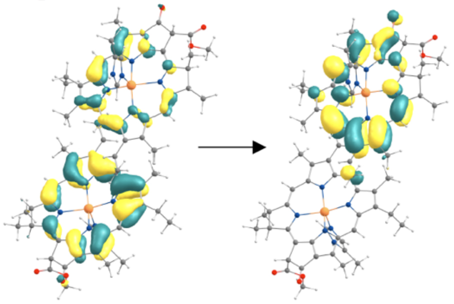
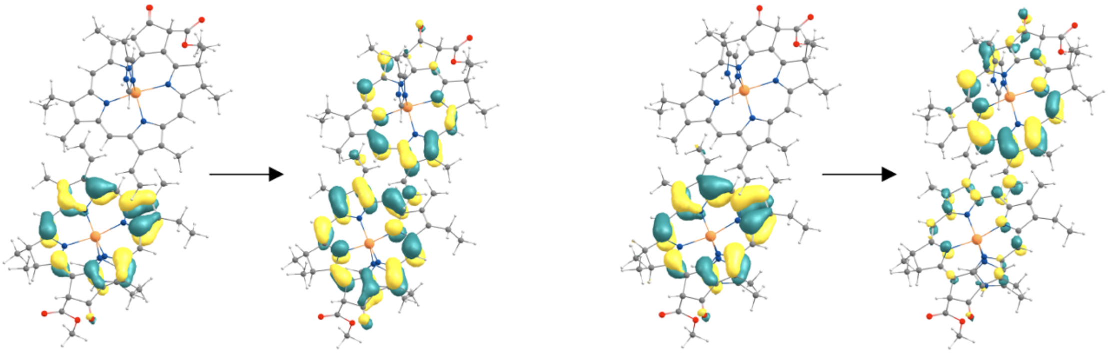


Fig. 6.10 Natural Transition Orbitals (NTOs) for the lowest excited state with significant CT character for the **C2-C10** dimer in snapshots 2-10.

Snapshot 9 (40 ns), S1 and S2



Snapshot 10 (45ns), S3

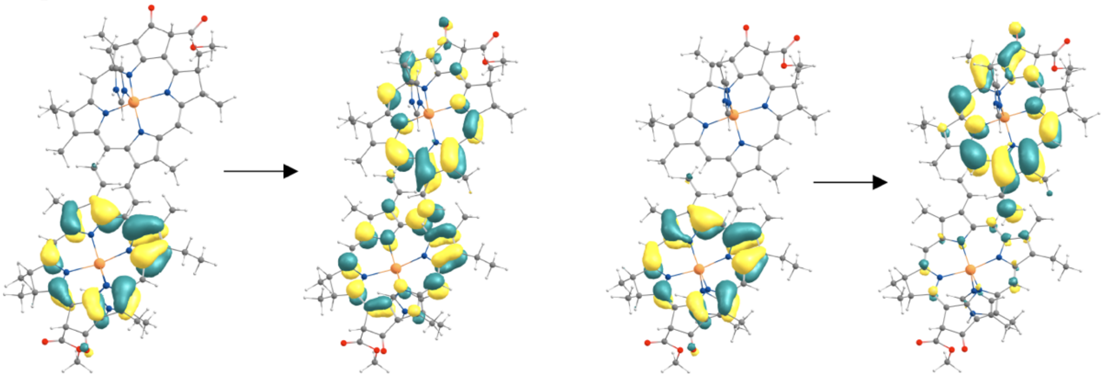


Fig. 6.10 (continued). Natural Transition Orbitals (NTOs) for the lowest excited state with significant CT character for the **C2-C10** dimer in snapshots 2-10.

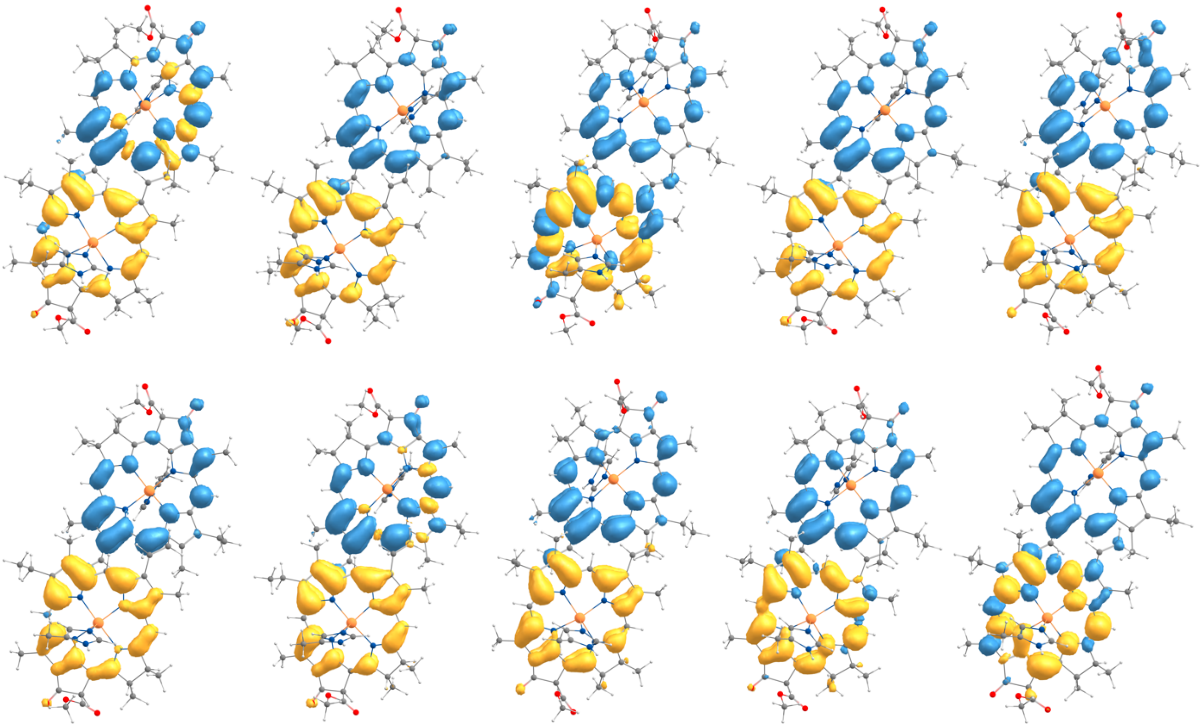


Fig. 6.11 TD-DFT difference densities for the lowest excited states with CT character for the **C2-C10** dimer in the 'crystal-like' (snapshot 1) and snapshots 2-10.

6.6 REFERENCES

1. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*, Wiley, Chichester, 2nd edn., 2014.
2. D. Shevela, L. O. Björn and Govindjee, *Photosynthesis: Solar Energy for Life*, World Scientific, Singapore, 2017.
3. R. Croce, R. v. Grondelle, H. v. Amerongen and I. v. Stokkum, eds., *Light Harvesting in Photosynthesis*, Taylor & Francis/CRC Press, Boca Raton, 2018.
4. J.-R. Shen, *Annu. Rev. Plant Biol.*, 2015, **66**, 23-48.
5. C. Pagliano, G. Saracco and J. Barber, *Photosynth. Res.*, 2013, **116**, 167-188.
6. T. M. Bricker, J. L. Roose, R. D. Fagerlund, L. K. Frankel and J. J. Eaton-Rye, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 121-142.
7. N. Cox, D. A. Pantazis and W. Lubitz, *Annu. Rev. Biochem.*, 2020, **89**, 795-820.
8. D. A. Pantazis, *ACS Catal.*, 2018, **8**, 9477-9507.
9. D. A. Pantazis, in *Hydrogen Production and Energy Transition*, ed. M. Van de Voorde, De Gruyter, 2021, vol. 1, pp. 427-468.
10. K. Yamaguchi, S. Yamanaka, H. Isobe, M. Shoji, T. Kawakami and K. Miyagawa, in *Photosynthesis: Molecular Approaches to Solar Energy Conversion*, eds. J.-R. Shen, K. Satoh and S. I. Allakhverdiev, Springer International Publishing, Cham, 2021, DOI: 10.1007/978-3-030-67407-6_2, pp. 39-80.
11. P. E. M. Siegbahn, *Biochim. Biophys. Acta Bioenerg.*, 2013, **1827**, 1003-1019.
12. F. Müh, C. Glöckner, J. Hellmich and A. Zouni, *Biochim. Biophys. Acta Bioenerg.*, 2012, **1817**, 44-65.
13. T. M. Bricker and L. K. Frankel, *Photosynth. Res.*, 2002, **72**, 131-146.
14. E. G. Andrizhiyevskaya, D. Frolov, R. Van Grondelle and J. P. Dekker, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4810.
15. F. L. De Weerd, I. H. M. Van Stokkum, H. Van Amerongen, J. P. Dekker and R. Van Grondelle, *Biophys. J.*, 2002, **82**, 1586-1597.
16. Y. Shibata, S. Nishi, K. Kawakami, J.-R. Shen and T. Renger, *J. Am. Chem. Soc.*, 2013, **135**, 6903-6914.
17. R. Croce and H. van Amerongen, *Science*, 2020, **369**, eaay2058.
18. F. Müh and A. Zouni, *Protein Sci.*, 2020, **29**, 1090-1119.
19. S.-J. Yang, E. A. Arsenault, K. Orcutt, M. Iwai, Y. Yoneda and G. R. Fleming, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2208033119.
20. H. van Amerongen and R. Croce, *Photosynth. Res.*, 2013, **116**, 251-263.
21. C. Kreisbeck and A. Aspuru-Guzik, *Chem. Sci.*, 2016, **7**, 4174-4183.
22. R. Croce and H. van Amerongen, *Nat. Chem. Biol.*, 2014, **10**, 492-501.
23. K. Broess, G. Trinkunas, C. D. Van Der Weij-De Wit, J. P. Dekker, A. Van Hoek and H. Van Amerongen, *Biophys. J.*, 2006, **91**, 3776-3786.
24. T. Mirkovic, E. E. Ostroumov, J. M. Anna, R. van Grondelle, Govindjee and G. D. Scholes, *Chem. Rev.*, 2017, **117**, 249-293.
25. C. König and J. Neugebauer, *ChemPhysChem*, 2012, **13**, 386-425.
26. A. Sirohiwal and D. A. Pantazis, *Acc. Chem. Res.*, 2023, **56**, 2921-2932.
27. P. Sarngadharan, S. Maity and U. Kleinekathöfer, *J. Chem. Phys.*, 2022, **156**, 215101.
28. F. Müh, M. Plöckinger, H. Ortmayer, M. Schmidt am Busch, D. Lindorfer, J. Adolphs and T. Renger, *J. Photochem. Photobiol., B*, 2015, **152**, 286-300.
29. F. Müh, M. E.-A. Madjet and T. Renger, *Photosynth. Res.*, 2012, **111**, 87-101.
30. A. P. Casazza, M. Szczepaniak, M. G. Müller, G. Zucchelli and A. R. Holzwarth, *Biochim. Biophys. Acta Bioenerg.*, 2010, **1797**, 1606-1616.

31. M.-L. Groot, R. N. Frese, F. L. De Weerd, K. Bromek, Å. Pettersson, E. J. G. Peterman, . . . J. P. Dekker, *Biophys. J.*, 1999, **77**, 3328-3340.
32. A. Sirohiwal, F. Neese and D. A. Pantazis, *Chem. Sci.*, 2021, **12**, 4463-4476.
33. B. Loll, J. Kern, W. Saenger, A. Zouni and J. Biesiadka, *Nature*, 2005, **438**, 1040-1044.
34. A. Guskov, J. Kern, A. Gabdulkhakov, M. Broser, A. Zouni and W. Saenger, *Nat. Struct. Mol. Biol.*, 2009, **16**, 334-342.
35. Y. Umena, K. Kawakami, J.-R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
36. F. H. M. Koua, Y. Umena, K. Kawakami and J.-R. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 3889-3894.
37. J. L. Hughes, R. Picorel, M. Seibert and E. Krausz, *Biochemistry*, 2006, **45**, 12345-12357.
38. R. Jankowiak, V. Zazubovich, M. Rätsep, S. Matsuzaki, M. Alfonso, R. Picorel, . . . G. J. Small, *J. Phys. Chem. B*, 2000, **104**, 11805-11815.
39. J. Hall, T. Renger, R. Picorel and E. Krausz, *Biochim. Biophys. Acta Bioenerg.*, 2016, **1857**, 115-128.
40. M. Di Donato, R. van Grondelle, I. H. M. van Stokkum and M. L. Groot, *J. Phys. Chem. B*, 2007, **111**, 7345-7352.
41. G. Raszewski and T. Renger, *J. Am. Chem. Soc.*, 2008, **130**, 4431-4446.
42. G. Raszewski, W. Saenger and T. Renger, *Biophys. J.*, 2005, **88**, 986-998.
43. T. Renger and E. Schlodder, *ChemPhysChem*, 2010, **11**, 1141-1153.
44. A. R. Holzwarth, M. G. Muller, M. Reus, M. Nowaczyk, J. Sander and M. Rogner, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6895-6900.
45. Y. Miloslavina, M. Szczepaniak, M. G. Müller, J. Sander, M. Nowaczyk, M. Rögner and A. R. Holzwarth, *Biochemistry*, 2006, **45**, 2436-2442.
46. M. Kaucikas, K. Maghlaoui, J. Barber, T. Renger and J. J. Van Thor, *Nat. Commun.*, 2016, **7**, 13977.
47. N. P. Pawlowicz, M. L. Groot, I. H. M. Van Stokkum, J. Breton and R. Van Grondelle, *Biophys. J.*, 2007, **93**, 2732-2742.
48. E. A. Arsenault, Y. Yoneda, M. Iwai, K. K. Niyogi and G. R. Fleming, *Nat. Commun.*, 2020, **11**, 1460.
49. J. Pan, A. Gelzinis, V. Chorošajev, M. Vengris, S. S. Senlik, J.-R. Shen, . . . J. P. Ogilvie, *Phys. Chem. Chem. Phys.*, 2017, **19**, 15356-15367.
50. E. Cignoni, V. Slama, L. Cupellini and B. Mennucci, *J. Chem. Phys.*, 2022, **156**, 120901.
51. C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294-343.
52. B. Mennucci, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6583-6594.
53. S. Maity and U. Kleinekathöfer, *Photosynth. Res.*, 2023, **156**, 147-162.
54. D. I. G. Bennett, K. Amarnath and G. R. Fleming, *J. Am. Chem. Soc.*, 2013, **135**, 9164-9173.
55. N. Liguori, R. Croce, S. J. Marrink and S. Thallmair, *Photosynth. Res.*, 2020, **144**, 273-295.
56. T. Renger and F. Müh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3348-3371.
57. S. Maity, P. Sarngadharan, V. Daskalakis and U. Kleinekathöfer, *J. Chem. Phys.*, 2021, **155**, 055103.
58. I. Schelter, J. M. Foerster, A. T. Gardiner, A. W. Roszak, R. J. Cogdell, G. M. Ullmann, . . . S. Kümmel, *J. Chem. Phys.*, 2019, **151**, 134114.
59. C. W. Kim, B. Choi and Y. M. Rhee, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3310-3319.
60. S. Maity, V. Daskalakis, M. Elstner and U. Kleinekathöfer, *Phys. Chem. Chem. Phys.*, 2021, **23**, 7407-7417.
61. K. Saito, K. Mitsuhashi, H. Tamura and H. Ishikita, *Biophys. J.*, 2023, **122**, 470-483.

62. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Am. Chem. Soc.*, 2020, **142**, 18174-18190.
63. S. Bhattacharjee, F. Neese and D. A. Pantazis, *Chem. Sci.*, 2023, **14**, 9503-9516.
64. A. Sirohiwal and D. A. Pantazis, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200356.
65. M. Capone, A. Sirohiwal, M. Aschi, D. A. Pantazis and I. Daidone, *Angew. Chem. Int. Ed.*, 2023, **62**, e202216276.
66. L. Zanetti-Polzi, S. Corni, I. Daidone and A. Amadei, *Phys. Chem. Chem. Phys.*, 2016, **18**, 18450-18459.
67. I. Daidone, L. Paltrinieri, A. Amadei, G. Battistuzzi, M. Sola, M. Borsari and C. A. Bortolotti, *J. Phys. Chem. B*, 2014, **118**, 7554-7560.
68. L. Zanetti-Polzi, M. Aschi, A. Amadei and I. Daidone, *J. Phys. Chem. Lett.*, 2017, **8**, 3321-3327.
69. L. Zanetti-Polzi, M. D. Smith, C. Chipot, J. C. Gumbart, D. L. Lynch, A. Pavlova, . . . I. Daidone, *J. Phys. Chem. Lett.*, 2021, **12**, 4195-4202.
70. S. Schott-Verdugo and H. Gohlke, *J. Chem. Inf. Model.*, 2019, **59**, 2522-2528.
71. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
72. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
73. C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.*, 2014, **10**, 865-879.
74. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
75. N. Sakashita, H. C. Watanabe, T. Ikeda, K. Saito and H. Ishikita, *Biochemistry*, 2017, **56**, 3049-3057.
76. F. Guerra, M. Siemers, C. Mielack and A.-N. Bondar, *J. Phys. Chem. B*, 2018, **122**, 4625-4641.
77. M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, 2003, **24**, 129-142.
78. F. Guerra, S. Adam and A.-N. Bondar, *J. Mol. Graphics Modell.*, 2015, **58**, 30-39.
79. P. Li and K. M. Merz, *Chem. Rev.*, 2017, **117**, 1564-1686.
80. I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B*, 2009, **113**, 13279-13290.
81. I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B*, 2008, **112**, 9020-9041.
82. R. J. Loncharich, B. R. Brooks and R. W. Pastor, *Biopolymers*, 1992, **32**, 523-535.
83. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
84. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
85. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
86. D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., . . . R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
87. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878-3888.
88. F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.
89. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
90. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
91. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396-1396.
92. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
93. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
94. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.

95. F. Neese, *J. Comput. Chem.*, 2003, **24**, 1740-1747.
96. F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057-1065.
97. N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904.
98. A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsák and D. A. Pantazis, *J. Phys. Chem. B*, 2020, **124**, 8761–8771.
99. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
100. R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775-4777.
101. H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198-1229.
102. M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708-2728.
103. S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife and D. R. Salahub, *Int. J. Quantum Chem.*, 2018, **118**.
104. S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes and M. J. Ramos, *WIREs Computational Molecular Science*, 2017, **7**, e1281.
105. L. Zanetti-Polzi, A. Amadei, R. Djemili, S. Durot, L. Schoepff, V. Heitz, . . . I. Daidone, *J. Phys. Chem. C*, 2019, **123**, 13094-13103.
106. E. P. Kenny and I. Kassal, *J. Phys. Chem. B*, 2016, **120**, 25-32.
107. C. Friedl, D. G. Fedorov and T. Renger, *Phys. Chem. Chem. Phys.*, 2022, **24**, 5014-5038.
108. C. G. Chen, A. N. Nardi, A. Amadei and M. D'Abramo, *J. Chem. Theory Comput.*, 2023, **19**, 33-41.
109. A. Sirohiwal and D. A. Pantazis, *J. Am. Chem. Soc.*, 2022, **144**, 22035-22050.
110. S. Barazzouk, L. Bekalé and S. Hotchandani, *J. Mater. Chem.*, 2012, **22**, 25316-25324.
111. A. Sirohiwal, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2021, **17**, 1858-1873.
112. F. Cardoso Ramos, M. Nottoli, L. Cupellini and B. Mennucci, *Chem. Sci.*, 2019, **10**, 9650-9662.
113. B. Mennucci, in *Green Chemistry and Computational Chemistry*, ed. L. Mammino, Elsevier, 2022, pp. 263-279.
114. L. Cupellini, M. Corbella, B. Mennucci and C. Curutchet, *WIREs Computational Molecular Science*, 2019, **9**, e1392.
115. M. Nottoli, L. Cupellini, F. Lipparini, G. Granucci and B. Mennucci, *Annu. Rev. Phys. Chem.*, 2021, **72**, 489-513.
116. H. H. Nguyen, Y. Song, E. L. Maret, Y. Silori, R. Willow, C. F. Yocum and J. P. Ogilvie, *Science Advances*, 2023, **9**, eade7190.
117. S.-T. Hsieh, L. Zhang, D.-W. Ye, X. Huang and Y.-C. Cheng, *Faraday Discuss.*, 2019, **216**, 94-115.
118. M. E.-A. Madjet, F. Müh and T. Renger, *J. Phys. Chem. B*, 2009, **113**, 12603-12614.
119. M. E. Madjet, A. Abdurahman and T. Renger, *J. Phys. Chem. B*, 2006, **110**, 17268-17281.
120. F. Caycedo-Soler, A. Mattioni, J. Lim, T. Renger, S. F. Huelga and M. B. Plenio, *Nat. Commun.*, 2022, **13**, 2912.
121. V. I. Novoderezhkin, R. Croce, M. Wahadoszamen, I. Polukhina, E. Romero and R. Van Grondelle, *Phys. Chem. Chem. Phys.*, 2016, **18**, 19368-19377.
122. M. Nottoli and F. Lipparini, *J. Chem. Phys.*, 2020, **153**, 224108.
123. S. Sen, V. Mascoli, N. Liguori, R. Croce and L. Visscher, *J. Phys. Chem. A*, 2021, **125**, 4313-4322.
124. M. Nottoli, S. Jurinovich, L. Cupellini, A. T. Gardiner, R. Cogdell and B. Mennucci, *Photosynth. Res.*, 2018, **137**, 215-226.

125. S. Bourne-Worster, O. Feighan and F. R. Manby, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2210811120.
126. T. P. J. Krüger and R. Van Grondelle, *J. Phys. B: At., Mol. Opt. Phys.*, 2017, **50**, 132001.
127. T. K. Ahn, T. J. Avenson, M. Ballottari, Y.-C. Cheng, K. K. Niyogi, R. Bassi and G. R. Fleming, *Science*, 2008, **320**, 794-797.
128. Y.-C. Cheng, T. K. Ahn, T. J. Avenson, D. Zigmantas, K. K. Niyogi, M. Ballottari, . . . G. R. Fleming, *J. Phys. Chem. B*, 2008, **112**, 13418-13423.
129. C. Ramanan, M. Ferretti, H. Van Roon, V. I. Novoderezhkin and R. Van Grondelle, *Phys. Chem. Chem. Phys.*, 2017, **19**, 22877-22886.
130. E. E. Ostroumov, J. P. Götze, M. Reus, P. H. Lambrev and A. R. Holzwarth, *Photosynth. Res.*, 2020, **144**, 171-193.
131. V. Sláma, L. Cupellini, V. Mascoli, N. Liguori, R. Croce and B. Mennucci, *J. Phys. Chem. Lett.*, 2023, **14**, 8345-8352.

7

Multiscale Modelling of the Soluble Methane Monooxygenase Hydroxylase Complex‡

7.1 Introduction

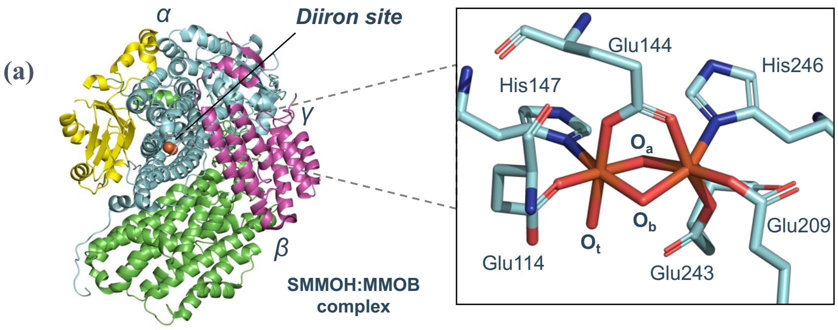
Methanotrophic bacteria thrive at the interface of aerobic and anaerobic environments, where they use a specialized enzyme called methane monooxygenase (MMO) to efficiently oxidize the methane (CH_4) obtained from anaerobic metabolism.[1-3] MMO activates molecular O_2, inserting one atom into the non-activated C–H bond of CH_4 to produce methanol (CH_3OH) without further oxidation. Subsequent enzymes in the pathway oxidize methanol in two-electron steps to yield CO_2 and reduced nicotinamide adenine dinucleotide (NADH), which provides energy. The resulting CO_2 is largely recycled back into anaerobic processes, completing a carbon cycle. Two main types of MMO are known: the membrane-bound particulate MMO (pMMO), found in almost all methanotrophs, and the iron-containing soluble form (sMMO). sMMO is generally easier to purify and characterize, making it an ideal candidate for structure-spectroscopic studies of O_2 activating enzymes.

The studies of sMMO catalysis have been principally focused on enzymes derived from two bacterial strains: *Methylosinus trichosporium* OB3b (*Mt* OB3b) and *Methylococcus capsulatus* Bath (*Mc* Bath). The sMMO enzyme is a dimeric multi-subunit complex[4] comprising four main protein components: the hydroxylase (MMOH, 245 kDa), which contains the diiron active site, a 16 kDa regulatory protein (MMOB), which aids substrate access to the active site, a 40 kDa reductase (MMOR) and an auxiliary protein MMOD, which is thought to inhibit MMOH.[5, 6] MMOH itself features an αβγ motif as shown in **Fig. 7.1a**. The diiron active site is situated near the surface of MMOH, and is accessible via a channel leading to a cavity at the active site. The sMMO active site (**Fig. 7.1a)** contains a nonheme diiron cluster coordinated by two histidine (His), four glutamate (Glu) residues and solvent-derived oxygen ligands. In the diferric resting state ($MMOH_{ox}$), the two high-spin Fe atoms are antiferromagnetically coupled, leading to a spin singlet ground state. Upon receiving two electrons from NADH-reduced MMOR, the enzyme transitions to the diferrous state

---

‡This Chapter is a part of an ongoing collaboration with Olivia McCubbin Stepanic and Serena DeBeer who is currently conducting the spectroscopic characterizations of the enzyme. The work is in progress at the time of submission of this thesis so a separate section for Results and Discussion have not been included. This chapter only includes my contribution to the project.

($MMOH_{red}$), which can react with O_2 to form a peroxo intermediate ($MMOH_p$).[2, 5] The cleavage of the O–O bond then produces the $MMOH_Q$ intermediate, a crucial component in the catalytic cycle responsible for methane (CH_4) activation. This intermediate has been the focus of numerous experimental and theoretical studies due to its pivotal role in the enzyme's function.



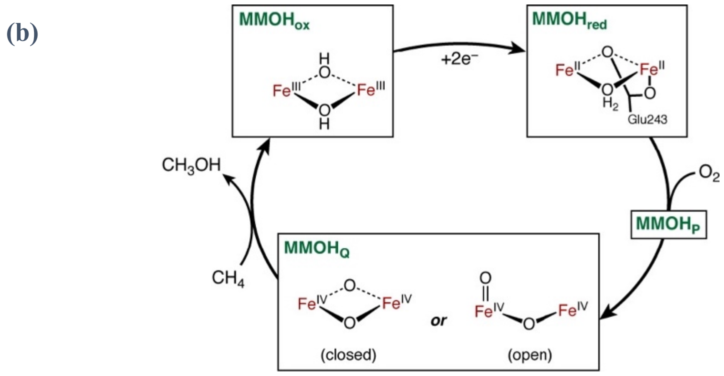


Fig. 7.1 **(a)** XFEL structure (6YD0.pdb) of the monomeric **MMOH+MMOB** complex, and its subunits (αβγ), MMOB (yellow) and the α subunit harboring the diiron active site (inset), **(b)** Abbreviated reaction cycle of sMMO, featuring the resting state $MMOH_{ox}$, $MMOH_{red}$, and the $MMOH_Q$ intermediate (reprinted in part with permission from *J. Am. Chem. Soc.* **2021**, 143, 17, 6560-6577).

Although the chemically reduced MMOH is catalytically active, the reaction rate is only physiologically significant when MMOH is complexed with MMOB. The most critical regulatory functions of MMOB include tuning redox potential of the diiron cluster, the acceleration of O_2 binding, a drastic increase in the turnover, and the selective binding and oxygenation of methane over other more easily oxidized hydrocarbons. Additionally, MMOB binding is thought to induce structural changes in the diiron core.[7] Crystallographic studies have reported that the active site is largely conserved in the MMOH+MMOB complex with the exception of Glu243 in the case of $MMOH_{red}$, where the side chain carboxylate rotates to coordinate with one iron in a bidentate manner and bridges to bind in a monodentate manner to the other iron (see **Fig 7.2**). The carboxylate shift dissociates the solvent bridge between the irons, creating a putative binding site for O_2. Recent crystal structures of $MMOH_{ox}$ collected from the same crystals at 4°C and –160°C revealed differences in the active site binding.[8, 9] However, it is also well known that freezing can impact crystal structures and potentially resulting in apparent substrate binding that does not occur above the freezing point.[10-14] Cutsail et al. reported variations in the K-edge white line of $\mathrm{MMOH_{ox}}$ under different freezing conditions.[15] Conversely, two good resolution structures of the $\mathrm{MMOH_{red}}$ state[7, 16] (1FYZ.pdb and 6YDI.pdb) differ in temperature, species, and the presence of MMOB and existing spectroscopic studies on this state also reveal ambiguities.[17, 18,19]All these factors have led to several key questions regarding the structural and spectroscopic characterization of the $\mathrm{MMOH_{red}}$+MMOB complex. Computational studies of the sMMO enzymes have aided experimental interpretation and proposed potential mechanisms in the past. However, a vast majority of these studies were based on QM clusters[20-29] and truncated QM/MM models[24, 30] neglecting the structural and long-range electrostatic effects of the protein matrix or lacking extensive connections to spectroscopic data. Recently, Schulz et al. used multiscale QM/MM approaches to explore possible conformations for $\mathrm{MMOH_{ox}}$ and $\mathrm{MMOH_{Q}}$ in *Mc* Bath,[31] however computational models to obtain structure-spectroscopic correlations of $\mathrm{MMOH_{red}}$ with bound MMOB have not been reported.[24, 32, 33]

In the current work, McCubbin Stepanic and DeBeer et al. used microfluidic mixers[34] to measure the X-ray Absorption Near Edge Structure (XANES) spectra[35] of $\mathrm{sMMO_{red}}$ and $\mathrm{sMMO_{ox}}$ in the absence and presence of MMOB at pH 7, and in the case of $\mathrm{MMOH_{ox}}$+MMOB, also at pH 7.5. Frozen solutions of the same samples were also measured for comparison with existing spectroscopic data on room temperature crystals.

The current experimental results on the MMOH+MMOB complex by DeBeer and co-workers necessitate the development of new computational models to provide a careful and thorough reevaluation of past spectroscopic interpretations. Here, we revisit the questions on the structure and properties of the $\mathrm{MMOH_{ox/red}}$ in the presence and absence of MMOB, using a multiscale QM/MM approach. Firstly, we explore the functional dynamics of resting state ($\mathrm{MMOH_{ox}}$) of the enzyme from *Mt* OB3b, with and without the complexed MMOB subunit using atomistic molecular dynamics simulations. Then we construct separate QM/MM models to describe the active site for $\mathrm{MMOH_{ox}}$ and $\mathrm{MMOH_{red}}$, examining how their geometric features and protonation states correlate with the full range of available spectroscopic properties. The computational protocol reported here is ensured to converge concerning the effects of the protein environment and incorporates available spectroscopic data. Our findings will allow us to narrow down the possible formulations of $MMOH_{ox/red}$ and to propose revised mechanistic insights for the regulatory influence of MMOB on the catalytic function of the sMMO enzyme.

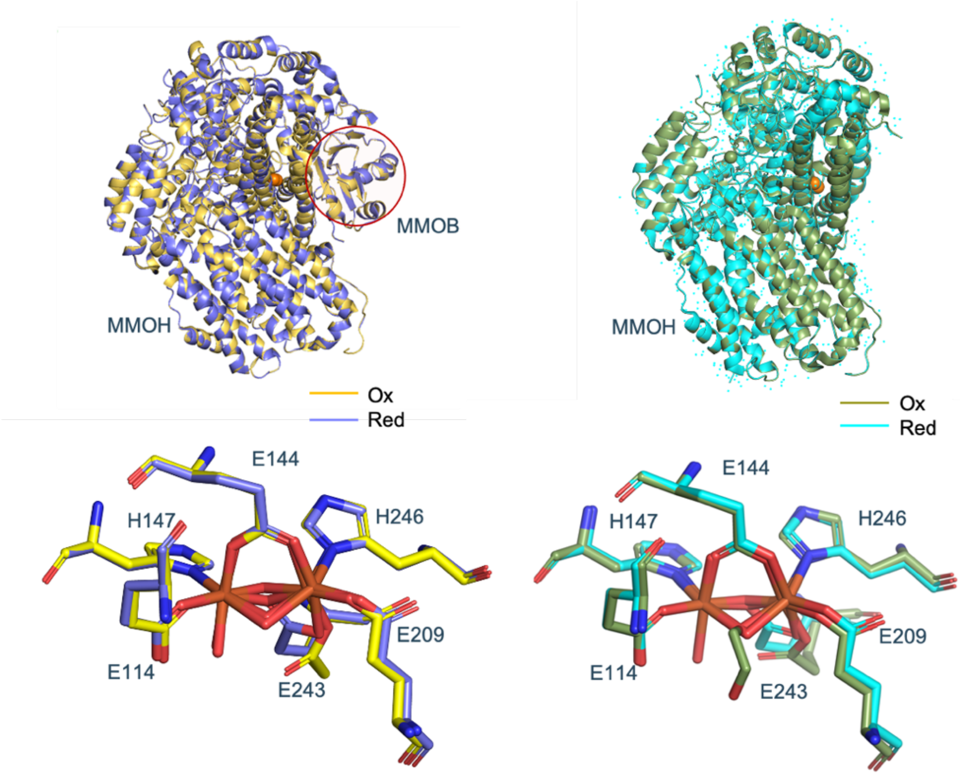


Fig. 7.2 Overall crystal structures of $MMOH_{ox}$ and $MMOH_{red}$ with MMOB bound (left) and MMOH without MMOB (right). The respective active sites are depicted along with the diiron core and protein residues.[7, 36]

7.2 Computational Methodology

7.2.1 System Preparation

We constructed two MM models starting from two different crystal structures for the $MMOH_{ox}$+MMOB complex and $MMOH_{ox}$ only (see **Fig. 7.2**). For the first model (**Model A**), we extracted one monomer from the high-resolution XFEL structure of $MMOH_{ox}$+MMOB from OB3b, by Srinivas et al. (6YD0.pdb, 1.95 Å).[7] For the second model (**Model B**), the X-ray crystal structure of $MMOH_{ox}$ reported by Jones et al. (6VK6.pdb, 1.52 Å)[36] was used as the starting point. The monomeric units were protonated using the H++ webserver under physiological conditions. Standard protonation states were assigned to the titratable residues (–1 for Asp, Glu and +1 for Arg, Lys) and the protonation states of the ligands coordinated to the diiron active site were manually assigned; His147 and His246 were protonated at the N_{ε} position. The active site ligands were renamed as shown in **Fig. 7.4** and the *antechamber* module was used to assign the atom types to all the ligand atoms. All the crystal waters were retained in the simulation setups.

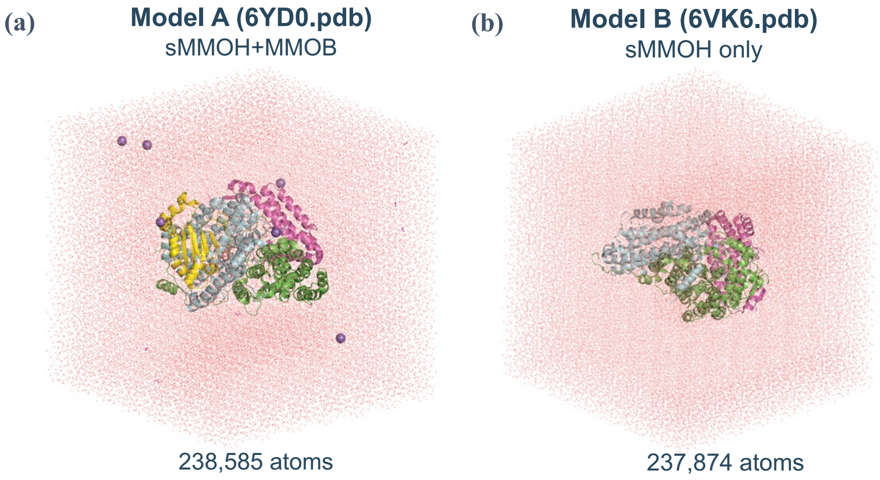


Fig. 7.3 Depiction of the complete setup of **(a)** MMOH with MMOB bound and **(b)** MMOH without MMOB (right), solvated within the TIP3P water box and counterions employed in classical MD simulations and QM/MM calculations throughout this work.

The complete protein system was solvated (in water) in a truncated octahedron box, with a minimum distance of 15 Å between the solute and box edge. Appropriate number of counterions were added randomly to maintain an overall neutral charge each setup. The final simulation setup consisted of 238,585 atoms for MMOH+MMOB and 237,874 atoms for MMOH only, as shown in **Fig. 7.3**.

7.2.2 Force Field Parameters

A bonded model is used for deriving the force-field parameters[37] of the diiron complex based on a small QM cluster extracted from the MMOH+MMOB protein (see **Fig. 7.4**). The cluster consisted of the diiron ions, along with the side-chains of 6 ligand residues (His147, His246, Glu144, Glu114, Glu209 and Glu243). The axial ligand to Fe1 and the bridging oxygens are modelled as waters. The sidechains of the ligand residues were terminated at the C_α–C_β bond, and capped with hydrogen atoms for the QM cluster calculations.

The charge on the QM system was computed using the MK-RESP[38] fitting methodology (see **Section 3.3**). The total charge of the QM cluster was –2 and the Fe(III) centers were considered in their high-spin (S=5/2,5/2) states. As a first step, the hydrogen atoms were optimized at B3LYP/def2-TZVP level of theory, whereas, the non-hydrogen atoms were kept fixed at their crystallographic positions. The RESP charge calculation was performed on the resulting optimized geometry, at B3LYP/def2-SVP level of theory using ORCA[39] and Multiwfn[40] programs. The total charge of the backbone atoms was restrained on the hydrogen link atom for each amino acid residue. During the generation of the topology file (*.prmtop*), the RESP charges of the backbone atoms (C, O, C_α, H_α, N, H) were assigned according to the original AMBER force field.

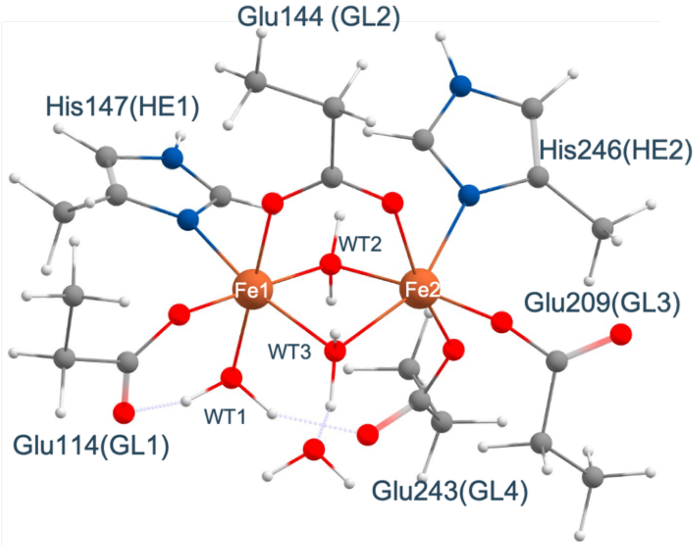


Fig. 7.4 QM cluster model used to derive the bonded parameters for the diiron site in sMMO (atomic coordinates derived from **6YD0.pdb**). The residue numbers are labelled according to the original crystal structure. The residue names used in the reconstructed MM model are labelled in parentheses.

The non-bonded parameters for Fe(III) were obtained from standard ion libraries in Amber compatible with the TIP3P water models.[41, 42] The bonded parameters involving Fe(III) are arbitrarily defined to maintain the structural integrity of the diiron active site. The bonds between Fe and the coordinated ligand atoms were restrained with force constants of 300 kcal $mol^{-1}Å^{-2}$, whereas, the angles, with 100 kcal $mol^{-1}rad^{-2}$. The equilibrium bond lengths and angles were fixed according to the crystal structure (6YD0.pdb, 6VK6.pdb) geometry in each case. The remaining protein residues are defined using the standard Amber14SB force field[43] while the TIP3P model is used for the solvent (water) molecules.

7.2.3 Molecular Dynamics Simulations

Step 1: Energy Minimization In the first step, all the hydrogen atoms of the system are energetically minimized for 2000 steps, which involves 1000 steps of steepest descent followed by 1000 conjugate gradient steps. The positions of heavy atoms were restrained with a force constant of 100 kcal $mol^{-1}Å^{-2}$. In the second minimization phase, all the hydrogen atoms, the solvent (waters) and ions were minimized for 2000 steps, keeping the rest of the protein restrained with a force constant of 100 kcal $mol^{-1}Å^{-2}$.

Step 2: Heating Dynamics (NVT) In the equilibration procedure, the systems are first heated from 0 to 298 K over 200 ps in the NVT ensemble. The temperature is controlled using the *Langevin* thermostat[44] with a collision frequency of 1.0 ps^{-1}. All the heavy atoms were harmonically restrained with a force constant of 50 kcal $mol^{-1}Å^{-2}$.

Step 3: Equilibration Dynamics (NPT) The final equilibration dynamics is performed in the NPT ensemble at physiological conditions (pressure ~1 atm, temperature ~298 K). The pressure was maintained using *Berendsen* barostat[45] with a relaxation time of 1 ps. The system was propagated in NPT for 500 ps until the average density of the simulation box was stable.

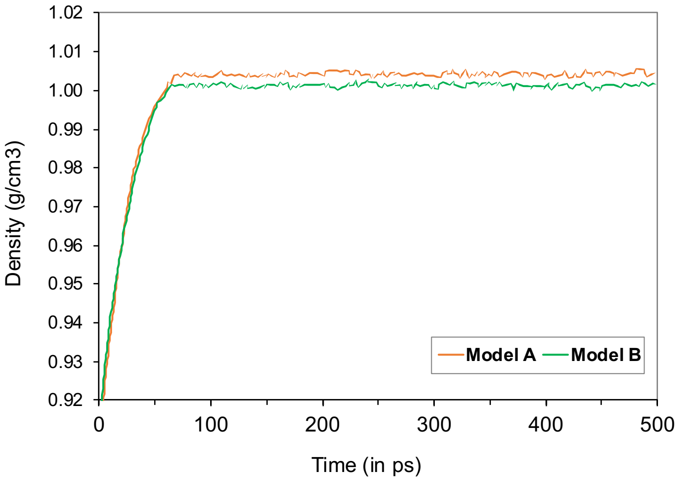
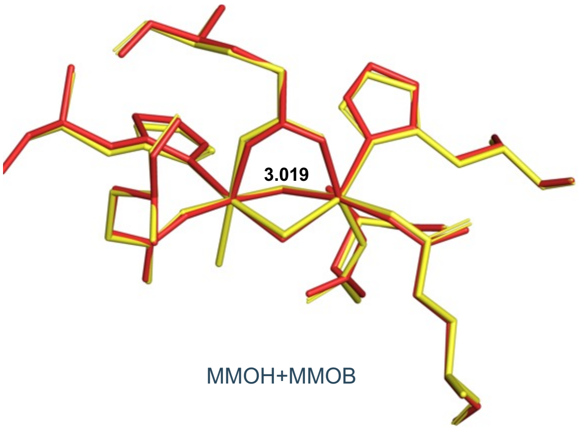


Fig. 7.5 Variation of system density over the equilibration simulations for each setup.

A time-step of 2 fs was used throughout for integration of the equation of the motions. All bonds involving the hydrogen atoms were constrained using the SHAKE algorithm. The Particle Mesh Ewald (PME) algorithm[46-48] is used to treat the long-range electrostatics with a non-bonded cut-off of 10 Å during the entire heating and simulation procedure. The protein RMSD and Fe–Fe distance during the simulations was monitored using the Cpptraj module[49] (**Fig. 7.5**).

The last frame of the equilibration MD simulations for each setup will be used for all QM/MM calculations in this work. The overall active site geometry (**Fig. 7.6**) and average Fe–Fe distance (**Fig. 7.7**) is conserved throughout the simulation process. All the MD and QM/MM setup, minimization and equilibration dynamics in this work are performed using the Amber20 program package.[50]



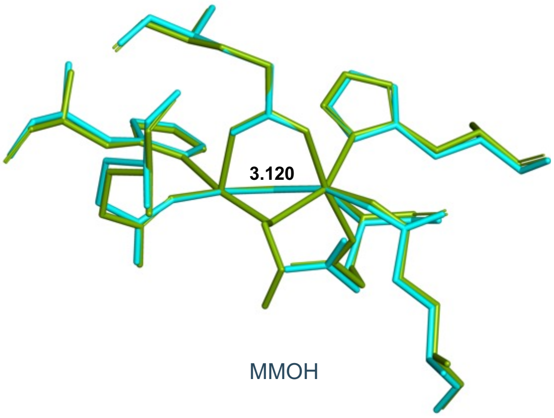


Fig. 7.6 Overlay of crystal structure and last frame of the equilibration simulations (to be used for QM/MM calculations). The MMOH+MMOB crystal structure is shown in yellow and MMOH in green. The active site from the last frame of MD simulation for Model A is represented in red and for Model B in cyan. The Fe–Fe distances indicated are from the last frame in each simulation.

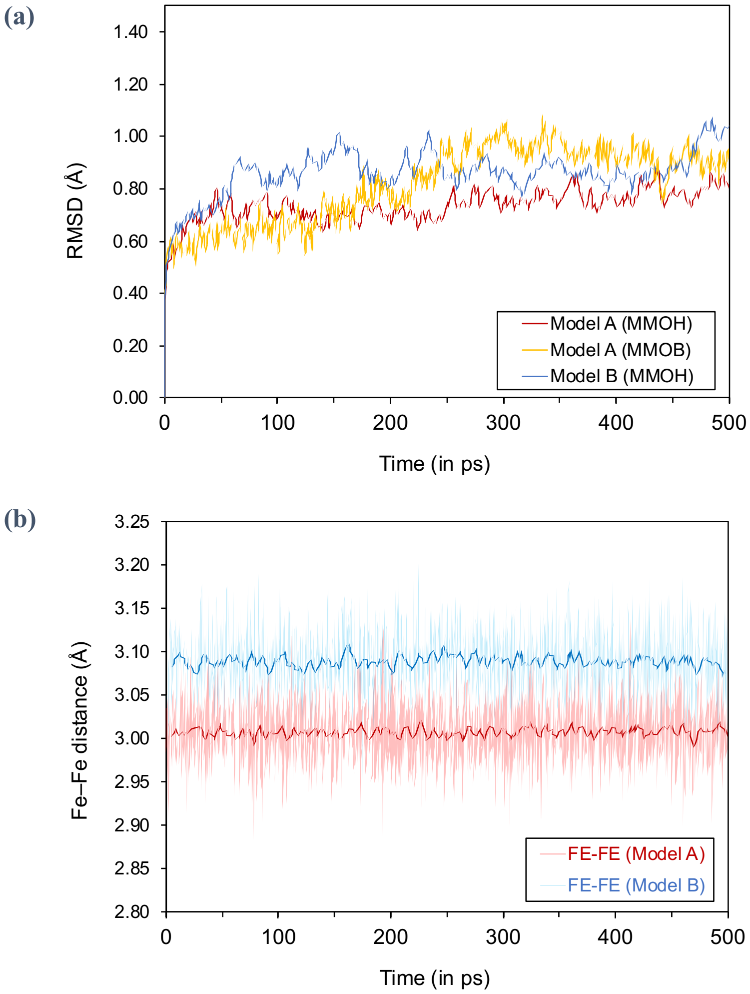
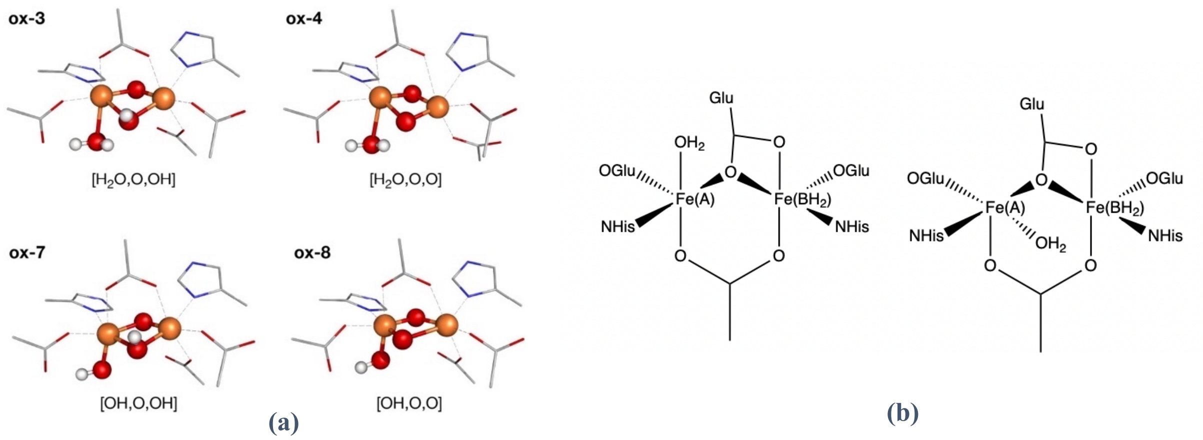


Fig. 7.7 **(a)** RMSD of the C_α atoms of MMOH with (Model A) and without MMOB (Model B). **(b)** Variation of the average Fe–Fe distance over equilibration MD simulations. The original crystal structure distances are 3.036 Å for Model A and 3.129 Å for Model B, respectively.

7.2.4 QM/MM Calculations

The first set of QM calculations are performed on the QM region described in **Fig 7.4**, while the rest of the protein and waters are treated as electrostatic point charges. Based on the various structures, protonation states and spectroscopy calculations (**Fig 7.8a**) recent work by Schulz et al.[31] suggested model ox-3 to be the best candidate for $MMOH_{ox}$ without MMOB bound. However, the current spectroscopic data on room-temperature $MMOH_{ox}$+MMOB do not seem to match most of the earlier proposed $MMOH_{ox}$ models and distinct differences have been observed on the MMOH data with and without MMOB. This led us to propose that the active site of $MMOH_{ox}$+MMOB could resemble ox-3, ox-4, ox-5, ox-7 or ox-8 and $MMOH_{ox}$ without MMOB could fit ox-4 or ox-7, however a number spectroscopic calculations have to be performed in order to verify the above hypothesis.



(a) **(b)**

Fig. 7.8 (a) Geometries of the Fe_2O_x core from optimized $MMOH_{ox}$ structures reported by Schulz et al. This figure is reproduced in part with permission from *J. Am. Chem. Soc.* **2021**, 143, 17, 6560-6577. **(b)** Proposed active site structures of $MMOH_{red}$, considered for QM/MM calculations in the current work.

The first set of QM/MM calculations will be performed on ox-3 for Model A and Model B, in order to verify whether the current spectroscopic characterization is consistent with the earlier proposed models. The active site in each case, is reconstructed based on the protonation states depicted in **Fig. 7.8a**. Specifically, the X-ray absorption spectra (XAS) for each structure will be computed using TD-DFT, with benchmarked functionals and def2-TZVP basis sets. Once we establish the most likely candidate for $MMOH_{ox}$+MMOB (Model A), and MMOH only (Model B) consistent with current spectroscopic data, we will proceed to perform spectroscopic calculations on $MMOH_{red}$+MMOB and without MMOB, based on the two alternate active site geometries as shown in **Fig. 7.8b**. We are currently performing multiscale QM/MM calculations on each of the systems described above using the multiscale module of ORCA 6.

7.3 REFERENCES

1. F. J. Tucci and A. C. Rosenzweig, *Chem. Rev.*, 2024, **124**, 1288-1320.
2. R. Banerjee, J. C. Jones and J. D. Lipscomb, *Annu. Rev. Biochem.*, 2019, **88**, 409-431.
3. C. E. Tinberg and S. J. Lippard, *Acc. Chem. Res.*, 2011, **44**, 280-288.
4. G. T. Gassner and S. J. Lippard, *Biochemistry*, 1999, **38**, 12768-12785.
5. R. Banerjee, Y. Proshlyakov, J. D. Lipscomb and D. A. Proshlyakov, *Nature*, 2015, **518**, 431-434.
6. J. C. Jones, R. Banerjee, K. Shi, M. M. Semonis, H. Aihara, W. C. K. Pomerantz and J. D. Lipscomb, *Biochemistry*, 2021, **60**, 1995-2010.
7. V. Srinivas, R. Banerjee, H. Lebrette, J. C. Jones, O. Aurelius, I.-S. Kim, . . . M. Högbom, *J. Am. Chem. Soc.*, 2020, **142**, 14249-14266.
8. A. C. Rosenzweig, P. Nordlund, P. M. Takahara, C. A. Frederick and S. J. Lippard, *Chem. Biol.*, 1995, **2**, 409-418.
9. A. C. Rosenzweig, C. A. Frederick, S. J. Lippard, P. Nordlund and auml, *Nature*, 1993, **366**, 537-543.
10. T. Doukov, D. Herschlag and F. Yabukarski, *Acta Crystallographica Section D*, 2023, **79**, 212-223.
11. S. Russi, A. Gonzalez, L. R. Kenner, D. A. Keedy, J. S. Fraser and H. van den Bedem, *J. Synchrotron Radiat.*, 2017, **24**, 73-82.
12. J. S. Fraser, H. van den Bedem, A. J. Samelson, P. T. Lang, J. M. Holton, N. Echols and T. Alber, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 16247-16252.
13. T. R. Stachowski, M. Vanarotti, J. Seetharaman, K. Lopez and M. Fischer, *Angew. Chem. Int. Ed.*, 2022, **61**, e202112919.
14. R. F. Tilton, Jr., J. C. Dewan and G. A. Petsko, *Biochemistry*, 1992, **31**, 2469-2481.
15. G. E. Cutsail Iii and S. DeBeer, *ACS Catal.*, 2022, **12**, 5864-5886.
16. D. A. Whittington and S. J. Lippard, *J. Am. Chem. Soc.*, 2001, **123**, 827-838.
17. J. G. DeWitt, A. C. Rosenzweig, A. Salifoglou, B. Hedman, S. J. Lippard and K. O. Hodgson, *Inorg. Chem.*, 1995, **34**, 2505-2515.
18. R. G. Castillo, R. Banerjee, C. J. Allpress, G. T. Rohde, E. Bill, L. Que, Jr., . . . S. DeBeer, *J. Am. Chem. Soc.*, 2017, **139**, 18024-18033.
19. G. E. Cutsail, III, R. Banerjee, A. Zhou, L. Que, Jr., J. D. Lipscomb and S. DeBeer, *J. Am. Chem. Soc.*, 2018, **140**, 16807-16820.
20. P. E. M. Siegbahn, *Chem. Phys. Lett.*, 2002, **351**, 311-318.
21. P. E. M. Siegbahn and R. H. Crabtree, *J. Am. Chem. Soc.*, 1997, **119**, 3103-3113.
22. T. Lovell, W.-G. Han, T. Liu and L. Noodleman, *J. Am. Chem. Soc.*, 2002, **124**, 5890-5894.
23. K. Yoshizawa, T. Ohta, T. Yamabe and R. Hoffmann, *J. Am. Chem. Soc.*, 1997, **119**, 12311-12321.
24. M. Torrent, T. Vreven, D. G. Musaev, K. Morokuma, Ö. Farkas and H. B. Schlegel, *J. Am. Chem. Soc.*, 2002, **124**, 192-193.
25. M. Torrent, D. G. Musaev, H. Basch and K. Morokuma, *J. Comput. Chem.*, 2002, **23**, 59-76.
26. M. Torrent, K. Mogi, H. Basch, D. G. Musaev and K. Morokuma, *J. Phys. Chem. B*, 2001, **105**, 8616-8628.
27. H. Basch, D. G. Musaev and K. Morokuma, *J. Phys. Chem. B*, 2001, **105**, 8452-8460.
28. B. F. Gherman, B. D. Dunietz, D. A. Whittington, S. J. Lippard and R. A. Friesner, *J. Am. Chem. Soc.*, 2001, **123**, 3836-3837.
29. B. D. Dunietz, M. D. Beachy, Y. Cao, D. A. Whittington, S. J. Lippard and R. A. Friesner, *J. Am. Chem. Soc.*, 2000, **122**, 2828-2839.

30. M. Hoffmann, I. V. Khavrutskii, D. G. Musaev and K. Morokuma, *Int. J. Quantum Chem.*, 2004, **99**, 972-980.
31. C. E. Schulz, R. G. Castillo, D. A. Pantazis, S. DeBeer and F. Neese, *J. Am. Chem. Soc.*, 2021, **143**, 6560-6577.
32. B. F. Gherman, S. J. Lippard and R. A. Friesner, *J. Am. Chem. Soc.*, 2005, **127**, 1025-1037.
33. D. Rinaldo, D. M. Philipp, S. J. Lippard and R. A. Friesner, *J. Am. Chem. Soc.*, 2007, **129**, 3135-3147.
34. K. A. Zielinski, A. M. Katz, G. D. Calvey, S. A. Pabit, S. K. Milano, C. Aplin, . . . L. Pollack, *IUCrJ*, 2023, **10**, 363-375.
35. J. K. Kowalska, F. A. Lima, C. J. Pollock, J. A. Rees and S. DeBeer, *Isr. J. Chem.*, 2016, **56**, 803-815.
36. J. C. Jones, R. Banerjee, K. Shi, H. Aihara and J. D. Lipscomb, *Biochemistry*, 2020, **59**, 2946-2961.
37. P. Li and K. M. Merz, *Chem. Rev.*, 2017, **117**, 1564-1686.
38. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
39. F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.
40. T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580-592.
41. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
42. P. Li, L. F. Song and K. M. Merz, Jr., *J. Phys. Chem. B*, 2015, **119**, 883-895.
43. X. He, V. H. Man, W. Yang, T.-S. Lee and J. Wang, *J. Chem. Phys.*, 2020, **153**, 114502.
44. R. J. Loncharich, B. R. Brooks and R. W. Pastor, *Biopolymers*, 1992, **32**, 523-535.
45. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
46. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878-3888.
47. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374-380.
48. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
49. D. R. Roe and T. E. Cheatham, III, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
50. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, . . . K. M. Merz, Jr., *J. Chem. Inf. Model.*, 2023, **63**, 6183-6191.

8

Summary and Perspective

In this dissertation, we have explored the multiscale computational modelling of the redox-active components and spectroscopic properties of biomolecular systems. The research highlights the complex aspects of the Photosystem II (PSII) complex, with particular emphasis on the light-driven electron transfer processes that are central to oxygenic photosynthesis. The following section provides a summary of the key objectives addressed in this thesis, along with suggestions for future research directions.

(1) **Modelling the Environment:** A central theme emphasized throughout this thesis is the crucial role of explicit solvation and protein matrix in describing (bio)chemical processes. We explored various chemical properties such as redox potentials, excitation energies, charge-transfer states and their coupling to molecular processes. This was accomplished using efficient multiscale QM methodologies (QM/QM and QM/MM) combined with large scale molecular dynamics (MD) to span different length and time scales. The methods are successfully applied to a range of biomolecular systems, from hydrated metal clusters to solvated metalloproteins and even larger pigment-protein complexes.

In **Chapter 2**, we stressed on the necessity of explicit consideration of a solvation sphere in addition to implicit solvation methods, to accurately determine redox potentials of transition metal clusters. However, a more refined computational protocol in future applications would need to account for the variability of the solvent shell, the dynamic nature of solvation, and the coupling of changes in the coordination number or geometry with the reorganization of the solvation layers.

In the subsequent chapters, our multiscale simulations on Photosystem II (PSII) demonstrated that the protein matrix electrostatics and its conformational dynamics play a pivotal role in inducing electrochromic shifts, localization/de-localization of excitons and access to low-energy charge transfer (CT) states both in the PSII-RC and CP43 antenna. All simulations discussed here utilized QM regions embedded within a molecular mechanics (MM) force field represented by electrostatic point charges. It is important to note that characterization of inter-molecular CT states and excitonic coupling within photosynthetic PPCs may necessitate advanced QM/MM methods with polarizable MM environments in future studies. Research on light harvesting complexes also requires extensive conformational sampling of the protein matrix. Although the MD-PMM approach demonstrated in **Chapter 6** partially addresses this, future studies of might need to employ more sophisticated QM/MM/MD methodologies.

(2) **Choice of QM Methods:** The second most critical factor in any QM based multiscale modelling approach is selecting an appropriate quantum chemical method. The QM method is generally determined by (i) the size of the QM region and (ii) the target chemical or spectroscopic property. Benchmarking DFT functionals for transition metal complexes is particularly challenging due to the variability in spin states and oxidation states, necessitating the use of ab-initio QM methods in some cases. In **Chapter 2**, we established a multilayer approach to DLPNO-CCSD(T) coupled with the PTE(S) solvation scheme to achieve chemically accurate redox potentials of complex open-shell systems. This cost-effective QM method can be applied both to systems involving implicit or explicit solvation and to redox-active molecules embedded within biological or inorganic matrices.

Even within the same (bio)chemical system, different QM methods are often necessary to compute different energies, making it essential to carefully benchmark these methods for each system. In **Chapter 4**, we exclusively utilized TD-DFT methods based on range-separated hybrid functionals (ωB97X-D3(BJ)) to compute vertical excitation energies and CT states of photosynthetic pigments, while PBE is used to obtain reliable QM/MM geometries and hybrid TPSSh functional to compute EPR parameters of triplet states in the PSII reaction center. In the past, the DLPNO–STEOM-CCSD method has also been used in a multiscale QM/MM framework to provide reference-quality electrochromic shifts of photosynthetic pigments. Multireference methods can also offer the most accurate description of excited states of photosynthetic pigments, which often exhibit multiple near-degenerate electronic states. Therefore, a possible methodological extension would involve a combined multireference-multiscale approach, particularly for higher energy excitations, as well as for describing of excited states of multiple pigments within PSII. In **Chapter 6**, we demonstrated the use of TD-DFT with QM/MM and large-scale MD-PMM calculations to describe the low-energy spectrum of the CP43 light harvesting antennae. However, studies of light harvesting protein complexes often require sampling of multiple configurations of the protein which in turn affects the geometry, extent of electronic coupling and spatial distribution of “red” Chls relevant for excitation energy transfer. The use of low-cost ab-initio QM methods combined with automatized MM based machine learning approaches to sample the conformational space, would therefore be a promising method in multiscale modelling of LHCs.

(3) **Mechanistic insights into Oxygenic Photosynthesis**: A significant portion of this thesis is dedicated towards the mechanistic understanding of oxygenic photosynthesis, primarily focusing on PSII. We investigated the primary charge-separation and charge-recombination pathways within the PSII-RC by calculating excitation energies, EPR properties and redox potentials using a QM/MM framework. In **Chapter 4**, we detailed the low-energy spectrum of excited singlet and triplet states, including both local and charge-transfer nature, and provided the first theoretical description of the "closed RC". This involved explicitly quantifying the electrostatic effect of the reduced quinone (Q_A^-) and its modulation of excited state properties of RC pigments. A logical extension of this study would be to investigate the electrostatic effects of other PSII cofactors on primary charge-separation pathways along both the D1 and D2 branches. In **Chapter 5**, we examined how natural genetic variations in the crucial D1 protein influence primary processes in photosynthesis through combined redox and spectral tuning of specific pigments. We identified the electrostatic effects of specific substitutions, which suggest potential targets for future mutagenesis experiments aimed at the spectral tuning of charge-transfer (CT) states. This work provides a foundation for engineered strategies to achieve precise tuning and functional optimization of photosynthetic systems. A potential extension of this research would involve studying primary charge separation in the far-red PSII variants, where organisms express alternative isoforms of the same RC protein in response to external light wavelengths. In **Chapter 6**, we identified a mixed excitonic charge-transfer state and characterized the distribution of Chl "domains" within the CP43 core antenna of PSII. The CT states may either spectrally tune the "red" region or act as photo-quenching intermediates in PSII, however this needs to be verified in future spectroscopic experiments. This work on CP43 could be extended to the investigation of CT states in the CP47 core antenna which would aid in future kinetic modelling of EET pathways from CP43/CP47 to the PSII-RC.

The multiscale models and methodological protocol outlined in this work can, in principle, be adapted to investigate light-induced and redox-active processes in other membrane-embedded protein complexes (Photosystem I, Cytochrome b_6f) with minor adjustments. Overall, this thesis provides a refined electronic-level understanding of primary and secondary electron transfer pathways, offering detailed insights into the electronic structure that serves as a foundation for exploring potential mechanisms of light-harvesting, primary charge separation and photoprotection mechanisms in oxygenic photosynthesis.