

UNIVERSITY OF CALIFORNIA, SAN DIEGO

An extended quantum mechanical molecular mechanics NWChem/AMBER interface
for estimating free energies and determining reaction paths in catalytic enzymes:
Application to cellulose degradation catalyzed by copper-dependent
polysaccharide monooxygenases

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

Teerapong Pirojsirikul

Committee in charge:

Professor John Weare, Chair
Professor Michael Galperin
Professor Michael Gilson
Professor Katja Lindenberg
Professor Ulrich Muller

2017

Copyright

Teerapong Pirojsirikul, 2017

All rights reserved.

The Dissertation of Teerapong Pirojsirikul is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

Signature Page	iii
Table of Contents	iv
List of Figures	vi
List of Tables	viii
Acknowledgements	ix
Vita	xi
Abstract of the Dissertation	xii
Chapter 1 Quantum Mechanical Molecular Mechanics (QM/MM) Methodology	1
1.1 Introduction	1
1.2 Overview of QM/MM method	2
1.2.1 Partitioning of QM/MM System	2
1.2.2 QM/MM Hamiltonian	3
1.2.3 QM/MM Boundary Treatment	5
1.3 Motivation and Objective	6
1.4 Organization of Chapters	7
1.5 References	8
Chapter 2 Extended NWChem-AMBER Interface for Large Scaled QM/MM Simulations	12
2.1 Introduction	12
2.2 QM/MM Methodology	12
2.3 QM/MM Optimization	13
2.3.1 Implementation of QM/MM Multiregion Optimization	15
2.3.2 User Interface for QM/MM Multiregion Optimization	16
2.4 Technical Details	20
2.4.1 NWChem	20
2.4.2 AMBER	21
2.4.3 Flow Control between QM and MM Optimization Cycles	22
2.5 Nudged Elastic Band (NEB) Method	24
2.5.1 Overview of NEB Method	24
2.5.2 NEB Method in ab-initio QM/MM Calculations	24
2.5.3 Implementation of ab-initio QM/MM NEB Interface	26
2.5.4 User Interface for QM/MM NEB Calculations	27
2.6 Free Energy Perturbation (FEP)	28
2.6.1 Overview of FEP in ab-initio QM/MM Calculations	28
2.6.2 Implementation of FEP Algorithm	30
2.6.3 User Interface for FEP Calculations	31
2.7 References	32
Chapter 3 Validation of QM/MM NWChem-AMBER Interface	34
3.1 Introduction	34

3.2 QM/MM Optimization Benchmarking	34
3.2.1 Methodology.....	34
3.2.2 Results and Discussion	35
3.3 QM/MM Method for SN2 Reaction of $\text{CH}_3\text{Br} + \text{OH}^-$ in Aqueous Solution	36
3.3.1 Overview.....	36
3.3.2 Computational Details	37
3.3.3 Results and Discussion	39
3.3.4 Conclusions	45
3.4 References	46
 Chapter 4 Combined Quantum-Mechanical Molecular Mechanics Calculations with NWChem and AMBER: Excited State Properties of Green Fluorescent	
Protein Chromophore Analogue in Aqueous Solution	49
4.1 Introduction	49
4.2 Computational Methodology and System Setup.....	53
4.2.1 Gas Phase	53
4.2.2 Aqueous Phase.....	54
4.3 Results and Discussion	59
4.3.1 Cationic Form	59
4.3.2 Anionic Form.....	61
4.3.3 Neutral Form.....	63
4.3.4 Zwitterionic Form	64
4.4 Conclusions	65
4.5 References	67
 Chapter 5 QM/MM Calculations for Catalysis of Cellulose Degradation by Copper-dependent Polysaccharide Monooxygenases.....	
5.1 Introduction	74
5.1.1 Cellulose and Its Decomposition	74
5.1.2 Polysaccharide Monooxygenases.....	74
5.2 Proposed Reaction Mechanisms of PMOs	76
5.2.1 Superoxo Mechanism	76
5.2.2 Oxyl Mechanism.....	77
5.3 Theoretical Investigations of PMO Reaction Mechanisms	78
5.4 Methodology.....	81
5.4.1 QM/MM Optimization of X-ray Structure	82
5.4.2 Development of Oxygen Binding Active PMO Models	82
5.4.3 QM/MM Optmization of Active PMOs.....	83
5.4.4 Determination of HAT Reaction Pathways.....	83
5.4.5 Free Energy Calculation.....	84
5.4.6 Open shell singlet state energy and Natural Bond Orbital (NBO) Analysis.....	86
5.5 Results and Discussion	87
5.5.1 Analysis of X-ray Structure.....	87
5.5.2 Enzyme Activation and O_2 Binding to the PMO Active Site	89
5.5.3 Hydrogen Abstraction with Copper-superoxo species	91
5.5.4 Hydrogen Abstraction with Copper-oxyl species	94
5.6 Conclusions	99
5.7 References	100

LIST OF FIGURES

Figure 1.1: An example of QM/MM system. GFP chromophore (sticks) is the QM subsystem and waters (lines) are the MM subsystem.....	3
Figure 2.1: Flow diagram for multiregion QM/MM optimization via NWChem-AMBER interface.....	15
Figure 2.2: Format of the file <i>crdparms</i> used in NWChem for a cycle of QM optimization.....	16
Figure 2.3: (Left) Example of an AMBER control file to generate <i>crdparms</i> . (Right) Example of <i>nwin</i> for a single cycle of QM optimization. Bold texts indicate required modifications.	18
Figure 2.4: Example of an AMBER <i>mdin</i> to perform a single cycle of MM optimization. Required modifications in the <i>qmmm</i> namelist is highlighted with bold texts.	19
Figure 2.5: Example of config.in to control the workflow of QM/MM multiregion optimization. The lines with # are comments and neglected.	20
Figure 2.6: Diagram for QM/MM NEB calculation via NWChem-AMBER interface.....	26
Figure 2.7: Example of <i>nwin</i> to perform QM/MM NEB calculation. Two <i>crdparms</i> are specified to generate a trial reaction path connecting 2 states. The total of 12 images is used to represent the path. The spring constant is 0.1 a.u. The NEB calculation is performed for 5 iterations with the initial step size of 0.5 a.u.	28
Figure 2.8: pseudocode for FEP algorithm.....	31
Figure 2.9: Example of AMBER <i>mdin</i> illustrating the FEP calculation between 2 states with the width of 0.2 between λ_i and λ_{i+1} . The MD simulation is performed for 100000 steps and the FEP calculation starts at the first step.	32
Figure 3.1: Water dimer with donor (left) and accepter (right)	35
Figure 3.2: Optimized structures (upper left) gas phase RS. (upper right) aqueous phase RS. (lower left) aqueous phase TS. (lower right) aqueous phase PS. Atoms in colors: green = carbon; crimson red = bromine; red = oxygen; white = hydrogen. Thick water sticks are solvation shell of either OH^- or Br^-	40
Figure 3.3: Charge distribution of each group of solute along the NEB reaction pathway.....	42
Figure 3.4: Comparison of QM internal energy between gas phase and aqueous phase along the NEB reaction pathway. The first NEB image in the gas phase is the reference point.	43
Figure 3.5: Free energy profile of the reaction. Solvation free energy obtained from FEP calculations are also shown. The reactant state is the reference point.....	44

Figure 4.1: Diagram for a single cycle of multiregion optimization using the AMBER/NWChem interface.....	55
Figure 4.2: GFP model chromophore (HBDI) in water with different protonation states (<i>cis</i> and <i>trans</i> conformers).....	57
Figure 5.1: Superoxo mechanism.....	77
Figure 5.2: Oxyl mechanism	78
Figure 5.3: Free energy cycle using multi-level perturbation method.....	85
Figure 5.4: Optimized structure aligned on reference X-ray structure (yellow) with hydrogen atoms omitted. X-ray parameters are in parentheses.....	88
Figure 5.5: RS _{SPO} structure (coordinating waters are transparent).....	92
Figure 5.6: TS _{SPO} structure (coordinating waters are transparent)	92
Figure 5.7: PS _{SPO} structure (coordinating waters are transparent)	93
Figure 5.8: RS _{OXL} structure	96
Figure 5.9: TS _{OXL} structure.....	96
Figure 5.10: PS _{OXL} structure.....	97
Figure 5.11: Free energy diagram for HAT via superoxo (SPO) and oxyl (OXL) mechanisms ...	98

LIST OF TABLES

Table 3.1: Binding energy (kcal/mol), distances (Å), and angle (°) of a water dimer at the full QM and QM/MM (with hydrogen bond donor D or acceptor A in the QM region) levels.....	35
Table 3.2: Collection of energy (kcal/mol) data and total free energies (forward and backward samplings) obtained from NEB and FEP calculations	44
Table 3.3: Comparison of free energies (kcal/mol) calculated along the NEB reaction pathway and smooth transformation.....	45
Table 4.1: Calculated excitation energy (in eV) of the GFP model chromophores.....	58
Table 4.2: Relative free energies of GFP model chromophores (kcal/mol)	59
Table 5.1: Geometrical parameters (in angstrom) of substrate bound PMO in different states from computations and experimental measurements	88
Table 5.2: Natural charges and spin densities from NBO analysis.....	90
Table 5.3: QM triplet and open shell singlet energies (Hartree) for each state along HAT with copper-superoxide species.....	90
Table 5.4: Free energy of HAT with Cu-superoxide species	93
Table 5.5: QM triplet and open shell singlet energies (Hartree) for each state along HAT with copper-oxyl species.....	94
Table 5.6: Free energy of HAT with Cu-oxyl species	97

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to Professor John Weare for being so supportive throughout my PhD life at UCSD. I was very fortunate to be part of his group and worked for him. John is a very kind, flexible and understanding person that a PhD student would hope for working with. Working with him not only made me become more expertise in computational chemistry but also allowed me to improve my programming skills, which I found very challenging but would be beneficial to my future career.

I also would like to thank NWChem developers from PNNL. I was very fortunate to have several opportunities to work and converse with them. I would like to thank Dr. Marat Valiev for his supports and guidance during code development in NWChem, Dr. Karol Kowalski for his contribution on high-level method calculations, Dr. Eric Bylaska who took care of me during my first visit, Dr. Niri Govind and Dr. Edo Apra for discussions about technical issues. In addition, I would like to thank Dr. Andreas Götz from SDSC and Dr. Ross Walker from GlaxoSmithKline, who are primary developers for QM/MM interface via AMBER and external QM packages. My work would not have succeeded without Dr. Götz's guidance and advice. I would also express my gratitude to my committee members, Prof. Katja Lindenberg, Prof. Michael Galperin, Prof. Ulrich Muller, and Prof. Michael Gilson for their time and guidance.

I would like to thank my former and current colleagues, Dr. Ying Chen, Mr. Duo Song, and Mr. Wasut 'Khun' Pornpatcharapong, for all their helps and kindnesses. Also, my life in San Diego would have not been lively without a group of Thai students/people. I would like to thank Dr. Wiroj Nantasetphong, Dr. Oraphin Chantarasriwong and Mr. Karu Chongsiripinyo for any fun moments during my PhD study, Dr. Tissana Kijasanayotin and Dr. Edward Dechaumphai and Thai soccer team for awesome soccer games together, Mr. Praween Siritanasak for his delicious dishes,

as well as expressing my gratitude towards all other Thai people who helped and shared great moments with me during my time in the US.

My PhD study would not have been completed without the funding from the Royal Thai Government. I would like to thank the organization for offering me a very great opportunity to study aboard. It was my deepest gratitude to all Thai people whose taxes were converted and spent towards me who earnestly expected to return their favors in my future career.

Specially, I would like to thank my family members, Mr. Pongsatorn Pirojsirikul, Ms. Khemjira Pichattinon, Ms. Patcharanun Pirojsirikul, and Ms. Onjira Pirojsirikul, who are always there for me. They are the unlimited fuel resource that keeps me driving forward to overcome the odds during my bad moments. I would like to dedicate this degree to them.

Chapter 4, in full, is a reprint of the material entitled “Combined Quantum-Mechanical Molecular Mechanics Calculations with NWChem and AMBER: Excited State Properties of Green Fluorescent Protein Chromophore Analogue in Aqueous Solution” by Pirojsirikul, T., Götz, A.W., Weare, J., Walker, R.C., Kowalski, K., and Valiev, M., which has been accepted for publication in the Journal of Computational Chemistry. The dissertation author is the primary author of this material, and co-authors have approved the use of the material for this dissertation.

VITA

2008	B.S. in Chemistry	Prince of Songkla University, Thailand
2011	M.S. in Chemistry	University of California, San Diego
2017	Ph.D. in Chemistry	University of California, San Diego

PUBLICATION

Pirojsirikul, T., Götz, A.W., Weare, J., Walker, R.C., Kowalski, K., and Valiev, M. “Combined Quantum-Mechanical Molecular Mechanics Calculations with NWChem and AMBER: Excited State Properties of Green Fluorescent Protein Chromophore Analogue in Aqueous Solution”, accepted

ABSTRACT OF THE DISSERTATION

An extended quantum mechanical molecular mechanics NWChem/AMBER interface
for estimating free energies and determining reaction paths in catalytic enzymes:
Application to cellulose degradation catalyzed by copper-dependent
polysaccharide monooxygenases

by

Teerapong Pirojsirikul

Doctor of Philosophy in Chemistry

University of California, San Diego, 2017

Professor John Weare, Chair

An extended QM/MM NWChem/AMBER interface has been developed and implemented to offer additional features in computations within the QM/MM framework. This includes the interface for the QM/MM multiregion optimization, nudged elastic band (NEB), and free energy perturbation (FEP). With these functionalities, it is feasible to apply *ab-initio* or density functional

(DFT) QM/MM methods to study various problems, for example, reaction mechanisms of enzymes, in which many degrees of freedom are involved and considered. The QM/MM multiregion optimization provides an efficient approach that reduces computational costs for optimization problems, e.g. geometry optimization. The minimum energy pathway of enzyme catalysis can be determined using the NEB method. Together with the FEP method, free energy profile of the reaction can be established.

A brief introduction of QM/MM methodology has been discussed in chapter 1. Chapter 2 covers the implementation of the extended QM/MM NWChem/AMBER interface with the related theoretical backgrounds also discussed. Chapter 3 includes the validation of the implemented interface using small test cases. Water dimer was chosen to validate the QM/MM multiregion optimization algorithm. The QM/MM NEB and FEP methods were applied to determine the reaction mechanism of a S_N2 reaction of $\text{CH}_3\text{Br} + \text{OH}^-$ in aqueous solution. Chapter 4 illustrates the use of the QM/MM multiregion optimization in combination with a high-level QM methodology (CR-EOMCCSD(T)) to investigate photophysical properties of a green fluorescent chromophore (GFP) analogue in aqueous solution. In chapter 5, the QM/MM method was applied to investigate the catalysis of cellulose degradation catalyzed by copper-dependent polysaccharide monooxygenases (PMOs). The computations were achieved through the utilization of the developed NWChem-AMBER interface. As a primary step, the hydrogen abstraction (HAT) process was examined. The corresponding proposed reaction mechanisms were investigated and activation free energies were estimated. This work also illustrates the use of the full explicit enzyme-substrate model complex in the study of reaction mechanism of PMOs.

Chapter 1

Quantum Mechanical Molecular Mechanics (QM/MM) Methodology

1.1 Introduction

Accurate analysis of chemical transformations in large-scale molecular systems is one of the universal problems that occur across different areas of chemistry, physics, and biology. Such cases are outside the scope of classical methods, but also not always appropriate for pure quantum mechanical (QM) simulations. One obvious reason is the computational expense related to quantum mechanical simulations. Another is the data size and complexity associated with electronic structure degrees of freedom. A potential solution to both issues can be found in multi-physics simulation where different levels of description are used for different parts of the system providing a balanced description of a particular chemical process. This can be achieved by a treatment of the reactive region (e.g. active site of an enzyme) with a QM method and of the environment (e.g. the protein matrix) with a molecular mechanic (MM) method. Such a hybrid approach is regarded as the QM/MM methodology. The principle of the approach was first introduced by Warshel and Levitt in 1976.^[1] Over years, this method has been further developed and gained more popularity in many research areas, and in 2013, the authors, along with Martin Karplus, won the Nobel Prize in Chemistry for “the development of multiscale models for complex chemical systems”.^[2]

One of the most challenging problems is the study of enzymatic reactions or other biomolecular processes. These problems usually involve with system sizes of hundreds of thousands of atoms. With increasing computer power in present days, however, calculations of such complex systems become feasible with the QM/MM method. Over a hundred atoms can be included in the QM active region to captures a transformation of the reaction. In order to acquire an accurate result, a selection of QM and MM methods is nontrivial. Since computational costs are mainly spent toward the QM calculations, hence, a compromise between efficiency and accuracy should

be considered. *ab-initio* QM methods generally provide accurate results with an expense for computational costs. On the other hand, the QM calculations with semi-empirical methods are more favorable in terms of efficiency but the accuracy may be suffered. In practice, density functional theory (DFT)^[3-5] becomes widely used within the QM/MM framework due to their efficiency/accuracy ratio. As far as the MM treatment is concerned, various choices of MM force fields are available to simulate biomolecular systems. These include AMBER,^[6-9] CHARMM,^[10-13] GROMOS,^[14,15] and OPLS-AA.^[16,17]

1.2 Overview of QM/MM method

The details of QM/MM methods and their applications for complex systems have been widely discussed in the literature.^[18-24] Only a brief overview of the methodology is outlined in this chapter.

1.2.1 Partitioning of QM/MM System

A QM/MM system constitutes of 2 subsystems—QM and MM subsystems. The QM subsystem is the part of the system associated with a chemical or electronic transformation (e.g. chemical reaction or electron excitation). Therefore, this part of the system is treated quantum mechanically. On the other hand, the MM subsystem is defined as a part of the system not directly involving in the chemical or electronic transformation but its contribution capable of modulating the change in QM subsystem. This part of the system is described with molecular mechanics (MM), also known as force field (FF). For example, Fig 1.1 shows a QM/MM system for the study of photophysical properties of a green fluorescent protein (GFP) chromophore analogue in the aqueous phase. In this case, the GFP chromophore analogue is treated as the QM subsystem and water molecules as the MM subsystem.

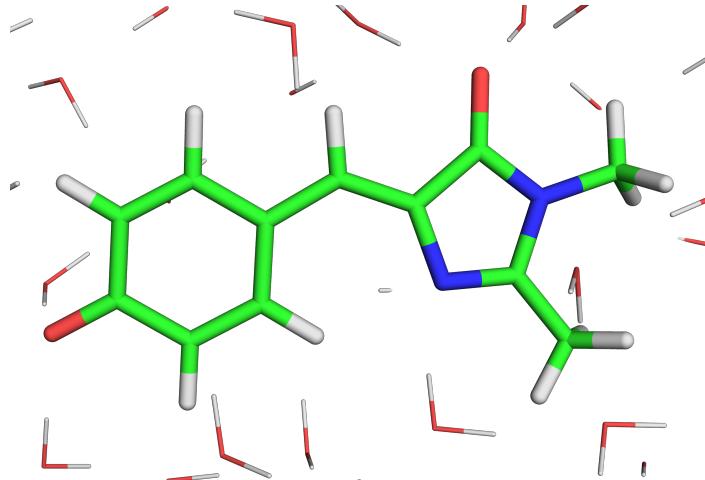


Figure 1.1: An example of a QM/MM system. GFP chromophore analogue (sticks) is the QM subsystem and waters (lines) are the MM subsystem.

1.2.2 QM/MM Hamiltonian

The total energy of a QM/MM system can be expressed as

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM}, \quad (1.1)$$

where E_{QM} is the quantum mechanical energy of the QM subsystem, E_{MM} the molecular mechanical interactions of the MM subsystem, and $E_{QM/MM}$ the interactions between the QM subsystem and the MM subsystem. E_{QM} is typically obtained from the calculation of the electronic wave function via a self-consistent field (SCF) procedure. E_{MM} is calculated from a choice of selected force field, which is expressed using an analytic functional form. The QM/MM interactions can be divided into three terms:

$$E_{QM/MM} = E_{QM/MM}^{electrostatics} + E_{QM/MM}^{VdW} + E_{QM/MM}^{bonded}. \quad (1.2)$$

The electrostatic term describes Coulomb interactions between QM and MM subsystems. In general, there are two schemes to compute this term—Mechanical and electrostatic embedding schemes. For the mechanical embedding scheme, the electron density of the QM subsystem is not

induced via the polarization of the MM charge during wave function evaluation. Instead, electrostatics between QM and MM subsystems is expressed by

$$E_{QM/MM}^{electrostatics} = \sum_i^{N_{QM}} \sum_k^{N_{MM}} \frac{Q_i Q_k}{|\mathbf{r}_i - \mathbf{R}_k|}, \quad (1.3)$$

where Q_i and Q_k are parameterized atomic charges of QM and MM atoms, respectively. Therefore, this evaluation is handled at the MM level. For electrostatic embedding, polarization of MM the MM subsystem is also taken into account during the computation of the wave function. Thus, we have

$$\langle \Psi | H_{eff} | \Psi \rangle = E_{QM} + E_{QM/MM}^{electrostatics} \quad (1.4)$$

and

$$E_{QM/MM}^{electrostatics} = - \sum_k^{N_{MM}} \int d\mathbf{r} \frac{\rho_{QM}(\mathbf{r}) Q_k}{|\mathbf{r} - \mathbf{R}_k|} + \sum_A^{N_{QM}} \sum_k^{N_{MM}} \frac{Z_A Q_k}{|\mathbf{R}_A - \mathbf{R}_k|}, \quad (1.5)$$

where ρ_{QM} is the electron density and Z_A are effective nuclear charges of the QM region, and Q_k are classical electrostatic charges.

Van der Waals interactions in the second term, $E_{QM/MM}^{VdW}$, are represented in terms of a Lennard-Jones potential:

$$E_{QM/MM}^{VdW} = \sum_A^{N_{QM}} \sum_k^{N_{MM}} \varepsilon_{Ak} \left[\left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^{12} - \left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^6 \right]. \quad (1.6)$$

ε_{Ak} is the well depth of the potential well and σ_{Ak} is the distance at which the potential between 2 atoms is zero.

QM-MM bonded interactions (the third term of Eq. (1.2)) takes into account any connections separating QM and MM atoms by 1-3 bonds. This can be expressed as

$$E_{QM/MM}^{bonded} = E_{QM/MM}^{bond} + E_{QM/MM}^{angle} + E_{QM/MM}^{dihedral}, \quad (1.7)$$

which is evaluated at the MM level using a force field. The functional form of Eq. (1.7) is

$$E_{QM/MM}^{bonded} = \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \delta)]. \quad (1.8)$$

r , θ , ϕ denote bond distances, angles, dihedral angles, respectively. r_0 , θ_0 correspond to equilibrium bond distances and angles. n and δ are dihedral multiplicity and phase, respectively. Note that each sum in Eq. (1.8) accounts only for the interaction between QM and MM atoms.

1.2.3 QM/MM Boundary Treatment

In some cases, QM-MM partitioning is performed on QM and MM subsystems connected by covalently chemical bonds. Cutting through these bonds creates unpaired electrons in the QM subsystem. Hence, a special treatment at the QM-MM boundary is required. This includes^[18]

- 1) The unsaturated bond of QM atoms at the boundary must be capped.
- 2) Overpolarization of the QM density by the MM charges at the boundary must be prevented in case of electrostatic embedding scheme.
- 3) Bonded terms that are associated with both subsystems have to be selected such that double-counting of interactions is avoided.

There are several approaches to cope with a QM/MM boundary. Among these, link atom approach^[25-37] is the simplest one. A monovalent link atom is introduced and placed along the bond vector between the QM and MM atoms. Most commonly used element is hydrogen atom although other elements (such as fluorine) may be used. Link atoms are treated as part of the QM subsystem and included in the QM calculation. At each step, the force acting on link atoms are distributed to the corresponding QM and MM atoms. When link atom method is used, overpolarization of classical MM atoms at the boundary must be treated properly.^[18] Some strategies may be used. 1) Deletion of classical MM charges at the boundary.^[30,32,34,35] 2) Shifting or redistributing classical charges to atoms connected to the boundary MM atoms.^[27,38-41] 3) Smearing the charges close to

the QM subsystem^[26,28,42] by using charge distributions. 4) Deletion of the one-electron integrals associated with link atoms.^[31,43-47]

Localized orbitals are another method for dealing with the QM/MM boundary. This involves using a frozen hybrid orbital to saturate the free valanced bond. For localized self-consistent field (LSCF) scheme,^[48-59] the atomic orbitals on the QM atoms at the QM/MM boundary are localized and hybridized. The doubly occupied hybridized orbital points towards the boundary MM atom but is frozen while other hybrid orbitals are optimized along with other QM orbitals during the SCF procedure. Generalized hybrid orbitals (GHO) scheme,^[60-66] also related to the LSCF and frozen orbital approaches, places the hybrid orbitals on the boundary MM atom and the orbital pointing the QM atom is active and included in SCF iterations. Overall, all localized orbital approaches require one or more parametrization steps.

1.3. Motivation and Objective

While various standalone software packages are available for QM and MM calculations, this is not the case for the computations within the QM/MM framework. Besides AMBER software package, which also provides semi-empirical QM/MM features apart from MM simulations, another example of the implementation of QM/MM methods into a single software package can be found in NWChem.^[67] This package provides functionalities to perform a QM/MM calculation using *ab-initio* method or DFT which is a parameter-free approach. However, it is not always possible to maintain both components to fulfill integrity of the code. Therefore, this may raise concerns in terms of credibility of results produced from the unstable code. Another approach is the implementation of the QM/MM methods using a programming model where the QM and MM components reside in two different and widely used software packages. In this case, maintenance issue can be less concerned. To expand functionalities for ab-initio or DFT approach, a MD package such as AMBER^[68] utilizes the second programming model to interface with external QM

packages, including NWChem. In general, MD engine evolves both the QM and MM subsystems concurrently. The QM forces can be retrieved from the QM calculation via an external QM package. This step of the calculation is costly and thus limiting size of the problem for a practical use. Therefore, further interface development has been inspired. In this dissertation, an extended QM/MM approaches via the interface between NWChem and AMBER packages has been implemented to provide additional functionalities to cope with the systems with many degrees of freedom and to be able to apply for chemical processes in enzyme complexes. The efforts include the implementation of QM/MM multiregion optimization, QM/MM nudged elastic band (NEB) calculation, and free energy perturbation (FEP). The QM/MM multiregion optimization provides an efficient algorithm for geometry optimization. The interface for the QM/MM NEB and FEP methods provides a means to study catalytic reactions of enzymes by determining the optimized reaction paths and estimating free energies of the reactions.

1.4 Organization of Chapters

Chapter 2 covers the implementation of the extended QM/MM NWChem/AMBER interface with related theoretical backgrounds also discussed. Chapter 3 includes the validation of the implemented interface using small test cases. Water dimer was chosen to validate the QM/MM multiregion optimization algorithm. NEB and FEP methods were applied to determine the reaction mechanism of a S_N2 reaction of CH₃Br + OH⁻ in aqueous solution. Chapter 4 illustrates the use of QM/MM multiregion optimization in combination with a high-level QM methodology (CR-EOMCCSD(T)) to investigate photophysical properties of a green fluorescent chromophore (GFP) analogue in aqueous solution. In chapter 5, the QM/MM method was applied to investigate the catalysis of cellulose degradation catalyzed by copper-dependent polysaccharide monooxygenases (PMOs). The computations were achieved through the utilization of the developed NWChem-AMBER interface.

1.5 References

- [1] A. Warshel, M. Levitt. *J. Mol. Biol.* **1976**, *103*, 227-249.
- [2] M. Karplus. *Angew. Chem. Int. Ed.* **2014**, *53*, 9992-10005.
- [3] P. Hohenberg, W. Kohn. *Phys. Rev. B* **1964**, *136*, 864-871.
- [4] W. Kohn, L. J. Sham. *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [5] R. G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules; Oxford University Press, New York, **1994**.
- [6] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [7] J. M. Wang, P. Cieplak, P. A. Kollman. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- [8] P. Kollman, Encyclopedia of Computational Chemistry, Vol. 1, **1998**.
- [9] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang, P. Kollman. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
- [10] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.
- [11] A. D. MacKerell, Encyclopedia of Computational Chemistry, Vol. 1, **1998**.
- [12] N. Foloppe, A. D. MacKerell. *J. Comput. Chem.* **2000**, *21*, 86-104.
- [13] A. D. MacKerell, N. K. Banavali. *J. Comput. Chem.* **2000**, *21*, 105-120.
- [14] W. F. van Gunsteren, Encyclopedia of Computational Chemistry, Vol. 2, **1998**.
- [15] W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, W. F. van Gunsteren. *J. Phys. Chem. A* **1999**, *103*, 3596-3607.
- [16] W. L. Jorgensen, Encyclopedia of Computational Chemistry, Vol. 3, **1998**.
- [17] W. L. Jorgensen, D. S. Maxwell, J. TiradoRives. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

- [18] H. M. Senn, W. Thiel. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229.
- [19] H. M. Senn, W. Thiel. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182-187.
- [20] H. Hu, W. Yang. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 17-30.
- [21] P. Banáš, P. Jurečka, N. G. Walter, J. Šponer, M. Otyepka. *Methods* **2009**, *49*, 202-216.
- [22] G. Groenhof. In *Biomolecular Simulations: Methods and Protocols*; Monticelli, L.; Salonen, E., Eds.; Humana Press: Totowa, NJ, **2013**; pp 43-66.
- [23] J. Gao, D. G. Truhlar. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467-505.
- [24] R. A. Friesner, V. Guallar. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389-427.
- [25] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, M. J. Frisch. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1-21.
- [26] D. Das, K. P. Eurenus, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček, B. R. Brooks. *J. Chem. Phys.* **2002**, *117*, 10534-10547.
- [27] A. H. De Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, G. J. Kramer. *J. Phys. Chem. B* **1999**, *103*, 6133-6141.
- [28] M. Eichinger, P. Tavan, J. Hutter, M. Parrinello. *J. Chem. Phys.* **1999**, *110*, 10452-10467.
- [29] U. Eichler, C. M. Kölmel, J. Sauer. *J. Comput. Chem.* **1997**, *18*, 463-477.
- [30] K. P. Eurenus, D. C. Chatfield, B. R. Brooks, M. Hodoscek. *Int. J. Quantum Chem.* **1996**, *60*, 1189-1200.
- [31] M. J. Field, M. Albe, C. Bret, P.-D. Martin, A. Thomas, others. *J. Comput. Chem.* **2000**, *21*, 1088-1100.
- [32] P. D. Lyne, M. Hodoscek, M. Karplus. *J. Phys. Chem. A* **1999**, *103*, 3462-3471.
- [33] F. Maseras, K. Morokuma. *J. Comput. Chem.* **1995**, *16*, 1170-1179.
- [34] U. Ryde. *J. Comput. Aided Mol. Des.* **1996**, *10*, 153-164.
- [35] U. C. Singh, P. A. Kollman. *J. Comput. Chem.* **1986**, *7*, 718-730.
- [36] T. K. Woo, L. Cavallo, T. Ziegler. *Theor. Chem. Acc.* **1998**, *100*, 307-313.
- [37] M. Swart. *Int. J. Quantum Chem.* **2003**, *91*, 177-183.
- [38] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, I. H. Hillier. *Faraday Discuss.* **1997**, *106*, 79-92.

- [39] P. H. König, M. Hoffmann, T. Frauenheim, Q. Cui. *J. Phys. Chem. B* **2005**, *109*, 9082-9095.
- [40] H. Lin, D. G. Truhlar. *J. Phys. Chem. A* **2005**, *109*, 3991-4004.
- [41] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, others. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1-28.
- [42] P. Amara, M. J. Field. *Theor. Chem. Acc.* **2003**, *109*, 43-52.
- [43] D. Bakowies, W. Thiel. *J. Phys. Chem.* **1996**, *100*, 10580-10594.
- [44] M. J. Field, P. A. Bash, M. Karplus. *J. Comput. Chem.* **1990**, *11*, 700-733.
- [45] J. Gao, M. A. Thompson, Combined quantum mechanical and molecular mechanical methods; ACS Publications, **1998**.
- [46] N. Ferré, M. Olivucci. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 71-82.
- [47] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus. *J. Phys. Chem. A* **2000**, *104*, 1720-1735.
- [48] X. Assfeld, J.-L. Rivail. *Chem. Phys. Lett.* **1996**, *263*, 100-106.
- [49] N. Ferré, X. Assfeld, J.-L. Rivail. *J. Comput. Chem.* **2002**, *23*, 610-624.
- [50] A. Fornili, P.-F. Loos, M. Sironi, X. Assfeld. *Chem. Phys. Lett.* **2006**, *427*, 236-240.
- [51] A. Fornili, Y. Moreau, M. Sironi, X. Assfeld. *J. Comput. Chem.* **2006**, *27*, 515-523.
- [52] A. Fornili, M. Sironi, M. Raimondi. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 157-172.
- [53] P.-F. Loos, X. Assfeld. *J. Chem. Theory Comput.* **2007**, *3*, 1047-1053.
- [54] G. Monard, M. Loos, V. Théry, K. Baka, J.-L. Rivail. *Int. J. Quantum Chem.* **1996**, *58*, 153-159.
- [55] R. B. Murphy, D. M. Philipp, R. A. Friesner. *Chem. Phys. Lett.* **2000**, *321*, 113-120.
- [56] D. M. Philipp, R. A. Friesner. *J. Comput. Chem.* **1999**, *20*, 1468-1494.
- [57] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, M. Ghitti. *Theor. Chem. Acc.* **2007**, *117*, 685-698.
- [58] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. G. Ferenczy. *J. Comput. Chem.* **1994**, *15*, 269-282.

- [59] R. B. Murphy, D. M. Philipp, R. A. Friesner. *J. Comput. Chem.* **2000**, *21*, 1442-1457.
- [60] P. Amara, M. J. Field, C. Alhambra, J. Gao. *Theor. Chem. Acc.* **2000**, *104*, 336-343.
- [61] J. Gao, P. Amara, C. Alhambra, M. J. Field. *J. Phys. Chem. A* **1998**, *102*, 4714-4721.
- [62] M. Garcia-Viloca, J. Gao. *Theor. Chem. Acc.* **2004**, *111*, 280-286.
- [63] J. Jung, C. H. Choi, Y. Sugita, S. Ten-No. *J. Chem. Phys.* **2007**, *127*, 204102.
- [64] J. Pu, J. Gao, D. G. Truhlar. *J. Phys. Chem. A* **2004**, *108*, 5454-5463.
- [65] J. Pu, J. Gao, D. G. Truhlar. *J. Phys. Chem. A* **2004**, *108*, 632-650.
- [66] J. Pu, J. Gao, D. G. Truhlar. *ChemPhysChem* **2005**, *6*, 1853-1865.
- [67] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [68] D. A. Case, R. M. Betz, W. Botello-Smith, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, Amber 2016, University of California, San Francisco, **2016**.

Chapter 2

Extended NWChem-AMBER Interface for Large Scaled QM/MM Simulations

2.1 Introduction

The presence two distinctly different models in the QM/MM methodology introduce unique challenges in its implementation. There are essentially two possible frameworks that can be followed. In the single code framework both QM and MM components reside within the same parent code. One example of such approach can be found in NWChem computational chemistry package.^[1] This significantly simplifies the management of data flow between QM and MM components and provides clear and structured approach to running QM/MM simulations. The main drawback is that the parent code needs to have native implementation of both QM and MM components, which is not always possible. The fact that QM and MM components are joined at the source code level and share the same parallel infrastructure also restricts the type of simulations that can be developed. In this work a different programming model where QM and MM components reside in different software packages, namely NWChem computational chemistry package^[1] and AMBER molecular mechanics code^[2,3] are utilized. Such approach allows much flexibility, but more upfront development is required to set up a simulation. These efforts are described in this chapter. NWChem can be downloaded free of charge from <http://www.nwchem-sw.org/index.php/Download>. While full functionalities of AMBER require licensing purchase, its companion AmberTools source code, which contains the main MD engine *sander*, is also available for free through <http://ambermd.org>. Both of which support parallel computations which are beneficial for models with many degrees of freedom within the QM/MM framework.

2.2 QM/MM Methodology

The total energy of a QM/MM system can be expressed as

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM}, \quad (2.1)$$

where E_{QM} is the quantum mechanical energy of the QM subsystem, E_{MM} the molecular mechanical interactions of the MM subsystem, and $E_{QM/MM}$ the interactions between the QM subsystem and the MM subsystem. The QM/MM interactions can be divided into three terms:

$$E_{QM/MM} = E_{QM/MM}^{electrostatics} + E_{QM/MM}^{VdW} + E_{QM/MM}^{bonded}. \quad (2.2)$$

The electrostatic term describes Coulomb interactions between QM and MM regions,

$$E_{QM/MM}^{electrostatics} = - \sum_k^{N_{MM}} \int d\mathbf{r} \frac{\rho_{QM}(\mathbf{r}) Q_k}{|\mathbf{r} - \mathbf{R}_k|} + \sum_A^{N_{QM}} \sum_k^{N_{MM}} \frac{Z_A Q_k}{|\mathbf{R}_A - \mathbf{R}_k|}, \quad (2.3)$$

where ρ_{QM} is the electron density and Z_A are effective nuclear charges of the QM region, and Q_k are classical electrostatic charges. Van der Waals interactions in the second term, $E_{QM/MM}^{VdW}$, are represented in terms of a Lennard-Jones potential:

$$E_{QM/MM}^{VdW} = \sum_A^{N_{QM}} \sum_k^{N_{MM}} \epsilon_{Ak} \left[\left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^{12} - \left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^6 \right]. \quad (2.4)$$

2.3 QM/MM Optimization

While QM/MM partitioning results in a significant reduction in computational load, this alone is not always sufficient to provide a practical approach for simulations other than single point energy calculations. The main problem here is that in addition to spatial extent, simulations of large systems have to contend with the problem of configurational complexity that comes with having many degrees of freedom. If we consider a simple problem of optimization, the number of steps required to achieve a converged solution would be at least that of number of degrees of freedom. The latter can easily approach tens of thousands for a biological system, and the cost of QM calculation at each step will render any simulations impractical. One possible solution to this

problem lies in the observation that small displacement of atoms far away from the QM region will have little effect on its properties. This forms the basis for multiregion optimization strategy where optimizations of QM and MM regions are alternated.^[4] During optimization of the MM region, QM region is represented by fixed electron density. The entire procedure contains the following steps:

- 1) Optimization of QM subsystem keeping MM subsystem fixed
- 2) Calculation of reduced electrostatic representation for the QM subsystem
- 3) Optimization of MM subsystem keeping QM subsystem fixed
- 4) The procedure is repeated in cycle from step 1 to 3 until converged.

Note that in step 2, instead of employing the direct charge density ρ_{QM} resulting from the wave function recalculation which is computationally expensive, the ESP charge fitting scheme can be used. A set of ESP charges representing QM subsystem will be calculated and used in the step 3 of MM optimization. This method greatly improves efficiency in QM/MM calculations for large systems. Hence, Eq. (2.3) for the MM optimization has a form

$$E_{QM/MM}^{electrostatics} = \sum_i^{N_{QM}} \sum_k^{N_{MM}} \frac{Q_i Q_k}{|\mathbf{r}_i - \mathbf{R}_k|} \equiv E_{ESP}, \quad (2.5)$$

where Q_i represents the ESP fitted charge of QM atom i .

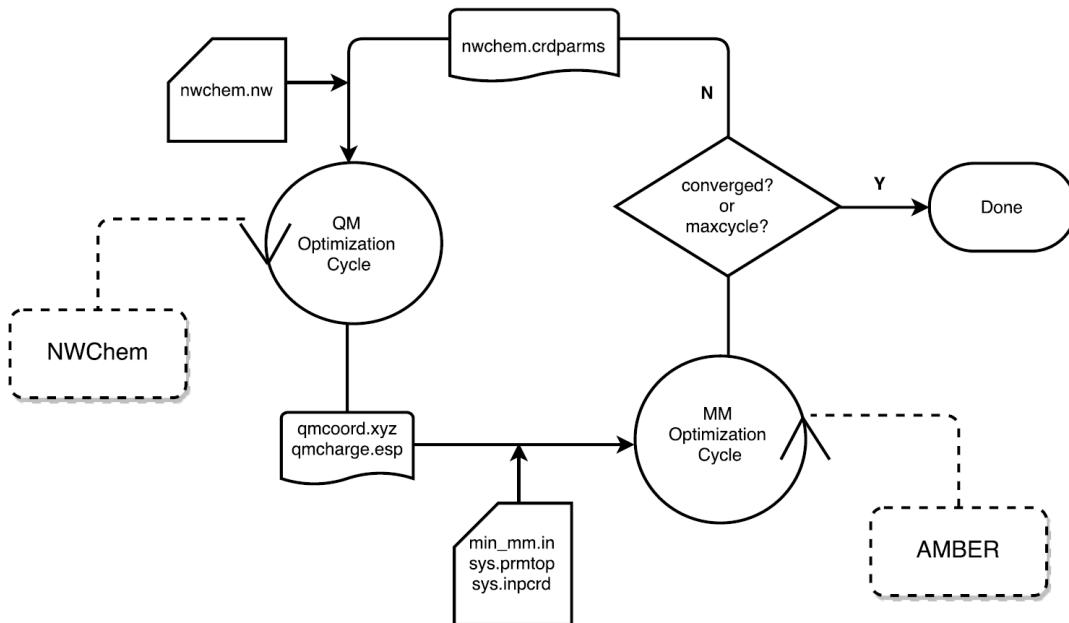


Figure 2.1: Flow diagram for multiregion QM/MM optimization via NWChem-AMBER interface

2.3.1 Implementation of QM/MM Multiregion Optimization

The flow diagram of the QM/MM multiregion optimization is shown in Fig. 2.1. The implementation was developed based on the existing NWChem-AMBER interface.^[5] The QM and MM optimizations are decoupled and performed alternatively. The optimization of the QM subsystem is performed with NWChem whereas the optimization of the MM subsystem is carried out in AMBER. In the current version, the work flow between the QM and MM optimizations is controlled through an external script. A new file *crdparms* is introduced here. This file contains information regarding QM and MM coordinates, charges of MM subsystem, and necessary MM parameters (both bonded and non-bonded) between the QM and MM subsystems. The format of *crdparms* is shown in Fig. 2.2. This file is required as a NWChem input. In a single cycle of the QM optimization, multiple iterations for optimizing the QM subsystem with the fixed MM subsystem are performed. Interactions between the QM and MM subsystems during this stage are based on the MM parameters from *crdparms*. Only electronic embedding scheme is supported in the current development. At the end of each QM optimization cycle, file *qmxyz* containing the

updated QM geometry and file *qmesp* consisting of the QM ESP fitted charges are written out. These files are used to represented the QM subsystem when the MM optimization is performed with AMBER MD engine *sander*. During a single cycle of the MM optimization, the QM subsystem is frozen. The alternative cycles between the QM and MM optimizations are carried out until the convergence is achieved.

```

qm
# indx tag      x      y      z
 1431  C -1.4961780  2.4782010 -8.2713899
...
end

mm
# indx tag      x      y      z      charge
 51   N  11.5012796  0.1519273  8.6941255 -0.4156966
...
end

bond
#   i     j     k_ij      r0
 1429  1431 3.1000E+02 1.5260E+00
...
end
(continued to angle)

```

```

angle
#   i     j     k     k_ijk      theta0
 1430  1429  1431 5.0000E+01 1.9111E+00
...
end

dihedral
#   i     j     k     l     k_ijkl      periodicity      phase
 1433  1431  1429  1442 1.5555E-01 3.0000E+00 0.0000E+00
...
end

vdw
#   i     j     A_coeff      B_coeff
 1431   79  9.2482E+05  5.9901E+02
...
end

scaled_vdw
#   i     j     A_coeff      B_coeff      one_scnb
 1433  1442 8.6154E+04 1.1252E+02 5.0000E-01
...
end

```

Figure 2.2: Format of the file *crdparms* used in NWChem for a cycle of QM optimization

2.3.2 User Interface for QM/MM Multiregion Optimization

As outlined in Fig. 2.1, the QM and MM calculations are decoupled in each cycle of the QM/MM optimization. However, these two steps are interdependent via the intermediate input files—*crdparms* for NWChem and *qmxyz* and *qmesp* for AMBER. The workflow between the two decoupled optimizations is handled via a shell script *qmmm_opt.py* written in Python language. This shell script requires a configuration file *config.in* in which the setup variables must be specified to drive the multiregion QM/MM optimization in more automatic fashion. This scheme is beneficial both in terms of implementation and performance.

The first step is a creation of a coordinate file *inpcrd* and a topology file *prmtop* that contain the system information and all MM parameters. This step is achieved via the standard AMBER preparation interface *LEap*. The second step involves the creation of an AMBER control file *mdin*, a NWChem file *nwin* and the *config.in* to run a QM/MM simulation. This can be separated into three subdivided steps.

1) For the QM optimization, *crdparms* and *nwin* are required. The initial *crdparms* can be generated through the standard AMBER *mdin* file. This includes a specification of the QM subsystem and the setting of *qm_theory* to ‘*WRT_CRDPARMS*’ in the *qmmm* namelist. An example of the *crdparms* creation is shown in Fig. 2.3. *nwin* contains all the detail of the QM calculation and requires a few modifications from the standard NWChem input format. First, the *geometry* directive is no longer needed since this section is read from *crdparms*. Second, the filename of *crdparms* must be specified in the *mm* input block. The name is prepended by the *crdparms load* keyword. Third, the keyword ‘*mm*’ must be specified in the *task* directive. A QM theory and basis set can be specified following the standard NWChem input format. Geometry optimization with NWChem can be invoked via the *driver* directive. Within the *driver* input block, user can specify maximum iterations per cycle for a single cycle of the QM optimization. Finally, *task esp* is specified to calculate the QM fitted ESP charges at the end of each QM optimization cycle. An example of *nwin* for the QM optimization is shown in Fig. 2.3.

```

# no geometry input block needed
# define charge, basis set and QM
# method here
...
mm
crdparms load nwchem.crdparms
end

driver
maxiter 20
end

task mm dft optimize ignore
task esp

```

```

&cntrl
...
imin=1, maxcyc=0,
ifqnt=1,
/
&qmmm
...
qm_theory='WRT_CRDPARMS',
/

```

Figure 2.3: (Left) Example of an AMBER control file to generate *crdparms*. (Right) Example of *nwin* for a single cycle of QM optimization. Bold texts indicate required modifications.

2) For the MM optimization, *mdin*, *inpcrd*, *prmtop*, *qmxyz* and *qmesp* are required as the input files. In the *cntrl* namelist of *mdin*, variables for minimization are specified in the standard way. In the *qmmm* namelist, however, a special setting is required to perform a single cycle of MM optimization using the multiregion scheme. The calculation is triggered when the variable *qm_theory* is set to ‘*READ_ESP*’ and the variable *task* is set to ‘*MM_ONLY*’. Additional setting includes a specification of the variable *qmcoordfile* to the filename of *qmxyz* and of the variable *espfile* to the filename of *qmesp* in the *qmmm* namelist. In addition, *crdparms* for the next QM optimization cycle can be created by setting the variable *wrt_crdparm* to 1. Fig. 2.4 depicts an example of *mdin* for a single cycle of the MM optimization in the multiregion scheme.

```

&cntrl
...
imin=1, maxcyc=2000,
ifqnt=1,
/
&qmmm
...
qm_theory='READ_ESP',
qmcoordfile='nwchem.opt.xyz'
espfile='nwchem.esp'
task='MM_ONLY',
wrt_crdparm=1,
/

```

Figure 2.4: Example of an AMBER *mdin* to perform a single cycle of MM optimization. Required modifications in the *qmmm* namelist is highlighted with bold texts.

- 3) The flow between the QM and MM optimization cycle is controlled via the script *qmmm_opt.py*. This script takes two arguments at the runtime—*config.in* and number of cpus *nprocs*. In the *config.in*, users specify the variables *topfile*, *crdfile*, *ambin*, *nwin* to the filenames of *prmtop*, *inpcrd*, *mdin* and *nwin*, respectively. The *istart* and *iend* variables are used to define the starting cycle *i-th* and the ending cycle *j-th* for the QM/MM optimization. An example of the *config.in* is shown in Fig. 2.5.

```
#####
# Configuration File for QM/MM optimization #
#####

#-- AMBER variables --
ambin = min_mm.in
topfile = sys.prmtop
crdfile = sys.inpcrd

#-- NWChem variables --
nwin = nwchem.nw

# --- start and end cycle --
istart = 1
iend = 5
```

Figure 2.5: Example of config.in to control the workflow of QM/MM multiregion optimization. The lines with # are comments and neglected.

2.4 Technical details

2.4.1 NWChem

The *mm* module has been implemented in NWChem for manipulating data from *crdparms* and computing all of the QM-MM interactions. The module was written in Fortran 77. This module is triggered through the specification of the *mm* input block and the ‘*mm*’ keyword in the task directive of the NWChem input (see Fig. 2.3). Data in *crdparms* is formatted in an input block-like pattern similar to those appearing in the standard NWChem input *nw*. Each block (see Fig. 2.2) begins with the corresponding keyword (*qm*, *mm*, *bond*, *angle*, *dihedral*, *vdw* or *scaled_vdw*) and is enclosed with the ‘*end*’ keyword. In the *bond*, *angle*, *dihedral*, *vdw* and *scaled_vdw* blocks, only those parameters involving the interactions of QM-MM, QM-MMlink, MMlink-MM are presented. A MMlink is referred to a MM atom directly bonded to at least one QM atom. Data allocation for each block is handled within the *mm* module. The QM coordinates are read and built from the *qm* block. In case there are crossing bonds between the QM and MM subsystems, indices in the *bond*

block are used to determine QM-MMlink pairs. Then hydrogen link atoms are placed along their corresponding bond vectors using the following expression

$$\mathbf{R}_{\text{H_link}} = \mathbf{R}_{\text{QM}} + 0.709(\mathbf{R}_{\text{MM}} - \mathbf{R}_{\text{QM}}). \quad (2.6)$$

The constant 0.709 is the ratio between the C-H and C-C equilibrium bond lengths. These link atoms are included in the QM subsystem. In case the electronic embedding scheme is applied, the polarization of the QM subsystem due to electric field of the MM subsystem via point charges are included in the wave function calculations. Gradient and energy calculations for the QM-MM interactions from the force field are performed within the *mm* module using the allocated data derived from *crdparms*. The final gradients and energy are the sum of those obtained from the SCF calculations and the *mm* module. However, the MM-MM interactions are ignored and not calculated in the QM calculation.

2.4.2 AMBER

crdparms file

The information of atom coordinates and force field parameters of a molecular system is contained in AMBER *inpcrd* and *prmtop* while the information of the QM subsystem resides in the *qmmm* namelist. AMBER utilizes these data to write out required data to create *crdparms*, which is required for QM computations. This file can be generated using one of the following methods:

- 1) Set the *qm_theory* variable to ‘*WRT_CRDPARMS*’. This method is commonly used prior to the first QM optimization is performed.
- 2) Set the *wrt_crdparm* variable to 1 when *qm_theory* is not ‘*WRT_CRDPARMS*’ and the variable *task* is not empty. An example for this usage is when a single cycle of the MM optimization is performed with the *qm_theory* = ‘*READ_ESP*’ and the *task* = ‘*MM_ONLY*’ are defined in *mdin* and *crdparms* is required for the next cycle of the QM optimization.

qmmm namelist

The setting of the *qm_theory* = ‘*READ_ESP*’ was designed to work specifically for the MM optimization within the multiregion scheme. Technically, the calculation in this step is performed at the MM level with the QM representation being reduced to a set of ESP charges. These charges are read through *qmesp* and used to evaluate electrostatics between the defined QM subsystem and MM atoms based on Eq. (2.5). In this case, the computation is performed within the *qm_mm* subroutine instead of the regular MM subroutine. Thus, the setup of the variables including the variable *qm_theory* in the *qmmm* namelist is still required to provide the information of the QM subsystem. Additionally, *qmesp* is a pre-requisite and the filename of which must be set in the *espfile* variable when the *qm_theory* = ‘*READ_ESP*’ is active.

2.4.3 Flow Control between QM and MM Optimization Cycles

The workflow between the two decoupled QM and MM calculations are controlled via the script *qmmm_opt.py* written in Python. The script takes two arguments—*config.in* and *nprocs*. The default values are used for the calculations unless user-defined information is provided. In any case, the filenames of the required inputs must be specified in *config.in*. The *qmmm_opt.py* can be regarded as an external driver that makes a series calls to NWChem and AMBER programs alternatively to performed the QM and MM optimization in more automatic fashion by utilizing a number of cpus *nprocs* defined by a user at the runtime. Both NWChem and AMBER support a parallel computation via the message passing interface (MPI).

Below is the list of all supporting user-defined variables in *config.in*.

<i>max_amber_nprocs</i>	=	<integer> no. of cpus used in a single cycle of the MM optimization
		(default is the no. of cpus defined at the runtime)
<i>runqm</i>	=	0 no QM optimization performed

	=	1 perform the QM optimization (<i>crdparms</i> required) (default)
<i>runmm</i>	=	0 no MM optimization performed
	=	1 perform the MM optimization (<i>qmxyz</i> and <i>qmesp</i> required) (default)
<i>chkconv</i>	=	1 write logfile (default)
		(This variable is superceded by the variable <i>runqm</i> , i.e. if <i>runqm</i> = 0, no logfile written)
		0 no logfile written (in case <i>runqm</i> = 1)
<i>istart</i>	=	<integer> beginning at <i>i-th</i> cycle (default 1)
<i>iend</i>	=	<integer> ending at <i>j-th</i> cycle (default 1)
<i>etol</i>	=	<float> energy tolerance for convergence (default value = 0.0001 Hartree)
<i>topfile</i>	=	<string> AMBER <i>prmtop</i> filename
<i>crdfile</i>	=	<string> AMBER <i>inpcrd</i> filename
<i>nwin</i>	=	<string> NWChem <i>nwin</i> file (e.g. <i>nwchem.nw</i>)
<i>ambin</i>	=	<string> AMBER <i>mdin</i> file (e.g. <i>min_mm.in</i>)
<i>qmout_prefix</i>	=	<string> prefix for QM output (default = <i>qm_opt</i>)
<i>mmout_prefix</i>	=	<string> prefix for MM output (default = <i>mm_opt</i>)
<i>nwdata_dir</i>	=	<string> In case a permanent directory is defined in NWChem input file this variable must be set. (default = . (working directory))
<i>nwscratch_dir</i>	=	<string> In case scratch directory is defined in NWChem input file this variable must be set. (default = . (working directory)).

Note that lines beginning with '#' or whitespace lines in *config.in* are ignored. If not specified, *runqm*, *runmm*, and *chkconv* are set to 1 by default.

2.5 Nudged Elastic Band (NEB) Method

2.5.1 Overview of NEB Method

The NEB method^[6-8] is an approach for finding a minimum energy path (MEP) between 2 end points. The MEP is identified by a construction of a set of images of the system between initial and final states. Adjacent images are connected by a spring to ensure the continuity of the path. The optimization of the band by minimizing the total force acting on the images leads to the MEP.

An elastic band with $N+1$ images denoted by $[\mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N]$ where the end points, \mathbf{R}_0 and \mathbf{R}_N are fixed and represent two local minima (initial and final states). The $N-1$ intermediate images are evolved by the optimization. A unit tangent $\hat{\mathbf{r}}_i$ of the image i is defined from two adjacent images along the path, \mathbf{R}_{i+1} and \mathbf{R}_{i-1} . The simplest form of the unit tangent can be expressed as

$$\hat{\mathbf{r}}_i = \frac{\mathbf{R}_{i+1} - \mathbf{R}_{i-1}}{|\mathbf{R}_{i+1} - \mathbf{R}_{i-1}|} . \quad (2.7)$$

The total force acting on an image is the sum of the spring force along the local tangent and the true force perpendicular the local tangent

$$\mathbf{F}_i = \mathbf{F}_i^s|_{\parallel} - \nabla E(\mathbf{R}_i)|_{\perp} , \quad (2.8)$$

where the true force is given by

$$\nabla E(\mathbf{R}_i)|_{\perp} = \nabla E(\mathbf{R}_i) - \nabla E(\mathbf{R}_i) \cdot \hat{\mathbf{r}}_i . \quad (2.9)$$

Here, E is the energy of the system. The spring force has a form

$$\mathbf{F}_i^s|_{\parallel} = k(|\mathbf{R}_{i+1} - \mathbf{R}_{i-1}| - |\mathbf{R}_i - \mathbf{R}_{i-1}|) \hat{\mathbf{r}}_i . \quad (2.10)$$

2.5.2 NEB Method in *ab-initio* QM/MM Calculations

To study chemical reactions in complex systems such as condensed phases or enzymes, combined QM/MM methodology may be used. In most cases, only a small number of atoms, so

called an active site, involve in bond breaking or bond forming process. Other atoms serve as a steric and electrostatic environment to influence the properties and reactivities of the active site. Thus, the combined QM/MM approach allows us to apply QM methods to study chemical process of the small active region while also including environmental effects from thousands of atoms toward the active site with MM methods.

To have an accurate description for the small active QM region, *ab-initio* molecular orbital theory or density functional theory (DFT) [9-11] is a common choice for the QM method. However, rigorous statistical mechanics sampling and reaction dynamics calculation with such a method are still not feasible. Hence, *ab-initio* QM/MM calculations are typically limited to the determination of the MEP that connects the reactant, transition, and product states.

The NEB method can be applied to determine the MEP on an *ab-initio* QM/MM potential surface (PES). In this scheme, it is assumed that the main contribution to the MEP reaction coordinate movement comes from QM active region. The NEB energy, therefore, has a form

$$E^{NEB} = \sum_n E[\{\mathbf{R}_{QM}^n\}, \{\mathbf{R}_{MM}^n\}, \psi] + \sum_n k |\mathbf{R}_{QM}^n - \mathbf{R}_{QM}^{n-1}|^2. \quad (2.11)$$

The index n in the equation denotes the number of images representing the reaction path.

2.5.3 Implementation of *ab-initio* QM/MM NEB Interface

To determine the MEP on an *ab-initio* QM/MM PES, [12-14] a strategy used in the multiregion QM/MM optimization can be applied to the NEB method. The initial step involves the construction of a trial reaction pathway. The simplest way to achieve this is by the linear interpolation of the reaction coordinate (QM coordinates) between two minima, in this case reactant and product states. Subsequent calculations consist of the following steps.

- 1) Calculation of force through the solution of the Schrodinger equation
- 2) Evolution of the QM region according to the NEB methodology

- 3) Calculation of ESP fitted charges for the QM subsystem
- 4) Optimization of the MM region with the QM subsystem represented by the ESP charges from step 3)
- 5) Repetition of step 1 to 4 until converged

The capability to perform NEB calculations has already been available in NWChem. To extend this capability within the QM/MM framework, the functionality implemented in the *mm* module has been integrated. The NEB calculations restricted to the QM subsystem that defines the reaction coordinate are carried out via the *neb* module while the interactions between the QM and MM subsystems during the NEB calculations are taken into account via *crdparms* parsed through the *mm* module in NWChem. On the other hand, the relaxations of the MM subsystems are handled via the MM optimization routine from the QM/MM multiregion scheme in AMBER. Fig. 2.6 is a diagram representing an *ab-initio* QM/MM NEB calculation via the NWChem-AMBER interface.

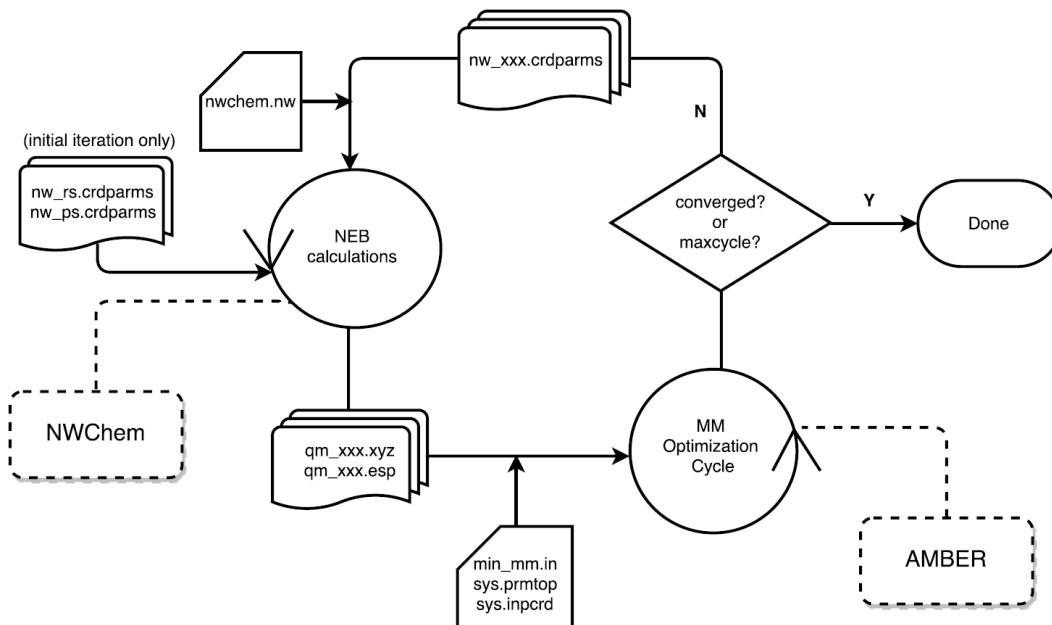


Figure 2.6: Diagram for QM/MM NEB calculation via NWChem-AMBER interface

2.5.4 User Interface for QM/MM NEB Calculations

To start a QM/MM NEB calculation, users are expected to have prior optimized structures for the reactant and product states. This can be achieved using QM/MM multiregion or other optimization methods. In the first NEB calculation, intermediate images are created by the linear interpolation between these 2 states and they will be used in subsequent NEB steps. The invocation of the NEB calculation in NWChem is via a specification of the *neb* input directive in *nwin*. The computational detail can be specified through the variables within the *neb* input block. For example, the number of images used to represent the reaction path can be set via the variable *nbeads*. The spring constant (in atomic unit) used for connecting adjacent images is controlled by the variable *kbeads*. The number of NEB optimization steps and the initial step size for evolution of the NEB images are defined via the variables *maxiter* and *stepsize*, respectively. Users are recommended to visit http://www.nwchem-sw.org/index.php/Release66:NWChem_Documentation for the full list of available variables used in the NEB calculation. Inclusion of the *mm* directive is mandatory for the QM/MM NEB calculation. The filename of *crdparms*, usually the one for the reactant state or the first image, must be specified in the *mm* input block via the keyword *crdparms load*. In addition, for the first NEB calculation, 2 *crdparms* files for the reactant and product states must be specified by setting

```
set qmmm:neb_path_limits <rs_crdparms> <ps_crdparms> ,
```

which are used to create the intermediate NEB images. Instead, to continue or restart the QM/MM NEB calculation, the following must be specified:

```
set qmmm:neb:restart .true.
```

To successfully continue the QM/MM NEB calculation, it requires that the *crdparms* and the wavefunction information (NWChem *movecs* files) for all NEB images have already existed. An example of the NWChem input for QM/MM NEB calculation is depicted in Fig. 2.7. The relaxation of the MM subsystem for the NEB images can be performed using the MM optimization

outlined in the QM/MM multiregion optimization section. The MM optimization can be carried out alternatively with the NEB calculation until the MEP is achieved.

```

# no geometry input block needed
# define charge, basis set and QM
# method here
...
mm
  crdparms load nwchem_rs.crdparms
end
set qm:mm:neb_path_limits nwchem_rs.crdparms \
  nwchem_ps.crdparms

neb
  nbeads 12
  kbeads 0.1
  maxiter 5
  stepsize 0.5
end

task mm dft neb ignore

```

Figure 2.7: Example of *nwin* to perform QM/MM NEB calculation. Two *crdparms* are specified to generate a trial reaction path connecting 2 states. The total of 12 images is used to represent the path. The spring constant is 0.1 a.u. The NEB calculation is performed for 5 iterations with the initial step size of 0.5 a.u.

2.6 Free Energy Perturbation (FEP)

2.6.1 Overview of FEP in *ab-initio* QM/MM Calculations

With the determined MEP, FEP calculations to estimate the free energy changes along the reaction path can be achieved.^[4,15-19] Suppose a reaction coordinate s is defined by $s = f(\mathbf{r}_{\text{QM}})$, where \mathbf{r}_{QM} represents the QM coordinates in a QM/MM system. Thus, the partition function can be expressed as

$$Z = \int \exp \{-\beta [E_{\text{qm}}(\mathbf{r}_{\text{QM}}) + E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) + E_{\text{mm}}(\mathbf{r}_{\text{MM}})]\} d\mathbf{r}_{\text{QM}} d\mathbf{r}_{\text{MM}}$$

$$= \int Z(s)ds, \quad (2.12)$$

where $\beta = 1/k_B T$ is the inverse temperature and k_B is the Boltzmann constant. \mathbf{r}_{MM} denotes the MM coordinates. Also, the partial partition function associated with the reaction coordinate can be defined as

$$\begin{aligned} Z(s) = & \int \delta(s - f(\mathbf{r}_{\text{QM}})) \times \exp\{-\beta[E_{\text{qm}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) \\ & + E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) + E_{\text{mm}}(\mathbf{r}_{\text{MM}})]\} d\mathbf{r}_{\text{QM}} d\mathbf{r}_{\text{MM}}, \end{aligned} \quad (2.13)$$

where δ is the Dirac delta function. Thus, the corresponding free energy of the reaction is

$$F(s) = -1/\beta \ln Z(s). \quad (2.14)$$

The free energy difference between two adjacent states A and B along the MEP is

$$\begin{aligned} \Delta F^{(A \rightarrow B)} &= F(s_B) - F(s_A) \\ &= -1/\beta \ln \left\langle \exp\{-\beta[E_{\text{total}}(\mathbf{r}_{\text{QM}}(s_B), \mathbf{r}_{\text{MM}}) - E_{\text{total}}(\mathbf{r}_{\text{QM}}(s_A), \mathbf{r}_{\text{MM}})]\} \right\rangle_A, \end{aligned} \quad (2.15)$$

where $\langle \dots \rangle_A$ denotes the ensemble average at the reaction coordinate s_A . In case the QM fluctuations are ignored and their coordinates are frozen to those on the MEP, $\mathbf{r}_{\text{QM}}^{\text{min}}$, this leads to

$$\begin{aligned} \Delta F^{(A \rightarrow B)} &= \Delta E_{\text{int}}^{(A \rightarrow B)} + \Delta F_{\text{mm}}^{(A \rightarrow B)} \\ &= E_{\text{int}}(\mathbf{r}_{\text{QM}}^{\text{min}}(s_B), \mathbf{r}_{\text{MM}}^{\text{min}}(s_B)) - E_{\text{int}}(\mathbf{r}_{\text{QM}}^{\text{min}}(s_A), \mathbf{r}_{\text{MM}}^{\text{min}}(s_A)) \\ &\quad - 1/\beta \ln \left\langle \exp\{-\beta[E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}^{\text{min}}(s_B), \mathbf{r}_{\text{MM}}) - E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}^{\text{min}}(s_A), \mathbf{r}_{\text{MM}})]\} \right\rangle_{\text{MM},A}, \end{aligned} \quad (2.16)$$

where $\langle \dots \rangle_{\text{MM},A}$ denotes the ensemble average over the MM subsystem. Additionally, E_{int} is the internal energy

$$E_{\text{int}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) = \langle \psi | H_{\text{eff}} | \psi \rangle - E_{\text{qm/mm}}^{\text{esp}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}). \quad (2.17)$$

The reaction coordinates in Eq. (2.16) is defined by QM coordinates and only the effects of MM fluctuations are included. If the QM fluctuations is assumed to be harmonic and the MM subsystem

is frozen, free energy contribution, denoted $F_{\text{qm-harmonic}}$, can be obtained from frequency calculations at those points on the MEP. The total free energy difference between 2 states A and B becomes

$$\Delta F^{(A \rightarrow B)} = \Delta E_{\text{int}}^{(A \rightarrow B)} + \Delta F_{\text{mm}}^{(A \rightarrow B)} + \Delta F_{\text{qm-harmonic}}^{(A \rightarrow B)}. \quad (2.18)$$

Thus, we only need the QM internal energies and the harmonic frequencies between these 2 states and perform the averaging over MM subsystem at the MM level to calculate the free energy difference between the 2 states. This provides an efficient and feasible approach for free energy estimation along the reaction pathway.

ESP charges (\mathbf{Q}) can be used to represent the fixed QM subsystem when sampling over the MM subsystem for the second term of Eq. (2.18). From a smooth transformation for the reaction coordinate between 2 states A and B,

$$\begin{aligned} \mathbf{r}_\lambda &= (1 - \lambda)\mathbf{r}_A + \lambda\mathbf{r}_B \\ \mathbf{Q}_\lambda &= (1 - \lambda)\mathbf{Q}_A + \lambda\mathbf{Q}_B, \end{aligned} \quad (2.19)$$

where λ is in the range between 0 and 1. We have that

$$\Delta F_{\text{mm}}^{(A \rightarrow B)} = -\sum_i 1/\beta \ln \left\langle \exp \left\{ -\beta \Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}} \right\} \right\rangle_{\text{MM}, \lambda_i}. \quad (2.20)$$

$$\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}} = E_{\text{qm/mm}}(\mathbf{r}_{\lambda_{i+1}}, \mathbf{Q}_{\lambda_{i+1}}, \mathbf{r}_{\text{MM}}) - E_{\text{qm/mm}}(\mathbf{r}_{\lambda_i}, \mathbf{Q}_{\lambda_i}, \mathbf{r}_{\text{MM}}). \quad (2.21)$$

Together with the MEP determined from the NEB method, we can apply the FEP approach to all adjacent images connecting reactant and product state to obtain the free energy profile along the reaction path.

2.6.2 Implementation of FEP Method

The FEP algorithm (Fig. 2.8) has been implemented in AMBER to perform the statistical averaging over the MM subsystem (the second term in Eq. (2.18)). This step requires the QM subsystem to be frozen at the $\mathbf{r}_{\text{QM}}^{\text{min}}$ along the MEP and to be represented by a set of ESP charges.

The electrostatics between the QM and MM subsystems is based on Eq. (2.5). The sampling is achieved via the standard MD protocol. During this stage, the FEP calculation can be concurrently computed. Given the QM coordinates and ESP charges for NEB images representing a MEP, free energy differences between adjacent images can be estimated. In practice, a smooth transformation between 2 states is applied and the estimation of the free energy differences is based on Eq. (2.20).

```

while i < nmax           ! nmax = max no. steps for FEP
  i_step = i_step + 1
  t_i = temp_i           ! temperature of step i
  tsum = tsum + t_i      ! sum of temperature over n step
  exp_i = exp(-(ener1-ener2)/(t_i*kb)) ! exponential term
  exp_sum = exp_sum + exp_i ! sum of exponential term
  taver = tsum/i_step    ! temperature average
  exp_aver = exp_sum/i_step ! exponential term average
  free_en = -kb*taver*log(exp_aver) ! free energy difference

```

Figure 2.8: pseudocode for FEP algorithm

2.6.3 User Interface for FEP Calculations

The entire process to obtain free energy differences as expressed in Eq. (2.20) is carried out in AMBER. Apart from the standard setup in the *cntrl* namelist for MD simulation, extra modifications in the *qmmm* namelist of AMBER *mdin* are necessary. First, both QM coordinates and ESP charges for 2 states are required. For the QM coordinates, filenames of *qmxyz1* and *qmxyz2* are defined through the variables *xyzfile1* and *xyzfile2*, respectively. For ESP charges, the variables *espfile1* and *espfile2* are set to the names of *qmesp1* and *qmesp2*, respectively. Second, the variable *qm_theory* must be specified to ‘*READ_ESP*’ and the *task* variable specified to ‘*FEP*’. Third, two values of lambda (λ_i and λ_{i+1}) in the range between 0 to 1 are specified via the variables *lambda1* and *lambda2*, respectively. The optional variable *fnstart* may be set to specify the starting point

where the free energy difference calculation begins. An example of AMBER *mdin* for FEP calculation is shown in Fig. 2.8.

```

&cntrl
...
nstlim = 100000,
ifqnt = 1,
/
&qmmm
...
qm_theory = 'READ_ESP',
task = 'FEP',
fnstart = 1,
lambda1 = 0.0 , lambda2 = 0.2 ,
xyzfile1 = 'qm1.xyz' , xyzfile2 = 'qm2.xyz' ,
espf1 = 'qm1.esp' , espf2 = 'qm2.esp',
/

```

Figure 2.9: Example of AMBER *mdin* illustrating the FEP calculation between 2 states with the width of 0.2 between λ_i and λ_{i+1} . The MD simulation is performed for 100000 steps and the FEP calculation starts at the first step.

2.7 References

- [1] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [2] R. Salomon-Ferrer, D. A. Case, R. C. Walker. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198-210.
- [3] D. A. Case, R. M. Betz, W. Botello-Smith, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, Amber 2016, University of California, San Francisco, **2016**.
- [4] Y. Zhang, H. Liu, W. Yang. *J. Chem. Phys.* **2000**, *112*, 3483-3492.
- [5] A. W. Götz, M. A. Clark, R. C. Walker. *J. Comput. Chem.* **2014**, *35*, 95-108.

- [6] H. Jonsson, G. Mills, K. W. Jacobsen. In Classical and Quantum Dynamics in Condensed Phase Simulations; World Scientific, **2011**; pp 385-404.
- [7] G. Henkelman, H. Jónsson. *J. Chem. Phys.* **2000**, *113*, 9978-9985.
- [8] D. Sheppard, R. Terrell, G. Henkelman. *J. Chem. Phys.* **2008**, *128*, 134106.
- [9] P. Hohenberg, W. Kohn. *Phys. Rev. B* **1964**, *136*, 864-871.
- [10] W. Kohn, L. J. Sham. *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [11] R. G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules; Oxford University Press, New York, **1994**.
- [12] H. Liu, Z. Lu, G. A. Cisneros, W. Yang. *J. Chem. Phys.* **2004**, *121*, 697-706.
- [13] L. Xie, H. Liu, W. Yang. *J. Chem. Phys.* **2004**, *120*, 8039-8052.
- [14] G. A. Cisneros, H. Liu, Z. Lu, W. Yang. *J. Chem. Phys.* **2005**, *122*, 114502.
- [15] J. Kästner, H. M. Senn, S. Thiel, N. Otte, W. Thiel. *J. Chem. Theory Comput.* **2006**, *2*, 452-461.
- [16] Z. Lu, W. Yang. *J. Chem. Phys.* **2004**, *121*, 89-100.
- [17] T. H. Rod, U. Ryde. *Phys. Rev. Lett.* **2005**, *94*, 138302.
- [18] T. H. Rod, U. Ryde. *J. Chem. Theory Comput.* **2005**, *1*, 1240-1251.
- [19] H. Hu, Z. Lu, W. Yang. *J. Chem. Theory Comput.* **2007**, *3*, 390-406.

Chapter 3

Validation of QM/MM NWChem-AMBER Interface

3.1 Introduction

To validate our developed QM/MM NWChem-AMBER interface, two test cases were chosen for benchmarking. First, water dimer was used to validate the QM/MM multiregion optimization. Second, the free energy and optimized reaction path of a biomolecular nucleophilic substitution (S_N2) reaction of $CH_3Br + OH^-$ in aqueous solution was determined using a combination of the QM/MM multiregion optimization, nudged elastic band, and free energy perturbation.

3.2 QM/MM Optimization Benchmarking

3.2.1 Methodology

Two water molecules were created by adopting an arbitrary geometry such that one water acts as a hydrogen bond donor (D) and another as hydrogen bond acceptor (A). In the first case, we treated D as the QM subsystem and A as the MM subsystem. The opposite was done in the second case. The multiregion QM/MM optimizations were carried out using TIP3P model^[1] for the classical water molecule and BLYP^[2,3]/aug-cc-pVQZ^[4] for the QM water molecule. The QM optimization cycles were carried using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm while the MM optimizations were performed using the steepest descent (SD) algorithm. In addition, full QM optimizations were performed at BP86/TZVP^[5,6] and B3LYP/TZVP levels. Binding energies were also calculated and reported. The results from the optimizations are shown in Table 3.1. The optimized structure of the water dimer is shown in Fig. 3.1.

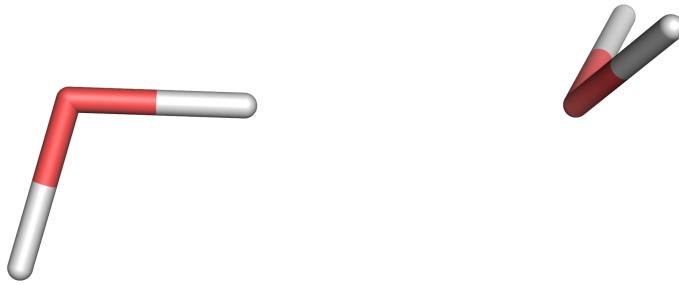


Figure 3.1: Water dimer with donor (left) and accepter (right)

Table 3.1: Binding energy (kcal/mol), distances (Å), and angle (°) of a water dimer at the full QM and QM/MM (with hydrogen bond donor D or acceptor A in the QM region) levels

D-A	E	d(H—O)	d(O—O)	a(H-O-H)	Refs.
Experiment	5.44	N/A	2.98	174±20	7-9
Full QM BP86	5.46	1.91	2.89	173.9	this work
	5.72	1.91	2.89	174.5	10
	5.68	1.89	2.86	165.5	11
Full QM B3LYP	5.71	1.93	2.90	174.5	this work
	5.99	1.93	2.90	175.6	10
QM/MM TIP3P- BLYP	5.38	1.82	2.78	177.9	this work
	5.56	1.86	2.82	176.7	10
	6.06	1.84	2.81	178.3	12
QM/MM BLYP- TIP3P	8.13	1.69	2.69	179.0	this work
	8.18	1.64	2.63	178.9	10
	8.34	1.69	2.68	179.0	12

3.2.2 Results and Discussion

Water dimers with C_s symmetry were obtained from all geometry optimizations. The full QM optimizations agree well with experimental values for binding energies. The hydrogen bond distances were underestimated by 0.1 Å. Regarding geometry optimizations using QM/MM scheme, the binding energies are close to the experimental value when the QM subsystems were treated as hydrogen acceptors. The overall hydrogen bond distances are underestimated by 0.2 Å. In spite of the agreement with the experimental value, the resulting binding energy was

underestimated (5.38 kcal/mol) with the QM/MM multiregion optimization in this work compared to other reference calculations at the same level.

Overall, the results obtained from the geometry optimizations in this work validate our implementation of NWChem-AMBER interface as well as the robustness of QM/MM multiregion optimization algorithm for a small test case such as water dimer.

3.3 QM/MM Method for S_N2 Reaction of CH₃Br + OH⁻ in Aqueous Solution

3.3.1 Overview

In this section, the determination of a biomolecular nucleophilic substitution (S_N2) reaction path in a solution has been performed to validate the implementation of QM/MM NEB and free energy perturbation (FEP) methods through the developed NWChem-AMBER interface. The S_N2 reaction of CH₃Br + OH⁻ in aqueous solution was chosen in this study. In the presence of a system with many degree of freedom, it is impractical to treat the whole system quantum mechanically. In this case, a hybrid QM/MM method can be used to alleviate the problem. The region involving bond forming and bond breaking can be defined with a QM description. On the other hand, the environment, whose effect is not directly associated to the chemical transformation, can be treated with a MM description. In this study, QM/MM NEB method was applied to determine the reaction path and the FEP calculations were used to determine the contribution of aqueous solution. In addition, the reaction barrier was estimated and compared with experimental data.

3.3.2 Computational Details

System Preparation

The solute CH₃Br and OH⁻ from gas phase optimization were solvated in a water box size of 40 Å³ using an AMBER interface LEaP.^[13] SPC/E water model^[14] was used to describe aqueous solution. Van der Waals (VdW) parameters for solute are based on the standard AMBER99 force

field.^[15,16] The solute was treated as the QM subsystem and water molecules were treated as the MM subsystem.

Prior to QM/MM optimization, molecular dynamics (MD) were carried out to relax the MM subsystem. During this stage, the QM subsystem was represented by a set of ESP charges and kept fixed. MD simulation was performed for 500 ps using NVT ensemble at 298 K. The last snapshot from MD trajectory was extracted for subsequent QM/MM optimizations.

QM/MM Optimization

The QM/MM multiregion optimization was adopted to obtain the reactant state (RS). The optimization was done via the newly developed QM/MM NWChem-AMBER interface. The QM calculations were performed using B3LYP/aug-cc-pVDZ. The net QM charge is -1. The cutoff of 15 Å was chosen to include MM charges in the wave function calculations via electronic embedding scheme and as a short-range cutoff in MM optimization cycles. Next, the product state (PS) was generated using the constrained QM/MM optimization. This was done by constraining the C-O and C-Br bonds of the reactant state. The same level of theory used in RS optimization was applied to obtain PS. Finally, the optimized RS and PS were used to determine the reaction path with the QM/MM NEB method.^[17]

QM/MM NEB Calculation

The determination of the reaction path was done via the QM/MM NEB method. The optimized RS and TS were adopted to generate an initial reaction path. A total of 10 images were used to represent the reaction path. Verification of the transition state (TS) was based on numerical frequency calculation and was identified by one imaginary frequency. The QM wave function was calculated using B3LYP/aug-cc-pVDZ with the net QM charge of -1. The MM optimization for each image was performed every 5 steps of the QM NEB calculation. During the MM optimization,

the QM subsystem for each image was kept fixed and represented by a set of fitted ESP charges which were calculated at the end of each NEB calculation.^[18]

Free Energy Calculation

The reaction barrier can be estimated via an approximation of free energy difference between RS and TS, which can be expressed as^[19]

$$\Delta F^{(A \rightarrow B)} = \Delta E_{\text{int}}^{(A \rightarrow B)} + \Delta F_{\text{mm}}^{(A \rightarrow B)} + \Delta F_{\text{qm-harmonic}}^{(A \rightarrow B)}, \quad (3.1)$$

where A and B represent RS and TS. $\Delta E_{\text{int}}^{(A \rightarrow B)}$ can be obtained by calculating the difference of QM internal energies between the 2 states. $\Delta F_{\text{qm-harmonic}}^{(A \rightarrow B)}$ is calculated through

$$\Delta F_{\text{qm-harmonic}}^{(A \rightarrow B)} = \Delta H_{\text{qm}}^{(A \rightarrow B)} - T \Delta S_{\text{qm}}^{(A \rightarrow B)}, \quad (3.2)$$

where H is enthalpy, S is entropy, and T is temperature. These values can be obtained via numerical frequency calculation for the corresponding states.

The second term of Eq. (3.1) is the contribution of the MM subsystem to the reaction and can be obtained through the FEP calculation,

$$\Delta F_{\text{mm}}^{(A \rightarrow B)} = - \sum_i 1/\beta \ln \left\langle \exp \left\{ -\beta \Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}} \right\} \right\rangle_{\text{MM}, \lambda_i}. \quad (3.3)$$

$$\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}} = E_{\text{qm/mm}}(\mathbf{r}_{\lambda_{i+1}}, \mathbf{Q}_{\lambda_{i+1}}, \mathbf{r}_{\text{MM}}) - E_{\text{qm/mm}}(\mathbf{r}_{\lambda_i}, \mathbf{Q}_{\lambda_i}, \mathbf{r}_{\text{MM}}). \quad (3.4)$$

$\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}}$ is the total MM energy difference between states λ_i and λ_{i+1} where the electrostatics between QM and MM subsystems can be expressed as

$$E_{\text{QM/MM}}^{\text{electrostatics}} = \sum_i^{N_{\text{QM}}} \sum_k^{N_{\text{MM}}} \frac{Q_i Q_k}{|\mathbf{r}_i - \mathbf{R}_k|}. \quad (3.5)$$

Q_i is the ESP fitted charge of the QM atom i and Q_k is an empirical charge of the MM atom k . The corresponding QM coordinates (\mathbf{r}) and charges (\mathbf{Q}) of the intermediate state λ_i is based on a smooth transformation between state A and state B,

$$\mathbf{r}_\lambda = (1 - \lambda)\mathbf{r}_A + \lambda\mathbf{r}_B$$

$$\mathbf{Q}_\lambda = (1 - \lambda)\mathbf{Q}_A + \lambda\mathbf{Q}_B, \quad (3.6)$$

where λ is in the range between 0 and 1.

In this study, the FEP calculations were carried out along the NEB reaction path. Free energy difference was calculated for each of 2 consecutive images. A total of 5 intermediate states with the λ increment of 0.2 were used for each pair of images. The samplings were performed in both forward and backward directions. Thus, this required a total of 90 FEP calculations. In addition, the free energy differences between RS and TS as well as between TS and PS were also performed through a direct smooth transformation. A total of 20 intermediate states (λ) were created between 2 corresponding states. The samplings were done through MD simulations using NVT ensemble at 298 K. The systems were first equilibrated for 30 ps and followed by a total of 300 ps of free energy calculations. The ESP charges for the QM subsystem were obtained using NWChem.^[20] All FEP calculations were performed using AMBER.^[13,21]

3.3.3 Results and Discussion

Reactant State (RS)

The optimized structure of RS in the gas and the aqueous solution are shown in Fig. 3.2. The hydrogen bond distance H---OH⁻ is 1.69 Å in the gas phase and 2.83 Å in the aqueous solution. The longer H---OH⁻ distance in the aqueous solution results from significant charge screening on OH⁻ by the surrounding waters, decreasing the H---O interaction. The C-O bond distance is 2.79 Å in the gas phase and 3.46 Å in the aqueous phase. However, the angle O-C-Br is $\sim 154^\circ$ in the aqueous phase compared to $\sim 124^\circ$ in the gas phase, indicating more in-line attack orientation. In addition, OH⁻ accepts 5 hydrogen bonds from waters with average O---OH⁻ distance of 2.59 Å. This solvation pattern is in consistent with previous experimental^[22,23] and theoretical studies^[24-26] of OH⁻ in the aqueous solution. The C-Br bond length is 1.98 Å in the aqueous solution compared to

2.03 Å in the gas phase while the C-H bond distances are 1.09 Å in the aqueous solution and 1.10-1.14 Å in the gas phase. The elongated C-H bond in the gas phase is caused by the resulting hydrogen bond with OH⁻.

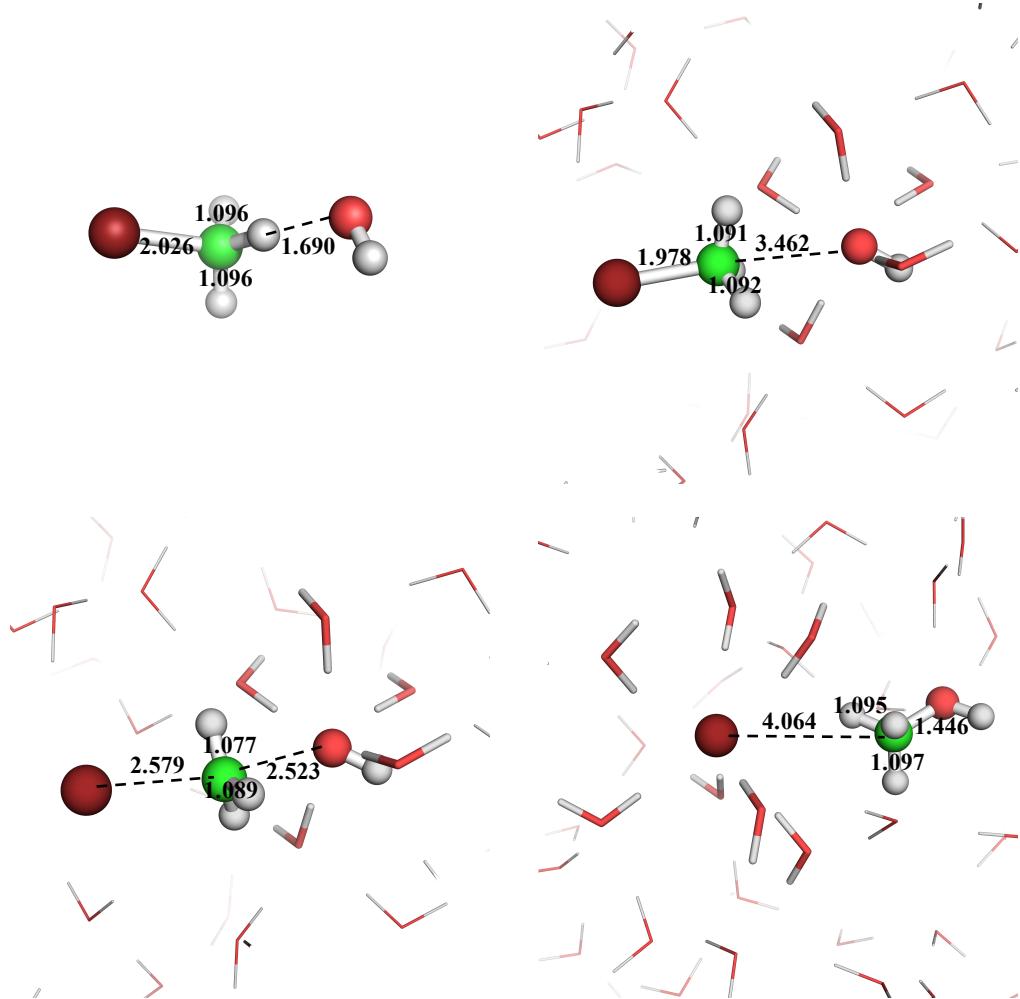


Figure 3.2: Optimized structures (upper left) gas phase RS. (upper right) aqueous phase RS. (lower left) aqueous phase TS. (lower right) aqueous phase PS. Atoms in colors: green = carbon; crimson red = bromine; red = oxygen; white = hydrogen. Thick water sticks are solvation shell of either OH⁻ or Br⁻.

Transition State (TS)

The structure of TS in aqueous solution is shown in Fig. 3.2. TS was identified by the top image from the NEB reaction path and confirmed with a single imaginary frequency at $175i\text{ cm}^{-1}$.

The formation of TS is via a rotation around a C-Br bond of CH_3Br , causing C to be more exposed to nucleophilic OH^- and breaking a hydrogen bond H--- OH^- . The C-O bond distance decreases to 2.52 Å from 3.46 Å in RS. On the other hand, The C-Br increases from 1.98 Å in RS to 2.58 Å. A nearly planar CH_3 group is formed in this state. Similar to the solvation pattern of OH^- in the reactant state, five hydrogen bonds are observed for the solvation around OH^- with average O--- OH^- distance of 2.63 Å.

Product State (PS)

The structure of PS in aqueous solution is shown in Fig. 3.2. In this state, the breaking of C-Br bond and the forming of C-OH are complete. C-Br bond distance is 4.06 Å and C-OH bond distance is 1.45 Å. The Br^- leaving group are solvated with seven waters with average Br^- -O distance of 3.21 Å. On the other hand, the newly formed OH group accepts three hydrogen bonds at average O---OH distance of 2.83 Å as well as donating an extra hydrogen bond at 2.62 Å.

Charge Distribution Along the Reaction Pathway

Fig. 3.3 displays charge distribution for each group along the NEB reaction pathway. At RS (image 1), the charge of CH_3Br is neutral and the negative charge of -1 is concentrated around OH group. At TS (image5), the excessive negative charge remains concentrated around OH group while the Br charge is shifted from -0.27 to -0.43. At PS (image 10), Br become a leaving group and carries the excessive charge of -1 while the total charge of CH_3OH is neutral.

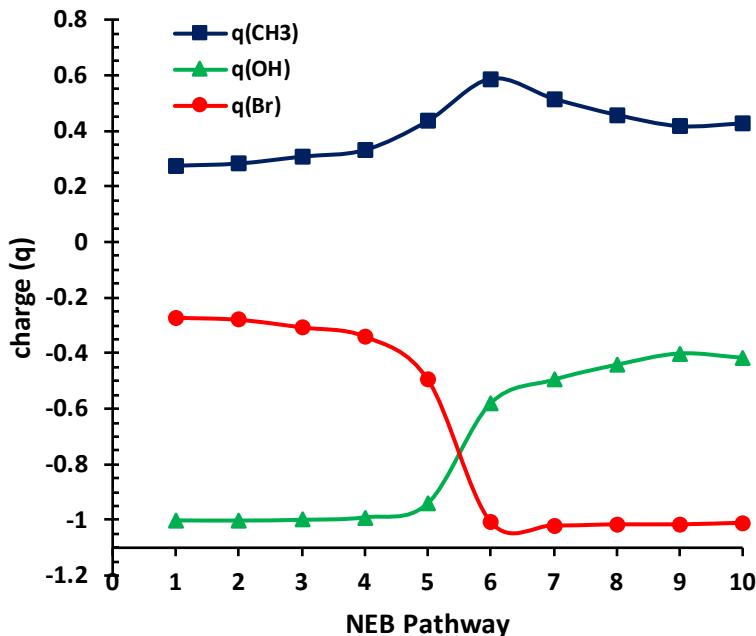


Figure 3.3: Charge distribution for each group of solute along the NEB reaction pathway

Free Energy Profile

The free energy profile along the NEB reaction pathway for the $\text{S}_{\text{N}}2$ reaction of $\text{CH}_3\text{Br} + \text{OH}^-$ in aqueous solution is shown in Fig. 3.5. In addition, the solvation free energy contribution from FEP calculation is also shown in the same graph. All free energy values are relative to that of the first NEB image, which represents RS. The free energy of activation was estimated to be 25.3 kcal/mol. This is consistent with the experimental value of 23.0 kcal/mol.^[27,28] Compared to the same theoretical study by Wang *et al.*,^[29] our calculated activation free energy is \sim 6 kcal/mol higher within the same level of theory. It is possible that different pathways were identified between the 2 studies, as indicated by the discrepancy of the imaginary frequencies of the transition states ($175i$ vs $398i \text{ cm}^{-1}$). The reaction free energy was calculated to be -31.5 kcal/mol.

The effect of solvent plays a significant role towards the activation free energy of the reaction. The free energy from solvation contributes 20.5 kcal/mol to the activation free energy and \sim 20.0 kcal/mol to the reaction free energy. The increase of solvation free energy can be explained

based on the difference between solvation free energy of OH^- (-104.6 kcal/mol)^[30] and Br^- (-85.56 kcal/mol).^[31] Together with solvation energies for CH_3Br (-0.8 kcal/mol)^[32] and CH_3OH (-3.19 kcal/mol),^[33] the solvation free energy of the reaction is 16.6 kcal/mol, which agrees well with our calculated value of \sim 20.0 kcal/mol. Thus, this causes the reaction become less favorable in the aqueous solution. The other solvent contribution along the reaction path is from polarization. Figure 3.4 shows the plots for the gas phase QM energies and the QM internal energies where the wave functions were evaluated in the presence of aqueous environment. The solvent polarization raises the energies by 14.7 kcal/mol for the reactant state, \sim 19.2 kcal/mol for the transition state, and \sim 14.4 kcal/mol for the product state. The polarization effect raises the reaction barrier by 4.4 kcal/mol and reduces the reaction energy by 0.3 kcal/mol.

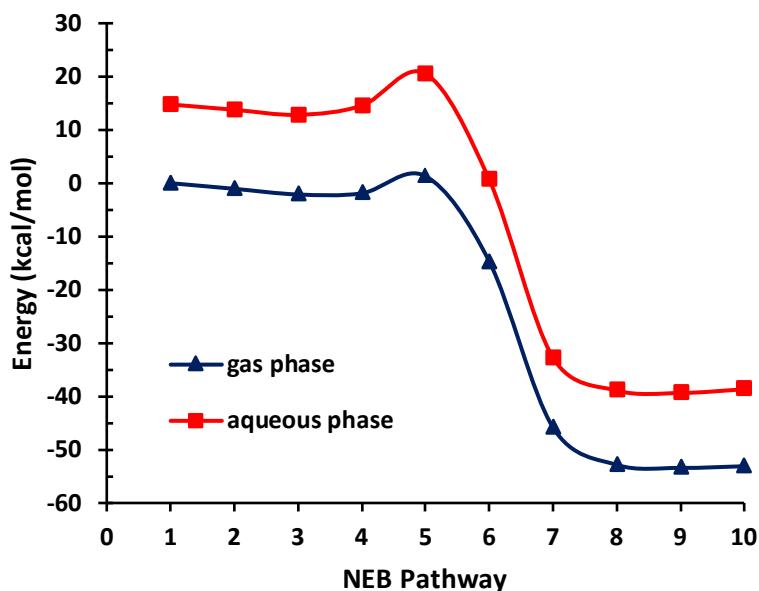


Figure 3.4: Comparison of QM internal energy between gas phase and aqueous phase along the NEB reaction pathway. The first NEB image in the gas phase is the reference point.

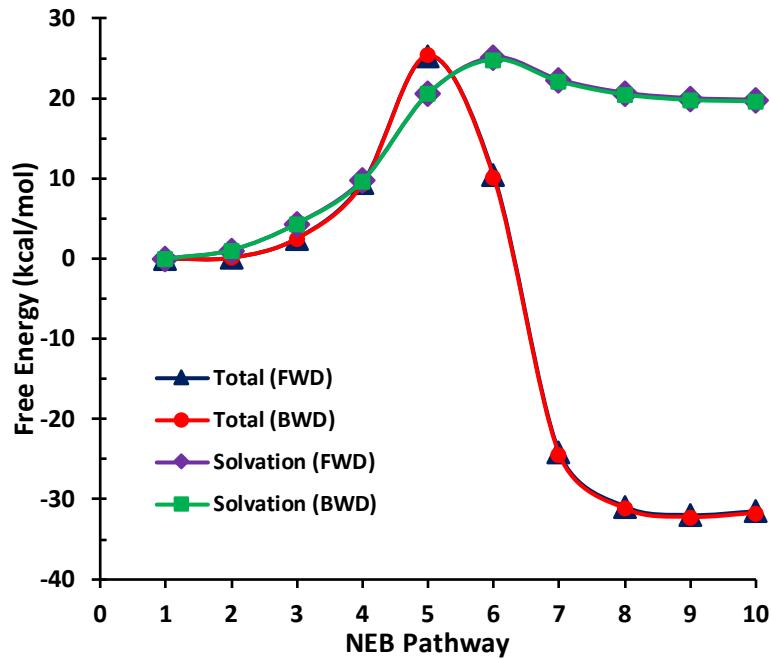


Figure 3.5: Free energy profile of the reaction. Solvation free energy obtained from FEP calculations are also shown. The reactant state is the reference point.

Table 3.2: Collection of energy (kcal/mol) data and total free energies (forward and backward samplings) obtained from NEB and FEP calculations.^[a]

NEB image	ΔE_{int}	$\Delta F_{\text{qm-harmonic}}$	ΔF_{mm}		ΔF_{total}	
			fwd	bwd	fwd	bwd
1	0.00	0.00	0.00	0.00	0.00	0.00
2	-1.00	0.04	1.07	1.07	0.10	0.10
3	-1.95	0.09	4.38	4.36	2.52	2.50
4	-0.23	-0.20	9.81	9.67	9.39	9.24
5	5.78	-1.00	20.53	20.60	25.30	25.36
6	-14.10	-0.58	25.09	24.84	10.42	10.17
7	-47.50	1.01	22.32	22.09	-24.17	-24.40
8	-53.67	2.08	20.68	20.45	-30.90	-31.14
9	-54.14	2.16	20.01	19.78	-31.96	-32.19
10	-53.44	2.14	19.81	19.61	-31.48	-31.69

[a] The first NEB image is the reference point.

Table 3.3: Comparison of free energies (kcal/mol) calculated along the NEB reaction pathway and smooth transformation^{[a],[b]}

state	ΔF_{mm}		ΔF_{total}	
	NEB pathway	Smooth transformation	NEB pathway	Smooth transformation
RS	0.00	0.00	0.00	0.00
TS	20.53 (20.60)	20.54 (20.48)	25.30 (25.36)	25.31 (25.24)
PS	19.81 (19.61)	20.05 (19.98)	-31.49 (-31.69)	-31.25 (-31.32)

[a] Numbers in parentheses were obtained from backward samplings.

[b] RS is the reference point.

Analysis of FEP calculations

In this study, FEP calculations along the NEB reaction pathway were performed both in forward and backward directions. As seen from Fig. 3.5, the forward and backward free energy profiles are almost identical. In addition, the activation free energy along the NEB reaction path agrees very well with that obtained via the direct smooth transformation between the reactant and the transition states (Table 3.3). The same is also true in case of the reaction free energy. Hence, this indicates that free energy is path independent and converges well based on the small difference between forward and backward samplings.

3.3.4 Conclusions

The S_N2 reaction CH₃Br + OH⁻ in aqueous solution was investigated using the QM/MM method. Together with the QM/MM NEB method and FEP calculations, the reaction free energy profile was determined. The activation and the reaction free energies were estimated to be 25.3 and -31.5 kcal/mol, respectively. The reaction barrier calculated in this work is consistent with the experimental value of 23.0 kcal/mol. The existence of solvent environment contributes

significantly to the overall reaction pathway and increases the height of the reaction barrier. This indicates unfavorability of the reaction in the aqueous solution.

In addition, the implementation of QM/MM optimization, NEB and FEP methods via NWChem-AMBER interface was also validated in this study. It is shown that the combination of the three methods within the QM/MM framework can be applied to determine the free energy profile for a simple reaction, in which the estimate reaction barrier is in good agreement with the experimental value. This provides a reference point for a further analysis of chemical reactions associated with significantly larger and more complex systems such as enzymes or RNAs.

3.4 References

- [1] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. *J. Chem. Phys.* **1983**, *79*, 926-935.
- [2] A. D. Becke. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [3] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- [4] T. H. D. Jr. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- [5] A. Schäfer, H. Horn, R. Ahlrichs. *J. Chem. Phys.* **1992**, *97*, 2571-2577.
- [6] A. Schäfer, C. Huber, R. Ahlrichs. *J. Chem. Phys.* **1994**, *100*, 5829-5835.
- [7] L. A. Curtiss, D. J. Frurip, M. Blander. *J. Chem. Phys.* **1979**, *71*, 2703-2711.
- [8] T. R. Dyke, J. S. Muentner. *J. Chem. Phys.* **1974**, *60*, 2929-2930.
- [9] K. S. Kim, B. J. Mhin, U. S. Choi, K. Lee. *J. Chem. Phys.* **1992**, *97*, 6649-6662.
- [10] A. W. Götz, M. A. Clark, R. C. Walker. *J. Comput. Chem.* **2014**, *35*, 95-108.
- [11] M. J. Loferer, H. H. Loeffler, K. R. Liedl. *J. Comput. Chem.* **2003**, *24*, 1240-1249.
- [12] B. Lev, R. Zhang, A. de la Lande, D. Salahub, S. Y. Noskov. *J. Comput. Chem.* **2010**, *31*, 1015-1023.
- [13] D. A. Case, R. M. Betz, W. Botello-Smith, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, P. Janowski, J. Kaus, A.

- Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, *Amber* 2016, University of California, San Francisco, **2016**.
- [14] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
- [15] T. Fox, P. A. Kollman. *J. Phys. Chem. B* **1998**, *102*, 8070-8079.
- [16] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [17] L. Xie, H. Liu, W. Yang. *J. Chem. Phys.* **2004**, *120*, 8039-8052.
- [18] H. Jonsson, G. Mills, K. W. Jacobsen. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, **2011**; pp 385-404.
- [19] Z. Lu, W. Yang. *J. Chem. Phys.* **2004**, *121*, 89-100.
- [20] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [21] R. Salomon-Ferrer, D. A. Case, R. C. Walker. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198-210.
- [22] M. E. Tuckerman, D. Marx, M. Parrinello. *Nature* **2002**, *417*, 925-929.
- [23] A. Botti, F. Bruni, S. Imberti, M. A. Ricci, A. K. Soper. *J. Mol. Liq.* **2005**, *117*, 81-84.
- [24] T. Wang, H. Yin, D. Wang, M. Valiev. *The Journal of Physical Chemistry A* **2012**, *116*, 2371-2376.
- [25] H. Yin, D. Wang, M. Valiev. *The Journal of Physical Chemistry A* **2011**, *115*, 12047-12052.
- [26] D. Wang, M. Valiev, B. C. Garrett. *The Journal of Physical Chemistry A* **2011**, *115*, 1380-1384.
- [27] E. A. Moelwyn-Hughes. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1949**, *196*, 540-553.
- [28] R. H. Bathgate, E. A. Moelwyn-Hughes. *Journal of the Chemical Society (Resumed)* **1959**, 2642-2648.
- [29] Y. Xu, T. Wang, D. Wang. *J. Chem. Phys.* **2012**, *137*, 184501.

- [30] D. M. Camaioni, C. A. Schwerdtfeger. *The Journal of Physical Chemistry A* **2005**, *109*, 10795-10797.
- [31] G. Hummer, L. R. Pratt, A. E. García. *J. Phys. Chem.* **1996**, *100*, 1206-1215.
- [32] V. Barone, M. Cossi, J. Tomasi. *J. Chem. Phys.* **1997**, *107*, 3210-3221.
- [33] C. Munz, P. V. Roberts. *Environ. Sci. Technol.* **1986**, *20*, 830-836.

Chapter 4

Combined Quantum-Mechanical Molecular Mechanics Calculations with NWChem and AMBER: Excited State Properties of Green Fluorescent Protein Chromophore Analogue in Aqueous Solution

4.1 Introduction

The combined quantum mechanics molecular mechanics (QM/MM)^[1-4] approach provides one of the simplest examples of a multi-physics methodology for the simulation of large systems. While it has been used successfully in a number of important applications there remains a significant barrier towards utilization of QM/MM methods in the general scientific community. Part of this problem can be found in the shortage of efficient, extensible, and readily accessible implementations of the methodology that can take advantage of existing highly supported software capabilities for quantum mechanical and molecular mechanics simulations. This is the issue that we are addressing in our work through the development of a generic QM/MM interface between two well-known community codes—NWChem^[12] and AMBER.^[13,14] In this paper we report on the use of this interface for the description of aqueous systems, using an analogue of the green fluorescent protein chromophore as an example application.

Green Fluorescent Protein (GFP)^[15] is a small protein, which is used extensively as a biomarker for gene expression and protein targeting.^[16] This widespread adoption of GFP stems from its unique photo-physical properties. It consists of a chromophore derived from an autocatalytic posttranslational cyclization among three amino acid residues, Ser65, Tyr66 and Gly67.^[17-26] The fluorescence is in the green region of the spectrum and pronounced when the chromophore covalently attached to the protein matrix. The latter ensures stability of the signal, which is important for imaging applications.^[27] The protein environment plays a key role in regulating the photo-physical properties of GFP.^[28] This is strongly supported by the inability of

GFP chromophore analogues to fluoresce in solutions.^[29,30] Additionally, since the discovery of wild-type GFP (wtGFP), a number of mutant GFP structures have been engineered that contain the same chromophore but exhibit different photo-physical properties.^[18,31-34] All these results point to the possibility for a rational design of GFP-like systems with desired properties through targeted changes in the protein environment. A fundamental understanding of GFP chromophore chemistry and its interactions with the protein matrix would be highly beneficial to the success of such efforts.

A natural starting point towards this goal can be found in studies of model GFP chromophores, such as the p-hydroxybenzylidene-imidazolidinone (HBDI) molecule. The analysis of HBDI in gas phase, cluster, and bulk aqueous environments provides a unique way to isolate intrinsic chemical properties of the GFP chromophore and track the impact of the environment in a controlled manner. One important question that can be addressed in these studies is how different protonation states of the chromophore impact the excited state properties. It is known that the absorption spectrum of the wtGFP, which has a strong A-band at 395 nm and the minor B-band at 477 nm, comes from two chemically distinct chromophore states. An increase in the pH level (~11-12) reduces the intensity of the A-band and enhances the low energy B-band at 470 nm.^[35] This suggests that the two bands are associated with neutral protonated and anion deprotonated states of the GFP chromophore.^[36-38] It has also been proposed that a doubly protonated cation state could be present in GFP.^[6] While probably unlikely in the wtGFP, such a possibility cannot be ruled out in the mutants of GFP. The important point here is that the protein environment has the capacity to control the protonation state of the chromophore, and proton transfer plays a critical role in controlling fluorescence properties of GFP.

The structural characteristics of the chromophore represent another important factor that can affect its spectroscopic properties. In either charge state (cation/neutral/anion/zwitterion) the chromophore can be found in a *cis* or *trans* conformation. The protein matrix restricts chromophore configuration enhancing fluorescence and preventing energy dissipation through internal

conversion (IC). In aqueous solutions such conformational restrictions are greatly reduced, which is thought to be the main reason for fluorescence suppression.

Experimental studies of GFP chromophore analogues such as HBDI in gas phase and aqueous solutions establish useful limits on the expected absorption spectrum and anticipated changes upon solvation. In the case of a cationic state the introduction of the aqueous environment results in a blue shift of the absorption spectrum of 3.05 eV^[5] in gas phase to 3.13-3.20 eV^[5-8] in aqueous solution. A similar effect is observed for the anion. There is, however, uncertainty in the actual value of the shift given several different estimates of the gas phase absorption values—2.59-2.84 eV.^[5-9] Experimental data for the neutral state is suggestive of a red shift, but the situation is not entirely clear. In gas phase the absorption value extrapolated from Kamlet-Taft fit by Dong *et al.* was determined to be 3.51 eV.^[6] However, the charged group technique by Rajput *et al.* places it at 3.35 eV for the neutral form.^[11] Recently, Greenwood *et al.* have applied an alternative method and reported a new maximum absorption value of the neutral form to be 3.65 eV.^[10] In aqueous phase the absorption value was determined to be 3.33-3.35 eV.^[5-8] Overall the available experimental data for HBDI provide an important reference point for detailed analysis of GFP photochemistry.

Theoretical studies have been instrumental in the analysis of the GFP chromophore and its analogues.^[39-60] While a lot of efforts have been put into these calculations using several model chromophores and various protonation states in the gas phase, this is not the case in the aqueous solution. Despite a number of calculations reported in the aqueous phase,^[49,51,56,59,61-65] the same analysis across all relevant protonation states and conformers remains incomplete. Our work contributes to these efforts by providing a systematic study among all relevant protonation states and conformers of the chromophore (HBDI) in gas phase and aqueous solution. Such study enables comparative analysis of excited states of various forms of the chromophore as a function of the environment and is important in understanding properties of the wtGFP and its mutants.

This type of analysis is difficult to accomplish using currently available computational data due to the large variations in calculated absorption values that are highly sensitive to the theoretical methodology. For example, for the neutral state in gas phase the semi-empirical models predict the absorption in a range of 2.77-3.65 eV,^[45,54,55,58] TD-DFT gives a range of 2.89-3.54 eV,^[47,53,59,60] and correlated methods 2.88-3.69 eV.^[11,43,44,50,53,60] Similar variations can be found in theoretical estimates for other GFP states. There is growing evidence that the accurate excited state description of the GFP chromophore requires explicit treatment of correlation effects. To address this issue, our excited state description relies on the completely renormalized EOMCCSD(T) approach (CR-EOMCCSD(T)), which further improves upon the high level EOMCCSD (Equation-of-motion CC with singles and doubles) approach by adding correction due to triple excitations.^[44]

Another important consideration is related to the solvent description. Commonly utilized continuum solvent models cannot properly capture directional hydrogen bonding effects^[66]—an important factor for both aqueous and protein phases. To address this problem, one needs to introduce explicit description of solvent degrees of freedom, which can be accomplished in a number of different ways. Our QM/MM framework focuses solely on interfacing with the standard AMBER force field.^[67] QM/MM methodologies based on empirical force fields have been proven to be a powerful tool in photophysical studies of the GFP chromophore or its variants in protein environment.^[51,64,65,68-76] It should be noted that other approaches are possible, such as those based on effective fragment potential (EFP)^[77,78] models. The EFP based method has been employed in a recent study by Ghosh *et al.*^[62] to determine solvatochromic shifts for the anionic form of a halogen-substituted HBDI model chromophore, 3,5-difluoro 4-hydroxybenzylidene imidazolinone (DFHBDI) in bulk aqueous phase. Our studies yield similar results for the relevant anionic HBDI species, indicating that simple QM/MM models provide a good starting point for the analysis of GFP chromophore analogues in aqueous solution. Our study also represents validity for the use of newly implemented algorithm via NWChem-AMBER interface in a condensed phase simulation.

4.2 Computational Methodology and System Setup

4.2.1 Gas Phase

All quantum-mechanical calculations were performed using the NWChem computational chemistry code.^[12] For all eight model chromophores (see Fig. 4.2) the optimizations were performed using density functional theory (DFT)^[79-81] with the B3LYP exchange-correlation functional^[82,83] and the 6-31G* basis set.^[84] The excitation energies for the optimized chromophore structures were calculated using the CR-EOMCCSD(T) method^[44] with the Def2-TZVP basis set.^[85] The CR-EOMCCSD(T) formalism^[86-88] provides perturbative extension of the excited-state EOMCCSD method (EOMCC with singles and doubles)^[89,90] to include the effect of triple excitation in a non-iterative manner. In the CR-EOMCCSD(T) approach the excitation energy ($\omega_K^{CR-EOMCCSD(T)}$) of the K-th state is calculated as a sum of EOMCCSD excitation energy ($\omega_K^{EOMCCSD}$) and a non-iterative correction ($\delta_K^{CR-EOMCCSD(T)}$) which captures correlation effects due to collective 3-body excitations, i.e.,

$$\omega_K^{CR-EOMCCSD(T)} = \omega_K^{EOMCCSD} + \delta_K^{CR-EOMCCSD(T)}. \quad (4.1)$$

The $\delta_K^{CR-EOMCCSD(T)}$ is expressed in terms of triply excited moments of the EOMCCSD equations and trial function, which includes single, double, and triple excitations. The triple amplitudes of the trial wavefunction are expressed perturbatively through the EOMCCSD amplitudes. The theoretical and implementation details of this formalism have been widely discussed in the literature and we refer readers to Refs. 86-88 for more details. Recent benchmark calculations with the CR-EOMCCSD(T) formalisms^[91] performed for 28 medium-size organic molecules have demonstrated that CR-EOMCCSD(T) non-iterative triples corrections to the EOMCCSD excitation energies provide significant improvements to the EOMCCSD estimations, while closely matching the results of the iterative and considerably more expensive CC3 and EOMCCSDT-3 calculations and their CASPT2 counterparts.

4.2.1 Aqueous Phase

Aqueous phase calculations utilized a QM/MM model where the total energy of a system is represented as

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM}, \quad (4.2)$$

where E_{QM} is the quantum mechanical energy of the QM subsystem, E_{MM} the molecular mechanical interactions of the MM subsystem, and $E_{QM/MM}$ the interactions between the QM subsystem and the MM subsystem. The QM/MM interactions can be divided into two terms:

$$E_{QM/MM} = E_{QM/MM}^{electrostatics} + E_{QM/MM}^{VdW}. \quad (4.3)$$

The electrostatic term describes Coulomb interactions between QM and MM regions,

$$E_{QM/MM}^{electrostatics} = - \sum_k^{N_{MM}} \int d\mathbf{r} \frac{\rho_{QM}(\mathbf{r}) Q_k}{|\mathbf{r} - \mathbf{R}_k|} + \sum_A^{N_{QM}} \sum_k^{N_{MM}} \frac{Z_A Q_k}{|\mathbf{R}_A - \mathbf{R}_k|}, \quad (4.4)$$

where ρ_{QM} is the electron density and Z_A are effective nuclear charges of the QM region, and Q_k are classical electrostatic charges. Van der Waals interactions in the second term, $E_{QM/MM}^{VdW}$, are represented in terms of a Lennard-Jones potential:

$$E_{QM/MM}^{VdW} = \sum_A^{N_{QM}} \sum_k^{N_{MM}} \epsilon_{Ak} \left[\left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^{12} - \left(\frac{\sigma_{Ak}}{R_{Ak}} \right)^6 \right]. \quad (4.5)$$

The above QM/MM description was deployed using the newly developed NWChem-AMBER interface (see Fig. 4.1), which builds on our extensible interface^[92] for QM/MM simulations with AMBER^[13,14] and external QM programs. This interface allows us to integrate the vast array of electronic structure methods available in NWChem^[12] including DFT, TD-DFT, and CC formulations with a classical description provided by AMBER. It should be noted that while the QM/MM partitioning results in a significant reduction in computational load, this alone is not always sufficient to provide a practical approach for simulations other than single point energy calculations. As a result for the optimization component of our calculations we have utilized a

multiregion optimization strategy where optimizations of QM and MM regions are alternated.^[93]

During optimization of the MM region, the QM region is represented by a fixed electron density.

The entire procedure contains the following steps:

- 1) Optimization of the QM subsystem keeping MM subsystem fixed.
- 2) Calculation of a reduced electrostatic representation for the QM subsystem.
- 3) Optimization of the MM subsystem keeping the QM subsystem fixed.
- 4) The procedure is repeated from step 1 to 3 as a cycle until converged.

Note that in step 2, instead of employing the direct charge density ρ_{QM} resulting from the wavefunction recalculation which is computationally expensive, the ESP charge fitting scheme can be used. A set of ESP charges representing the QM subsystem are calculated and used in step 3 of MM optimization. This method greatly improves the efficiency of QM/MM calculations for large systems. Hence, Eq. (4.4) for the MM optimization has a form:

$$E_{QM/MM}^{electrostatics} = \sum_i^{N_{QM}} \sum_k^{N_{MM}} \frac{Q_i Q_k}{|\mathbf{r}_i - \mathbf{R}_k|} \equiv E_{ESP}, \quad (4.6)$$

where Q_i represents the ESP fitted charge of QM atom i .

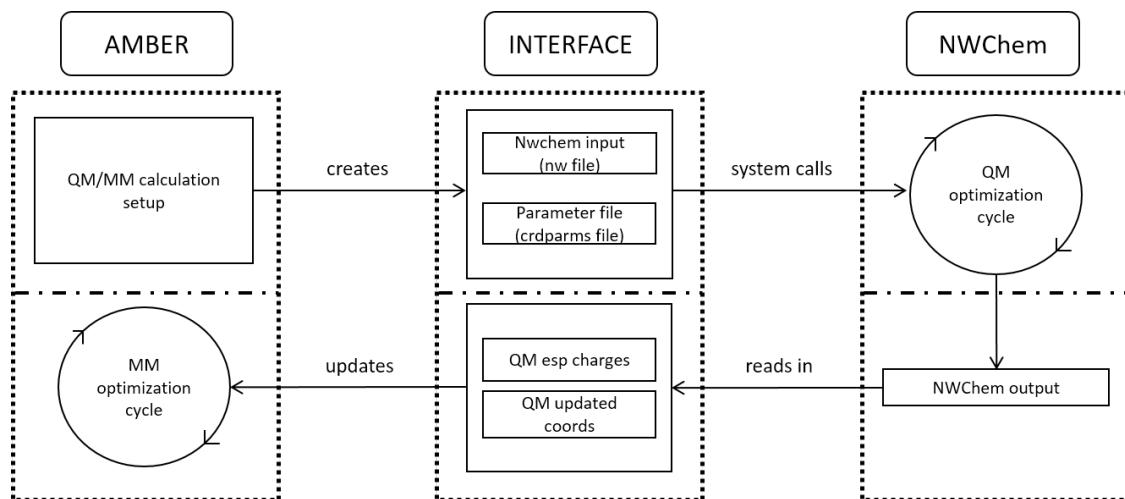


Figure 4.1: Diagram for a single cycle of multiregion optimization using the AMBER/NWChem interface.

To define the aqueous solution model the chromophore (HBDI) was solvated within a 22 Å radius spherical droplet of classical triangulated water molecules (TIP3P model).^[94] A soft half-harmonic potential was imposed beyond 22 Å radius from the model chromophore. The model chromophores were included in the QM subsystem and the water molecules were treated at the MM level. The VdW parameters for the chromophores were taken from the General AMBER Force Field (GAFF).^[95]

All systems were initially optimized and equilibrated for 500 ps at 300 K using the PMEMD engine of the AMBER v16 code.^[14] During the course of the MD equilibration, the chromophore was held fixed at the center of the water droplet. The equilibrated systems were optimized using the QM/MM multiregion optimization protocol described above. Mirroring gas phase calculations, the QM description at this step was based on a DFT/B3LYP/6-31G* level of theory because forces are not yet available from the CR-EOMCCSD(T) methods. Following optimization, the excitation energies were calculated using CR-EOMCCSD(T) level of theory with the Def2-TZVP basis set.^[85]

To estimate relative stabilities of conformers, we used a procedure where solvation energies (E_{solv}) were calculated using the self-consistent reaction field theory of Klamt and Schüürmann (COSMO).^[96] The atomic radii were set at 1.172 for H, 1.576 for O, and 2.126 Å for N.^[97] COSMO calculations were performed on QM/MM optimized structures with all the solvent removed. Thermal corrections of enthalpy (dH) and total entropies (S) at 298.15K were calculated using QM/MM vibrational calculations at DFT/B3LYP/6-31G* level of theory. These were combined with CCSD internal QM (ΔE_{int}^{CCSD}) to arrive at the approximate values for relative free energies,

$$\Delta G = (\Delta E_{int}^{CCSD} + \Delta E_{solv}) + \Delta H - T\Delta S \quad (4.7)$$

The calculated relative free energies are listed in the Table 4.2.

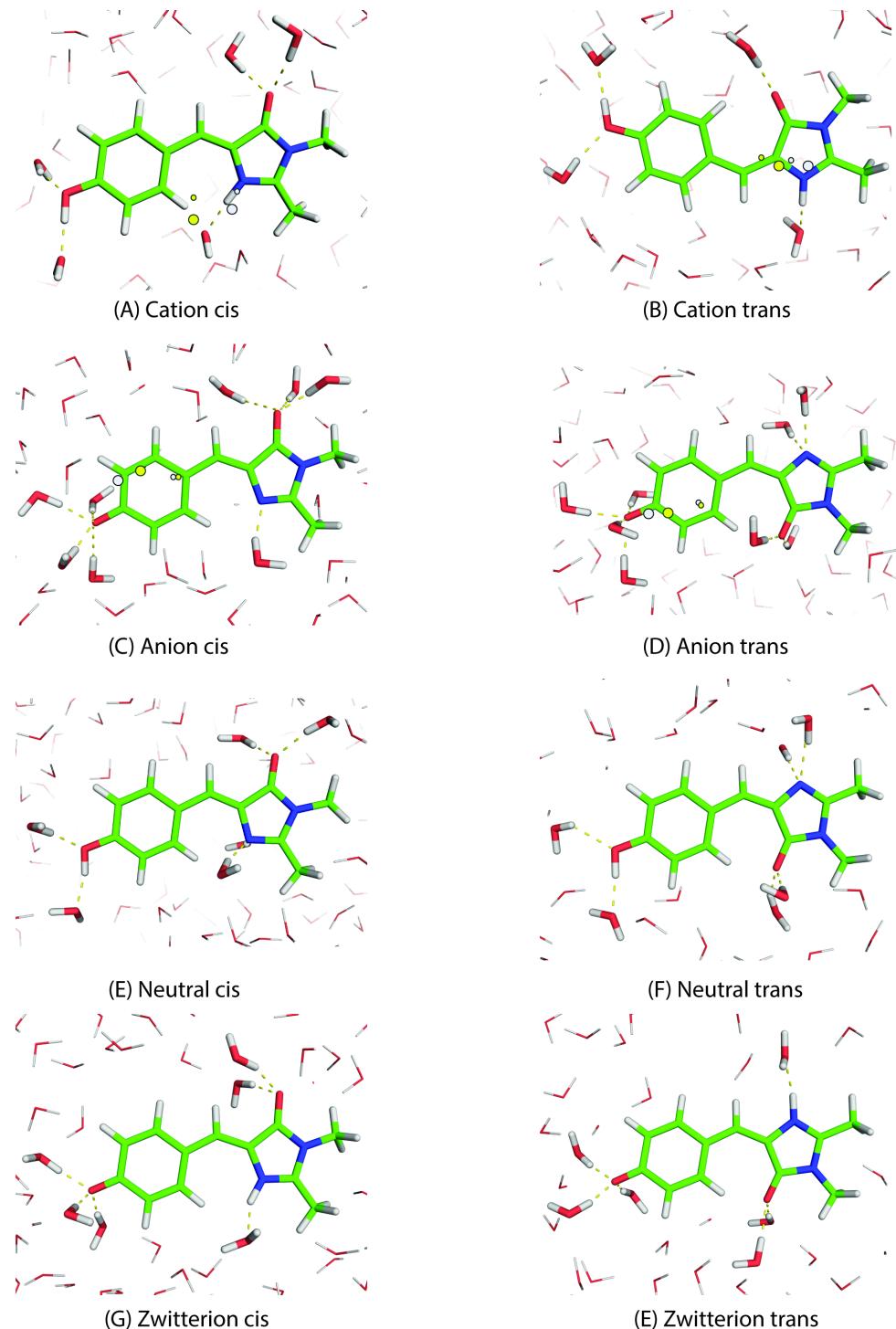


Figure 4.2: GFP model chromophore (HBDI) in water with different protonation states (*cis* and *trans* conformers). Carbon atoms are green, nitrogen atoms blue, oxygen atoms red and hydrogen atoms white. For cationic and anionic forms, the large white and yellow circles represent the effective center of charge for the ground and excited states correspondingly. Small circles refer to the same quantities in the gas phase.

Table 4.1: Calculated excitation energy (in eV) of the GFP model chromophores

Chromophore	Gas				Water	
	TD-DFT	CR-EOMCCSD(T)	Experiment	TD-DFT	CR-EOMCCSD(T)	Experiment
Cation (<i>cis</i>)	3.16	3.15 (3.47)	3.05 ^[a]	3.22	3.15 (3.46)	3.13-3.20 ^[b]
Cation (<i>trans</i>)	3.18	2.95 (3.26)		3.29	3.20 (3.51)	
Anion (<i>cis</i>)	3.09	2.70 (3.03)	2.59-2.84 ^[c]	3.18	2.93 (3.25)	2.90-2.92 ^[b]
Anion (<i>trans</i>)	3.11	2.70 (3.03)		3.16	2.80 (3.12)	
Neutral (<i>cis</i>)	3.45	3.69 (3.99)	3.35-3.65 ^[d]	3.37	3.43 (3.74)	3.33-3.35 ^[b]
Neutral (<i>trans</i>)	3.39	3.61 (3.92)		3.40	3.39 (3.70)	
Zwitterion (<i>cis</i>)	2.67	2.73 (3.06)	n/a	3.11	2.69 (3.01)	n/a
Zwitterion (<i>trans</i>)	2.65	2.97 (3.31)	n/a	3.08	2.62 (2.94)	n/a

[a] Taken from Ref. 33. [b] Taken from Refs. 32-35. [c] Taken from Refs. 32-36. [d] Taken from Refs. 32, 37, 38. Numbers in the parentheses are based on EOMCCSD calculations.

Table 4.2: Relative free energies of GFP model chromophores (kcal/mol)^{[a],[b]}

	Gas phase	Aqueous phase	
		Calculation	Experiment ^[c]
$\Delta G_{cis \rightarrow trans}^{cat}$	0.45	0.74	0.79
$\Delta G_{cis \rightarrow trans}^{ani}$	3.00	2.59	2.29
$\Delta G_{cis \rightarrow trans}^{neu}$	3.33	2.55	2.10
$\Delta G_{cis \rightarrow trans}^{zwit}$	1.03	3.59	n/a
$\Delta G_{cis}^{neu \rightarrow zwit}$	26.22	22.72	n/a
$\Delta G_{trans}^{neu \rightarrow zwit}$	23.91	23.76	n/a

[a] *cat* = cation; *ani* = anion; *neu* = neutral; *zwit* = zwitterion

[b] The estimate Gibbs energies are based on Eq. (4.7)

[c] Taken from Ref. 98.

4.3 Results and Discussion

4.3.1 Cationic Form

The optimized structures of the cationic *cis* and *trans* conformers in aqueous solution are shown in Fig. 4.2. Overall, the presence of aqueous solution does not impact the structures in a significant way. The only discernible change is found for the *cis* conformer, which deviates from the planar structure of the *trans* conformer due to steric hindrance introduced by the nitrogen protonation. The angle between the phenol and the imidazolinone plane increases from ~16 to 19 degrees when going from gas to aqueous phase. For both conformers, we observe two water molecules coordinating the phenol OH group, and one water coordinating the NH group of imidazolinone. The imidazolinone ring oxygen in the *trans* conformer is more solvent exposed and has two coordinating waters, compared to one in the *cis* conformer. In gas phase the two conformers appear to have nearly the same ground state energies—at the CCSD level of theory the *cis* conformer is slightly more stable than the *trans* conformer by 0.45 kcal/mol. In solution, our

QM/MM COSMO estimates show similar slight preference for the *cis* conformer by 0.74 kcal/mol. The result agrees well with the observed value of 0.79 kcal/mol reported by Tonge *et al.*^[98] Overall, these numbers indicate that either conformer can be potentially accommodated in aqueous phase.

In the gas phase, we calculate the absorption values to be 3.15 and 2.95 eV for *cis* and *trans* conformers, respectively. These values are in agreement with the experimental value of 3.05 eV^[5] and in line with previous calculations using aug-MCQDPT2 (3.06 eV)^[40] and XMCQDPT2 (3.12 eV).^[50] For aqueous phase, we find that the *cis* conformer is characterized by the absorption value of 3.15 eV, and the *trans* conformer at 3.20 eV. Interestingly, both *cis* and *trans* absorption numbers agree with existing experimental data. In acidic aqueous solution (pH < 1), Andersen *et al.*^[5,7] reported the maximum absorption of HBDI to be 3.13 eV. However, a slightly different absorption value of 3.17-3.20 eV^[6,8] has been obtained by other groups. Looking at the differences between gas phase and aqueous values, we observe that they are sensitive to the solute conformation. Indeed, for the *cis* conformer no difference is observed and for the *trans* conformer we notice a slight blue shift.

To gain further insight into the impact of an aqueous environment on the excited state properties of *cis* and *trans* conformers it is instructive to analyze the changes in charge distribution that occur upon transfer to aqueous phase and during excitation. For charged systems, such as a cation or anion, a useful metric for such analysis can be found by defining an effective center of charge as a point where the dipole moment of the system (calculated at CCSD level of theory) vanishes. Given a total dipole (\mathbf{d}) and charge of the system (Q), it is given by

$$\mathbf{R}_c = \frac{\mathbf{d}}{Q}. \quad (4.8)$$

The location of this point for the cationic system is shown in Fig. 4.2 (white/yellow circles), both for the unpolarized system (QM calculation at QM/MM geometry without inclusion of MM water molecules) and the fully polarized system (QM/MM calculation). In the ground state, the

center of (overall positive) charge in both structures is near the NH bond on the imidazolinone ring, suggesting that this is the primary localization point of the electron hole in the system. Comparing the two conformers, it is evident that the center of charge is pushed further inside the molecule in the *trans* conformer. We observe that the impact of the aqueous environment is to pull the charge closer to the solvent boundary, which maximizes the solvation energy. Given the above mentioned difference between the two structures, the center of charge in the *cis* case is located much closer to the solvent compared to *trans*, which is consistent with the larger solvation energy that stabilizes the *cis* isomer. Transition to the excited state shifts the center of charge closer to the phenol group. Our calculations indicate that in the *trans* configuration, this move brings the electron hole to the area of higher (more positive) electrostatic potential, while the opposite is true for the *cis* configuration. This is consistent with the increased gap between the ground and excited state in the *trans* conformer.

4.3.2 Anionic Form

The optimized structures of the anionic state are also shown in Fig. 4.2. The imidazolinone nitrogen is no longer protonated and we no longer have a twisting distortion in the *cis* conformer as observed in the cationic state. Other than a slight puckering in the *cis* conformer, which pulls the opposite ends of the molecule towards each other, the structural impact of aqueous solution remains small. This is consistent with previous QM/EFP calculations^[62] of the related anionic form of DFHBDI chromophore where near planar structures formed the main population of the sampled structures. The anionic state exhibits much stronger solvent coordination than the cationic state. In both *cis* and *trans* conformers, we can identify four water coordinating phenol groups. The solvent accessibility of the imidazolinone nitrogen and oxygen depends on the conformation, which impacts the coordination. For the *cis* conformer we observe one water interacting with a nitrogen and three waters interacting with an oxygen, and for the *trans* conformer both centers are

coordinated by two waters. The CCSD gas phase calculations also indicate that the *cis* conformer is relatively more stable than the *trans* conformer by 3.0 kcal/mol. Similar situation persists in aqueous phase, where our estimates show the relative stability of the *cis* conformer at 2.6 kcal/mol. This is in line with the previously calculated value of 2.1 kcal/mol obtained using a combined CASSCF/EFP model,^[61] and agrees well with the experimental value of 2.3 kcal/mol from Tonge *et al.*^[98]

In the gas phase, several hypotheses have been put forward regarding the absorption value of the anionic form of the GFP chromophore. Measurements from action spectroscopy revealed the absorption peak at 2.59 eV.^[5,7,8] On the other hand, extrapolation from Kamlet-Taft fit by Dong *et al.*^[6] suggested a more energetic absorption value of 2.84 eV. In addition, studies from photo-destruction spectroscopy propose the absorption value above 2.61 eV.^[9] Our calculations show that in gas phase both conformers have absorption values at 2.70 eV. This falls squarely within the range of experimental values and is also consistent with previous calculations from CASPT2 (2.67 eV),^[46] SA-CASSCF (2.69 eV)^[48] and MRMP2 (2.61 eV)^[42] approaches. To determine the effect of geometry on the absorption values, we have performed gas phase excitation energy calculations for anionic HBDI chromophores (both conformers) obtained from QM/MM optimizations. The absorption values for both conformers are estimated to be 2.65 eV, which is red-shifted from the gas phase optimized structure by -0.05 eV. A Similar red shift of -0.13 eV in the absorption energy was reported in previous QM/EFP calculations^[62] for the anionic DFHBDI.

Our calculations place the aqueous phase absorption values at 2.93 and 2.80 eV for the *cis* and *trans* conformers, respectively. Previous CASPT2 calculation of HBDI in water through PCM model estimated the absorption value for the *cis* conformer to be 2.69 eV.^[39] Additional calculated absorption value of 2.86 eV was also reported for HBI model using CASPT2 method and explicit TIP3P water model.^[51] Experimentally, the absorption spectrum for HBDI at high pH (> 11) was measured to be 2.90-2.92 eV,^[5-9] which agrees very well with our 2.93 eV calculated value for the

cis conformer. Similar to the cationic state, the shift in absorption values is sensitive to the conformation, and these effects are amplified in the anionic state—the absorption energy increases by 0.23 eV in the *cis* conformer and by 0.1 eV in the *trans* conformer. This is in line with the increase of absorption value of 0.1 eV for DFHBDI in bulk aqueous phase as reported by Ghosh *et al.*^[62] In addition, a very small solvent effect of polarization on the change of the excitation energy (0.03 eV) is suggested. Therefore, this supports the choice of non-polarizable MM force field used in our study.

Performing the same charge analysis as we did for the cationic state, we observe that the center of negative charge is located on the phenol group side. Its position is not affected much by transition to the excited state, which is different to what was observed for the cation. This indicates that excitation in the anionic case incurs little charge rearrangement. In both structures the transition to aqueous environment pulls the charge closer to the solvent. It also increases separation between the ground and excited state, particularly so in the *cis* case.

4.3.3 Neutral Form

The neutral form of the GFP model chromophore is characterized by a protonated phenol group. The optimized structures of both *cis* and *trans* conformers are shown in Figure 2. Similar to the anionic state, the structure of the *trans* conformer remains virtually unchanged upon transfer to aqueous solution while the *cis* state exhibits slight puckering. We observe that the phenol OH group is coordinated by two water molecules in both *cis* and *trans* conformers. The coordination of the imidazolinone is very similar to that observed in the anionic state. The CCSD calculations suggest that the *cis* conformer of the neutral form of the GFP model chromophore is relatively more stable than the *trans* conformer by 3.3 kcal/mol. Our estimation in the aqueous phase also reveals similar *cis* conformer stability over the *trans* conformer by 2.55 kcal/mol. This is in line with the observed value of 2.1 kcal/mol from experiment.^[98]

In the gas phase discrepancies between the maximum absorption of the neutral forms have been reported. from experiments. The extrapolation from a Kamlet-Taft fit suggests the absorption value at 3.51 eV.^[6] The experiment using the charge group by Rajput *et al.* proposed the absorption at 3.35 eV.^[11] Recently, Greenwood *et al.*^[10] applied an alternative approach and reported the absorption value to be 3.65 eV. Our calculated absorption values are 3.69 and 3.61 eV for the *cis* and *trans* conformers, respectively, and seem to confirm to be within these measurements. Our numbers are also in line with previously calculated values of 3.71 eV (CAM-B3LYP)^[10] and 3.58 eV (CASPT2)^[60] for the *cis* conformer of HBDI.

In aqueous solution our calculated excitation energies of the neutral form were calculated to be 3.43 and 3.39 eV for the *cis* and *trans* isomers, respectively. These values are in good agreement with the experimentally measured absorption energy of 3.33-3.35 eV^[5-8] for neutral pH solutions of the chromophore. Previous calculations by other groups have been performed at the TD-DFT level using different model chromophores,^[47,59] but no consensus has been reached. Xie and Zeng, using a combined explicit-implicit type model,^[99,100] reported absorption values of 3.56 and 3.49 eV for the *cis* and *trans* conformers of HBMIA, respectively.^[59] On the other hand, Nemukhin *et al.*, using a polarizable continuum model (PCM), estimated the absorption values for HBI at 2.74 (*cis*) and 2.84 eV (*trans*).^[47] Additionally, the absorption peak of HBI from SA-2-CAS approach using Monte Carlo samples of the ground state chromophore in TIP3P water was reported to be 4.27 eV.^[64]

4.3.4 Zwitterionic Form

In addition to the neutral form discussed above, a number of computational studies related to the zwitterionic form have been reported. The latter still carries net zero charge but instead of the phenol group the proton is now located on the imidazolinone nitrogen group. In gas phase, such a state is most certainly higher in energy than the neutral state. Our calculations show at the CCSD

level of theory that the transfer of the proton from phenol to imidazolinone raises the energy by ~ 26 kcal/mol. There is little difference between the two conformations of the zwitter-ionic form. The *cis* conformer is slightly more stable than the *trans* conformer by 1.03 kcal/mol (CCSD). While the aqueous solution stabilizes the zwitterionic form, it remains ~ 20 kcal/mol above the neutral state.

In gas phase, our calculated absorption values of the zwitterionic HBDI are 2.73 and 2.97 eV for *cis* and *trans* conformers, respectively. These numbers are consistent with the previous calculated values in a range of 2.46-3.13 eV reported by Polyakov *et al.*^[50] and with the calculations from post-HF and TD-DFT approaches reported for HBI and HBMIA.^[54,58,59] In the aqueous phase we estimate absorption values of 2.69 and 2.62 eV for the zwitterionic form. This is in reasonable agreement with the observed absorption value of 2.55 eV reported for the zwitterionic form of the methoxyl derivative (pHBDIMe⁺) species.^[6] Previous calculations showing less agreement were based on the TD-DFT model and varied depending on the solvent treatment. PCM estimates^[47] give values of 2.31 (*cis*) and 2.48 eV (*trans*) for the HBI model, and the combined explicit-implicit type model^[99,100] of Xie and Zeng gives 3.33 eV for the *trans* form of HBMIA.^[59] Overall, there seems to be no compelling reason to assume that the zwitterionic form should contribute to the observed spectrum at neutral pH values.

4.4 Conclusions

We have implemented QM/MM approach by combining NWChem and AMBER codes. This initial implementation is geared towards studies of aqueous systems allowing efficient optimization of large systems and calculations of excited state properties at various levels of theory. The interface has been tested through a systematic study of the photophysical properties for the GFP model chromophore analogue (HBDI) in aqueous phase using the high level QM method and explicit solvent treatment. Our calculations for the excitation energies of HBDI show good

agreement with experimental data for both gas phase and aqueous environment. As indicated in Table 1, the performance of less computationally extensive TD-DFT methods is mixed at best, and a high level quantum-mechanical treatment appears to be essential to reliably capture variations of excited states across different protonation states and conformations. To fully approach experimental absorption numbers, inclusion of triples excitations is necessary. As Table 1 indicates, CR-EOMCCSD(T) leads to a sizable (~0.3 eV) correction to EOMCCSD numbers. The triples corrections, however, are uniform, which hints at the possibility of adding it as a post-processing correction. The MM treatment of the solvent is clearly an approximation, but overall agreement with experimental aqueous absorption energies suggests that it may be sufficient to capture the essential features of the aqueous environment and its influence on excited states.

Our results show that impact of aqueous environment on cationic and anionic states generally results in blue shifts, i.e. increase in the separation between ground and excited energy levels upon solvation. The effect is much less pronounced in the cationic state. Center of charge analysis shows that the cationic state exhibits greater charge rearrangement during excitation than the anion. The absorption energy shifts upon solvation are sensitive to the conformation of the chromophore. These effects can be rationalized based on charge movement into the area of higher/lower external electrostatic potentials, providing a potential pathway for control of GFP properties in the protein environment.

Our results indicate that at neutral pH levels the experimentally observed absorption signal of the GFP chromophore is most likely coming from the phenol protonated form of GFP. The other possibility suggested in previous works, a zwitterionic form, is higher in energy and does not match the experimental absorption data.

The presented results provide baseline and insights for further analysis of the GFP chromophore in protein matrix. Dealing with such a complex system within the QM/MM framework requires a treatment for bond crossing and better electrostatic description between QM

and MM subsystems. In addition, more sophisticated free energy calculations are necessary. These effects will be taken into account in our future work to offer a more versatile QM/MM interface with capabilities that encompass a vast array of simulations types for large complex systems.

Acknowledgements

Chapter 4, in full, is a reprint of the material entitled “Combined Quantum-Mechanical Molecular Mechanics Calculations with NWChem and AMBER: Excited State Properties of Green Fluorescent Protein Chromophore Analogue in Aqueous Solution” by Pirojsirikul, T., Götz, A.W., Weare, J., Walker, R.C., Kowalski, K., and Valiev, M., which has been accepted for publication in the Journal of Computational Chemistry. The dissertation author is the primary author of this material, and co-authors have approved the use of the material for this dissertation.

This work was supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, U. S. Department of Energy (DOE), and performed using EMSL, a national scientific user facility sponsored by DOE’s Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the DOE. RCW acknowledges support from Intel Corp and Exxact Corp. AWG acknowledges funding from the Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the DOE and support by the National Science Foundation (Grant No. CHE-1416571).

4.5 References

- [1] M. J. Field, P. A. Bash, M. Karplus. *J. Comput. Chem.* **1990**, *11*, 700-733.
- [2] H. M. Senn, W. Thiel. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229.
- [3] U. C. Singh, P. A. Kollman. *J. Comput. Chem.* **1986**, *7*, 718-730.

- [4] A. Warshel, M. Levitt. *J. Mol. Biol.* **1976**, *103*, 227-249.
- [5] L. H. Andersen, A. Lapierre, S. B. Nielsen, I. B. Nielsen, S. U. Pedersen, U. V. Pedersen, S. Tomita. *Eur. Phys. J. D* **2002**, *20*, 597-600.
- [6] J. Dong, K. M. Solntsev, L. M. Tolbert. *J. Am. Chem. Soc.* **2006**, *128*, 12038-12039.
- [7] S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, L. H. Andersen. *Phys. Rev. Lett.* **2001**, *87*, 228102.
- [8] M. W. Forbes, R. A. Jockusch. *J. Am. Chem. Soc.* **2009**, *131*, 17038-17039.
- [9] K. Chingin, R. M. Balabin, V. Frankevich, K. Barylyuk, R. Nieckarz, P. Sagulenko, R. Zenobi. *Int. J. Mass spectrom.* **2011**, *306*, 241-245.
- [10] J. B. Greenwood, J. Miles, S. D. Camillis, P. Mulholland, L. Zhang, M. A. Parkes, H. C. Hailes, H. H. Fielding. *J. Phys. Chem. Lett.* **2014**, *5*, 3588-3592.
- [11] J. Rajput, D. B. Rahbek, L. H. Andersen, T. Rocha-Rinza, O. Christiansen, K. B. Bravaya, A. V. Erokhin, A. V. Bochenkova, K. M. Solntsev, J. Dong, J. Kowalik, L. M. Tolbert, M. Axman Petersen, M. Brondsted Nielsen. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9996-10002.
- [12] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [13] R. Salomon-Ferrer, D. A. Case, R. C. Walker. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198-210.
- [14] D. A. Case, R. M. Betz, W. Botello-Smith, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, Amber 2016, University of California, San Francisco, **2016**.
- [15] R. Y. Tsien. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
- [16] M. Zimmer. *Chem. Rev.* **2002**, *102*, 759-782.
- [17] R. M. Wachter. *Acc. Chem. Res.* **2007**, *40*, 120-127.
- [18] R. Heim, D. C. Prasher, R. Y. Tsien. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501-12504.
- [19] C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast, W. W. Ward. *Biochemistry* **1993**, *32*, 1212-1218.

- [20] B. G. Reid, G. C. Flynn. *Biochemistry* **1997**, *36*, 6786-6791.
- [21] D. P. Barondeau, C. J. Kassmann, J. A. Tainer, E. D. Getzoff. *Biochemistry* **2005**, *44*, 1960-1970.
- [22] D. P. Barondeau, J. A. Tainer, E. D. Getzoff. *J. Am. Chem. Soc.* **2006**, *128*, 3166-3168.
- [23] D. P. Barondeau, C. J. Kassmann, J. A. Tainer, E. D. Getzoff. *J. Am. Chem. Soc.* **2006**, *128*, 4685-4693.
- [24] L. Zhang, H. N. Patel, J. W. Lappe, R. M. Wachter. *J. Am. Chem. Soc.* **2006**, *128*, 4766-4772.
- [25] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier. *Gene* **1992**, *111*, 229-233.
- [26] N. P. Lemay, A. L. Morgan, E. J. Archer, L. A. Dickson, C. M. Megley, M. Zimmer. *Chem. Phys.* **2008**, *348*, 152-160.
- [27] S. R. Meech. *Chem. Soc. Rev.* **2009**, *38*, 2922-2934.
- [28] A. Acharya, A. M. Bogdanov, B. L. Grigorenko, K. B. Bravaya, A. V. Nemukhin, K. A. Lukyanov, A. I. Krylov. *Chem. Rev.* **2017**, *117*, 758-795.
- [29] S. Kojima, H. Ohkawa, T. Hirano, S. Maki, H. Niwa, M. Ohashi, S. Inouye, F. I. Tsuji. *Tetrahedron Lett.* **1998**, *39*, 5239-5242.
- [30] H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi, F. I. Tsuji. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13617-13622.
- [31] M. V. Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov, S. A. Lukyanov. *Nat Biotech* **1999**, *17*, 969-973.
- [32] R. Heim, A. B. Cubitt, R. Y. Tsien. *Nature* **1995**, *373*, 663-664.
- [33] R. M. Wachter, M.-A. Elsliger, K. Kallio, G. T. Hanson, S. J. Remington. *Structure* **1998**, *6*, 1267-1277.
- [34] R. M. Wachter, B. A. King, R. Heim, K. Kallio, R. Y. Tsien, S. G. Boxer, S. J. Remington. *Biochemistry* **1997**, *36*, 9759-9765.
- [35] W. W. Ward, H. J. Prentice, A. F. Roth, C. W. Cody, S. C. Reeves. *Photochem. Photobiol.* **1982**, *35*, 803-808.
- [36] A. F. Bell, X. He, R. M. Wachter, P. J. Tonge. *Biochemistry* **2000**, *39*, 4423-4431.

- [37] A. F. Bell, D. Stoner-Ma, R. M. Wachter, P. J. Tonge. *J. Am. Chem. Soc.* **2003**, *125*, 6919-6926.
- [38] K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormö, S. J. Remington. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2306-2311.
- [39] P. Altoe, F. Bernardi, M. Garavelli, G. Orlandi, F. Negri. *J. Am. Chem. Soc.* **2005**, *127*, 3952-3963.
- [40] K. B. Bravaya, A. V. Bochenkova, A. A. Granovskii, A. V. Nemukhin. *Russ. J. Phys. Chem. B* **2008**, *2*, 671-675.
- [41] A. K. Das, J.-Y. Hasegawa, T. Miyahara, M. Ehara, H. Nakatsuji. *J. Comput. Chem.* **2003**, *24*, 1421-1431.
- [42] E. Epifanovsky, I. Polyakov, B. Grigorenko, A. Nemukhin, A. I. Krylov. *J. Chem. Theory Comput.* **2009**, *5*, 1895-1906.
- [43] V. Helms, C. Winstead, P. W. Langhoff. *J. Mol. Struct.: THEOCHEM* **2000**, *506*, 179-189.
- [44] K. Kowalski, S. Krishnamoorthy, O. Villa, J. R. Hammond, N. Govind. *J. Chem. Phys.* **2010**, *132*, 154103.
- [45] T. Laino, R. Nifosi, V. Tozzini. *Chem. Phys.* **2004**, *298*, 17-28.
- [46] M. E. Martin, F. Negri, M. Olivucci. *J. Am. Chem. Soc.* **2004**, *126*, 5452-5464.
- [47] A. V. Nemukhin, I. A. Topol, S. K. Burt. *J. Chem. Theory Comput.* **2006**, *2*, 292-299.
- [48] S. Olsen, S. C. Smith. *J. Am. Chem. Soc.* **2008**, *130*, 8677-8689.
- [49] A. Petrone, P. Caruso, S. Tenuta, N. Rega. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20536-20544.
- [50] I. V. Polyakov, B. L. Grigorenko, E. M. Epifanovsky, A. I. Krylov, A. V. Nemukhin. *J. Chem. Theory Comput.* **2010**, *6*, 2377-2387.
- [51] A. Sinicropi, T. Andruniow, N. Ferré, R. Basosi, M. Olivucci. *J. Am. Chem. Soc.* **2005**, *127*, 11534-11535.
- [52] A. Toniolo, S. Olsen, L. Manohar, T. J. Martinez. *Faraday Discuss.* **2004**, *127*, 149-163.
- [53] O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch. *Chem. Phys. Lett.* **2004**, *396*, 202-207.
- [54] A. A. Voityuk, A. D. Kummer, M.-E. Michel-Beyerle, N. Rösch. *Chem. Phys.* **2001**, *269*, 83-91.
- [55] A. A. Voityuk, M.-E. Michel-Beyerle, N. Rösch. *Chem. Phys. Lett.* **1997**, *272*, 162-167.

- [56] A. A. Voityuk, M.-E. Michel-Beyerle, N. Rösch. *Chem. Phys.* **1998**, *231*, 13-25.
- [57] S. Wan, S. Liu, G. Zhao, M. Chen, K. Han, M. Sun. *Biophys. Chem.* **2007**, *129*, 218-223.
- [58] W. Weber, V. Helms, J. A. McCammon, P. W. Langhoff. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6177-6182.
- [59] D. Xie, J. Zeng. *J. Comput. Chem.* **2005**, *26*, 1487-1496.
- [60] C. Filippi, M. Zaccheddu, F. Buda. *J. Chem. Theory Comput.* **2009**, *5*, 2074-2087.
- [61] I. Polyakov, E. Epifanovsky, B. Grigorenko, A. I. Krylov, A. Nemukhin. *J. Chem. Theory Comput.* **2009**, *5*, 1907-1914.
- [62] S. Bose, S. Chakrabarty, D. Ghosh. *J. Phys. Chem. B* **2016**, *120*, 4410-4420.
- [63] K. Bhaskaran-Nair, M. Valiev, S. H. M. Deng, W. A. Shelton, K. Kowalski, X.-B. Wang. *J. Chem. Phys.* **2015**, *143*, 224301.
- [64] A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko, T. J. Martínez. *J. Phys. Chem. B* **2009**, *113*, 3280-3291.
- [65] K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, A. I. Krylov. *Acc. Chem. Res.* **2012**, *45*, 265-275.
- [66] B. Mennucci. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583-6594.
- [67] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [68] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, A. I. Krylov. *J. Am. Chem. Soc.* **2013**, *135*, 11541-11549.
- [69] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, A. I. Krylov. *J. Phys. Chem. B* **2011**, *115*, 8296-8303.
- [70] S. Faraji, A. I. Krylov. *J. Phys. Chem. B* **2015**, *119*, 13052-13062.
- [71] K. B. Bravaya, O. M. Subach, N. Korovina, V. V. Verkhusha, A. I. Krylov. *J. Am. Chem. Soc.* **2012**, *134*, 2807-2814.
- [72] P. K. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. I. Krylov, L. V. Slipchenko. *J. Phys. Chem. B* **2016**, *120*, 6562-6574.
- [73] P. Armengol, R. Gelabert, M. Moreno, J. M. Lluch. *J. Phys. Chem. B* **2015**, *119*, 2274-2291.

- [74] V. R. I. Kaila, R. Send, D. Sundholm. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491-4495.
- [75] C. Daday, C. Curutchet, A. Sinicropi, B. Mennucci, C. Filippi. *J. Chem. Theory Comput.* **2015**, *11*, 4825-4839.
- [76] V. A. Mironov, K. B. Bravaya, A. V. Nemukhin. *J. Phys. Chem. B* **2015**, *119*, 2467-2474.
- [77] M. S. Gordon, L. Slipchenko, H. Li, J. H. Jensen. In Annual Reports in Computational Chemistry; Spellmeyer, D. C.; Wheeler, R., Eds.; Elsevier, **2007**; pp 177-193.
- [78] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko. *Chem. Rev.* **2012**, *112*, 632-672.
- [79] P. Hohenberg, W. Kohn. *Phys. Rev. B* **1964**, *136*, 864-871.
- [80] W. Kohn, L. J. Sham. *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [81] R. G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules; Oxford University Press, New York, **1994**.
- [82] A. D. Becke. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [83] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- [84] R. Ditchfield, W. J. Hehre, J. A. Pople. *J. Chem. Phys.* **1971**, *54*, 724-728.
- [85] F. Weigend, R. Ahlrichs. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
- [86] K. Kowalski, P. Piecuch. *J. Chem. Phys.* **2004**, *120*, 1715-1738.
- [87] K. Kowalski, P. Piecuch. *J. Chem. Phys.* **2002**, *116*, 7411-7423.
- [88] P. Piecuch, J. R. Gour, M. Włoch. *Int. J. Quantum Chem.* **2009**, *109*, 3268-3304.
- [89] J. Geertsen, M. Rittby, R. J. Bartlett. *Chem. Phys. Lett.* **1989**, *164*, 57-62.
- [90] J. F. Stanton, R. J. Bartlett. *J. Chem. Phys.* **1993**, *98*, 7029-7039.
- [91] P. Piecuch, J. A. Hansen, A. O. Ajala. *Mol. Phys.* **2015**, *113*, 3085-3127.
- [92] A. W. Götz, M. A. Clark, R. C. Walker. *J. Comput. Chem.* **2014**, *35*, 95-108.
- [93] Y. Zhang, H. Liu, W. Yang. *J. Chem. Phys.* **2000**, *112*, 3483-3492.
- [94] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. *J. Chem. Phys.* **1983**, *79*, 926-935.

- [95] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- [96] A. Klamt, G. Schüürmann. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799-805.
- [97] E. V. Stefanovich, T. N. Truong. *Chem. Phys. Lett.* **1995**, *244*, 65-74.
- [98] X. He, A. F. Bell, P. J. Tonge. *FEBS Lett.* **2003**, *549*, 35-38.
- [99] J. Zeng, N. S. Hush, J. R. Reimers. *J. Chem. Phys.* **1993**, *99*, 1508-1521.
- [100] N. S. Hush, J. R. Reimers. *Chem. Rev.* **2000**, *100*, 775-786.

Chapter 5

QM/MM Calculations for Catalysis of Cellulose Degradation

by Copper-dependent Polysaccharide Monooxygenases

5.1 Introduction

5.1.1 Cellulose and Its Decomposition

Cellulose^[1] is one of the most abundant substances on earth. It is a polysaccharide consisting of β -1,4-glycosidic linkage between D-glucose units. These sugar units locked in cellulose polymers are the sole energy and carbon sources that can be used for sustainable biofuel and biochemical production through fermentation. Due to inter- and intra-molecular hydrogen bond network formed within and between glucan chains coupled with high degree of polymerization, this confers high tensile strength of cellulose and renders the recalcitrance of crystalline cellulose for decomposition. Despite a recent focus on their use and conversion to fuels and chemicals, the lignocellulosic resource remains largely underutilized due to the lack of cost effective technologies.

Hydrolytic cellulases were enzymes known to be primarily responsible for cellulose decomposition. However, the hydrolysis is slow due to the hypothesis of rate limiting step being glucan chain separation of crystalline cellulose that has to be threaded into active site cleft of the enzymes. This is suggested from the faster rate of hydrolysis of amorphous forms compared to crystalline substrates.^[2] However, this conventional hydrolysis framework has recently been challenged by a new class of enzymes known as polysaccharide monooxygenases (PMOs).

5.1.2 Polysaccharide Monooxygenases

Polysaccharide monooxygenases (PMOs) or lytic polysaccharide monooxygenases (LPMOs) are monomeric copper-dependent enzymes. These enzymes were recently found to enhance the decomposition of cellulose in synergy with hydrolytic enzymes in biomass

degradation. Due to their promising activities in conversion of biomass to sugars, a number of studies and research regarding this class of enzyme have been increasing during recent years.

Secreted from cellulolytic fungi and chitinolytic bacteria, PMOs were originally classified into Family 61 glycosidic hydrolases (GH61) or Family 33 carbohydrate-binding modules (CBM33) that can enhance cellulose and chitin decomposition, respectively. However, they were lack of noticeable hydrolytic activities. It was later discovered that these enzymes exhibit oxidative activities in cellulose and chitin degradation.^[3] Therefore, the enzymes were reclassified in the carbohydrate-active enzyme (CAZy) database from GH61 and CBM33 to Family 9 and Family 10 auxiliary active (AA9 and AA10), respectively. Since cellulose degradation is a primary focus in this study, only PMOs from AA9 will be mentioned for the rest of the chapter.

PMO active site consists of monomeric copper (Cu) bound with conserved amino acids, two of which are histidine residues providing three coordinating sites. A “histidine brace” is a distinct coordinating character found in these enzymes by that a terminal His1 provides two coordination sites to Cu. These enzymes also contain a binding flat surface that can directly interact with cellulose without prior depolymerization of glucan chain. PMOs can be categorized into three types. Type-1 PMOs hydroxylate the C1 position of pyranose residues and produce an aldonolactone. Type-2 PMOs hydroxylate the C4 position of pyranose and yield a-ketoaldehyde. Type-3 PMOs are less specific, giving an aldonolactone and/or nonreducing end oxidized products. Upon binding with cellulose, it has been hypothesized that atmospheric diatomic oxygen is recruited into the active site and reduce Cu(I) to Cu(II) via an internal electron transfer. The resulting reactive Cu-O-O[·] species may directly undergo hydrogen abstraction (HAT) at C1 or C4 or it is further reduced to become more reactive species (e.g. Cu-O[·]) followed by HAT. Subsequent reactions involve oxygen rebound (hydroxylation) onto cellulosic C1 or C4 carbon and the cleavage of 1,4-glycosidic bond. However, the detailed mechanism of this process is unknown and still a subject of debate. To unravel the mystery of these unusual enzymes, both experimental and

theoretical investigations are required. The elucidation of how these copper-metalloenzymes play a role in the oxidation of C-H bond would provide a significant impact for the future of renewable resources.

5.2 Proposed Reaction Mechanisms of PMOs

It has been hypothesized that PMOs are activated once atmospheric O₂ is recruited into the active site and bind with a reduced form of Cu(I). Whether O₂ recruitment happens prior to or after cellulose binding is still a subject of debate. Walton *et al.*^[4] suggested that the production of reactive oxygen species is regulated by the presence of cellulose in a PMO active site. This implication is also supported by the fact that rapid release of O₂ after the formation of a Cu-superoxide enzyme complex without a substrate.^[5] Several catalytic reaction pathways for PMOs have been proposed and detailed mechanisms have been discussed.^[6-9] Here, two main mechanisms used in this work are discussed.

5.2.1 Superoxide Mechanism

This mechanism has been proposed by groups of Philips and Beeson.^[10,11] Upon enzyme-substrate binding, O₂ is recruited to bind with Cu(I). The oxidation of Cu(I) is achieved through an internal electron transfer from Cu(I) to O₂, forming a superoxide species (Cu(II)-O-O[·]). Fig. 5.1 outlines possible pathways for cellulose decomposition via the superoxide intermediate. The next step is HAT from either C1 or C4 of the substrate, creating a Cu(II) peroxy (Cu(II)-O-OH) intermediate and a substrate radical. Subsequently, second electron transfer coupling with protonation cleaves the O-O bond of the Cu(II)-O-OH intermediate, releasing H₂O and generating Cu(II)-O[·], namely the Cu(II) oxyl intermediate. Next, the hydroxylation of the substrate occurs via the abstraction of oxygen by the radical substrate coupled with a proton transfer. The hydroxylated

substrate becomes unstable and can be easily decomposed by breaking of glycosidic bonds. This pathway corresponds to the lower route in Fig. 5.1

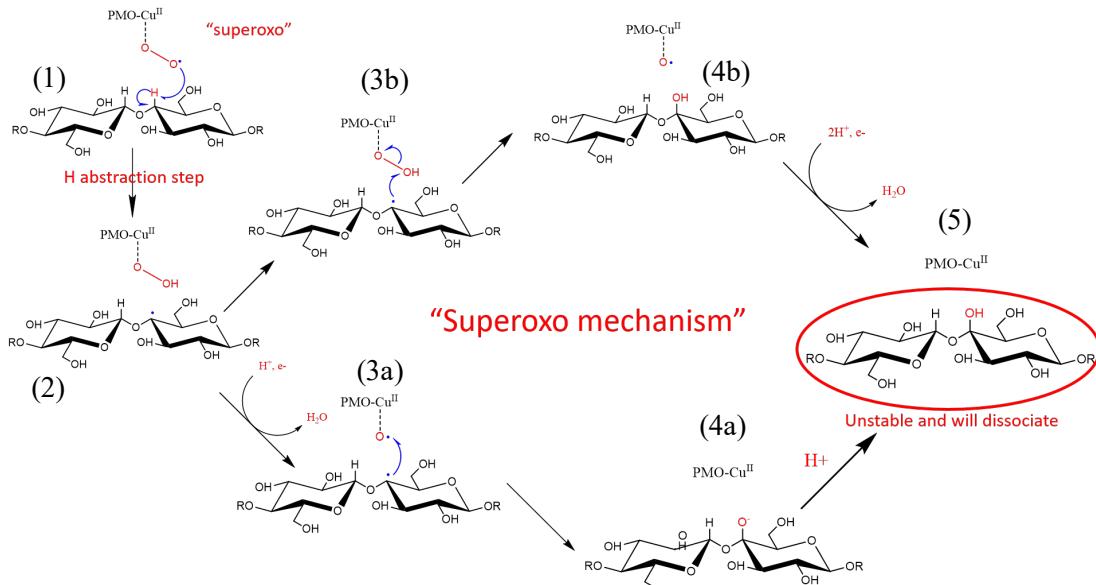


Figure 5.1: Superoxime mechanism

Another pathway of the superoxime intermediate (the upper route of Fig. 5.1) has been proposed based on the computations by Kim *et al.*^[12] Instead of O-O bond being cleaved via the second electron transfer, the radical substrate is directly hydroxylated through the breaking of O-OH bond of the Cu(II)-O-OH species. The resulting substrate then undergoes the decomposition. The second electron transfer coupled with protonation occurs to eject O[·] out of the PMO-Cu(II) as H₂O. Finally, Cu(II) is reduced to Cu(I) through an external electron transfer prior to the next catalytic cycle.

5.2.2 Oxy Mechanism

In this mechanism (Fig. 5.2), the Cu(II)-O-O[·] intermediate which results from O₂ binding with PMO-Cu(I), is further reduced by an external reducing agent such as ascorbic acid to form

Cu(II)-O[·], namely the copper oxyl intermediate, and a water as a byproduct. This intermediate is suggested to be more reactive than Cu(II)-O-O[·]. The next step involves HAT from the substrate, followed by hydroxylation of the substrate via a rebound mechanism. Then the reaction proceeds to decomposition of the hydroxylated cellulose and recycling step to obtain Cu(I) for the next catalytic step.

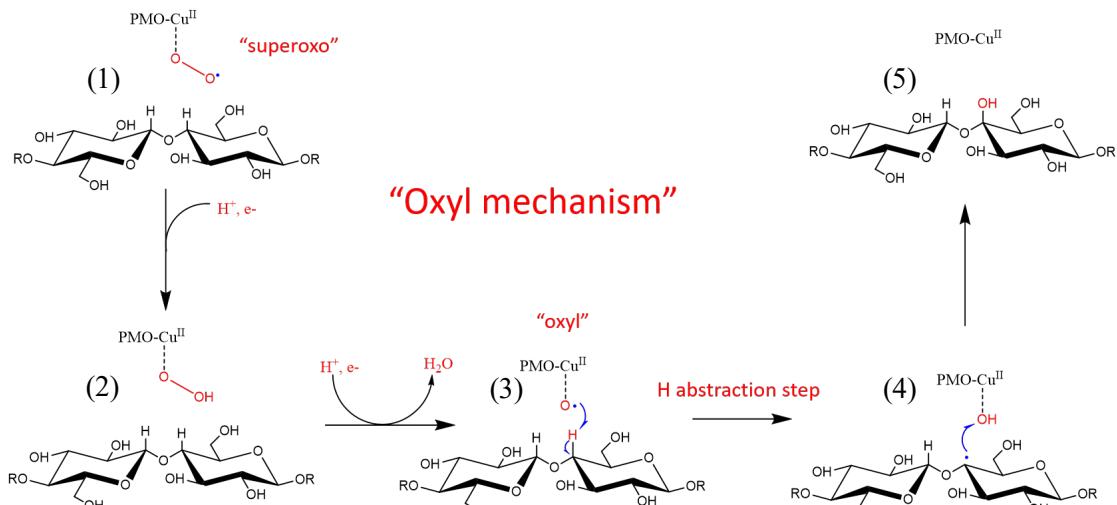


Figure 5.2: Oxyl mechanism

5.3 Theoretical Investigations of PMO Reaction Mechanisms

While there have been a growing number of experimental studies to unravel the mystery of PMO reaction mechanism, this is not the case for theoretical studies. Kim *et al.*^[12] employed open shell DFT^[13-15] calculations to investigate the reaction pathways for cellobiose decomposition using PMO characterized from *T. aurantiacus*^[16] as their model. The simulations were performed based on a so-called active site model (ASM), which included an optimum number of active site's residues, Cu, O₂, water, and substrate explicitly described through a quantum mechanical (QM) description. The rest of the system was treated implicitly using a conductor-like polarization continuum model (CPCM). The calculated reaction barriers for the rate determining steps of the

superoxo and oxyl mechanisms using B3LYP^[17,18]/6-311++G**//B3LYP/6-31G* were reported to be ~40.0 and ~24.0 kcal/mol, respectively. Thus, they suggested that fungal PMOs possibly catalyzes the cellulose decomposition via the oxyl mechanism. The role of N-terminal methylation towards the reaction mechanism was also examined. However, no distinct reaction barriers impact in the absence of N-terminal methylation could be drawn at the employed level of theory.

To date, no experimental measurements have been reported regarding the strength of C1-H or C4-H bond in polysaccharides. However, the estimation of these bond strengths was performed with glucose and reported to be ~95 kcal/mol.^[19] Hence, it is likely that the C-H bond strength would be higher in crystalline lattice of a polysaccharide. How PMOs catalyze such a high energetic process is still questionable. For this reason, investigation of the HAT step of PMOs was performed in this work to provide more insightful information for understanding the chemistry of PMOs.

Earlier study by Kim *et al.*^[12] was limited to an active site model (ASM) of the PMO active site and the effect from protein environment was approximated using an implicit solvent model (CPCM) of ether. In addition, finite temperature fluctuations of the protein environment were not considered and the reaction energetics was described based on total energy differences. The total energy is not always an appropriate metric in a characterization of a reaction process for a complex system due to rugged potential surface. Rather, free energy is a quantity suited to determine the driving force of a chemical process. For this reason, more realistic treatment of the system was developed in this study using a hybrid quantum mechanical molecular mechanics (QM/MM) method. This allows us to investigate the system with a full representation of PMOs. Within the QM/MM framework, the reactive region associated with bond breaking or bond forming is treated using a QM method while the protein environment, which is not directly involved in the enzyme chemistry, is treated with a molecular mechanical (MM) scheme. In addition, it is also suggested that hydrogen bonding network between water molecules and enzyme-substrate plays an important

role in enzyme-substrate binding and could also provide a potential proton transfer channel.^[4] Thus, it is more appropriate to represent the system with explicit solvent model to capture hydrogen bonding network.

To obtain an unbiased view of reaction process, nudged elastic band (NEB) method^[20-22] can be applied to obtain the minimum energy path (MEP) on the QM/MM potential surface (PES). The NEB method optimizes a trial reaction pathway between two minima (usually reactant and product states). The overall pathway includes intermediate images connecting 2 end states via harmonic spring forces. Finite temperature fluctuations of the protein environment are handled via free energy perturbation (FEP) to investigate the environmental effect towards the reaction process. With the NEB method and FEP calculations combined, free energy profile of the reaction can be obtained and the activation free energy can be estimated.

Recently, Walton *et al.*^[4] characterized the PMO structures from *L. similis* using X-ray crystallography. These PMOs show cleavage activity on soluble oligosaccharides such as cellobiose (G3) or cellohexose (G6). The site of oxidation was found to be specific at C4. The resolved X-ray structures also include substrate bound PMOs. In this work, one of these structures was developed and used for the study the C-H activation step of PMO within QM/MM framework. Estimation of the reaction barrier was achieved through the NEB method and FEP calculations. Also presented here is the application of newly developed NWChem-AMBER interface which allows us to utilize *ab-initio* QM/MM methodologies with large complex systems. This study provides primary insights about catalytic functionality of PMOs on a soluble polysaccharide substrate that can further hint us to unravel the mystery of cellulose degradation catalyzed by these enzymes.

5.4 Methodology

5.4.1 QM/MM Optimization of X-ray Structure

Our first step was to provide detailed analysis of an X-ray crystallographic structure to establish a baseline for subsequent simulations. The resolved X-ray structure of *L. similis* (PDB 5ACF)^[4] with a bound substrate (G3) has been used as our reference structure. The model system was prepared with AMBER preparation interface LEaP. The entire PMO enzyme, crystallographic waters, ions (Cu and Cl⁻), and G3 were included in our model. Additional waters and Na⁺ ions were added into the system to obtain a proper simulation box and ensure neutrality of the system. The size of the simulation box was 79 x 79 x 81 Å³. In this model system, the ff14SB force field^[23,24] was applied for amino acids and the GLYCAM06 force field^[25,26] for G3. TIP3P water model was used to describe bulk aqueous solution.^[27] Here, an active region was defined for further QM/MM simulations. This includes part of His1, His78, Gln 162, Tyr164 amino acids, part of G3, one Cu atom, one exogenous Cl ligand and one water molecule (107 atoms in total). Bad contacts from preparation step were eliminated by initial minimization. Subsequent molecular dynamics (MD) using NVT ensemble at 298 K was performed for 500 ps using the MD engine of Amber.^[28,29] During these steps, the defined active region was held fixed and represented with ESP fitted charges throughout the simulation.

A snapshot from MD trajectory was selected for geometry optimization. This was performed using the multiregion optimization algorithm via the developed NWChem-AMBER interface. The defined active region (107 atoms) was treated as the QM subsystem and described using open-shell DFT with B3LYP exchange-correlation functional. 6-311G* basis set was applied for Cu^[5] and 6-31G* basis set for all other QM atoms. The total QM charge is 1 and the electronic state is doublet. The rest of the system was defined as the MM subsystem and described using AMBER and GLYCAM06 force fields. Hydrogen link atom approach was employed to saturate electron valency at the QM-MM boundary and was also treated as part of the QM subsystem.

Through the multiregion optimization, the QM subsystem and the MM subsystem were decoupled and alternatively optimized. Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was used during the QM optimization cycles while steep descent (SD) algorithm was applied during the MM optimization cycles.

5.4.2 Development of Oxygen Binding Active PMO Models

It has been hypothesized that the enzyme is activated through the Cu-O₂ binding and forms the PMO-Cu(II)-O-O[·] enzyme-substrate complex. To date, only X-ray structures from *N. crassa*^[30] has been reported with the presence of dioxygen species in the PMO's active site. However, those are substrate-unbound characterized enzymes. Hence, the model of the PMO-Cu(II)-O-O[·] complex with a bound substrate in this work was developed based on the reference X-ray structure by replacement of Cl⁻ ligand with O₂ in the equatorial position adopting an end-on configuration. This is consistent with a theoretical study by Kjaergaard *et al.*^[5] They suggested that this O₂ binding coupled with inner sphere electron transfer is slightly favorable ($\Delta G^\circ = -2.6$ kcal/mol). Nevertheless, the model used in this work differs from the one adopted by Kim *et al.* where O₂ is bound to Cu in the axial position.^[12] This results from two different characterized PMO X-ray structures being used. In addition, Kim's group developed their model based on the substrate-unbound enzyme while the structure used in this work contains a substrate occupying the axial coordinating site to Cu, thus rendering this binding site infeasible for O₂ binding. The same strategy was also employed to create a Cu-oxyl enzyme-substrate complex for the oxyl mechanism. In this case, a monoatomic O replaces Cl⁻. The resulting structures were then solvated in water boxes and neutralized. The same AMBER^[23,24] and GLYCAMS06^[25,26] force fields from the previous section were used. Subsequent MD simulations were carried out using the same protocol applied to the X-ray system model.

5.4.3 QM/MM Optimization of Active PMOs

An equilibrated structure of the active Cu(II)-O-O[·] enzyme-substrate complex was extracted from MD simulation. From now on, this structure is denoted RS_{SPO}. RS_{SPO} was optimized using the QM/MM multiregion optimization algorithm. The QM subsystem includes atoms defined in the active region (except Cl⁻) plus O₂ and 2 additional water molecules entering the PMO's active site during MD step. A total of 114 atoms (including H-link atoms) were treated as the QM subsystem. The same basis sets as used in previous section were applied. The total QM charge is 1 and the electronic state is triplet. The wavefunction was evaluated using the open-shell DFT with B3LYP exchange-correlation functional. The MM subsystem was described using AMBER^[23,24] and GLYCAM06^[25,26] force fields. The optimization protocol in the previous section was used to obtain RS_{SPO}.

The above strategy was also applied to obtain the active Cu(II)-O[·] enzyme-substrate complex, denoted RS_{OXL}. Similarly, the QM subsystem of RS_{OXL} contains those atom pre-defined as the active region (except Cl⁻) plus monoatomic O and 1 water molecule. This results in a total of 110 QM atoms. The same QM/MM method and setup were used to acquire the optimized RS_{OXL}.

5.4.4 Determination of HAT Reaction Pathways

In this section, reaction mechanisms for HAT by active PMO models were investigated. It has been proposed that a radical substrate and an active PMO with abstracted hydrogen are yielded from this process (See Fig. 5.1 and 5.2). Since the structures of such intermediates cannot be obtained from experimental data, the models were developed based on the RS states of active PMO models. This was achieved through the QM/MM constrained optimization. Starting from RS states, the O2-H4 distance (in case of Cu(II)-O-O[·]) or O1-H4 (in case of Cu(II)-O[·]) was constrained at ~1.0 Å. Other degrees of freedom were allowed to move during the optimization. Next, the

constraints were lifted after H4 was transferred to oxygen followed by the other unconstrained QM/MM optimizations until the the product states (PS_{SPO} and PS_{OXL}) were obtained.

The optimized RS and PS structures from both SPO and OXL systems were then used to determine optimized reaction paths for HAT via active PMOs. This was achieved through nudged elastic band (NEB) method,^[20-22] which optimizes a trial reaction pathway between 2 fixed end states (in this case RS and PS). The reaction pathway is represented by the images of the system connected via harmonic spring forces. Within the QM/MM framework, the reaction coordinate is usually dominated by the QM coordinates. Hence, the NEB energy function is expressed as

$$E^{NEB} = \sum_n E[\{\mathbf{R}_{QM}^n\}, \{\mathbf{R}_{MM}^n\}, \psi] + \sum_n k |\mathbf{R}_{QM}^n - \mathbf{R}_{QM}^{n-1}|^2. \quad (5.1)$$

The index n in the equation denotes the number of images representing the reaction path.

A total of 10 images were used to represent each reaction path. The QM wave function was calculated using open-shell DFT/B3LYP with the net QM charge of 1 at the triplet state. The same basis sets used in the QM/MM optimization were also used in this step. The MM optimization for each image was performed every 5 steps of the the QM NEB calculation. During the MM optimization, the QM subsystem for each image was kept fixed and represented by a set of fitted ESP charges calculated at the end of each NEB calculation. The QM/MM NEB calculations were performed through the NWChem^[31]-AMBER^[28,29] interface. Once the optimized energy paths were obtained, the top image for each system was used to search for the corresponding transition state (TS). Each optimized TS structure was verified by one imaginary frequency through a numerical frequency calculation.

5.4.5 Free Energy Calculation

The reaction barrier can be estimated through the calculation of free energy difference between RS and TS along the optimized NEB reaction path. This can be achieved via the multilevel

purturbation method^[32] shown in Fig. 5.3. Therefore, the total free energy difference can be expressed as

$$\Delta F_{A \rightarrow B} = \Delta F_{A \rightarrow A}^{QM \rightarrow MM} - \Delta F_{B \rightarrow B}^{QM \rightarrow MM} + \Delta F_{A \rightarrow B}^{MM}. \quad (5.2)$$

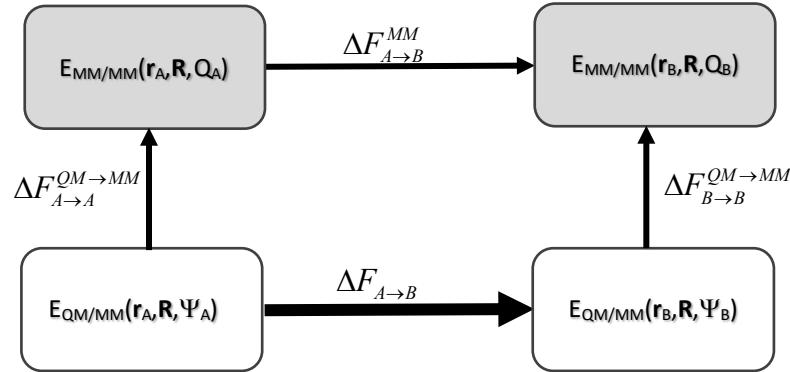


Figure 5.3: Free energy cycle using multi-level perturbation method

The first two terms are the free energy differences for changing the QM description to fixed coordinates (\mathbf{r}_A or \mathbf{r}_B) and electrostatic potential (ESP) charges (\mathbf{Q}_A or \mathbf{Q}_B). The QM fluctuations are assumed to be harmonic. Therefore, Eq. (5.2) can be formulated as^[33]

$$\Delta F_{A \rightarrow B} = \Delta E_{A \rightarrow B}^{\text{int}} + \Delta F_{A \rightarrow B}^{QM-\text{harmonic}} + \Delta F_{A \rightarrow B}^{MM}. \quad (5.3)$$

The first two term in Eq. (5.3) can be obtained from the QM/MM single point energy and frequency calculations, respectively. The third term is finite temperature fluctuations of the MM subsystem, which can be approximated using FEP method. From a smooth transformation for the reaction coordinate between 2 states A and B,

$$\begin{aligned} \mathbf{r}_\lambda &= (1 - \lambda)\mathbf{r}_A + \lambda\mathbf{r}_B \\ \mathbf{Q}_\lambda &= (1 - \lambda)\mathbf{Q}_A + \lambda\mathbf{Q}_B \end{aligned} \quad (5.4)$$

where λ is in the range between 0 and 1. We have that

$$\Delta F_{A \rightarrow B}^{MM} = -\sum_i 1/\beta \ln \left\langle \exp \left\{ -\beta \Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{ESP} \right\} \right\rangle_{MM, \lambda_i}. \quad (5.5)$$

$$\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{ESP} = E_{qm/mm}(\mathbf{r}_{\lambda_{i+1}}, \mathbf{Q}_{\lambda_{i+1}}, \mathbf{r}_{MM}) - E_{qm/mm}(\mathbf{r}_{\lambda_i}, \mathbf{Q}_{\lambda_i}, \mathbf{r}_{MM}). \quad (5.6)$$

Therefore, FEP calculations were performed between RS and TS and between TS and PS for both SPO and OXL mechanisms. A total of 20 λ values with the width of 0.05 were used for smooth transformation between the 2 corresponding states. The samplings were done though MD simulations using NVT ensemble at 298 K. The ESP charges for the QM subsystem can be obtained using NWChem^[31] and all FEP calculations were carried out using AMBER.^[28,29]

5.4.6 Open shell singlet energy and Natural Bond Orbital (NBO) Analysis

The oxidative mechanisms of oxygen binding PMOs may involve spin-crossing events. Therefore, open-shell singlet or triplet states were considered to obtain the lowest energy spin state along the reaction process. In this work, the triplet spin states were used in the QM/MM optimizations and NEB calculations. The open-shell singlet state energies were evaluated based on the triplet optimized structures. Yamaguchi broken-spin-symmetry (BS) formula^[34,35] was used to compute the energy of the spin-purified low spin state (LS):^[36]

$$E^{LS} = \frac{E^{BS}(\langle S^2 \rangle^{HS} - \langle S^2 \rangle^{LS}) - E^{HS}(\langle S^2 \rangle^{BS} - \langle S^2 \rangle^{LS})}{\langle S^2 \rangle^{HS} - \langle S^2 \rangle^{LS}}. \quad (5.7)$$

$\langle S^2 \rangle^{HS}$ and $\langle S^2 \rangle^{BS}$ are expectation values of the total spin operators of the triplet and open-shell singlet, respectively. $\langle S^2 \rangle^{LS}$ is 0 and assumed for the closed-shell singlet.

In addition, spin densities and charge distribution of copper-oxygen bound complexes for each spin state were analyzed using Natural bond orbital (NBO) analysis.^[37]

5.5 Results and Discussion

5.5.1 Analysis of X-ray Structure

In the PMO active site, Cu is bound tightly with 4 donating atoms from amino acid residues plus one exogenous Cl^- ligand. Tyr164 coordinates with Cu and occupies the axial position with the distance of 2.47 Å. Methylated His1 and His78 provides 3 coordinating sites to Cu in the equatorial positions with the distance of ~2.0 Å, with Cu-Nδ(His1) distance a little shorter at 1.88 Å. Cl^- ligand occupies the last equatorial coordinating site with the distance of 2.34 Å. Since another axial position is occluded by part of the bound substrate, this results in a total of 5 ligands binding to Cu. The optimized structure obtained from QM/MM multiregion optimization is shown in Fig. 5.4. Overall, the organization of the active site is in very good agreement with the reference X-ray structure (see Table 5.1). The RMSD of heavy atoms of the defined active region including exogenous Cl^- ion is 0.131 Å. Hence, this validates the selected optimization algorithm as well as the level of theory employed with the enzyme-substrate model. The oxidation state of Cu is 2 as implicated by the natural charge of 1.32 from NBO analysis (see Table 5.2), consistent with characterized X-ray data of 5ACF.

Table 5.1: Geometrical parameters (in angstrom) of substrate bound PMO in different states from computations and experimental measurements^{[a],[b]}

distance	5ACF (expt)	Opt	Cu-superoxo (SPO)			Cu-oxyl (OXL)		
			RS _{SPO}	TS _{SPO}	PS _{SPO}	RS _{OXL}	TS _{OXL}	PS _{OXL}
Cu-O(Tyr164)	2.47	2.46	2.56	2.68	2.53	2.62	2.88	2.65
Cu-Ne(His78)	2.06	2.03	2.03	2.04	2.02	2.02	2.02	2.02
Cu-N δ (His1)	1.88	1.99	2.01	2.00	2.02	1.89	1.97	2.00
Cu-N(His1)	2.16	2.05	2.03	2.06	2.05	2.09	2.13	2.09
Cu-XA	2.34	2.39	1.97	1.97	1.96	1.89	1.99	1.94
XB-C1	3.91	4.03	3.34	2.94	3.25	3.99	3.29	3.74
XB-C4	3.74	3.68	3.26	2.56	2.84	3.45	2.57	3.48
O1-O2(SPO)	-	-	1.32	1.42	1.47	-	-	-
RMSD	-	0.131	0.298	0.286	0.299	0.216	0.277	0.254

[a] expt = experiment; Opt = oxygen unbound optimized structure; XA = Cl, O1_{SPO}, or O1_{OXL}; XB = Cl, O2_{SPO}, O1_{OXL}.

[b] RMSDs were calculated based on the alignment of common heavy QM atoms of the corresponding structures onto those of the reference X-ray structure.

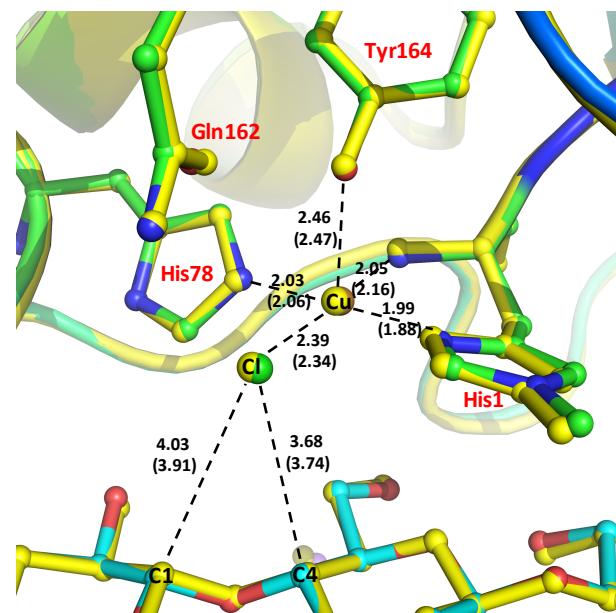


Figure 5.4: Optimized structure aligned on reference X-ray structure (yellow) with hydrogen atoms omitted. X-ray parameters are in parentheses

5.5.2 Enzyme Activation and O₂ Binding to the PMO Active Site

Upon O₂ binding, the Cu(II)-O-O[·] enzyme-substrate complex is formed. The oxidation state of Cu is 2 as implicated from the natural charge of 1.34 from NBO analysis. This is consistent with the suggestion that PMOs are activated by the oxidation of PMO-Cu(I) via an internal electron transfer to O₂. The spin densities of 0.62(Cu), 0.58(O1), and 0.57(O2) from NBO analysis (Table 5.2) indicates biradical character of the Cu(II)-O-O[·] species. The optimized structure of RS_{SPO} is shown in Fig. 5.5. The RMSD of 0.298 Å indicates partial rearrangement of the active site when O₂ is in place of Cl⁻ as an exogenous ligand in the equatorial position. A notable change is the shorter Cu-O distance of 1.97 Å compared to the Cu-Cl distance of 2.34 Å. O₂ coordinates with Cu with an end-on configuration. There are little changes in distances between Cu and coordinating amino acids. The O1-O2 bond distance is 1.32 Å, resembling the O-O distance of superoxide species (1.33 Å). The O2-C1 and O2-C4 bond distances are 3.34 and 3.26 Å, respectively. These distances are in close proximity for HAT. Two water molecules entering the active site were observed during the equilibration step and included in the QM subsystem. These water molecules stabilize O₂[·] through hydrogen bonds. The triplet (T) is the ground state of RS_{SPO}, which is ~3.4 kcal/mol more stable than the open-shell singlet (S) (Table 5.3). This S-T gap is consistent with similar end-on Cu-O₂ end-on complexes.^[38]

Table 5.2: Natural charges and spin densities from NBO analysis^{[a]-[c]}

Structure	Natural Charge				Spin density			
	Cu	X	O2	C4	Cu	X	O2	C4
RS _{X-ray}	1.32	-0.76	-	0.05	0.65	0.09	-	0.00
RS _{SPO}	1.34	-0.36	-0.33	0.05	0.62 (0.63)	0.58 (-0.33)	0.57 (-0.51)	0.01 -0.01
TS _{SPO}	1.36	-0.57	-0.46	0.19	0.64 (0.64)	0.24 (0.01)	0.28 (-0.25)	0.53 (-0.54)
PS _{SPO}	1.37	-0.63	-0.54	0.34	0.65 (0.65)	0.14 (0.11)	0.01 (-0.01)	0.84 (-0.84)
RS _{OXL}	1.36	-0.70	-	0.05	0.65 (0.64)	1.11 (-0.85)	-	0.00 (0.00)
TS _{OXL}	1.35	-0.90	-	0.15	0.65 (-0.63)	0.71 (0.40)	-	0.40 (0.40)
PS _{OXL}	1.36	-1.12	-	0.35	0.66 (0.66)	0.15 (0.13)	-	0.85 (-0.85)

[a] X = Cl_{X-ray} or O1_{SPO}

[b] spin multiplicity: X-ray = doublet; SPO, OXL = open-shell singlet and triplet

[c] spin densities of open-shell singlet are in parentheses. Note open-shell singlet and triplet natural charges are almost identical and only single values were reported.

Table 5.3: QM triplet and open shell singlet energies (Hartree) for each state along HAT with copper-superoxo species.^[a]

State	Triplet		Singlet (open shell)			$\Delta E_{S \rightarrow T}$ (kcal/mol)
	E	$\langle S^2 \rangle$	E	$\langle S^2 \rangle$	E_{corr}	
RS _{SPO}	-4502.39163	2.0071	-4502.38891	1.0055	-4502.38618	-3.42
TS _{SPO}	-4502.33526	2.0096	-4502.33474	1.0056	-4502.33423	-0.64
PS _{SPO}	-4502.34852	2.0058	-4502.34849	1.0045	-4502.34846	-0.04

[a] E_{corr} was calculated based on Eq. (5.7)

5.5.3 Hydrogen Abstraction with Copper-superoxo Species

After the Cu(II)-O-O[·] enzyme-substrate complex is formed through the activation of O₂ binding, the reaction proceeds through HAT step. At the transition state (TS_{SPO}), O₂ of Cu(II)-O-O[·] moves closer to H4 of the substrate at the distance of 1.12 Å. The C4-H4 bond is partially broken and the bond length changes from 1.09 at RS_{SPO} to 1.52 Å at TS_{SPO}. The O1-O2 bond of Cu(II)-O-O[·] is lengthened to 1.42 Å, indicating that Cu(II)-O-OH is partially formed. The Cu-O(Tyr164) distance is a little longer at 2.68 Å. This results from Cu moving along with O₂ for HAT while maintaining the square planar configuration for 4 equatorial ligands. The distances between Cu and the equatorial ligands are almost identical to those at RS_{SPO}. TS_{SPO} was verified through a numerical frequency calculation with one imaginary frequency at 1581*i* cm⁻¹. At the product state (PS_{SPO}), the C4-H4 bond is completely broken and Cu(II)-O-OH is fully formed, giving a radical substrate. The O1-O2 bond becomes 1.47 Å, which is close to the ideal O-O bond length of peroxide species at 1.49 Å. Cu and 4 equatorial ligands move up a little from the substrate, causing Cu close to O(Tyr164) like in the case of RS_{SPO} with the distance of 2.53 Å. Along the reaction, two water molecules keep coordinating with O₂. While the triplet is the ground state for RS_{SPO}, the S-T gaps of TS_{SPO} and of PS_{SPO} are very small. The optimized structures of TS_{SPO}, and PS_{SPO} states are illustrated in Fig. 5.6-5.7.

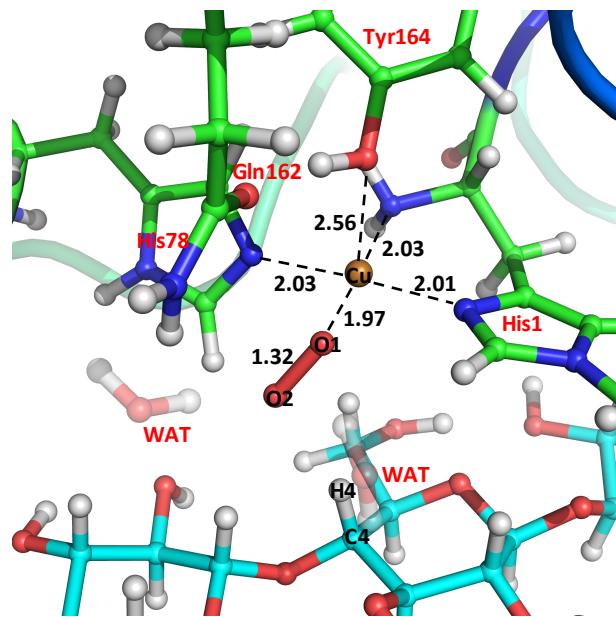


Figure 5.5: RS_{SPO} structure (coordinating waters are transparent)

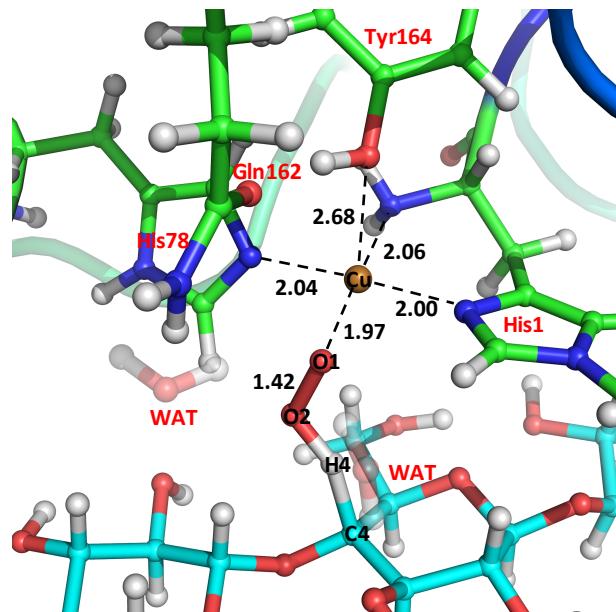


Figure 5.6: TS_{SPO} structure (coordinating waters are transparent)

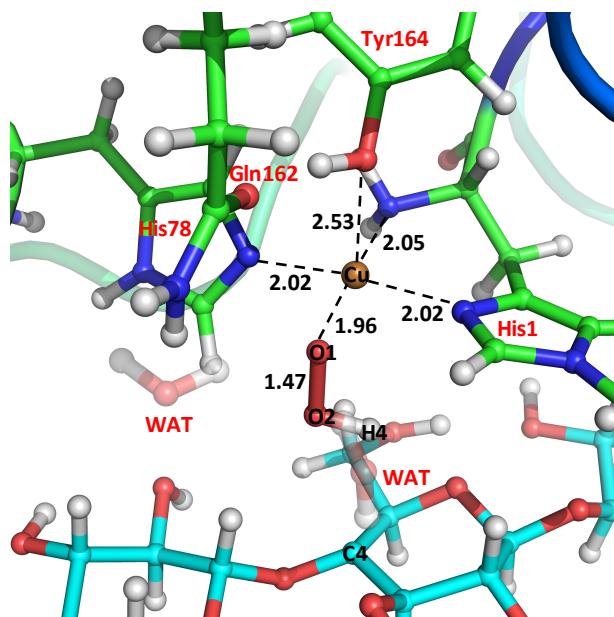


Figure 5.7: PS_{SPO} structure (coordinating waters are transparent)

Table 5.4: Free energy of HAT with Cu-superoxo species

State	ΔE_{int}	$\Delta F_{\text{qm-harmonic}}$	ΔF_{mm}	ΔF_{total}
RS _{SPO}	0.00	0.00	0.00	0.00
TS _{SPO}	31.66	-3.35	1.92	30.24
PS _{SPO}	25.56	-1.05	1.36	25.87

The activation free energy of HAT through the Cu(II)-O-O[·] intermediate was estimated to be 30.2 kcal/mol. The main contribution of the total activation free energy comes from ΔE_{int} between RS_{SPO} and TS_{SPO} while the fluctuations of environment in the MM subsystem have little effect towards the total free energy. Our calculated reaction barrier is comparable to ~34 kcal/mol of the activation energy for HAT calculated by Kim *et al.*^[12] The relative free energy of RS_{SPO} and PS_{SPO} was calculated to be 25.9 kcal/mol, suggesting an endothermic process for HAT through Cu-O-O[·]. However, The Michaelis-Menten kinetics of cleavage of a soluble substrate by *L. similis* gives $K_m = 43 \mu\text{M}$ and $k_{\text{cat}} = 0.11 \text{ s}^{-1}$ at 310 K,^[4] which corresponds to the activation energy of ~15

kcal/mol in the rate determining step. Hence, this implicates that HAT is unlikely activated through Cu(II)-O-O[·] due to the high estimated reaction barrier.

Table 5.5: QM triplet and open shell singlet energies (Hartree) for each state along HAT with copper-oxyl species.^[a]

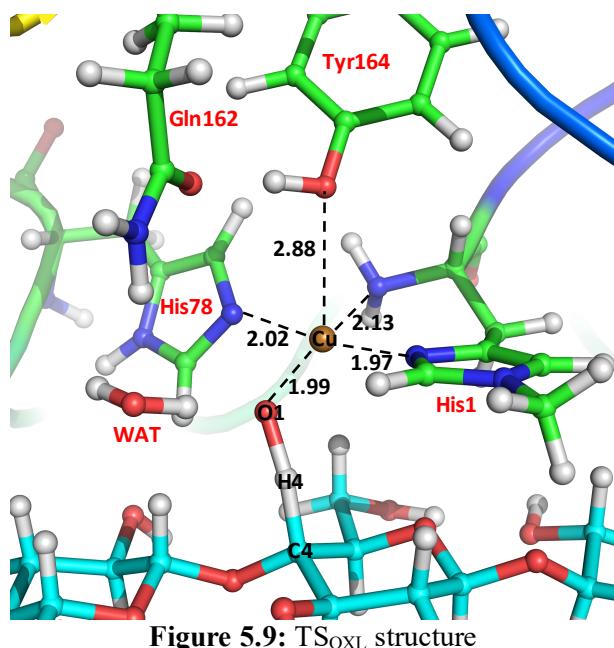
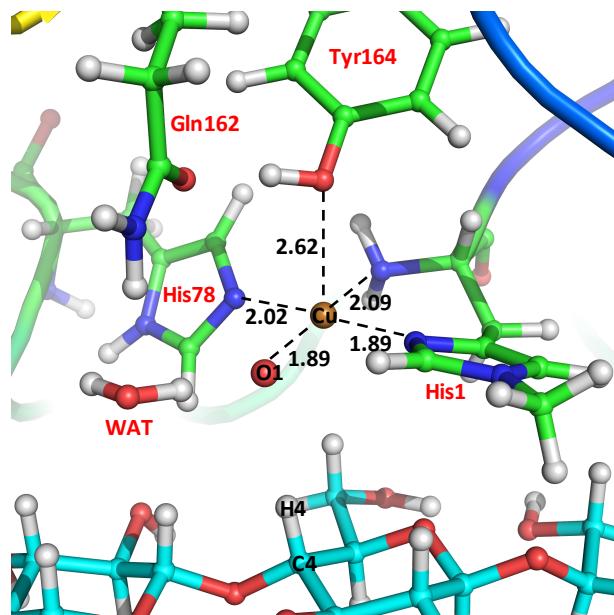
State	triplet (T)		open shell singlet (S)			$\Delta E_{S \rightarrow T}$ (kcal/mol)
	E	$\langle S^2 \rangle$	E	$\langle S^2 \rangle$	E_{corr}	
RS _{OXL}	-4350.56914	2.0063	-4350.56482	0.9990	-4350.56055	-5.39
TS _{OXL}	-4350.66689	2.0118	-4350.66678	0.9529	-4350.66668	-0.13
PS _{OXL}	-4350.69869	2.0055	-4350.69879	1.0026	-4350.69890	0.13

[a] values of E_{corr} were calculated based on Eq. (5.7)

5.5.4 Hydrogen Abstraction with Copper-oxyl species

Due to the high activation free energy of HAT through the Cu(II)-O-O[·] species, it is suggested that more reactive species is required. One hypothesis is the Cu(II)-O-O[·] intermediate can be further reduced by a small electron donor such as ascorbic acid to form an oxyl Cu(II)-O[·] species prior to HAT. The presence of this reactive species has been implicated in various bioinorganic mechanism.^[39-41] Kim *et al.*^[12] calculated the transformation of Cu(II)-O-O[·] to Cu(II)-O[·] with ascorbic acid as an electron and proton donor using B3LYP/6-31G* and reported that Cu(II)-O[·] is ~3.1 kcal/mol more reactive. However, no barriers are considered since the order of electron and proton transfers is unknown. In this work, we investigated HAT step presuming that the Cu(II)-O[·] is already formed and any prior steps are not rate determining steps. This corresponds to step (3) in Fig. 5.2. The optimized structure of the Cu(II)-O[·] enzyme-substrate complex, denoted RS_{OXL}, is shown in Fig. 5.8. The oxidation state of 2 for Cu(II)-O[·] was implicated by the natural charge of 1.36. In addition the spin density of 1.11 for O1 reflects a highly reactive radical character for Cu(II)-O[·]. The RMSD of 0.216 Å suggests that there are slight changes in the active site

compared to that of the X-ray structure. The distances between Cu and amino acid ligands in the equatorial positions are almost identical to those in the X-ray structure. The Cu-O distance, however, is 1.89 Å shorter than 2.34 Å of the Cu-Cl distance in the X-ray structure and than 1.97 Å of the Cu-O₂ distance in RS_{SPO} structure. The Cu-O(Tyr164) distance of RS_{OXL} is 2.62 Å, a little longer than 2.48 Å of the X-ray structure. The O1-C1 and O1-C4 distance is in close proximity for HAT at 3.99 and 3.49 Å, respectively. One water molecule entering the active site and coordinating with the oxy O1 was observed during MD simulation and also included in the QM subsystem. RS_{OXL} is a triplet in its ground state, with ~5.4 kcal/mol more stable than the open-shell singlet state. At the transition state (TS_{OXL}), the O1-H4 bond is partly formed with the distance of 1.24 Å and the C4-H4 bond is partly broken with the bond distance changing from 1.10 Å to 1.35 Å. Cu and the other equatorial amino acid ligands move along with O1 to keep the square planar during HAT, elongating the Cu-O(Tyr164) distance from 2.62 Å at RS_{OXL} to 2.88 Å at TS_{OXL}. The distances between Cu and the equatorial ligands are almost identical to those at RS_{OXL} with the average value of ~2.0 Å. TS_{OXL} was verified through numerical frequency calculation with one imaginary frequency at 1964*i* cm⁻¹. At the product state (PS_{OXL}), H4 is completely transferred to O1 and a radical substrate is formed. Cu and equatorial ligands move up a little, reducing the C-O(Tyr164) distance from 2.88 Å at TS_{OXL} to 2.65 Å. Along the reaction, the water molecule stabilizes O1 through a hydrogen bond. The triplet and open-shell singlet for TS_{OXL} and PS_{OXL} are almost isoenergetic (Table 5.3). The optimized structures of TS_{OXL}, and PS_{OXL} are shown in Fig. 5.9-5.10.



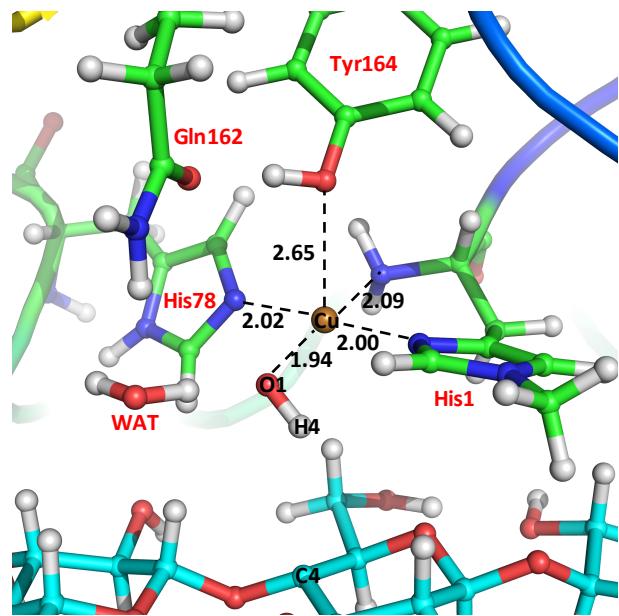


Figure 5.10: PS_{OXL} structure

Table 5.6: Free energy of HAT with Cu-oxyl species

State	ΔE_{int}	$\Delta F_{\text{qm-harmonic}}$	ΔF_{mm}	ΔF_{total}
RS _{OXL}	0.00	0.00	0.00	0.00
TS _{OXL}	17.12	-3.57	-0.18	13.36
PS _{OXL}	-4.45	-1.21	1.27	-4.40

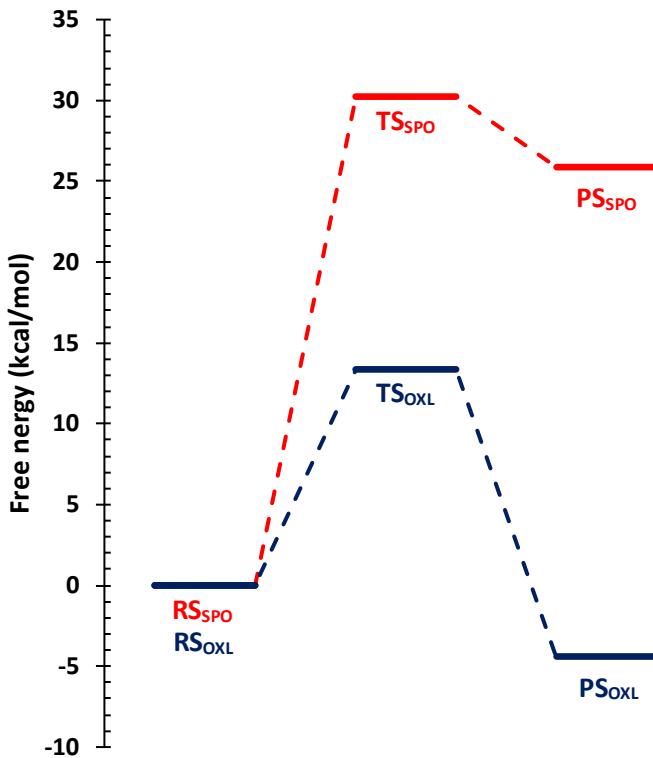


Figure 5.11: Free energy diagram for HAT via superoxo (SPO) and oxyl (OXL) mechanisms

The activation free energy of HAT through the Cu(II)-O[·] species was estimated to be 13.36 kcal/mol. Similar to the case of superoxo mechanism, the difference in the internal energy between RS_{OXL} and TS_{OXL} contributes greatly to the reaction barrier and fluctuations from the MM environment are little. The reaction HAT through Cu(II)-O[·] is slightly exothermic with the reaction free energy calculated to be -4.40 kcal/mol. The calculated activation energy for HAT by Kim *et al.* was reported to be ~24 kcal/mol with B3LYP/6-311++G**//B3LYP/6-31G*.^[12] This step is also suggested to be the rate determining step in their overall oxyl mechanism. Our estimated activation free energy is comparable to activation energy of ~15 kcal/mol in the rate determining step for *L. similis* used for models in this work. Thus, this suggests that C-H activation by PMOs requires a strong oxidative species such as the Cu(II)-O[·] intermediate. However, this is proposed based on the assumption that prior reduction step of Cu(II)-O₂ to Cu(II)-O[·], in which the detail of electron

and proton transfers is still unknown, is favorable and not the rate determining step. In addition, Cu(II)-oxyl intermediate has been only detected in the gas phase but has been implicated from theoretical studies to play an important role other enzymatic reaction.

5.6 Conclusions

The hydrogen abstraction of a soluble oligosaccharide via an oxidative mechanism of a polysaccharide monooxygenase from *L. similis* was investigated using the QM/MM methodology. Two reaction mechanisms were examined—superoxo and oxyl mechanism. The high estimate reaction barrier (~30 kcal/mol) and reaction free energy (~26 kcal/mol) given by the superoxo mechanism suggests that the reaction is thermodynamically unfavorable. Therefore, the activation of the C-H bond does not likely occur through the Cu(II)-O-O[·] intermediate. On the other hand, the hydrogen abstraction via the oxyl mechanism gives the reaction barrier of ~13 kcal/mol, which is comparable to the reaction barrier of ~15 kcal/mol from the experiment. The calculated reaction free energy of ~4.0 kcal/mol also indicates a slight thermodynamically favorable reaction. Thus, a more favorable reaction pathway for the hydrogen abstraction is implicated from the oxyl mechanism. The interpretation from the free energy calculations also suggests that the main contribution of free energy comes from the internal energy part of the QM subsystem while little effects are contributed by the MM subsystem.

The calculations in this work suggest that a more reactive species such as Cu(II)-O[·] is likely required for the hydrogen abstraction of PMOs, consistent with the results from the theoretical study by Kim *et al.*^[12] However, we cannot preclude the possibility that another reactive oxygen binding Cu that can be formed and activate C-H bond using a different mechanism. Recently, the study by Tolman *et al.*^[42] indicated that the hydrocarbon C-H bond of up to 90 kcal/mol can be oxidized through a Cu(III)-OH intermediate of a small molecule with histidine brace. It is also possible that PMOs may adopt multiple catalytic pathways. Additionally, there are still many

unanswered questions for these fascinating enzymes such as; whether O₂ binding of PMOs happens prior to or after or synergistically with the substrate binding or whether this is an important factor for reactivity; what is the external electron source and the electron transfer mechanism to create a more reactive Cu-oxygen intermediate such as Cu-O[·]; what is the selectivity factors for C1-H or C4-H abstraction; or whether the N-methylation of His1 impacts the catalytic activity of PMOs. All these problems require a synergistic contribution of experimental and computational studies for more insightful information to unravel the mystery behind these enzymes. This study provides a primary insight about the catalytic activity of PMOs on soluble polysaccharides which can be used as a baseline for a further analysis of cellulose decomposition.

Also presented in this work is the application through the newly developed QM/MM NWChem-AMBER interface to gain more insights into a chemical process of PMOs. The entire systems used in this work were treated with explicit models. The QM/MM multiregion optimization was validated through a good agreement between the optimized and reference X-ray structures. The combination of the QM/MM NEB and multi-level free energy perturbation provide an efficient tool to determine the optimized reaction paths and estimate the free energies of enzymatic reactions.

5.6 References

- [1] Y. Nishiyama, P. Langan, H. Chanzy. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.
- [2] D. B. Wilson. *Curr. Opin. Biotechnol.* **2009**, *20*, 295-299.
- [3] G. Vaaje-Kolstad, B. Westereng, S. J. Horn, Z. Liu, H. Zhai, M. Sørlie, V. G. H. Eijsink. *Science* **2010**, *330*, 219-222.
- [4] K. E. H. Frandsen, T. J. Simmons, P. Dupree, J.-C. N. Poulsen, G. R. Hemsworth, L. Ciano, E. M. Johnston, M. Tovborg, K. S. Johansen, P. von Freiesleben, L. Marmuse, S. Fort, S. Cottaz, H. Driguez, B. Henrissat, N. Lenfant, F. Tuna, A. Baldansuren, G. J. Davies, L. Lo Leggio, P. H. Walton. *Nat. Chem. Biol.* **2016**, *12*, 298-303.
- [5] C. H. Kjaergaard, M. F. Qayyum, S. D. Wong, F. Xu, G. R. Hemsworth, D. J. Walton, N. A. Young, G. J. Davies, P. H. Walton, K. S. Johansen, K. O. Hodgson, B. Hedman, E. I. Solomon. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8797-8802.

- [6] P. H. Walton, G. J. Davies. *Curr. Opin. Chem. Biol.* **2016**, *31*, 195-207.
- [7] G. R. Hemsworth, E. M. Johnston, G. J. Davies, P. H. Walton. *Trends Biotechnol.*, **33**, 747-761.
- [8] W. T. Beeson, V. V. Vu, E. A. Span, C. M. Phillips, M. A. Marletta. *Annual Review of Biochemistry*, *Vol 84* **2015**, *84*, 923-946.
- [9] J. Y. Lee, K. D. Karlin. *Curr. Opin. Chem. Biol.* **2015**, *25*, 184-193.
- [10] W. T. Beeson, C. M. Phillips, J. H. D. Cate, M. A. Marletta. *J. Am. Chem. Soc.* **2012**, *134*, 890-892.
- [11] C. M. Phillips, W. T. Beeson, J. H. Cate, M. A. Marletta. *ACS Chemical Biology* **2011**, *6*, 1399-1406.
- [12] S. Kim, J. Ståhlberg, M. Sandgren, R. S. Paton, G. T. Beckham. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 149-154.
- [13] P. Hohenberg, W. Kohn. *Phys. Rev. B* **1964**, *136*, 864-871.
- [14] W. Kohn, L. J. Sham. *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [15] R. G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules; Oxford University Press, New York, **1994**.
- [16] R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J.-C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, P. H. Walton. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 15079-15084.
- [17] A. D. Becke. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [18] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- [19] J. Gao, D. A. Thomas, C. H. Sohn, J. L. Beauchamp. *J. Am. Chem. Soc.* **2013**, *135*, 10684-10692.
- [20] H. Jonsson, G. Mills, K. W. Jacobsen. In Classical and Quantum Dynamics in Condensed Phase Simulations; World Scientific, **2011**; pp 385-404.
- [21] D. Sheppard, R. Terrell, G. Henkelman. *J. Chem. Phys.* **2008**, *128*, 134106.
- [22] G. Henkelman, H. Jónsson. *J. Chem. Phys.* **2000**, *113*, 9978-9985.

- [23] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *J. Am. Chem. Soc.* **1996**, *118*, 2309-2309.
- [24] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.
- [25] M. B. Tessier, M. L. DeMarco, A. B. Yongye, R. J. Woods. *Mol. Simulat.* **2008**, *34*, 349-363.
- [26] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, R. J. Woods. *J. Comput. Chem.* **2008**, *29*, 622-655.
- [27] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. *J. Chem. Phys.* **1983**, *79*, 926-935.
- [28] D. A. Case, R. M. Betz, W. Botello-Smith, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, Amber 2016, University of California, San Francisco, **2016**.
- [29] R. Salomon-Ferrer, D. A. Case, R. C. Walker. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198-210.
- [30] X. Li, William T. Beeson Iv, Christopher M. Phillips, Michael A. Marletta, Jamie H. D. Cate. *Structure* **2012**, *20*, 1051-1061.
- [31] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [32] M. Valiev, B. C. Garrett, M.-K. Tsai, K. Kowalski, S. M. Kathmann, G. K. Schenter, M. Dupuis. *J. Chem. Phys.* **2007**, *127*, 051102.
- [33] Z. Y. Lu, W. T. Yang. *J. Chem. Phys.* **2004**, *121*, 89-100.
- [34] K. Yamaguchi, F. Jensen, A. Dorigo, K. N. Houk. *Chem. Phys. Lett.* **1988**, *149*, 537-542.
- [35] T. Soda, Y. Kitagawa, T. Onishi, Y. Takano, Y. Shigeta, H. Nagao, Y. Yoshioka, K. Yamaguchi. *Chem. Phys. Lett.* **2000**, *319*, 223-230.
- [36] A. Kunishita, M. Z. Ertem, Y. Okubo, T. Tano, H. Sugimoto, K. Ohkubo, N. Fujieda, S. Fukuzumi, C. J. Cramer, S. Itoh. *Inorg. Chem.* **2012**, *51*, 9465-9480.
- [37] E. D. Glendening, C. R. Landis, F. Weinhold. *J. Comput. Chem.* **2013**, *34*, 1429-1437.

- [38] A. d. l. Lande, V. Moliner, O. Parisel. *J. Chem. Phys.* **2007**, *126*, 035102.
- [39] R. A. Himes, K. D. Karlin. *Curr. Opin. Chem. Biol.* **2009**, *13*, 119-131.
- [40] A. Kunishita, J. Teraoka, J. D. Scanlon, T. Matsumoto, M. Suzuki, C. J. Cramer, S. Itoh. *J. Am. Chem. Soc.* **2007**, *129*, 7248-7249.
- [41] P. Comba, S. Knoppe, B. Martin, G. Rajaraman, C. Rolli, B. Shapiro, T. Stork. *Chemistry – A European Journal* **2008**, *14*, 344-357.
- [42] D. Dhar, W. B. Tolman. *J. Am. Chem. Soc.* **2015**, *137*, 1322-1329.