Computational studies of mechanisms underlying the reactions of enzymatic and nonenzymatic systems

Xu, Kai

2017

Xu, K. (2017). Computational studies of mechanisms underlying the reactions of enzymatic and nonenzymatic systems. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/72138>

<https://doi.org/10.32657/10356/72138>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**Computational Studies of Mechanisms Underlying the
Reactions of Enzymatic and Nonenzymatic Systems**

XU KAI

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2017

**Computational Studies of Mechanisms Underlying the
Reactions of Enzymatic and Nonenzymatic Systems**

XU KAI

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the
Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2017

Acknowledgement

First and foremost, I would like to thank my supervisor Professor Hajime Hirao. His patient guidance, thorough support and valuable advice have been very helpful during my Ph.D. study. I am also thankful to him for the example he sets as an outstanding and productive computational chemist.

My special thanks are due to Professor Yong Wang from Chinese Academy of Sciences with whom we worked on the CYP19A1 project together. His constructive comments and deep knowledge broadened my understanding of computational chemistry a lot.

The members of the Hirao group have contributed tremendously to my academic and private time here at NTU. My thanks go to Cathy, Mike, Sareeya, Hung, Nandun, Mai, António, Adam, Adhitya, Wilson, Zhengyu, Kazuki, Xu, Zhi xiang, Ruixing, Binh, Pablo, Xiaoqing, and Anil for our friendship, for the knowledge and skills they imparted to me, for the discussions we had, for the projects we worked on together, for the suggestions they gave, for the happy hours we shared and above all, for making my time at NTU enjoyable.

Last but not least, I thank my family and friends for their continuous support, willingness to listen, and encouragement, which have helped me throughout my PhD study so much!

Contents

List of Figures	iv
List of Tables.....	ix
List of Schemes	x
List of Abbreviations.....	xi
Abstract	xiii
Chapter 1 Introduction	1
1.1 Density Functional Theory	2
1.2 Molecular Mechanics	4
1.3 ONIOM.....	6
1.3.1 Energy	6
1.3.2 Treatment of the boundary	9
1.3.3 Energy gradients.....	11
1.3.4 Geometry optimization.....	12
1.3.5 Embedding schemes.....	16
1.3.6 Setup for ONIOM calculations.....	20
1.3.7 Preparation of a decent initial orbital guess for the model system.....	27
1.4 References.....	28
Chapter 2 Mechanism Scenario in the Conversion of Androgens into Estrogens by CYP19A1	32
2.1 Introduction.....	32
2.2 Methodology.....	37
2.2.1 Protein setup and MM parameters.....	37
2.2.2 ONIOM calculations.....	37
2.3 Results and Discussion.....	39
2.3.1 The first step	39
2.3.2 The second step.....	40
2.3.3 The third step.....	46
2.4 References.....	52

Chapter 3 Ferric-Superoxide Intermediate in the Verdoheme Formation Step of Heme Oxygenase.....	57
3.1 Introduction.....	57
3.2 Methodology.....	60
3.3 Results and Discussion.....	62
3.4 Conclusion.....	69
3.5 References.....	69
Chapter 4 The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase	74
4.1 Introduction.....	74
4.2 Methodology.....	77
4.3 Results and Discussion.....	78
4.3.1 Protonated Glu763: Conf I.....	81
4.3.2 Protonated Glu763: Conf II.....	83
4.3.3 Ionized Glu763	83
4.3.4 Oxidation of <i>n</i> -butylisocyanide.....	85
4.4 Conclusion.....	90
4.5 References.....	90
Chapter 5 Mechanism of an Asymmetric Ring-Opening Reaction of Epoxide with Amine Catalyzed by a Metal-Organic Framework: Insights from Combined Quantum Mechanics and Molecular Mechanics Calculations.....	93
5.1 Introduction.....	93
5.2 Methodology.....	96
5.2.1 Model setup.....	96
5.2.2 QM/MM calculations	97
5.3 Results and Discussion.....	99
5.3.1 Stereochemistry in the ring-opening reaction of cyclohexene oxide.....	99
5.3.2 Epoxide ring-opening step	101
5.3.3 Proton-transfer step	113

5.4 Conclusion	118
5.5 References.....	119
Appendix	127
Appendix A.....	127
Appendix B.....	134
Appendix C	138
Appendix D	140
List of Publications	144

List of Figures

Chapter 1 Introduction

Figure 1. Comparison of the hierarchical regions used in (a) the ONIOM2(QM:MM) method and (b) the additive QM/MM method. QM and MM calculations are denoted by red and green colors. 9

Figure 2. Schematic illustration of several key atoms at the core–environment boundary in the ONIOM2(QM:MM) method. 11

Figure 3. A possible setup procedure for an ONIOM calculation of an enzyme–substrate complex. 25

Figure 4. Example of connectivity specification in Gaussian 09 software for the first 20 atoms in an enzyme–substrate complex. 27

Chapter 2 Mechanism Scenario in the Conversion of Androgens into Estrogens by CYP19A1

Figure 1. Energy profiles (in kcal/mol) for the first step. Key bond distances in transition states are shown in Å. 40

Figure 2. Energy profiles (in kcal/mol) for path I & II in the second step. Key bond distances in transition states are shown in Å. 43

Figure 3. Energy profile (in kcal/mol) for path III in the second step. Key bond distances in transition states are shown in Å. 45

Figure 4. Energy profiles (kcal/mol) for path III in the second step with two water molecules added. Key bond distances in transition states are shown in Å. 46

Figure 5. Energy profiles (kcal/mol) for path A-C in the third step. Key bond distances in transition states are shown in Å. 48

Figure 6. SNOs for (a) TS1a and (b) TS1b. 50

Figure 7. Energy profiles (kcal/mol) for path D-E in the third step. Key bond distances in transition states are shown in Å. 51

Chapter 3 Ferric-Superoxide Intermediate in the Verdoheme Formation Step of Heme Oxygenase

Figure 1. Geometries and schematic illustrations of the different spin states considered for **1bA** (a) and **1bB** (b). Values in parentheses are relative energies in kcal/mol, and values in square brackets are $r(\text{C-O})$ for **1bA** or $r(\text{Fe-O})$ for **1bB** in Å. 63

Figure 2. Energy diagram (in kcal/mol) for the reaction of **1bB** by DFT calculations. 65

Figure 3. (a) Energy diagram (in kcal/mol) for the reaction of **1bB** by ONIOM calculations. (b) Transition-state geometries. Key bond distances are shown in Å. 66

Figure 4. Geometries of Int_o-o and Int_o-o' derived by ONIOM calculations. 67

Figure 5. CO-coordinated **1bA** and the complex of **Int_o-o** and CO. The values underneath the structures are relative energies in kcal/mol. 69

Chapter 4 The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase

Figure 1. Reactant complexes (RCs) for ONIOM(QM:MM) calculations. (a-c) RCs for the oxidation of carbon monoxide in cases where protonated Glu763 adopts Conf I (a) and Conf II (b), and Glu763 is ionized (c). (d) RC''' for the oxidation of *n*-butylisocyanide. 80

Figure 2. Superimposition of RC and RC'. Carbon atoms in RC are colored in light sea green while carbon atoms in RC' are colored in dark gray. Key atoms in RC' are represented in ball-stick style. 81

Figure 3. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is protonated and adopts Conf I. Key bond distances are shown in Å. 82

Figure 4. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is protonated and adopts Conf II. Key bond distances are shown in Å. 84

Figure 5. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is ionized. Key bond distances are shown in Å. 84

Figure 6. Energy diagram (in kcal/mol) and transition-state geometries for the oxidation of *n*-butylisocyanide by Mo-Cu CODH. Key bond distances are shown in Å. 85

Chapter 5 Mechanism of an Asymmetric Ring-Opening Reaction of Epoxide with Amine Catalyzed by a Metal-Organic Framework: Insights from Combined Quantum Mechanics and Molecular Mechanics Calculations

Figure 1. (a) MOF catalyst **1** examined in this study.²⁰ (b) Model of MOF **1** used for QM/MM calculations (3 × 3 layer). Cyan carbon atoms were kept fixed during geometry optimization. (c) QM region for QM/MM calculations. Carbon atoms of the substrates (cyclohexene oxide and aniline) are colored magenta. 95

Figure 2. (a) Two possible steps in the reaction of cyclohexene oxide with aniline. (b) Formation of (*R,R*)- and (*S,S*)-β-amino alcohols. 100

Figure 3. Influence of the ring conformation on the stereochemistry in the ring-opening reaction of cyclohexene oxide with aniline. The resultant cyclohexane ring prefers the chair conformation. 101

Figure 4. Four possible structures of RC. Cyclohexene oxide has conformation **A** in (*S,S*)-1-RC and (*S,S*)-2-RC and conformation **B** in (*R,R*)-1-RC and (*R,R*)-2-RC. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 103

Figure 5. Energy diagram and optimized structures of TS1 and Int1 for the (*S,S*)-1 and (*R,R*)-1 pathways. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 105

Figure 6. Energy diagram and optimized structures of TS1 and Int1 for the (*R,R*)-2 pathway. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 106

Figure 7. AIM analysis on the structure of (*R,R*)-2-TS1. The carboxylate group of a linker was replaced with a hydrogen atom. Bond paths and bond critical points are denoted by black lines and light green dots, respectively. 107

Figure 8. QM/MM-optimized structures of (*R,R*)-1-TS1 and (*R,R*)-1-TS1 obtained with the B3LYP functional. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 109

Figure 9. Several RC species optimized using the extended QM region; RCext1 in which epoxide coordinates to the Cu(II) ion; RCext2 and RCext3 in which aniline coordinates to the Cu(II) ion and epoxide forms a hydrogen bond with the BINOL moiety. Relative energies shown in kcal/mol are with respect to RCext1. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. QM atoms are shown in stick. 111

Figure 10. Energy diagram and optimized structures of key species for the hydrogen bond assisted ring-opening reaction (DFT). Relative energies shown in kcal/mol are with respect to RC1. Key distances are indicated in Å. B3LYP-D3/6-311+G**//B3LYP-D3/6-31G* was used. 112

Figure 11. Energy diagram and optimized structures of key species for the hydrogen bond assisted ring-opening reaction (QM/MM). Relative energies shown in kcal/mol are with respect to RC_{BINOL}. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. QM atoms are shown in stick. 113

Figure 12. Energy diagram for the proton transfer step. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. 115

Figure13. QM/MM-optimized structures of (*S,S*)-1-Int3-2 and (*R,R*)-2-Int3-2. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 116

Figure 14. Full energy diagram for the reaction. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. 118

List of Tables

Chapter 1 Introduction

Table 1. Example of the progress of forces, displacements, and energies during geometry optimization (the final three steps) with Gaussian 09. 14

Chapter 4 The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase

Table 1. Key distances in the thiocarbonate intermediate obtained by Siegbahn *et al.* and our optimized Int03b. 87

Chapter 5 Mechanism of an Asymmetric Ring-Opening Reaction of Epoxide with Amine Catalyzed by a Metal-Organic Framework: Insights from Combined Quantum Mechanics and Molecular Mechanics Calculations

Table 1. The energy differences between the triplet state and the open-shell singlet state in RCs. Relative energies shown in kcal/mol are with respect to the triplet state in each RC. 99

Table 2. Relative QM/MM energies (in kcal/mol) of RC, TS1, and Int1 in the four possible pathways, which were evaluated by using the B3LYP-D3 or B3LYP functional for the QM part. Relative energies are measured from (*R,R*)-1-RC. 108

List of Schemes

Chapter 2 Mechanism Scenario in the Conversion of Androgens into Estrogens by CYP19A1

Scheme 1. A possible reaction sequence for the CYP19A1-catalyzed estrogen formation. The red carbon denotes C19. 33

Scheme 2. Possible reaction mechanisms for the first step (a) and the second step (b). 34

Scheme 3. Electronic reorganizations during the hydroxylation in the first step. a): the one reported by Martí *et al.*¹⁴; b): the conventional scheme.¹⁵ 35

Scheme 4. QM region for ONIOM calculations. 38

Scheme 5. Possible mechanisms for the second step when Cpd I is the reactive species. 41

Scheme 6. Proposed mechanism, path III, for the second step when Cpd 0 is the reactive species. 44

Chapter 3 Ferric-Superoxide Intermediate in the Verdoheme Formation Step of Heme Oxygenase

Scheme 1. HO-catalyzed reaction with emphasis on the second step. 59

Scheme 2. QM region for ONIOM calculations. 61

Chapter 4 The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase

Scheme 1. Reaction mechanisms for Mo-Cu CODH proposed by Hofmann *et al* (a), by Siegbahn *et al* (b) and by Kirk *et al* (c). 76

Scheme 2. Our proposed mechanism for the Mo-Cu CODH catalyzed carbon monoxide oxidation. 86

List of Abbreviations

QM/MM	Quantum mechanics and molecular mechanics
QM	Quantum mechanics
MM	Molecular mechanics
DFT	Density functional theory
LDA	Local density approximation
GGA	Generalized gradient approximation
LSDA	Local spin density approximation
UHF	Unrestricted Hartree-Fock method
AMBER	Assisted Model Building with Energy Refinement
RESP	Restricted electrostatic potential
gaff	General AMBER force field
ONIOM	Our own N-layer Integrated molecular Orbital and molecular Mechanics
EE	Electronic embedding
ME	Mechanical embedding
PE	Polarizable embedding
LA	Link atom
LAC	Link atom connection
LAH	Link atom host
SCF	Self-consistent field
vdW	van der Waals
PDB	Protein Data Bank
MD	Molecular dynamics

AIIs	Aromatase inhibitors
Cpd I	Compound I
KIE	Kinetic isotope effect
Cpd 0	Compound 0
SNO	Spin natural orbital
PCET	Proton-coupled electron transfer
HO	Heme oxygenase
Mo-Cu CODH	Mo-Cu carbon monoxide dehydrogenase
FAD	Flavin adenine dinucleotide
MOFs	Metal-organic frameworks
PDW	Paddlewheel
AROs	Asymmetric ring-opening reactions
AIM	Atoms-in-Molecules
ZPE	Zero-point energy
<i>R</i> -5,5'-H ₂ BDA	<i>R</i> -2,2'-dihydroxyl-1,1'-binaphthalene-5,5'-dicarboxylic acid
UFF	Universal force field
SBU	Secondary building units

Abstract

Computational chemistry has been proven to be increasingly helpful and thus important when investigating chemical properties such as chemical reactivity. The hybrid quantum mechanics and molecular mechanics (QM/MM) method that exploits the advantages of both quantum mechanics (QM) and molecular mechanics (MM) methods is a promising tool to gain detailed insights into complex systems such as enzymes and metal-organic frameworks (MOFs). Herein, QM/MM calculations are applied to three enzyme systems (cytochrome P450 Family 19 Subfamily A Member 1 (CYP19A1), heme oxygenase (HO), and Mo-Cu carbon monoxide dehydrogenase (MoCu CODH)) and one MOF (Cu-PDW) system, to elucidate the reaction mechanisms involved therein. For CYP19A1, the substrate was found able to act as proton source in the second step and Cpd I-driven mechanisms for the third step that well explain latest experimental observations were proposed. Ferric superoxide was found as a likely reactive species for ferric verdoheme formation in HO in the absence of a reducing equivalent and the reaction mechanism was also studied. For Mo-Cu CODH, the release of CO₂ from its thiocarbonate intermediate, which was energetically difficult in DFT calculations, was found plausible in our QM/MM calculations, implying the importance of the protein environment. For the Cu-PDW MOF, CH- π interactions between substrates and the organic ligand in the MOF were identified as a key factor that leads to the enantioselectivity in the reaction studied. The mechanistic scenarios derived could provide guidelines for the development of drugs targeting CYP19A1, the engineering of the aforementioned enzymes, and the design of more efficient Cu-PDW MOFs.

Chapter 1 Introduction

The ever-growing intellectual curiosity of chemists, both experimental and computational, to understand the chemistry of complex molecular systems such as enzymes at the quantum mechanical level has led to the formulation of the hybrid quantum mechanics and molecular mechanics (QM/MM) method.¹⁻⁴ Although the Schrödinger equation cannot be solved exactly for molecules containing many electrons, if an accurate approximation of quantum mechanics (QM) is applied, their electronic properties can be described in a reliable manner that accounts for electron correlation effects. Many QM methods have been put forward over the years, and through the application of these methods, quantum mechanical aspects of molecules have been explored and elucidated. However, in general, the more accurate a QM approximation is, the more demanding and thus more time-consuming the computation tends to be. In fact, the number of atoms that accurate QM methods can handle within a short time has not increased so rapidly as to reach that of enzymes (i.e., $\sim 10^3$) at the time of writing. Therefore, even with the fast growth of computer power achieved so far, a quantum mechanical treatment of large molecules is still a difficult challenge. The QM/MM approach exploits the advantages of QM and MM methods: the former is more accurate while the latter is computationally much less demanding. In essence, it is with this feature that the QM/MM method can describe quantum mechanical events of large molecular systems such as enzymatic reactions at practically manageable computational costs. In my thesis work, the QM/MM method was applied to several molecular systems. Therefore, the theoretical background of the QM method (especially, density functional theory (DFT)), the MM method, and the QM/MM method is briefly reviewed first.

1.1 Density Functional Theory

Density functional theory (DFT) is a quantum mechanical modeling approach widely used in the communities of physics and chemistry. It became very popular in the past several decades since it provides a good balance between accuracy and computational cost. DFT uses the electron density, a physical observable that depends on three spatial coordinates, rather than a wave function that is defined by N spin and $3N$ spatial coordinates, to determine the energy and other properties of the target system and thus becomes less computational demanding.

The history of DFT conceptually started in 1920s when the Thomas-Fermi model was established.^{5,6} Hohenberg and Kohn proved the Hohenberg-Kohn existence theorem, which states that “ $v(r)$ is (to within a constant) a unique functional of $n(r)$; since, in turn, $v(r)$ fixes H we see that the full many-partical ground state is a unique functional of $n(r)$ ”.⁷ Here $v(r)$ is the external potential and $n(r)$ stands for the electron density. They also provided the Hohenberg-Kohn variational theorem, which shows that the density obeys a variational principle similar to that in molecular orbital theory. These two theorems together with the work done by other researchers legitimately established modern DFT.⁷ However, DFT became practically useful only after the key breakthrough made by Kohn and Sham in the way of solving DFT equations. In their protocol, the density is determined by solving the problem of a fictitious non-interacting system of electrons.⁸ In the Kohn-Sham scheme, the final electronic energy is the sum of the kinetic energy of electrons (the 1st term in eq. 1.1), the nuclear-electron interaction energy (the 2nd term in eq. 1.1), the classical electron-electron repulsion energy (the 3rd term in eq. 1.1), and the exchange-correlation energy (the 4th term in eq. 1.1).

$$E[\rho] = -\frac{1}{2} \sum_i^N \langle \Psi_i | \nabla^2 | \Psi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \iint |\Psi_i(\mathbf{r}_1)|^2 \frac{1}{r_{12}} |\Psi_j(\mathbf{r}_2)|^2 d\mathbf{r}_1 d\mathbf{r}_2 - \sum_i^N \int \sum_A^{\text{nuclei}} \frac{Z_A}{r_{1A}} |\Psi_i(\mathbf{r}_1)|^2 d\mathbf{r}_1 + E_{XC}[\rho] \quad (1.1)$$

where Ψ_i are the Kohn-Sham orbitals and N is the number of electrons. Minimizing the energy by variational principle results in one-electron Kohn-Sham equations which are given as:

$$\left(-\frac{1}{2} \nabla^2 + \sum_j^N \int \frac{|\Psi_j(\mathbf{r}_2)|^2}{r_{12}} d\mathbf{r}_2 - \sum_A^{\text{nuclei}} \frac{Z_A}{r_{1A}} + V_{XC}(\mathbf{r}_1) \right) \Psi_i = \epsilon_i \Psi_i \quad (1.2)$$

where V_{XC} is the exchange-correlational potential. Since the exact form of the exchange-correlation energy is unknown except for a free electron gas, many approximation methods such as local density approximation (LDA),⁸ generalized gradient approximation (GGA),^{8,9} and hybrid methods such as B3LYP¹⁰⁻¹³ have been proposed to perform DFT calculations at various levels of accuracy. In LDA, if the electron density changes very slowly at any given point then $E_{XC}[\rho]$ is calculated as follows

$$E_{XC}[\rho] = \int \rho(r) \epsilon_{XC}(\rho(r)) dr \quad (1.3)$$

where $\epsilon_{xc}(\rho)$ is the exchange-correlation energy per electron in a homogeneous electron gas of density ρ . $\epsilon_{xc}(\rho)$ is further split into two parts, $\epsilon_x(\rho)$ and $\epsilon_c(\rho)$. The correlation part $\epsilon_c(\rho)$ could be represent by a known function $\epsilon_c^{VWN}(\rho)$ while the exchange part $\epsilon_x(\rho)$ is defined as follows

$$\epsilon_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} (\rho(r))^{1/3} \quad (1.4)$$

When spin is considered in LDA, a method called local spin density approximation (LSDA) is established. Similar to unrestricted Hartree-Fock method (UHF), LSDA allows different spatial Kohn-Sham orbitals for electrons with different spins. Obviously, LSDA improves the calculation results for open-shell species and

dissociation processes. GGA, compared to LDA, includes the gradient of the electron density and thus makes its functional gradient-corrected and further improves the calculation accuracy by a factor 3-5. This improvement is somehow believed to be the key factor leading to the popularity of DFT. Beyond GGA, the hybrid method is also very popular. For example, the exchange-correlation energy for the widely used B3LYP hybrid functional is defined as follows.

$$E_{xc}^{B3LYP} = (1 - a_0 - a_x)E_x^{LSDA} + a_0E_x^{HF} + a_xE_x^{B88} + (1 - a_0)E_c^{VWN} + a_cE_c^{LYP} \quad (1.5)$$

where E_x^{LSDA} is the LSDA exchange functional, E_x^{HF} is the Hartree-Fock exchange functional based on Kohn-Sham orbitals, E_x^{B88} is the Becke 88 exchange functional in GGA, E_c^{VWN} is the correlation functional mentioned above in LDA, and E_c^{LYP} is the Lee-Yang-Parr correlation functional. The parameters are defined as: $a_0=0.20$, $a_x=0.72$, and $a_c=0.81$.

1.2 Molecular Mechanics

Molecular mechanics (MM) is a molecular simulation method in which every atom is treated as a single particle, while bonded interactions are described as springs. This method calculates the sum of bonded and non-bonded energy terms to obtain the potential energy of the system (eq. 1.6).¹⁴

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{VDW} \quad (1.6)$$

There are parameters such as atomic charges and van der Waals radii assigned to individual atoms, while parameters like equilibrium bond lengths and bond force constants are assigned to bonds. Similar parameters are also given to angles, dihedral angles, and improper angles. These parameters are called “force field” collectively. They could be derived from quantum mechanical calculations or experimental data. Currently, there are many types of force fields developed for different purposes,

examples of which include MM3¹⁵ for the conformation analysis of organic compounds, ECEPP¹⁶ for peptide and protein simulations, and AMBER¹⁷ and CHARMM¹⁸ for molecular dynamics (MD) simulations of macromolecules. Note that different force fields have different expressions of energies.

AMBER (Assisted Model Building with Energy Refinement) is a series of force fields developed by the Kollman group for studies of biomolecules.^{17,19-21} The software package developed for MD simulations with these force fields has the same name.²² The basic energy expression employed in the AMBER force field is shown in eq. 1.7.

$$V(r) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \left(\frac{V_n}{2} \right) [1 + \cos(n\phi - \delta)] \\ + \sum_{nonbij} \left[\left(\frac{A_{ij}}{r_{ij}^{12}} \right) - \left(\frac{B_{ij}}{r_{ij}^6} \right) + \left(\frac{q_i q_j}{r_{ij}} \right) \right] \quad (1.7)$$

This is the simplest form of energy and other terms such as the polarization term and the solvation term could also be added to this equation in the AMBER software package. The latest AMBER force field is capable of treating proteins, nucleotides, carbohydrates, and lipids. However, there are always non-standard molecules and/or residues, e.g., heme and substrates in our research, not included in the AMBER force field. Thus, a parameterization step is necessary for such molecules. Charges can be obtained from restricted electrostatic potential (RESP)²³ fitting using ANTECHAMBER in the AMBER software package, whereas other necessary parameters can be taken from the general AMBER force field (gaff).²⁴ If parameters are not included in the gaff, then other approaches such as QM calculations may be adopted to obtain a complete and accurate set of force fields for these non-standard residues and/or molecules.

1.3 ONIOM

There are two major ways of evaluating the total energy of an entire molecular system in the QM/MM method, namely, the ‘additive’ and ‘subtractive’ schemes. What we utilized is the subtractive scheme called ONIOM (Our own N-layer Integrated molecular Orbital and molecular Mechanics).²⁵⁻³³ ONIOM encompasses not only ONIOM2(QM:MM), but also hybrid QM/QM (ONIOM(QM:QM’)) and other types of multilayer, multilevel calculations. Here, QM’ stands for a lower-level QM method. Thus, ONIOM is a powerful and flexible method that offers many different ways of dealing with large molecular systems such as enzymes. We herein use the two-layer QM/MM treatment (ONIOM2(QM:MM)) method, which is particularly useful in exploring enzymatic reactions computationally.

1.3.1 Energy

As in any other QM or MM method, the potential energy or (total) energy of a molecule is the most practically useful quantity in ONIOM. When applying the ONIOM method to calculate the energy, one must first define “layers” so that the entire system can be described by multilevel approximations.²⁵⁻³² In the case of ONIOM2(QM:MM), two layers are defined that are usually called “model” and “real” systems (Figure 1a). A more accurate but more expensive QM method and a less accurate but much faster MM method are used in combination. A QM method is used to describe the model system quantum mechanically, whereas an MM method is used for classical descriptions of both the real and model systems. The energy of the system is defined in ONIOM2(QM:MM) as

$$E_{\text{ONIOM}} = E_{\text{QM,model}} + E_{\text{MM,real}} - E_{\text{MM,model}} \quad (1.8)$$

This formula may be compared with the energy expression for the additive QM/MM:

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \quad (1.9)$$

One notices a formal similarity between the two energy expressions, in that they both consist of three terms. However, there is a clear conceptual distinction between the subtractive and additive schemes of QM/MM; this may be understood intuitively by comparing how hierarchical regions or systems are defined in the two schemes (see Figure 1). As shown in Figure 1a, the real system in ONIOM2(QM:MM) is not the complement of the model system in describing the entire system. As such, just applying QM and MM methods to the model and real systems, respectively, to get the first two terms in eq. 1.8, necessarily leads to double counting of the energy of the model system, which is evaluated in both the QM and MM calculations. However, just by adding the third term in eq. 1.8, the computational demand of which is generally much lower than that for the first two terms, the double-counting problem is eliminated. In other words, out of the two types of energies calculated for the model system by the QM and MM methods, the unnecessary and less important MM energy is “subtracted” here. This way of defining two layers has a great technical advantage because it does not require the preparation of a hollowed molecule (shown in green on the right-hand side in Figure 1b) that is necessary in the additive QM/MM (*vide infra*). As such, ONIOM allows for straightforward exploitation of a variety of QM and MM methods that have already been implemented in a software package. In addition, this feature allows the hybridization of two different QM methods, i.e., ONIOM2(QM:QM’), to be achieved in a straightforward manner, and more than two methods can also be combined more easily than in the additive scheme of QM/MM.

On the other hand, when applying the additive QM/MM, again the atoms in the system are prioritized first to define two regions, but here, the entire system is split into, say, “model” and “environment” systems (see Figure 1b). It is clear that the environment is essentially the complement to the model system within the entire

system, and thus using the first two terms in eq. 1.9 does not lead to overcounting of the energy contribution of the model system. However, these two terms do not account for the interaction energy between the two systems, and to compensate for this “shortage”, we must add the third term in eq. 1.9. At this point, readers may wonder how this sort of interaction between the two regions is accounted for in the ONIOM2(QM:MM) method. As far as the mechanical embedding (ME) scheme of ONIOM is concerned, this effect is included in the second term of eq. 1.8, which evaluates the energy of the real system using MM, whereas strictly speaking, when applying the electronic embedding (EE) scheme of ONIOM2(QM:MM), it is not accurate to say that only the second term accounts for the interaction between the model and environment systems. More details about the EE energy will be discussed later. Given that the model and environment systems complement each other in making up the entire system in the additive QM/MM scheme, and that different methods (QM and MM) are applied to these systems, it is no wonder that these systems are often called the “QM region” and “MM region”, respectively. It is indeed a common practice to use these terms, especially in additive QM/MM methods. However, some care needs to be exercised in the case of ONIOM2(QM:MM). The model system may safely be called the “QM region” in the same way as in the additive QM/MM. However, addressing the outer environment part as the “MM region” could be misleading because MM is applied to both the model and real systems (the second and third terms in eq. 1.8); neither of these systems is the same as the environment system or the MM region of the additive QM/MM (Figure 1). Therefore, the term “MM region” should be used with some care to mean the environment system in ONIOM with this caveat in mind.

(a) ONIOM2(QM:MM)

(b) Additive QM/MM

Figure 1. Comparison of the hierarchical regions used in (a) the ONIOM2(QM:MM) method and (b) the additive QM/MM method. QM and MM calculations are denoted by red and green colors.

1.3.2 Treatment of the boundary

The essential meanings of the layers used in ONIOM2(QM:MM) have already been explained well in Figure 1a, but in practice, one must care about another technical issue that emerges when defining a model system. Because the model system is a limited portion within the entire system, in many applications of ONIOM to enzymatic systems, a few covalent bonds must be cut to create a distinct model system, which in turn leads to the generation of unsaturated bonds. Therefore, the boundary between the model and environment systems must be treated with some care. Several different methods have been proposed to make such boundaries

chemically natural. In the standard approach adopted in ONIOM, a link atom (LA)³⁴ is placed according to

$$\mathbf{R}_{\text{LA}} = \mathbf{R}_{\text{LAC}} + g(\mathbf{R}_{\text{LAH}} - \mathbf{R}_{\text{LAC}}) \quad (1.10)$$

where LAC and LAH stand for “link atom connection” and “link atom host”, respectively (see Figure 2), and g is a scale factor. Here, the LAC–LAH bond must be cut to define a model system; consequently, the LAC atom in the model system (or the QM region) will become a dangling bond. In other words, the LAC atom will not satisfy the valency requirement unless some special amendment is made, and thus a QM calculation on the resultant model system (say, the “core” system that has not undergone the amendment of LA addition) may give a spurious result, especially in terms of the electronic distribution within the model system. To rectify this problem, a LA, which is typically a hydrogen (H-link) atom, is placed between the LAC and LAH according to eq. 1.10. The LA is now included in the model system, although not in the real system, which leads to a minor inconsistency between the number of atoms in the original, real system and the number of atoms in the ONIOM model. Nevertheless, the position of the LA is restricted by eq. 1.10, and thus this inconsistency does not lead to a redundancy of the internal degrees of freedom in the coordinates of the real system $\mathbf{R}$. Therefore, even frequency calculations can be performed meaningfully within $\mathbf{R}$.

Figure 2. Schematic illustration of several key atoms at the core–environment boundary in the ONIOM2(QM:MM) method.

1.3.3 Energy gradients

Derivatives of potential energy with respect to certain variables play indispensable roles in computational studies of molecules. In particular, geometry optimization and frequency analysis are two major types of calculation in the study of chemical reactions, and these calculations can be run efficiently if derivatives of energy with respect to the atomic coordinates of the real system $\mathbf{R}$ can be calculated analytically. In ONIOM2(QM:MM), the first derivative of the energy of the entire system with respect to $\mathbf{R}$ is simply

$$\frac{\partial E_{\text{ONIOM}}}{\partial \mathbf{R}} = \frac{\partial}{\partial \mathbf{R}} (E_{\text{QM,model}} + E_{\text{MM,real}} - E_{\text{MM,model}}) \quad (1.11)$$

This equation suggests that the $\partial E / \partial \mathbf{R}$ values can be calculated separately for the three different steps, i.e., the QM calculation on the model system “QM, model” and the MM calculations on the real system “MM, real” and the model system “MM, model”. However, for this purpose, it is practically more convenient if the derivatives of $E_{\text{QM,model}}$ and $E_{\text{MM,model}}$ can be calculated with respect to the coordinates of the model system ($\mathbf{R}_{\text{model}}$) instead of $\mathbf{R}$, because many software packages can calculate such

quantities. This is indeed possible by introducing the Jacobian ($\mathbf{J}$) for a coordinate transformation:

$$\frac{\partial E_{\text{ONIOM}}}{\partial \mathbf{R}} = \frac{\partial E_{\text{QM,model}}}{\partial \mathbf{R}_{\text{model}}} \cdot \mathbf{J} + \frac{\partial E_{\text{MM,real}}}{\partial \mathbf{R}} - \frac{\partial E_{\text{MM,model}}}{\partial \mathbf{R}_{\text{model}}} \cdot \mathbf{J} \quad (1.12)$$

The Hessian can also be calculated similarly as

$$\frac{\partial^2 E_{\text{ONIOM}}}{\partial \mathbf{R}^2} = \mathbf{J}^T \cdot \frac{\partial^2 E_{\text{QM,model}}}{\partial \mathbf{R}_{\text{model}}^2} \cdot \mathbf{J} + \frac{\partial^2 E_{\text{MM,real}}}{\partial \mathbf{R}^2} - \mathbf{J}^T \cdot \frac{\partial^2 E_{\text{MM,model}}}{\partial \mathbf{R}_{\text{model}}^2} \cdot \mathbf{J} \quad (1.13)$$

With these derivatives in hand, efficient geometry optimization and frequency calculations are possible with the ONIOM2(QM:MM) method.

1.3.4 Geometry optimization

Geometry optimization on potential energy surfaces of molecules is one of the most frequently used techniques in ONIOM³⁵ as well as pure QM and MM calculations. Technically, geometry optimization boils down to a nonlinear function minimization problem, in which the minimized target function is the potential energy and the variables are nuclear coordinates. Thus, it is made more efficient if analytical derivatives of energy with respect to the nuclear arrangements are available. In general, the optimization converges more rapidly if a vector containing the exact first derivatives ($\mathbf{g}$, i.e., negative of the forces) and a matrix containing the exact second derivatives ($\mathbf{H}$, i.e., Hessian) of energy with respect to some appropriate coordinates $\mathbf{q}$ for the molecule are available. In particular, second-order methods, which make use of the Hessian, permit a faster convergence of geometry optimization, because the energy of the system after a new step ($\Delta \mathbf{q}$) is taken can be predicted more accurately by

$$E(\mathbf{q} + \Delta \mathbf{q}) \approx E(\mathbf{q}) + \mathbf{g} \Delta \mathbf{q} + \frac{1}{2} \Delta \mathbf{q}^T \mathbf{H} \Delta \mathbf{q} \quad (1.14)$$

However, exact evaluation of the Hessian matrix elements is much more time-consuming than that of energy and gradient calculations; this is therefore often sidestepped. When optimizing a minimum-energy geometry, the Hessian is usually not

calculated exactly even once throughout the entire process; instead, an approximate Hessian matrix is prepared and updated by using the gradient information. Geometry optimization of a transition state is more challenging, and necessitates location of the geometry at a saddle point of a potential energy surface. In this case, it is necessary to use a more reliable Hessian matrix than in the case of energy minimization. In a commonly used procedure for the geometry optimization of transition states in Gaussian 09 software,³³ the Hessian is calculated exactly at the first step of geometry optimization, and it is updated without executing exact evaluation of the Hessian in subsequent steps. It should also be noted that in general, a good initial geometry must be used to ensure that the geometry optimization of a transition state is successful. To this end, a series of relaxed energy scanning calculations is usually performed before starting transition-state optimization, with a certain bond distance or bond angle used as a reaction coordinate.

As we have seen above, energies and energy gradients for the three different calculations can be computed separately in the ONIOM2(QM:MM) method. When only the energy E_{ONIOM} is required, the three calculations usually proceed from “MM, real”, to “QM, model”, and then to “MM, model”. A primitive approach to geometry optimization is therefore to do the same for gradient calculations in a sequential manner and obtain the ONIOM energy gradients for the coordinate system $\mathbf{R}$. However, the computational demands for energies and gradients are typically much lower for the MM calculations on the real system than for the QM calculations on the model system. Given these differences, sequentially evaluating the gradients for the three steps to move the geometry of the entire system one step forward is not a sensible procedure. Instead, in ONIOM2(QM:MM) calculations, geometry optimization is usually performed by employing the so-called microiteration scheme,³⁵

in which the geometry of the environment system is fully optimized at a given geometry of the model system. After achieving zero forces for the environment system, the atoms in the model system can be moved one step forward (macroiteration). These are iterated until the system reaches a stationary point on the potential energy surface. For example, in Gaussian 09 software, the progress of geometry optimization can be monitored by constantly checking the values shown in Table 1. At every geometry optimization step, Gaussian makes a judgment about whether the calculation has converged or not. As shown in the table, the atomic forces and displacements are of particular importance for the judgment. Once a geometry is obtained that has maximum force, root-mean-square (RMS) force, maximum displacement, and RMS displacement values smaller than the threshold values, the geometry optimization is regarded as converged.

Table 1. Example of the progress of forces, displacements, and energies during geometry optimization (the final three steps) with Gaussian 09.

Cycle	Item	Value	Threshold	Converged?
N-2	Maximum Force	0.000084	0.000450	YES
	RMS Force	0.000018	0.000300	YES
	Maximum Displacement	0.003625	0.001800	NO
	RMS Displacement	0.000600	0.001200	YES
	ONIOM: extrapolated energy	-2411.92895189751		
N-1	Maximum Force	0.000704	0.000450	NO
	RMS Force	0.000056	0.000300	YES
	Maximum Displacement	0.003091	0.001800	NO
	RMS Displacement	0.000591	0.001200	YES
	ONIOM: extrapolated energy	-2411.92895152356		
N	Maximum Force	0.000168	0.000450	YES
	RMS Force	0.000019	0.000300	YES
	Maximum Displacement	0.001292	0.001800	YES
	RMS Displacement	0.000267	0.001200	YES
	ONIOM: extrapolated energy	-2411.92895225989		

Despite the proven efficiency of the microiteration/macroiteration method, one sometimes encounters cases in which geometry optimization does not converge, with unwanted oscillation behavior in the energy, even after running many optimization cycles. One simple, practically useful solution to this problem may be to extract the XYZ coordinates that have the smallest “Maximum Force” value (see Table 1) from the output file and then restart a new geometry optimization calculation, because having small forces is a good indication that the geometry is close to the targeted stationary point. In addition, the current updated Hessian may not be very reliable anymore and thus may have to be reset. This is a convenient recipe for resolving the issue of slow convergence, especially in optimizing a transition-state geometry.

The so-called quadratically coupled optimization method improves the reliability of optimization steps, thereby expediting the convergence of geometry optimization.³⁶ The above-mentioned alternating microiteration/macroiteration procedure works fine in many cases. However, this method does not necessarily guarantee a well-behaved change in the Hessian during updates. In addition, the cross terms of the Hessian, which represent the coupling between the coordinates of the model and environment systems, are neglected. This omission sometimes gives rise to unpleasant oscillation behavior in the energy during geometry optimization. The quadratic coupling method amends the deficiencies of the updated Hessian, thereby allowing a more reliable geometry optimization step to be taken. In many cases, the oscillation behavior during geometry optimization can be resolved significantly by this method, and a very fast convergence can be attained.

1.3.5 Embedding schemes

There are several different ways to deal with the electrostatic interactions between the model and environment systems in QM/MM methods. Two major approaches are the mechanical embedding (ME) and electronic embedding (EE) schemes. Furthermore, a polarizable embedding (PE) version of ONIOM has also been examined recently.³⁷ ONIOM2(QM:MM)-ME is the original form of ONIOM2(QM:MM), in which the core–environment electrostatic interaction is evaluated fully classically. Fully classical evaluation of the interaction means that the electrostatic interactions are evaluated using the Coulomb formula that is included within the second term of eq. 1.15. Consequently, the total energy, $E_{\text{ONIOM-ME}}$, is affected significantly by the atomic point charges used for the model system. When conducting reaction pathway analyses, point charges of the model system may be determined for a reactant state, and this charge set is usually used throughout for all subsequent species on the reaction pathway. However, the electronic state of the model system could essentially be altered depending on its geometry, and accordingly, the point charges should differ at different geometries. Therefore, a more sophisticated procedure for ONIOM-ME is to determine charges for different states. However, even for one particular intermediate state, determining a fully consistent charge set cannot be achieved straightaway. For example, certain initial charges may be assigned to the model system, and a new charge set may be obtained for the optimized geometry. However, another run of geometry optimization with the new charge set may lead to a slightly different optimized geometry, which in turn could have somewhat different optimal point charges. In this context, the best procedure is to repeat the charge update and geometry optimization alternately until the energy does not change anymore,²⁹

although this procedure requires additional computational costs compared with the primitive ONIOM-ME calculation.

ONIOM-ME is a computationally very efficient approach for studying large molecular systems. However, it does not account for the polarization of the model system, which occurs under the perturbation exerted by the environment system, or more specifically, by the point charges placed on atoms in the environment system. By contrast, the EE version of ONIOM (ONIOM-EE) can describe this sort of polarization effect.³² In the ONIOM2(QM:MM)-EE method, the energy of the system is approximated as

$$E_{\text{ONIOM-EE}} = E_{\text{QM,model}}^{\text{V}} + E_{\text{MM,real}} - E_{\text{MM,model}}^{\text{V}} \quad (1.15)$$

which looks similar to eq. 1.15 for the ONIOM-ME method; however, the first and third terms are modified so that the effect of the polarization of the model system is included effectively. The first term is calculated using the Hamiltonian:

$$H_{\text{QM,model}}^{\text{V}} = H_{\text{QM,model}} - \sum_i \sum_{A_{\text{env}}} \frac{s_{A_{\text{env}}} q_{A_{\text{env}}}}{r_{iA_{\text{env}}}} + \sum_{A_{\text{QM}}} \sum_{A_{\text{env}}} \frac{Z_{A_{\text{QM}}} s_{A_{\text{env}}} q_{A_{\text{env}}}}{r_{A_{\text{QM}}A_{\text{env}}}} \quad (1.16)$$

Here, q is a point charge in the environment system, r is the distance between an electron in the model system and a point charge in the environment system or between a nucleus in the model system and a point charge in the environment system, and the subscripts QM and env refer to the QM (model) and environment systems, respectively. The first term on the right-hand side in eq. 1.16 is also used in the ONIOM-ME method, but here, two additional terms are used. In fact, it is due to the second term of eq. 1.16 that the electronic state of the model system undergoes polarization. For example, when a density functional theory (DFT) method is used for the “QM, model” calculation, this term is appended to the one-electron operator of the Kohn–Sham equations. Iterative self-consistent field (SCF) optimization of Kohn–Sham orbitals results in polarized electronic distributions within the model system, as

compared with the case in which this term is not included (i.e., ONIOM-ME or gas-phase calculation of the model system). One notices that a scale factor is used in the second term. The scale factor is essentially 1 (no scaling) for most of the point charges, but it could also be 0. The use of the scale factor originates from the fact that the resultant electronic state of the model system is likely to suffer from overpolarization if all of the point charges from the environment system are included in the QM calculation. Thus, if no scaling is applied, the interaction between the electrons in the model system and a nearby atom, for example, LAH in Figure 2, could be too large. In particular, the LAH and LA are very close to each other, and thus inclusion of the point charge of the LAH could affect the QM region too strongly, but not meaningfully. Therefore, some of the point charges in the environment system, typically those for atoms that are within three bonds away from the LAC, are left out by using a scale factor of 0.³⁰ The third term in eq. 1.16 accounts for the classical electrostatic interactions between the nuclei in the model system and the point charges in the environment system. For consistency, the same scale factor as in the second term is used here. We can now realize that the second and third terms in eq. 1.16 describe the electrostatic interaction between the model system and the environment system.

van der Waals (vdW) interactions constitute another important form of nonbonded interactions in ONIOM2(QM:MM). Where are the vdW interactions between the two systems calculated in the EE scheme? This is done in the second term of eq. 1.15, which evaluates all of the MM energy contributions within the entire system. Thus, the vdW term is calculated in the same way as in ONIOM-ME. However, this also means that the (coulombic) electrostatic interactions between the two systems are again calculated here, despite the fact that the electrostatic

interactions have already been calculated semiclassically in the first term of eq. 1.15. Thus, the electrostatic interactions have been calculated twice up to this point. Out of these two electrostatic contributions, we want to eliminate the somewhat less accurate coulombic electrostatic interactions. This is done during the MM calculation on the model system by the third term of eq. 1.15. More specifically, this term has the form

$$E_{\text{MM,model}}^{\text{V}} = E_{\text{MM,model}} + \sum_{A_{\text{QM}}} \sum_{A_{\text{env}}} \frac{q_{A_{\text{QM}}} s_{A_{\text{env}}} q_{A_{\text{env}}}}{r_{A_{\text{QM}}A_{\text{env}}}} \quad (1.17)$$

where the first term is the same as that used in eq. 1.8, while the second term has the role of rectifying the double counting of electrostatic interactions.

We have seen above that the electrostatic interactions in ONIOM-EE are calculated in the first term of eq. 1.15, and in so doing, the point charges for the nearby atoms in the environment system are zeroed out. It is important to note, however, that this does not mean that the contributions of these nearby atoms to the electrostatic interactions are completely neglected, because when calculating the second term in eq. 1.15, such charge elimination is not applied to these nearby atoms. In addition, these contributions are not eliminated by eq. 1.17. Therefore, the nearby atoms are still allowed to have nonzero electrostatic interactions with the assigned point charges of the model system in a classical sense through the Coulomb formula.

We have also seen above that geometry optimization can be performed efficiently by using the technique of microiteration. In an ONIOM2(QM:MM)-EE calculation, during the microiteration step, the positions of the atoms in the environment system change, and this in turn should alter the electronic state of the model system through the second term of eq. 1.16. However, determining the electronic state of the model system quantum mechanically for every different configuration of the environment system will cause a significant increase in the number of QM calculations, and thus microiteration will not accelerate the geometry

optimization as markedly as it does in ONIOM2(QM:MM)-ME. Therefore, instead of running too many QM calculations, point charges are determined and assigned to the model system at a given geometry of the model system, to approximate the core–environment interaction classically during the microiteration for the geometry optimization with the ONIOM2(QM:MM)-EE method.

1.3.6 Setup for ONIOM calculations

ONIOM2(QM:MM) is frequently used when studying enzymatic reactions, and its fundamental theories have already been explained above. Once an input file for an enzyme–substrate system is prepared, the ONIOM calculations can be carried out more or less routinely to investigate reaction mechanisms, just by following the theories described above. However, in practice, one of the most time-consuming steps in ONIOM calculations is often the prior setup of a structure model that is necessary before starting ONIOM calculations. Even though the setup process may not involve any demanding calculations, it often requires careful checking. As such, the setup is technically a very important step in ONIOM calculations and has been explained in detail in a recently published book chapter.³⁸ As mentioned there, different research groups have developed their own workflows for the setup, and there is no single perfect method. Thus, what we explain below is just one of many choices.

Typically, an initial structure of the enzyme of interest is taken from the Protein Data Bank (PDB) webpage, where a large number of protein structures are deposited that were determined mainly by X-ray diffraction techniques.³⁹ The XYZ coordinates in the downloaded PDB files sometimes have missing atoms or amino acid residues. Missing atoms may be appended by use of software such as Modeller.⁴⁰ Hydrogen atoms of proteins are not easily determined by X-ray diffraction and thus are usually

not included in PDB files. Although the positions of hydrogen atoms on carbon atoms can be straightforwardly determined on the basis of hybridization states and chemical intuitions, the positions of hydrogen atoms on titratable residues must be determined with care. For this purpose, web tools for predicting local pK_a values such as PROPKA are useful.⁴¹ Nevertheless, it is advised that visual inspection be carried out after any automated protonation of an enzyme.

The organic substrate molecule that undergoes a chemical reaction in the active site of an enzyme can be built with molecular visualization software such as GaussView.⁴² In contrast to amino acid residues for which atomic charges have already been determined and implemented in force-field libraries, substrates usually do not have predetermined charges. To fill this gap, the geometry of a substrate in its isolated state may be optimized first using a QM method, and the geometry-optimized substrate may then be subjected to an additional calculation to determine atomic charges with, e.g., the Restrained Electrostatic Potential (RESP) method.²³ The Antechamber module of AMBER software can be used to determine RESP atomic charges based on the electrostatic potential as calculated by Gaussian 09.^{22,33} As we have seen above, the atomic charges determined here will have a significant influence on the energies in the ONIOM2(QM:MM)-ME method. vdW parameters should also be given to the substrate so that the nonbonded vdW interactions between the core and environment systems can be calculated by the second term of eq. 1.8 or eq. 1.15. The vdW parameters may be taken, for example, from the existing parameter libraries for a certain version of AMBER. Parameters for bonded interactions should also be given to the substrate. However, if the entire substrate is included in the model system in ONIOM calculations, the bonded parameters used for the substrate will not affect the ONIOM energy, because the energies for bonded interactions will be calculated twice

in the two MM calculations in ONIOM2(QM:MM), and the opposite signs of these energy terms result in their mutual cancellation. Therefore, the result will remain unaffected even if some spurious bonded parameters are given to the atoms in the model system for which parameters are not available.

The geometry of the substrate in a certain pose may already be included in a PDB file. In such cases, this geometry may be used directly in the ONIOM calculation because the PDB geometry is likely to reflect the plausible configuration of the substrate during reactions. Still, this substrate geometry is not complete because hydrogen atoms are not included in the PDB file. Therefore, hydrogen atoms must be added to the substrate, e.g., by using GaussView, before starting any geometry optimization calculations for the substrate. When a crystal structure is not available for an enzyme in complex with the substrate, the geometry of the substrate must be built with some software like GaussView, as mentioned above. Importantly, however, in this case, there will be uncertainties regarding its configuration within the active site of the enzyme. Therefore, a few plausible configurations of the substrate may be produced first by performing docking simulations using, e.g., AutoDock Vina.⁴³

The resultant protein–substrate structure may be subjected either to a molecular dynamics (MD) simulation or directly to an ONIOM calculation. By determining the reaction potential energy profile in the latter approach, one can already gain valuable insight into the enzyme reactivity. However, enzymes generally have many more degrees of freedom than organic molecules, and thus using potential energies as calculated by eq. 1.8 or eq. 1.15 for assessing the relative stability of different states does not necessarily provide sufficient information about the reactivity. A more reliable quantity in such situations is free energy, which can be evaluated by taking into account a large number of possible conformations of the enzyme and substrate. To

generate many different conformational samples for free energy calculations, MD is the method of choice. However, combining MD sampling and ONIOM2(QM:MM) energy calculation could lead to prohibitively high computational demands, although calculations become feasible if appropriate approximations are introduced to reduce the cost.⁴⁴ Therefore, in many cases, MD simulation is used to produce a handful of “snapshots” of different enzyme-substrate geometries. Potential energy calculations may be performed for each of the selected snapshots, to see how the shapes of the reaction potential energy surfaces can differ for different snapshots. Such calculations, which involve additional MD simulations and ONIOM calculations on different snapshots, are expected to provide more reliable enzyme–substrate geometries than the direct ONIOM calculation using a crystal structure, and a more reliable picture of the reaction pathway may be obtained. However, it should be kept in mind that here, MD simulation is performed with MM force fields that have rather limited accuracy, which means that MD simulation might not always improve the enzyme–substrate geometry compared with the original crystal structure. Thus, there is no guarantee that the snapshots reflect likely enzymatic configurations. In addition, a number of snapshots in the order of 10^0 or 10^1 is not at all sufficient for obtaining statistically meaningful interpretations. Therefore, even if very different barrier heights are obtained for two different snapshots, one cannot immediately generalize from this result to argue, for instance, that the system will encounter these distinct situations frequently. Thus, these calculations should be carried out with the inherent limitations in mind. A possible setup procedure is summarized in Figure 3, which one may follow when running ONIOM calculations of an enzyme–substrate complex.

Thus, ONIOM2(QM:MM) calculations are performed, in many cases, to calculate potential energy surfaces of enzymes. Although this procedure is regarded as

a practically reasonable compromise, one must be aware of one problematic consequence that could be observed during determination of reaction pathways. When using potential energy for the analysis of chemical reactions, it is important to make sure that this potential energy is smoothly connected along the reaction coordinate. One typical case in which this necessary requirement is violated is encountered when the conformation of an amino acid residue in the environment system changes significantly at some point on the reaction pathway. This kind of conformational change could result in a decrease in the total energy (i.e., E_{ONIOM}) by, say, 10 kcal/mol, but this energy decrease should not be regarded as being associated with the exothermicity of the reaction. For example, this sort of conformational change could make a transition state even lower in energy than the prior intermediate state, but this energy diagram does not allow one to make a meaningful interpretation of the enzyme reactivity. When such an energy drop has happened, the energy of the prior intermediate should be newly determined. If the combination of the second and third terms of eq. 1.8 ($E_{\text{MM,real}} - E_{\text{MM,model}}$), which is also called the S_{low} value,³² changes significantly on a reaction pathway, it is possibly a telltale sign of this kind of unwanted conformational change in the environment system.

Figure 3. A possible setup procedure for an ONIOM calculation of an enzyme–substrate complex.

When starting ONIOM calculations, one must make sure that the system's "connectivity" is defined correctly for all atoms. In pure QM calculations, the connectivity does not have to be defined, because such information is not used when solving the Kohn–Sham equations or any other QM equations. However, in ONIOM2(QM:MM) calculations, one must declare clearly how atoms in the system are covalently linked. The reason is that such connectivity information is necessary when applying an MM force-field method for the second and third terms of eq. 1.8 or eq. 1.15. An example is given in Figure 4, which describes the connectivity for the first few atoms in an enzyme–substrate complex in Gaussian 09 format. The first line states that the first atom is connected to the third, fourth, second, and fifth atoms. The number "1.0" denotes the bond order, but as long as an AMBER-type force field is used, the bond order does not affect the resultant energy because the energy expression does not contain any contributions from bond order. Again, GaussView helps determine the connectivity of large enzymes. The second to fourth lines just mention the atom ID for a single atom, but this does not necessarily mean that these atoms are not connected to any atom. In this particular case, these atoms are connected to the first atom, as described in the first line. In other words, the connectivity specification can be done only once.

```

1 3 1.0 4 1.0 2 1.0 5 1.0
2
3
4
5 6 1.0 7 1.0 12 1.0
6
7 10 1.0 8 1.0 9 1.0
8
9
10 11 1.0
11
12 13 2.0 14 1.5
13
14 15 1.0 16 1.0

```

```

15
16 17 1.0 23 1.0 18 1.0
17
18 19 1.0 20 1.0 21 1.0
19
20

```

Figure 4. Example of connectivity specification in Gaussian 09 software for the first 20 atoms in an enzyme–substrate complex.

1.3.7 Preparation of a decent initial orbital guess for the model system

Especially when a transition metal such as iron is included in the model system, the QM calculation on the model system, namely, the first term in eq. 1.8 or eq. 1.15, does not always converge to the desired electronic state. This is particularly true when a DFT method is used in a broken-symmetry fashion to describe the electronic state of the model system. When the calculation has converged to an undesired electronic state, a conventional approach for rectifying the problem is to alter the electronic occupancy of orbitals. When a Kohn–Sham DFT method is used, a single determinant of occupied orbitals (Kohn–Sham orbitals) is used to describe the noninteracting electrons. To change the electronic occupancy of Kohn–Sham orbitals, one must first identify those orbitals that are supposed to be occupied and unoccupied, by using graphics software such as GaussView. Subsequently, the ordering of these orbitals is altered so that some of the initially unoccupied and occupied orbitals are made occupied and unoccupied, respectively. To prepare a decent initial orbital guess for ONIOM calculations, one reasonable procedure may be to extract only the model system from the entire system and perform pure DFT calculations to get a desired orbital occupancy, because for the orbital manipulation, thousands of MM atoms could be distracting. Once a desired orbital occupancy is achieved, the orbital data may be printed as a text file by using the “punch=mo” keyword of Gaussian 09. The text-format orbital guess may be read

by the “guess=cards” keyword when performing an ONIOM calculation, so that the specified initial guess is read in the QM calculation step on the model system. A more convenient approach that has become available recently is the fragment guess option of Gaussian 09, which allows intuitive constructions of initial orbitals and is thus highly useful.

1.4 References

1. Honig, B.; Karplus, M. *Nature*, **1971**, 229, 558-560.
2. Warshel, A.; Levitt, M. *J. Mol. Biol.*, **1976**, 103, 227-249.
3. Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.*, **2009**, 48, 1198-1229.
4. van der Kamp, M. W.; Mulholland, A. J. *Biochemistry*, **2013**, 52, 2708-2728.
5. Thomas, L. H. *Math. Proc. Cambridge Phil. Soc.* **1927**, 23, 542-548.
6. Fermi, E. *Rend. Accad. Naz. Lincei.* **1927**, 6, 602-607.
7. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136, B864-B871.
8. Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, 140, A1133-A1138.
9. Rasolt, M.; Davis, H. L. *Phys. Lett.* **1981**, 86A, 45-47.
10. Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D. *Curr. Pharm. Des.* **2014**, 20, 3281-3292.
11. Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648-5652.
12. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785-789.
13. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 1200-1211.
14. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623-11627.
15. Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, 111, 8551-8566.

16. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361-2381.
17. Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077-13082.
18. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187-217.
19. Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
20. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
21. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955-1970.
22. Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M. and Kollman, P. A. *AMBER 2015*, University of California, San Francisco, CA, 2015.
23. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
24. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
25. Maseras, E.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.
26. Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959-1967.

27. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.
28. Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1-21.
29. Tao, P.; Fisher, J. F.; Shi, Q.; Vreven, T.; Mobashery, S.; Schlegel, H. B. *Biochemistry* **2009**, *48*, 9839-9847.
30. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery Jr., J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815-826.
31. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. *WIREs Comput. Mol. Sci.* **2012**, *2*, 327-350.
32. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. -B.; Ding, L.; Morokuma, K. *Chem. Rev.* **2015**, *115*, 5678-5790.
33. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.;

- Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. *J. Gaussian 09 Revision E.01*, Gaussian, Inc., Wallingford CT, 2009.
34. Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718-730.
 35. Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760-769.
 36. Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701-704.
 37. Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2014**, *10*, 1588-1598.
 38. Foresman J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 3rd ed., Gaussian Inc. Pittsburgh, PA, 2015.
 39. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
 40. Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779-815.
 41. Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704-721.
 42. Dennington, R.; Keith, T.; Millam, J. *GaussView, Version 5*, Semichem Inc., Shawnee Mission, KS, 2009.
 43. Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455-461.
 44. Kawatsu, T.; Lundberg, M.; Morokuma, K. *J. Chem. Theory Comput.* **2011**, *7*, 390-401.

Chapter 2 Mechanism Scenario in the Conversion of Androgens into Estrogens by CYP19A1

2.1 Introduction

CYP19A1, also called aromatase, is a member of the cytochrome P450 enzyme family, which contains a heme cofactor. CYP19A1 converts androgens into estrogens via a sequential, three-step reaction utilizing one molecular oxygen and two electrons from NADPH in each step (Scheme 1).¹⁻³ This conversion is the last and pivotal step in the biosynthesis of estrogens.⁴ Estrogens, though they are the primary female sex hormone, exist in both men and women and involve in multiple physiologic functions.⁵ For example, estrogens not only promote the development of the female reproductive system and secondary sexual characteristics and also involve in their regulation. Another example is that estrogens also help both men and women to maintain their bone mass.⁵⁻⁶ Beyond their physiological functions, estrogens could also affect the progress of several diseases, viz. hormone-sensitive breast cancer.⁷⁻⁸ Thus the inhibition of CYP19A1 becomes a very promising therapeutic strategy for the treatment of breast cancer and other estrogen-related diseases. Currently several inhibitors of CYP19A1, well known as aromatase inhibitors (AIs), have already been approved by US FDA for the treatment of breast cancer and ovarian cancer.

Scheme 1. A possible reaction sequence for the CYP19A1-catalyzed estrogen formation. The red carbon denotes C19.

The mechanism of this conversion might provide useful information for the development of novel AIs; thus it should be understood clearly. For example, Martí *et al.*⁹ utilized QM/MM calculations to examine whether or not the inhibition of CYP19A1 by exemestane, a third-generation aromatase inhibitor, takes place in the first step. The first step is believed to occur through hydroxylation on the 19-methyl group, which consists of the initial hydrogen abstraction from C-19 to form an alkyl radical intermediate and the rebound of the newly produced hydroxyl radical to the alkyl radical to generate 19-hydroxymethyl product (Scheme 2a).^{2,10-11} In this step, compound I (Cpd I) is widely accepted as the reactive species.¹²⁻¹³ Martí *et al.*¹⁴ have recently reported a QM/MM study of the first step of the CYP19A1 reaction triggered by Cpd I. In their results, when the hydrogen abstraction occurs, its electron rearrangement occurs in such a way that the beta electron from the substrate is shifted into the a_{2u} orbital in the doublet state or into the π^*_{xz} orbital in the quartet state. However, as shown in Scheme 3, there is a widely accepted electron rearrangement

scheme,¹⁵ in which the alpha electron of the substrate in the doublet state and the beta electron of the substrate in the quartet state moves into the a_{2u} orbital, respectively. More surprisingly, in the rebound process in the quartet state, the electronic rearrangement proposed by Martí *et al.*¹⁴ involves a two-electron transfer process which is absent in the conventional scheme.¹⁵ We, therefore, would like to investigate the possibility of the conventional electronic rearrangement in the first step.

Scheme 2. Possible reaction mechanisms for the first step (a) and the second step (b).

Scheme 3. Electronic reorganizations in the first step. a): the one reported by Martí *et al.*¹⁴; b): the conventional scheme.¹⁵

The reactive species in the second step is also believed to be Cpd I.¹²⁻¹³ There are three very interesting experimental findings for this step. The first observation is that the 19-*pro-R* hydrogen in the 19-hydroxymethyl analogs of the androgens is selectively removed in the second step.¹⁶⁻¹⁷ The second finding is that there is no observed kinetic isotope effect (KIE) for this 19-*pro-R* hydrogen.¹⁷ The third one shows that the O₂ consumed in the second step is not incorporated into the resulting aldehyde product of the second step.¹⁸ If this step follows a typical hydroxylation mechanism for P450s, its mechanism would be the one shown in Scheme 2b. The reaction is initiated by a hydrogen abstraction from the 19-hydroxymethyl group in the substrate by Cpd I, followed by a rebound process of the generated hydroxyl radical. The first and third findings suggest that either the produced *gem*-diol intermediate in Scheme 2b undergoes a selective dehydration, which only removes the newly formed

hydroxyl group, or the 19-*pro-R* hydrogen is removed without a rebound process, which is obviously quite different from the mechanism shown in Scheme 2b. The second finding indicates that the remove of the 19-*pro-R* hydrogen should not be the rate-limiting process in the second step. After checking the crystal structures of human placental CYP19A1¹⁹⁻²⁰, no potential proton donor or acceptor is found close to the 19-methyl group in the substrate, which makes a selective dehydration at the active site not that possible. Therefore, an alternative mechanism, which differs from a conventional hydroxylation by P450s and could explain this selectivity, is highly demanding. Thus, we would like to investigate mechanistic scenario of the second step.

In contrast to the first and second steps, the nature of the active species in the third step has been under debates for a long time. Both Cpd I and ferric superoxide have been proposed as likely candidates. ¹⁸O-labeling experiments performed by Akhtar *et al.*²¹ and Caspi *et al.*²² show that one oxygen atom from O₂ is incorporated into the formic acid produced in the third step. This implies the involvement of ferric superoxide as the active species and theoretical studies about the ferric-peroxide-driven mechanism have already been done²³. However, recently three groups reported new experimental results utilizing resonance Raman spectroscopy²⁴, kinetic solvent isotope effect²⁵ and ¹⁸O-labeling experiments²⁶. All these results support Cpd I as the active species. More importantly, this new ¹⁸O-labeling experiment using more sensitive techniques and instruments indicates that no oxygen atom from ¹⁸O₂ is incorporated into the produced formic acid and thus overrules the ferric-peroxide-driven mechanism and makes Cpd I the only reasonable reactive species for this step. We, therefore, would like to elucidate mechanistic details about the Cpd I-driven mechanism in the third step.

2.2 Methodology

2.2.1 Protein setup and MM parameters

The initial structure used in our QM/MM study started from the X-ray crystal structure of human aromatase reported by Ghosh *et al.*¹⁹ (PDB code 3EQM, 2.90 Å resolution). Two phosphate ions included in this crystal structure were removed. All aspartic acid residues except Asp309 and all glutamic acid residues were assumed ionized, while all lysine and arginine residues were assumed protonated. The protonation states of histidine residues were decided as follows: HIE62, HIE105, HIE109, HIP111, HID128, HIP171, HID325, HID402, HID459, HIP475, and HIE480. HIE represents the histidine residue that is protonated at its ϵ position while HID stands for the one protonated at its δ position. Both ϵ position and δ position are protonated in HIP. The methyl group at the C19 position of androstenedione in 3EQM was modified to a hydroxymethyl group for the second step and to a *gem*-diol group and an aldehyde group for the third step. Hydrogen atoms were added by the AMBER10 software package²⁷. MM parameters for standard amino acid residues were taken from the ff03.r1 force field²⁸ of AMBER. The MM parameters for heme were those reported by Shahrock *et al.*²⁹ while MM parameters for substrates were taken from the general AMBER force field (gaff)³⁰ with the atomic charges determined at the B3LYP/6-31G* level using the restrained electrostatic potential (RESP) method³¹.

2.2.2 ONIOM calculations

The ONIOM(QM:MM) method implemented in the Gaussian 09 software package³² was used for all QM/MM calculations. The QM region for QM/MM calculations was defined as illustrated in Scheme 4. It includes Cpd I (porphine with HS⁻ ligand), side chains of THR310 and ASH309, and the substrate (two rings out of

the four rings). Hydrogen link atoms were utilized to make the boundary chemically natural during the ONIOM calculations. The total charge of the QM region is 0.

Scheme 4. QM region for ONIOM calculations.

A five-step MM geometry optimization was performed before the ONIOM calculations to refine the initial structure for the complex of CYP19A1 and the *gem*-diol substrate in the third step. First, all hydrogen atoms were relaxed while the rest atoms were fixed. Then an energy minimization was performed with all backbone atoms fixed, followed by an energy minimization with the QM heavy atoms fixed. Finally, the initial geometry for subsequent ONIOM calculations was obtained after a full MM optimization. By modifying the structure of the 19-*gem*-diol substrate in the ONIOM-optimized geometry for CYP19A1 in complex with the 19-*gem*-diol substrate, initial geometries of 19-methyl, 19-hydroxymethyl and 19-aldehyde substrates in complex with CYP19A1 were obtained.

In the ONIOM calculations, only residues within 10 Å away from the Fe ion were allowed to move. The AMBER03 force field (GAFF was used for substrates) was used in ONIOM calculations by the “softfirst” function.^{28,30} Geometry optimizations and frequency calculations were performed at the ONIOM(B3LYP/[SDD(Fe),6-

31G*(rest)]:AMBER03)³³⁻³⁶ level of theory with the mechanical embedding scheme. Single-point energy calculations were done at the ONIOM(B3LYP/6-311+G(d,p):AMBER03) level of theory with the electronic embedding scheme on the optimized geometries. The reported energy data were obtained from these single-point energy calculations with zero-point energy (ZPE) corrections.

2.3 Results and Discussion

2.3.1 The first step

Only one pathway was considered with its electronic reorganization following Scheme 3b in both doublet and quartet states. Figure 1 shows the reaction energy profiles and optimized geometries of transition states. The initial step is H-abstraction from the C-H bond on C-19 in the 19-methyl substrate by the oxo group in Cpd I via TS1d and TS1q for the doublet state and quartet state respectively, which leads to Cpd II and a radical intermediate (Int1d and Int1q, respectively). The energy barriers for this H-abstraction are not high (13.8 kcal/mol in the doublet state and 16.2 kcal/mol in the quartet state). Similar to the reported hydroxylation processes by P450s, the rate-limiting step in this hydroxylation is the C-H bond activation. In the doublet state, H-abstraction is followed by a barrierless rebound process resulting in a 19-hydroxymethyl product while in the quartet state, there is a very small barrier (0.5 kcal/mol) for this rebound process. All these results indicate that the conventional electronic rearrangement mechanism¹⁵ is plausible and is consistent with previously reported theoretical results.

Figure 1. Energy profiles (in kcal/mol) for the first step. Key bond distances in transition states are shown in Å.

2.3.2 The second step

As shown in Scheme 5, two pathways, path I and path II, were taken into consideration for the second step when Cpd I is the reactive species. Path I starts from H-abstraction of the 19-*pro-R* hydrogen by the oxo group in Cpd I, leading to a radical intermediate and Cpd II. Then a rebound process leads to a 19-*gem*-diol product. As mentioned previously, there is no potential proton source or proton acceptor near this newly generated 19-*pro-R* hydroxyl group in the 19-*gem*-diol product; thus how it could be selectively removed is unclear and not considered here. Another possibility that the radical intermediate undergoes one more H-abstraction, instead of a rebound process, from the O-H bond on C-19 to produce the desired 19-aldehyde product was tried but calculations led to the 19-*gem*-diol product only. In path II, the first step is H-

abstraction from the O-H bond in the 19-hydroxymethyl group by Cpd I, followed by one more H-abstraction that abstracts the 19-*pro-R* hydrogen atom to generate the desired aldehyde product. Unlike path I, path II does not need a further selective dehydration process and thus seems more reasonable and likely. And if H-abstraction from the O-H bond in the 19-hydroxymethyl group has a higher energy barrier than H-abstraction of the 19-*pro-R* hydrogen, it would further explain the absence of KIE for the 19-*pro-R* hydrogen.

Scheme 5. Possible mechanisms for the second step when Cpd I is the reactive species.

According to the reaction energy profiles shown in Figure 2, path II is favored over path I by 1.5 kcal/mol in energy (TS02II vs TS01I). One thing worth noting is that, in path I, when the 19-*pro-R* hydrogen in RCII becomes increasingly closer to the oxo oxygen in Cpd I, the steric hindrance between Thr310 and the substrate induces conformational changes. TS01I, therefore, connects RCI, instead of RCII, and Int01I. RCI is 5.3 kcal/mol higher than RCII in energy. In RCI, the 19-hydroxyl group on C-

19 acts as a hydrogen bond donor and the hydroxyl oxygen in Thr310 is the hydrogen bond acceptor while their roles are switched in RCII. Due to the fact that the remove of the 19-*pro-R* hydrogen is the rate-limiting step in path II, Path II could only explain two of the experimental observations, the selective remove of the 19-*pro-R* hydrogen and the retention of the 19-hydroxyl oxygen in the product. KIE for the 19-*pro-R* hydrogen in path II was calculated using the frequency data according to the semi-classical Eyring equation and the equation with the Wigner quantum correction.³⁷⁻³⁸ The values obtained are 6.02/7.61 (Eyring/Wigner) and it obviously contradicts the experimentally observed absence of KIE for the 19-*pro-R* hydrogen. This might be justified by the assumption that the rate-limiting process in the second step is not the reaction itself but other processes, such as the formation of Cpd I and the release of the product. However, KIE was observed for the 19-methyl hydrogen in the first step, indicating that the reaction itself is at least partially rate-limiting in the first step.¹⁷ Considering the structural similarity of the 19-methyl substrate and the 19-hydroxymethyl substrate, we believe that the reaction itself should also be at least partially rate-limiting in the second step. Thus, both Path I and II fail to fit all the experimental observations.

Figure 2. Energy profiles (in kcal/mol) for path I & II in the second step. Key bond distances in transition states are shown in Å.

It is well known that two protons are consumed during the formation of Cpd I in P450 enzymes and this proton transfer typically involves a threonine residue near the active site,³⁹⁻⁴¹ which in CYP19A1 should be Thr310. Since the hydroxyl group of the 19-hydroxymethyl substrate rests close to the active site and Thr310, thus it is highly possible that there is a competing pathway in which the proton transfer occurs with the assistance of this hydroxyl group in the 19-hydroxymethyl substrate. A new reaction mechanism, path III shown in Scheme 6, was proposed. In this pathway, Cpd 0 is the reactive species and the reaction is initiated by a homolytic cleavage of the O-O bond to form a hydroxyl radical. Then this hydroxyl radical abstracts a proton from the hydroxyl group in the substrate to generate Cpd I and a substrate anion. Finally, this anion is converted to the desired aldehyde product after losing one hydride to Cpd I.

Scheme 6. Proposed mechanism, path III, for the second step when Cpd 0 is the reactive species.

As shown in Figure 3, the O-O homolytic cleavage has a barrier of 21.3 kcal/mol in energy while the proton transfer from the substrate to the newly generated hydroxyl radical has a higher barrier (24.2 kcal/mol in energy). Interestingly, after the proton transfer, instead of a third step for the abstraction of the 19-*pro-R* hydride by Cpd I as suggested in Scheme 6, the aldehyde product is formed and the 19-*pro-R* hydride is abstracted without any barrier. The 19-*pro-R* hydrogen is not involved in the rate-limiting step and thus could explain the absence of KIE for the 19-*pro-R* hydrogen. Moreover, in this mechanism, the 19-*pro-R* hydrogen is selectively removed and no O₂ oxygen is incorporated into the aldehyde product. Path III, therefore, is consistent with all the three aforementioned experimental observations. In addition, contrary to the typical Cpd I formation mechanism³⁹⁻⁴¹, the second proton in path III comes from the substrate and does not involve any water molecule, and thus may explain the kinetic

solvent isotope effect results reported by Sligar *et al.*²⁵, in which the second step has a smaller kinetic solvent isotope effect (1.2) than the first (2.7) and third (~2.5) steps.

Figure 3. Energy profile (in kcal/mol) for path III in the second step. Key bond distances in transition states are shown in Å.

However, there is still one question remained unsolved for path III, which is whether this pathway is preferred over the typical Cpd I formation pathway that involves threonine. In current system, there is no hydrogen bond network that could transfer protons from protonated Asp309, the proton source, to Thr310. Based on the distance between Asp309 and Thr310, two water molecules were added to the system to connect them. Reaction energy profiles and optimized geometries of transition states for these pathways are shown in Figure 4. From the energy profiles, TS02III'_a, in which the second proton comes from the substrate, and TS02III'_a, a transition state involved in the typical proton transfer pathway, have comparable energy barrier heights. Interestingly, path III is very exothermic while Cpd 0 and Cpd I were reported

to be very close in energy³⁹⁻⁴¹. Thus, path III is favored over the typical Cpd I formation pathway thermodynamically.

In summary, three different pathways were investigated for the second step. Both path II and III give the desired aldehyde product. However, in path II, the remove of 19-*pro-R* hydrogen is the rate-limiting process, which contradicts the experimentally observed absence of KIE for the 19-*pro-R* hydrogen. Path III that fits all the experimental observations, therefore, becomes the only plausible pathway.

Figure 4. Energy profiles (kcal/mol) for path III in the second step with two water molecules added. Key bond distances in transition states are shown in Å.

2.3.3 The third step

For the third step, both 19-*gem*-diol and 19-aldehyde substrates were considered. For 19-*gem*-diol substrate, three pathways (path A-C) were considered. The initial step in these pathways are H-abstraction from the O-H bond on C-19 (path A), H-abstraction from the C-H bond on 19-carbon (path B), and H-abstraction from the 1β-carbon of the cyclohexenone ring (path C). Reaction energy profiles and optimized

geometries of transition states obtained for these pathways are shown in Figure 5. Compared to path C, path A and B are seen to have low energy barriers (10.7 and 10.3 kcal/mol, respectively) for the first step. Thus, both H-abstraction from the O-H bond via TS1a and from the C-H bond via TS1b should be plausible in the first step in the reaction. In path A, H-abstraction is followed by a barrierless, homolytic C-C bond cleavage to generate an intermediate (Int1a) that includes HCO₂H, a substrate radical, and Cpd II. As a result of one electron transfer from prophyrin to Fe and a proton transfer from formic acid to Cpd II, a somewhat more stable intermediate (Int1a') containing a porphyrin- π -cation-radical-ligated Fe(III)-H₂O species and HCO₂⁻ can be formed. The H-abstraction from the C-H bond on the 1 β -carbon via TS2a establishes a new C-C double bond within the 6-membered steroid-A ring of the substrate (Int2a). The proton, which was abstracted by Cpd II from HCOOH temporarily, is donated back to HCOO⁻ in Int2a. The resulting cyclodienone moiety in Int2a can easily undergo keto-enol tautomerization to form phenol. This is supported by DFT calculations, which is done with two water molecules in the vicinity of the carbonyl group. (Figure 1S in Appendix A) Therefore, it is probable that the produced cyclodienone species may be released first from the active site, and that a phenol-type product may be formed through keto-enol tautomerization in a water-rich environment outside the enzyme.

Interestingly, after the H-abstraction from the C-H bond via TS1b in path B, barrierless H-abstraction from the one of the two O-H bonds on 19-C leads to the formation of a molecule bearing a carboxylic acid group (Int2b). The relatively low energy barrier for path B suggests that this route is one possible channel leading to a carboxylic acid product. This fits the observation by Yoshimoto and Guengerich that a

carboxylic acid as another product of the aromatase-catalyzed reaction of 19-oxo-androstenedione.

In path C, the initial H-abstraction from the 1 β -carbon on the cyclohexenone ring via TS1c has a quite high barrier (17.3 kcal/mol). Beyond this H-abstraction, a facile radical rebound step leads to an intermediate containing a hydroxylated cyclohexenone species. H-abstraction from the 1 β -carbon is often postulated as a viable mechanism for the third step effected by Cpd I.³⁴⁻³⁶ However, according to our comparative analyses, path C is apparently a less favorable pathway. The reason for the higher barrier for path C should be the steric restriction imposed on the substrate within the active-site pocket.

Figure 5. Energy profiles (kcal/mol) for path A-C in the third step. Key bond distances in transition states are shown in Å.

Generally speaking, O-H bonds of alcohols are stronger than their C-H bonds.³⁷ Contrary to this, our calculations predicted that TS1a and TS1b are almost equally stable. To gain insight into this somewhat counterintuitive outcome, an additional spin natural orbital (SNO) analysis³⁸ was performed. As shown in Figure 6, the SNO for TS1a is not localized on the breaking O-H bond but has a lone-pair distribution perpendicular to the breaking bond. This is a typical orbital signature of proton-coupled electron transfer (PCET), and the PCET effect could stabilize transition state in the reactions of oxoiron(IV) species.³⁸⁻⁴⁰ By contrast, the SNO for TS1b is firmly localized on the forming and breaking bonds; thus H-abstraction via TS1b is characterized as hydrogen atom transfer. The trend observed here is also reminiscent of the experiment results reported by Wang *et al.*, who showed that a synthetic diiron(IV) complex cleaved the O-H bonds in alcohols, rather than the weaker C-H bonds.⁴¹ In addition, low energy barriers for H-abstraction from an O-H bond were obtained computationally for other reactions of P450s.⁴²⁻⁴⁴

Figure 6. SNOs for (a) TS1a and (b) TS1b.

Two pathways (path D and E) were considered for the reaction of the aldehyde substrate. The initial step is H-abstraction from 19-carbon via TS1d and 1 β -carbon on the cyclohexenone ring via TS1e, respectively. As shown in Figure 6, path D has a very low energy barrier (6.0 kcal/mol) for H-abstraction while path E has a relatively high barrier (20.5 kcal/mol). The carboxylic acid product that produced in path B could also be obtained in path D through a barrierless rebound step; thus provides an alternative channel for the formation of a carboxylic acid product. This mechanism explains the experimental results reported by Yoshimoto and Guengerich that the amount of the carboxylic acid product significantly decreased when the aldehyde hydrogen on 19-carbon was replaced by deuterium. Further steps in path E were not examined due to the high barrier.

As mentioned above, five pathways were investigated for two substrates in the third step. Path A, which starts from a *gem*-diol substrate and leads to a cyclodienone species that could be further converted into desired estrogen, is identified as the sole

low-energy-barrier channel to the aromatized product. The absence of O₂-derived oxygen in the produced HCOOH in Int2a is consistent with the ¹⁸O-labeling experiment by Yoshimoto and Guengerich and thus further supports our argument.

The carboxylic product observed by Yoshimoto and Guengerich could be formed either through path B from the *gem*-diol substrate or path D from the aldehyde substrate. Though these two pathways result in the same product, the source of the oxygen in the produced carboxylic acid group is quite different. In path D, one of the oxygen atom in the carboxylic acid group is O₂-derived oxygen while in path B, none of the two oxygen atoms is. Interestingly, these two different types of the carboxylic acid product also somehow supports ¹⁸O-labeling experiment by Yoshimoto and Guengerich, in which they found two carboxylic acid products, one containing ¹⁸O while the other without ¹⁸O, and the ratio of the amount of the one containing ¹⁸O to the amount of that without ¹⁸O was 3:1. Thus, combining experimental data and computational data, path D should be more favorable than path B.

Figure 7. Energy profiles (kcal/mol) for path D-E in the third step. Key bond distances in transition states are shown in Å.

2.4 References

1. Thompson, E. A.; Siiteri, P. K. *J. Biol. Chem.* **1974**, *249*, 5364-5372.
2. Cole, P. A.; Robinson, C. H. *J. Med. Chem.* **1990**, *33*, 2933-2942.
3. Simpson, E. R.; Mahendroo, M. S.; Means, G. D.; Kilgore, M. W.; Hinshelwood, M. M.; Graham-Lorence, S.; Amarneh, B.; Ito, Y.; Fisher, C. R.; Michael, M. D. *Endor. Rev.* **1994**, *15*, 342-355.
4. Bulun, S. E.; Lin, Z.; Imir, G.; Amin, S.; Demura, M.; Yilmaz, B.; Martin, R.; Utsunomiya, H.; Thung, S.; Gurates, B.; Tamura, M.; Langoi, D.; Deb, S. *Pharmacol. Rev.* **2005**, *57*, 359-383.
5. Nelson, L. R.; Bulun, S. E. *J. Am. Acad. Dermatol.* **2001**, *45*, S116-S124.
6. Morishima, A.; Grumbach, M. M.; Simpson, E. R.; Fisher, C.; Qin, K. *J. Clin. Endocrinol. Metab.* **1995**, *80*, 3689-3698.
7. Miller, W. R. *J. Steroid Biochem. Mol. Biol.* **1991**, *39*, 783-790.
8. Bulun, S. E.; Price, T. M.; Mahendroo, M. S.; Aitken, J.; Simpson, E. R. *J. Clin. Endocrinol. Metab.* **1993**, *77*, 1622-1628.
9. Viciano, I.; Martí, S. *J. Phys. Chem. B* **2016**, *120*, 3331-3343.
10. Meyer, A. S. *Biochim. Biophys. Acta* **1955**, *17*, 441-442.
11. Meyer, A. S. *Experientia* **1995**, *11*, 99-102.
12. Groves, J. T. *J. Inorg. Biochem.* **2006**, *100*, 434-447.
13. Rittle, J.; Green, M. T. *Science* **2010**, *330*, 933-937.
14. Viciano, I.; Castillo, R.; Martí, S. *J. Comput. Chem.* **2015**, *36*, 1736-1747.
15. Hirao, H.; Kumar, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2005**, *127*, 13007-13018.
16. Osawa, Y.; Shibata, K.; Rohrer, D.; Weeks, C.; Duax, W. L. *J. Am. Chem. Soc.* **1975**, *97*, 4400-4402.

17. Arigoni, D.; Battaglia, R.; Akhtar, M.; Smith, T. *J. Chem. Soc., Chem. Commun.* **1975**, 185-186.
18. Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. *J. Chem. Soc., Chem. Comm.* **1981**, 129-130.
19. Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. *Nature* **2009**, *457*, 219-223.
20. Ghosh, D.; Lo, J.; Morton, D.; Valette, D.; Xi, J.; Griswold, J.; Hubbell, S.; Edbuta, C.; Jiang, W.; An, J.; Davies, H. M. *J. Med. Chem.* **2012**, *55*, 8464-8476.
21. Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. *Biochem. J.* **1982**, *201*, 569-580.
22. Caspi, E.; Wicha, J.; Arunachalam, T.; Nelson, P.; Spiteller, G. *J. Am. Chem. Soc.* **1984**, *106*, 7282-7283.
23. Sen, K.; Hackett, J. C. *Biochemistry* **2012**, *51*, 3039-3049.
24. Mak, P. J.; Luthra, A.; Sligar, S. G.; Kincaid, J. R. *J. Am. Chem. Soc.* **2014**, *136*, 4825-4828.
25. Khatri, Y.; Luthra, A.; Duggal, R.; Sligar, S. G. *FEBS Lett.* **2014**, *588*, 3117-3122.
26. Yoshimoto, F. K.; Guengerich, F. P. *J. Am. Chem. Soc.* **2014**, *136*, 15016-15025.
27. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. (2008), AMBER 10, University of California, San Francisco.

28. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
29. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. III. *J. Comput. Chem.* **2012**, *33*, 119-133.
30. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
31. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Rev. D.01*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
33. Beck, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

34. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.
35. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.
36. Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; Vol. 67.
37. Melander, L.; Saunders, W. H., Jr. In *Reaction rates of isotopic molecules.*; Robert E. Krieger Publishing Company: Malabar, FL, 1987; Chapter 2.
38. Wang, Y.; Kumar, D.; Yang, C.; Han, K.; Shaik, S. *J. Phys. Chem.* **2007**, *111*, 7700-7710.
39. Kumar, D.; Hirao, H.; de Visser, S. P.; Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Phys. Chem.* **2005**, *109*, 19946-19951.
40. Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 13204-13215.
41. Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, *110*, 949-1017.
42. Korzekwa, K. R.; Trager, W. F.; Mancewicz, J.; Osawa, Y. *J. Steroid Biochem. Mol. Biol.* **1993**, *44*, 367-373.
43. Hackett, J. C.; Bruggemeier, R. W.; Hadad, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 5224-5237.
44. Kramos, B.; Olah, J. *Struct. Chem.* **2015**, *26*, 279-300.
45. Blanksby, S. J.; Ellison, G. B. *Acc. Chem. Res.* **2003**, 255-263.
46. Usharani, D.; Lacy, D. C.; Borovik, A. S.; Shaik, S. *J. Am. Chem. Soc.* **2013**, *135*, 17090-17104.
47. Hirao, H.; Chuanprasit, P. *Chem. Phys. Lett.* **2015**, *621*, 188-192.

48. Hirao, H.; Ng, W. K. H.; Moeljadi, A. M. P.; Bureekaew, S. *ACS Catal.* **2015**, *5*, 3287-3291.
49. Wang, D.; Farquhar, E. R.; Stubna, A.; Munck, E.; Que, L., Jr. *Nat. Chem.* **2009**, *1*, 145-150.
50. Wang, Y.; Yang, C.; Wang, H.; Han, K.; Shaik, S. *ChemBioChem* **2007**, *8*, 277-281.
51. Schyman, P.; Lai, W.; Chen, H.; Wang, Y.; Shaik, S. *J. Am. Chem. Soc.* **2011**, *133*, 7977-7984.
52. Hirao, H.; Thellamurege, N.; Chuanprasit, P.; Xu, K. *Int. J. Mol. Sci.* **2013**, *14*, 24692-24705.

Chapter 3 Ferric-Superoxide Intermediate in the Verdoheme Formation Step of Heme Oxygenase

3.1 Introduction

Heme oxygenase (HO) is a membrane-bound hemoprotein that has a variety of physiological implications.¹⁻⁵ The heme-binding active site of HO is defined by two helices,⁶ wherein heme acts as the prosthetic group and the substrate and undergoes oxidative degradation into biliverdin that is a precursor of antioxidant bilirubin.⁷⁻¹² Besides biliverdin, physiologically important iron and CO are produced by the reaction. This oxidative transformation proceeds via a few sequential steps, as illustrated in Scheme 1, and during the reaction, three molecules of O₂ and seven electrons are used in total. Over the past few decades, the reaction mechanism of HO has attracted considerable attention. Particularly in recent years, computational methods have been fruitfully used to gain a clearer mechanistic understanding. For example, for the first step of the HO reaction, i.e., α -*meso*-hydroxylation that yields α -*meso*-hydroxyheme (**1**), a mechanism involving O–O homolysis of Fe^{III}OOH¹³⁻¹⁵ and a water-assisted oxo mechanism¹⁶⁻¹⁸ have been proposed based on computational studies. Insights into the mechanism of conversion from verdoheme (**2**) to biliverdin have also been derived.^{19,20}

We are concerned with the second step of HO reaction that is associated with the conversion of **1** into **2** (Scheme 1). This verdoheme formation step, in which CO is eliminated arguably from the α -*meso* position, remains somewhat mysterious despite a few computational studies^{21,22} because the lifetime of **1** is short and intermediates have not been well characterized. Nevertheless, some clues have been obtained on the mechanism from spectroscopy analyses. Particularly important is the report by Liu *et al.* who showed that **1** exhibits EPR signals corresponding to a ferric ion and an

organic radical.²³ This provided a strong piece of evidence that the hydroxyl group of **1** is deprotonated, and the real ligand should therefore be oxophlorin (Por-O). The dichotomous EPR signals of **1** indicated that the resultant species should exist as a mixture of three resonance structures: ferric phenolate **1a**, ferrous keto π neutral radical **1b**, and ferric keto anion **1c** (Scheme 1).²⁴ For the conversion from **1** to **2**, an electron should be supplied to the system, even though its timing has been a matter of debate. Matera *et al.* suggested that an electron is introduced before verdoheme formation.²⁵ By contrast, Liu *et al.* demonstrated that ferric verdoheme can be formed even in the absence of a reducing equivalent,²³ much in line with the earlier observation for the reaction of the ferric α -*meso*-hydroxyheme-myoglobin complex with O₂.²⁶ These somewhat different views and results probably reflect the fact that the timing of reduction may not be narrowly restricted. Rather, **2** may be formed via several different mechanisms in which an electron could be supplied at different stages. Herein, we focus on the latter mechanism, i.e. the one without prior reduction of **1**.

Scheme 1. HO-catalyzed reaction with emphasis on the second step.

It is usually posited that O_2 binds to the ferrous keto π neutral radical **1b** at pyrrole α -carbon to give a ferrous peroxy radical **1bA** (Scheme 1) that supposedly triggers the subsequent steps en route to verdoheme.⁷⁻⁹ The possible formation of **1bA** is mechanistically interesting. However, Safari *et al.* reported that theoretically **1a**, in the protein environment, could be easily converted into **1b**, which is reactive toward O_2 to generate **1bB**.²⁷ Our ongoing interest in the nature and reactivity of ferric-

superoxide species in iron enzymes²⁸⁻³¹ made us rather curious about whether ferric superoxide could be the reactive species and the sequential mechanistic scenario driven by ferric superoxide once it is formed. Herein, DFT and QM/MM calculations were applied to elucidate mechanistic details of the conversion from α -meso-hydroxyheme (1) to verdoheme (2) in HO.

3.2 Methodology

For DFT calculations, the methyl, vinyl, and propionate substituents of Por-O were replaced by hydrogen, and imidazole was used to model the His25 axial ligand. Geometry optimizations and frequency calculations were performed using the B3LYP functional in conjunction with the [SDD(Fe),6-31G*(others)] basis set (B1).³²⁻³⁶ Single-point energy calculations were performed on the optimized geometries with the 6-311+G(d,p) basis set (B2), taking into account the somewhat polar nature of the enzyme active site by the IEFPCM SCRF method ($\epsilon = 5.6968$).³⁷ The B3LYP/B2(SCRF)//B3LYP/(B1) energies with zero-point energy (ZPE) corrections are used for the discussion.

For QM/MM calculations, the X-ray crystal structure of HO used in our ONIOM simulations is the one determined by Noguchi *et al.* (PDB code 4G7L, 1.80 Å resolution).³⁸ The formic acid included in this crystal structure was removed. All aspartic acid residues and glutamic acid residues were assumed to be ionized, while all lysine and arginine residues were assumed protonated. Protonation states of histidine residues were determined as follows: HID25, HIE56, HIE84, HIE100, HIE113, HID119, HIE125, HIE132, and HIE197. HIE stands for a neutral histidine residue in which the N ϵ atom is protonated while HID means its N δ atom is protonated. Hydrogen atoms were added to the crystal structure using AMBER10 software.³⁹ MM parameters

for standard amino acid residues were taken from the ff03.r1⁴⁰ force field of AMBER. The MM parameters reported by Shahrock *et al*⁴¹ and the general amber force field (gaff)⁴² were used for the meso-hydroxyheme and dioxygen, respectively. The atomic charges for the heme-dioxygen and proximal HID25 were determined at the B3LYP/6-31G* level using the restrained electrostatic potential (RESP)⁴³ method.

The QM region in ONIOM calculations was defined as illustrated in Scheme 2. It includes α -meso-hydroxyheme, dioxygen, and HID25. Hydrogen link atoms were added to the QM atoms at the boundary to make the QM region saturated. The total charge of the QM region was 0.

Scheme 2. QM region for ONIOM calculations.

To refine the initial structure, preparatory MM energy minimizations were performed. First, all hydrogen atoms were relaxed while all the heavy atoms were fixed. Then an energy minimization was performed for the whole system except all backbone atoms. After this, the whole system was further relaxed with heavy atoms in the QM region fixed to obtain the initial geometry for subsequent ONIOM calculations.

In ONIOM geometry optimization calculations, all residues within 15 Å away from the Fe ion in the geometry obtained in preparatory MM calculations were allowed to move freely. The AMBER03 force field was used in ONIOM calculations

using the “softfirst” function in Gaussian 09. Geometry optimizations were performed using the ONIOM(B3LYP/B1:AMBER) level of theory.^{32-36,44-46} Zero-point energy (ZPE) corrections were also obtained by frequency calculations at this level. Single-point energy calculations using the ONIOM(B3LYP/B2:AMBER) level of theory with the electronic embedding scheme were performed on the optimized geometries. These single point energies with ZPE corrections are reported for the discussion of the mechanistic scenario. Gaussian 09 was used for all calculations,⁴⁶ and UCSF Chimera was used to visualize molecules.⁴⁷

3.3 Results and Discussion

DFT calculations on **1bA** and **1bB** in different spin states were performed to evaluate their relative stabilities. Figure 1 shows the geometries and spin states of **1bA** and **1bB**. For both intermediates, end-on geometries were obtained. Several different spin states were described by unrestricted DFT calculations. Of the eight examined spin states, doublet **1bB** with a ferromagnetically coupled Fe–OO pair (²**1bB**) was the most stable. Its antiferromagnetic counter-part (²**1bB'**) could not be obtained because its geometry optimization resulted in dissociation of dioxygen. The quartet ferric superoxide (⁴**1bB**) was almost as stable as ²**1bB**, as its relative energy was higher than that of ²**1bB** only slightly. In the ferrous peroxy radical intermediate (**1bA**), the iron ion is pentacoordinate, and thus higher spin states were somewhat more stable than the monoradical doublet state (²**1bA**). Interestingly, however, all spin states of **1bA** were >16 kcal/mol higher in energy than ferric superoxide. These results indicate that ferric superoxide may be a more likely reactive species in the verdoheme formation step of HO. Matera *et al.* observed an EPR signal at $g = 2.004$ when **1** was exposed to O₂, and this signal was tentatively attributed to a peroxy radical.²³ However, the

relatively high stability of **1bB** leads us to suspect that the signal may be attributable to a phenoxy radical of Por-O (see **1bB** in Scheme 1).

Figure 1. Geometries and schematic illustrations of the different spin states considered for **1bA** (a) and **1bB** (b). Values in parentheses are relative energies in kcal/mol, and values in square brackets are $r(\text{C}-\text{O})$ for **1bA** or $r(\text{Fe}-\text{O})$ for **1bB** in $\text{\AA}$.

To assess the capability of ferric superoxide to produce verdoheme, we performed reaction pathway analyses from **1bB** using DFT and ONIOM calculations. Obtained reaction energy profiles are shown in Figure 2 and Figure 3a, respectively. Unlike the conversion from heme to **1** where the distal oxygen (O_d) of $\text{Fe}^{\text{III}}\text{OOH}$ attacks the meso carbon atom, O_d of **1bB** was found to attack an adjacent pyrrole α -carbon via a transition state, **TS_ooc**. This leads to the formation of a bridged $\text{Fe}^{\text{III}}\text{OOR}$ -type intermediate **Int_ooc**. While $^4\text{1bB}$ is almost as stable as $^2\text{1bB}$ at the outset, the $\text{Fe}^{\text{III}}\text{OOR}$ formation in the quartet state has a barrier of 30.0 kcal/mol in DFT results and is thus highly unfavorable. Therefore, the reaction should preferentially choose the doublet-state energy surface over the quartet one. The reaction is followed by a homolytic cleavage of the O–O bond in **Int_ooc** to form a ferryl alkoxy radical intermediate (**Int_o-o**). Ortiz de Montellano and coworkers

previously proposed that this kind of ferryl intermediate should be involved in the reaction.⁷⁻⁹ Much to our surprise, this ferryl alkoxy radical triggers cyclization by attacking the α -carbon on another pyrrole; this step has a very low energy barrier (only 1.1 kcal/mol in DFT results and barrierless in ONIOM results). As a consequence, a unique intermediate is formed that possesses a four-membered ring (**Int_cyc**). As far as we are aware, involvement of such a cyclic intermediate in the HO reaction has never been suggested. There is a subsequent cleavage of two C–C bonds that allows CO to dissociate on the proximal side of the ligand. At the transition state of this C–C bond cleavage (**TS_dis**), the distances of the two breaking C–C bonds are not identical in both DFT (2.11 and 1.71 Å) and ONIOM results (2.07 Å and 1.69 Å in **TS_dis** while 2.06 Å and 1.69 Å in **TS_dis'**). **Int_dis** is finally formed as a result of the dissociation of CO. Thus, the calculations show that the verdoheme core can indeed be formed from the ferric-superoxide state. Afterward, protonation of the oxo moiety of **Int_dis** will facilitate transfer of the radical electron on the ligand to Fe to make the iron oxidation state +3, ultimately leading to the formation of a molecule of water and ferric verdoheme. Ferrous verdoheme (**2**) will be formed upon reduction of the ferric verdoheme by an electron. The proposed reaction scenario, unlike the first and last step in which the involvement of a water cluster in the distal pocket is critical, does not need this distal water cluster. This reaction scenario, therefore, indicates that the stereoselectivity observed in the first and last steps caused by this distal water cluster should be absent in the second step. This fits the experimental observation that the conversion from meso-hydroxyheme to verdoheme has no stereoselectivity toward different meso-hydroxyhemes.⁴⁸ Moreover, though the distal water cluster is suggested to be capable to stabilize radicals in theoretical studies for the other two steps,^{15,19} our ONIOM results show that the pathway in which O_d of **1bB** attacks the adjacent pyrrole

α -carbon C, which is closer to the distal water cluster, and the pathway in which O_d of **1bB** attacks the adjacent pyrrole α -carbon C' are both plausible. (Figure 3a)

Figure 2. Energy diagram (in kcal/mol) for the reaction of **1bB** by DFT calculations.

Figure 3. (a) Energy diagram (in kcal/mol) for the reaction of **1bB** by ONIOM calculations. (b) Transition-state geometries. Key bond distances are shown in Å.

There are several differences between the mechanistic scenarios found in DFT calculations and ONIOM calculations. First, DFT calculations indicate that the rate-

limiting step is the C-C bond cleavage via **TS_dis** while ONIOM calculations point out that it shifts to the formation of the bridged $\text{Fe}^{\text{III}}\text{OOR}$ -type intermediate **Int_ooc**. Second, there is a small barrier for the cyclization after **Int_o-o** in DFT calculations while this cyclization is barrierless in ONIOM calculations. It could be justified by the fact that this barrier in DFT calculations is very small (1.1 kcal/mol); thus, when protein environment is taken into consideration, this cyclization becomes barrierless. Third, the protein environment is taken into consideration in QM/MM calculations, resulting in two types of adjacent pyrrole α -carbon, C and C' as shown in Figure 3a, for O_d of **1bB** to attack in the QM/MM derived scenarios while there is only one type of adjacent pyrrole α -carbon in DFT calculations. Consequently, two different energy profiles were obtained in QM/MM calculations. For example, there are two ferryl alkoxy radical intermediates in the QM/MM derived scenarios while only one found by DFT calculations. **Int_o-o** and **Int_o-o'** in ONIOM derived scenarios are shown in Figure 4. The meso-hydroxyheme ring in **Int_o-o** is bent toward the distal pocket while in **Int_o-o'** it is somehow more flat. It is worth noting that **Int_o-o** is 10.7 kcal/mol higher than **Int_o-o'** in energy. The factor that causes this bent **Int_o-o** might be the water cluster in the distal pocket.

Figure 4. Geometries of **Int_o-o** and **Int_o-o'** derived by ONIOM calculations.

Despite the relatively low stability of **1bA** compared with **1bB**, when the iron is inhibited by a bound CO ligand, such a ferrous peroxy radical species could be formed. Interestingly, Liu *et al.* found that CO-complexed **1** also reacts with O₂ to generate ferric verdoheme.²³ One possible interpretation of this observation might be that O₂ does not have to bind to iron, and it may be **1bA** that always triggers the verdoheme formation. However, in the reaction in the absence of CO, **1bA** will not be preferentially formed because it is much higher in energy than **1bB**. To gain insight into the mechanism of the reaction in the presence of bound CO (Figure 5a), we calculated the binding energy (*BE*) between **1bA** and CO. *BE* is defined as

$$BE = E(^2[\mathbf{1bA-CO}]) - \{E(^4\mathbf{1bA}') + E(^1\text{CO})\} \quad (1)$$

where $E(^2[\mathbf{1bA-CO}])$, $E(^4\mathbf{1bA}')$, and $E(^1\text{CO})$ are the energies of CO-complexed **1bA**, **⁴1bA'**, and CO, respectively. *BE* was calculated as -7.7 kcal/mol. Since this *BE* is not very large, it should be possible that the peroxy radical in **1bA** displaces the coordinating CO to produce an Fe^{III}OOR intermediate. Indeed, as shown in Figure 5, the complex of **²Int_o-o** and CO, was more stable than $^2[\mathbf{1bA-CO}]$ by -3.5 kcal/mol. Therefore, even if a **1bA**-CO complex is formed, CO can dissociate and more stable **Int_o-o** can later be formed. Once **Int_o-o** is formed, a homolytic O-O bond cleavage will take place to follow the pathway described in Figure 2 and Figure 3a. Thus, in this particular case, the reaction will not pass through the ferric-superoxide species for the verdoheme formation.

Figure 5. CO-coordinated **1bA** and the complex of **Int_o-o** and CO. The values underneath the structures are relative energies in kcal/mol.

3.4 Conclusion

In summary, our calculations provided a clear and consistent view on the mechanism of ferric verdoheme formation in the absence of a reducing equivalent. We found that ferric superoxide is more stable than a ferrous peroxy radical and can act as a reactive species for the reaction. Although it is out of the scope of this work, the reaction of reduced **1**²⁵ will offer another interesting conundrum for computational studies. The important role of ferric superoxide may be underscored even more in this reaction because the porphyrin has a closed-shell character and thus the formation of a porphyrin-bound peroxy radical should be highly unfavorable. Our results also suggest even in the presence of CO, a CO-bound ferrous peroxy radical might be formed but it then would possibly undergo a similar pathway as the one suggested herein for ferric superoxide.

3.5 References

1. Verma, A.; Hirsch, D. J.; Glatt, C. E.; Ronnett, G. V.; Snyder, S. H. *Science*

- 1993**, 259, 381-384.
2. Maines, M. D. *Annu. Rev. Pharmacol. Toxicol.* **1997**, 37, 517-554.
 3. Markesbery, W. R. *Free Radical Biol. Med.* **1997**, 23, 134-147.
 4. Yachie, A.; Niida, Y.; Wada, T.; Igarashi, N.; Kaneda, H.; Toma, T.; Ohta, K.; Kasahara, Y.; Koizumi, S. *J. Clin. Invest.* **1999**, 103, 129-135.
 5. Otterbein, L. E.; Bach, F. H.; Alam, J.; Soares, M.; Lu, H. T.; Wysk, M.; Davis, R. J.; Flavell, R. A.; Choi, A. M. K. *Nat. Med.* **2000**, 6, 422-428.
 6. Schuller, D. J.; Wilks, A.; Ortiz De Montellano, P. R.; Poulos, T. L. *Nat. Struct. Biol.* **1999**, 6, 860-867.
 7. Ortiz De Montellano, P. R. *Acc. Chem. Res.* **1998**, 31, 543-549.
 8. Ortiz De Montellano, P. R. *Curr. Opin. Chem. Biol.* **2000**, 4, 221-227.
 9. Colas, C.; Ortiz De Montellano, P. R. *Chem. Rev.* **2003**, 103, 2305-2332.
 10. Matsui, T.; Unno, M.; Ikeda-Saito, M. *Acc. Chem. Res.* **2010**, 43, 240-247.
 11. Unno, M.; Matsui, T.; Ikeda-Saito, M. *Nat. Prod. Rep.* **2007**, 24, 553-570.
 12. Unno, M.; Matsui, T.; Ikeda-Saito, M. *J. Inorg. Biochem.* **2012**, 113, 102-109.
 13. Sharma, P. K.; Kevorkiants, R.; de Visser, S. R.; Kumar, D.; Shaik, S. *Angew. Chem. Int. Ed.* **2004**, 43, 1129-1132.
 14. Kumar, D.; De Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.* **2005**, 127, 8204-8213.
 15. Chen, H.; Moreau, Y.; Derat, E.; Shaik, S. *J. Am. Chem. Soc.* **2008**, 130, 1953-1965.
 16. Kamachi, T.; Shestakov, A. F.; Yoshizawa, K. *J. Am. Chem. Soc.* **2004**, 126, 3672-3673.
 17. Kamachi, T.; Yoshizawa, K. *J. Am. Chem. Soc.* **2005**, 127, 10686-10692.
 18. Kamachi, T.; Nishimi, T.; Yoshizawa, K. *Dalton Trans.* **2012**, 41, 11642-11650.
 19. Lai, W. Z.; Chen, H.; Matsui, T.; Omori, K.; Unno, M.; Ikeda-Saito, M.; Shaik,

- S. J. Am. Chem. Soc.* **2010**, *132*, 12960-12970.
20. Gheidi, M.; Safari, N.; Zahedi, M. *J. Mol. Model.* **2010**, *16*, 1401-1413.
 21. Gheidi, M.; Safari, N.; Zahedi, M. *Inorg. Chem.* **2012**, *51*, 7094-7102.
 22. Gheidi, M.; Safari, N.; Zahedi, M. *Inorg. Chem.* **2012**, *51*, 12857-12866.
 23. Liu, Y.; Moënne-Loccoz, P.; Loehr, T. M.; Ortiz De Montellano, P. R. *J. Biol. Chem.* **1997**, *272*, 6909-6917.
 24. Morishima, I.; Fujii, H.; Shiro, Y.; Sano, S. *Inorg. Chem.* **1995**, *34*, 1528-1535.
 25. Matera, K. M.; Takahashi, S.; Fujii, H.; Zhou, H.; Ishikawa, K.; Yoshimura, T.; Rousseau, D. L.; Yoshida, T.; Ikeda-Saito, M. *J. Biol. Chem.* **1996**, *271*, 6618-6624.
 26. Sano, S.; Sano, T.; Morishima, I.; Shiro, Y.; Maeda, Y. *Proc. Nat. Acad. Sci. U.S.A.* **1986**, *83*, 531-535.
 27. Gheidi, M.; Safari, N.; Zahedi, M. *Inorg. Chem.* **2014**, *53*, 2766-2775.
 28. Hirao, H.; Morokuma, K. *J. Am. Chem. Soc.* **2009**, *131*, 17206-17214.
 29. Hirao, H. *J. Phys. Chem. B* **2011**, *115*, 11278-11285.
 30. Hirao, H.; Morokuma, K. *J. Am. Chem. Soc.* **2010**, *132*, 17901-17909.
 31. Hirao, H.; Morokuma, K. *J. Am. Chem. Soc.* **2011**, *133*, 14550-14553.
 32. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
 33. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.
 34. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.
 35. Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866-872.
 36. Hehre, W.; Radom, L.; Schleyer, P. v. R.; Pople, J. *Ab Initio Molecular Orbital Theory*, John Wiley & Sons: New York, 1986.
 37. Cancès, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032-3041.
 38. Sugishima, M.; Moffat, K.; Noguchi, M. *Biochemistry* **2012**, *51*, 8554-8562.

39. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. (2008), AMBER 10, University of California, San Francisco.
40. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
41. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. III. *J. Comput. Chem.* **2012**, *33*, 119-133.
42. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
43. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
44. Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.
45. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327–350.
46. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K.; Burant, J. C.; Millam, J. M.; Iyengay, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.;

- Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Comperts, R.; Startmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenbuerg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 09, revision B.01; Gaussian, Inc.: Wallingford, CT, 2010.
47. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
48. Zhang, X.; Fujii, H.; Matera, K. M.; Migita, C. T.; Sun, D.; Sato, M.; Ikeda-Saito, M.; Yoshida, T. *Biochemistry* **2003**, *42*, 7418-7426.

Chapter 4 The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase

4.1 Introduction

Oligotropha carboxidovorans is a Gram-negative bacterium that can be isolated from waste water and utilizes carbon monoxide as the source of carbon and energy.¹ In this bacterium, Mo-Cu carbon monoxide dehydrogenase (Mo-Cu CODH) catalyzes the conversion of carbon monoxide to carbon dioxide:²

The two reducing equivalents generated in this oxidation reaction are transferred to a flavin adenine dinucleotide (FAD) through two [2Fe-S] clusters.³ About 16% of the carbon dioxide produced is assimilated by the bacterium.¹ CODHs have attracted special attention because of their important role in removing hazardous carbon monoxide from atmosphere.⁴ Furthermore, a unique Mo-Cu diheteronuclear active site observed in Mo-Cu CODH aroused great interest in the structure and mechanism of the enzyme.

In 1999, the first crystal structure of Mo-Cu CODH from *Oligotropha carboxidovorans* was reported by Dobbek *et al.*³ In that crystal structure, the active site contained a Mo(=O)₂OH cluster, along with a S-selanylcysteine residue. In 2002, they reported an improved structure of the enzyme containing a diheteronuclear CuSMo(=O)OH cluster and a deprotonated cysteine (Cys388), but without a Se atom.⁵ A crystal structure of Mo-Cu CODH in complex with *n*-butylisocyanide was also determined in the latter study, and interestingly, a thiocarbamate-like structure was observed in the structure, which led to the hypothesis that a thiocarbonate intermediate is formed during the oxidation of carbon monoxide. In 2005, Hofmann *et al.*⁶ and Siegbahn *et al.*⁷ reported density functional theory (DFT) studies of the oxidation

reaction of carbon monoxide catalyzed by Mo-Cu CODH. Hofmann *et al.* suggested that the catalytic cluster in the active site should be CuSMo(=O)_2 instead of CuSMo(=O)OH (Scheme 1a). According to their proposed mechanism, after the binding of carbon monoxide to Cu, oxidation of the carbon monoxide to CO_2 by the equatorial Mo=O group takes place, followed by one-step isomerization to form a S-bound intermediate. Finally, with the assistance of a water molecule that interacts with both Cys388 and CO_2 , CO_2 is released. Siegbahn *et al.* assumed that the active species should contain a CuSMo(=O)OH unit (Scheme 1b). Nevertheless, after the binding of carbon monoxide to Cu, the proton in the CuSMo(=O)OH cluster is removed from the active site via a glutamic acid residue (Glu763), to form a CuSMo(=O)_2 cluster. This cluster appears in Scheme 1a, and the following steps are similar to the steps in Scheme 1a. However, after the formation of the S-bound intermediate, one more isomerization leads to the thiocarbonate intermediate then a water molecule coordinates to Mo, which makes the pathway in Scheme 1b distinct from that in Scheme 1a. This water molecule induces one-step isomerization that leads to an intermediate in which CO_2 binds to Cu, and then CO_2 is released. In 2010, Zhang *et al.* reported kinetic and spectroscopic studies of Mo-Cu CODH.⁸ Their EPR and resonance Raman studies showed that the active site of Mo-Cu CODH should have a CuSMo(=O)_2 unit, instead of CuSMo(=O)OH , thus corroborating the aforementioned theoretical predictions. Another interesting finding in their study was the near pH-independence of the oxidation of carbon monoxide by Mo-Cu CODH, which implied that the oxidation process should not be affected significantly by the ionization states of the residues near the reaction center. However, in the mechanism proposed by Kirk *et al.*, the involvement of a hydroxide ion is critical (Scheme 1c). In their mechanism, once CO_2 is generated, a hydroxide ion is introduced and coordinates to Cu, which

then attacks CO₂ to form a bicarbonate intermediate. Finally, the product is released in the form of bicarbonate.⁹

Scheme 1. Reaction mechanisms for Mo-Cu CODH proposed by Hofmann *et al* (a), by Siegbahn *et al* (b) and by Kirk *et al* (c).

In this study, we performed combined quantum mechanics and molecular mechanics (QM/MM) calculations using the ONIOM(QM:MM) method to address the following questions regarding the reaction mechanism of Mo-Cu CODH. We first attempted to answer how CO₂ is released, through QM/MM calculations with the effect of protein environment taken into account. Second, we examined the effect of different protonation states of Glu763 on the reaction. This residue was assumed to be protonated in Siegbahn's study, whereas it was not included in Hofmann's model. Very recently, Rokhsana *et al.* theoretically proposed a CuSMo(=O)₂ active site with protonated Glu763 for Mo-Cu CODH at its oxidized state, and they also pointed out

that an ionized Glu763 is possible during the reaction.¹⁰ The protonation state of Glu763 during the reaction, therefore, remains unclear. Because systematic studies do not seem to have been performed to address this issue, we herein examined different protonation states of Glu763. Third, we attempted to answer why the reaction outcomes for two substrates, carbon monoxide and *n*-butylisocyanide, are different. In other words, why is CO₂ generated when carbon monoxide is the substrate, and why does *n*-butylisocyanide inhibit Mo-Cu CODH? Since the essential role of hydroxide in the mechanism of Kirk *et al.* appears to contradict the pH-independence observed in kinetic studies, this mechanism is not examined herein.

4.2 Methodology

A high-resolution crystal structure of Mo-Cu CODH determined by Dobbek *et al.*⁵ at 1.5 Å resolution (PDB code 1N5W, oxidized form) was used to construct models for computational studies. To reduce the computational costs, only chain B, where oxidization of carbon monoxide occurs, was used. All aspartic acid residues and glutamic acid residues except Glu763 were treated as ionized. All lysine and arginine residues were protonated. The protonation states of histidine residues were determined based on their pKa values predicted by the PDB2PQR website interface¹¹ and visual inspection of the crystal structure. Histidine residues 59, 93, 177, 178, 210, 213, and 788 were protonated at their ε position (HIE), whereas residues 61, 180, 339, 766, and 793 were protonated at their δ position (HID). The other histidine residues were protonated at both ε and δ positions. Carbon monoxide was added with carbon positioned toward Cu. Hydrogen atoms were added to the crystal structure using AMBER10 software.¹²

MM energy minimization calculations for all hydrogen atoms, for all atoms except the backbone atoms, and then for all atoms in the MM region were performed sequentially to refine the initial structure. Subsequently, ONIOM calculations were performed with all residues within 15 Å away from the Cu ion in the MM optimized structure allowed to move. The initial geometry for oxidation of *n*-butylisocyanide was constructed by superimposing RC', a reactant complex for carbon monoxide with Mo-Cu CODH when Glu763 is protonated, with a crystal structure of Mo-Cu CODH in complex with *n*-butylisocyanide determined by Dobbek *et al.*⁵ at 1.09 Å resolution (PDB code 1N62).

The ONIOM(QM:MM) method implemented in the Gaussian 09 software package was used for all QM/MM calculations.¹³⁻¹⁵ The B3LYP/[SDD(Cu,Mo),6-31G*(others)] method was applied to the QM region in all geometry optimizations and frequency calculations,¹⁶⁻¹⁹ while the AMBER03 force field was used for MM calculations (GAFF was used for carbon monoxide, CuSMo(=O)₂ cluster, and the cofactor) through the “softfirst” option of Gaussian 09 in all calculations^{20,21}. The mechanical embedding (ME) scheme was used for geometry optimizations and frequency calculations. Single-point energy calculations were carried out using the ONIOM(B3LYP/def2-TZVP:AMBER03) method with the electronic embedding (EE) scheme.^{22,23} The energy profiles were obtained by using these single point energies and zero-point energy (ZPE) corrections estimated with the lower-level of theory.

4.3 Results and Discussion

In this section, we discuss the reaction pathways for the oxidation of carbon monoxide and *n*-butylisocyanide in Mo-Cu CODH obtained by ONIOM(QM:MM) calculations. As mentioned above, the activity of Mo-Cu CODH is essentially pH-

independent, and thus different protonation states of Glu763 were examined to verify it computationally. It is also worth noting that two protons are generated and transferred away from the active site during the oxidation. Close examination of the crystal structure of Mo-Cu CODH led us to assume that this proton transfer might occur through a hydrogen bond network comprising Glu763, Ser389, Glu488, Ser764, and Asp684. We considered two different conformations of Glu763. In one conformation (Conf I in Figure 1a, RC), the protonated Glu763 residue forms a hydrogen bond with the equatorial Mo=O group, whereas it acts as a hydrogen bond donor to form a hydrogen bond with Ser389 in the other conformation (Conf II in Figure 1b, RC'). In calculations for the oxidation of *n*-butylisocyanide, Glu763 was treated as protonated and adopted Conf II. QM regions of our ONIOM(QM:MM) calculations are summarized in Figure 1. The QM region includes the substrate (carbon monoxide or *n*-butylisocyanide), the CuSMo(=O)₂ unit, Cys388, side chains of residues close to the CuSMo(=O)₂ unit (Glu763, Gln240, Ser389, and Ser570), backbone atoms from C-terminal of Val384 to Ser389, and one water molecule. All heavy atoms included in QM region are displayed in the ball-stick style in Figure 1.

Figure 1. Reactant complexes (RCs) for ONIOM(QM:MM) calculations. (a-c) RCs for the oxidation of carbon monoxide in cases where protonated Glu763 adopts Conf I (a) and Conf II (b), and Glu763 is ionized (c). (d) RC''' for the oxidation of *n*-butylisocyanide.

As shown in Figure 2, when protonated Glu763 adopts different conformations, several differences are noticed in the RCs. First, Glu763 itself slightly rotates. Second, due to the hydrogen bond between Glu763 and Mo=O group, Mo=O group moves closer to Glu763 in Conf I. This movement also induces the change of position of the cofactor. Third, CO in Conf I is closer to the water cluster near the active site, leading

to different positions for the water cluster. Though RC is 16.5 kcal/mol higher in energy than RC', ONIOM calculations are applied to both RCs to explore the reaction mechanism.

Figure 2. Superimposition of RC and RC'. Carbon atoms in RC are colored in light sea green while carbon atoms in RC' are colored in dark gray. Key atoms in RC' are represented in ball-stick style.

4.3.1 Protonated Glu763: Conf I

Reaction energy profiles and optimized structures of transition states for the reaction in Conf I are shown in Figure 3. Similar to the pathways in Scheme 1, the first step involves the attack on carbon monoxide by the equatorial Mo=O group in the CuSMo(=O)₂ cluster, which has an energy barrier of 0.7 kcal/mol. The first intermediate, Int01, is lower in energy than RC, by 7.4 kcal/mol, and in Int01, the newly formed CO₂ binds to Cu. This CO₂ is released via TS02a with an energy barrier of 13.8 kcal/mol. After the release of CO₂, the water molecule included in the QM

region can coordinate to Mo to obtain Int02a_2, which is 19.8 kcal/mol lower than Int02a in energy. This Int02a_2 could be viewed as a starting point for MoCu-CODH to regenerate its active species CuSMo(=O)_2 .

Figure 3. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is protonated and adopts Conf I. Key bond distances are shown in Å.

However, from Int01, there is another lower-barrier pathway that involves two-step isomerization. This pathway is similar to what were proposed by Hofmann *et al.* and Siegbahn *et al.* and leads to the formation of a thiocarbonate intermediate, which is 10.2 kcal/mol more stable than Int01. The first isomerization leading to Int02b has a higher barrier than the second step in this two-step isomerization. The barrier for two-step isomerization is 6.4 kcal/mol lower than that for the CO_2 release, which thus implies that the formation of the thiocarbonate intermediate is favored.

4.3.2 Protonated Glu763: Conf II

Reaction energy profiles and optimized structures of transition states for the reaction in Conf II are shown in Figure 4. The energy barrier of the first step is 5.0 kcal/mol. The following steps are also similar to those in Conf I. The barrier for the release of CO₂ to give the final product is 6.0 kcal/mol. Two-step isomerization from Int01' leads to the formation of a thiocarbonate intermediate (Int03b'), which is 15.8 kcal/mol more stable than Int01'. The energy barrier for this isomerization is 3.0 kcal/mol, which also indicates that this pathway is favored over the pathway that releases CO₂.

4.3.3 Ionized Glu763

Reaction energy profiles and optimized structures of transition states for the reaction involving ionized Glu673 are shown in Figure 5. The first step has an energy barrier of 5.8 kcal/mol, which does not differ significantly from the corresponding barrier in the Conf II when Glu763 is protonated. The following CO₂ release has a barrier of 6.2 kcal/mol in energy with respect to Int01''. Again, two-step isomerization that has a barrier of 2.6 kcal/mol is more favorable than the release of CO₂. The thiocarbonate intermediate Int03b'' is lower in energy than Int01'' by 15.8 kcal/mol.

Figure 4. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is protonated and adopts Conf II. Key bond distances are shown in Å.

Figure 5. Energy diagram (in kcal/mol) and optimized transition-state geometries for the reaction in which Glu763 is ionized. Key bond distances are shown in Å.

4.3.4 Oxidation of *n*-butylisocyanide

Reaction energy profiles and optimized structures of transition states for the oxidation of *n*-butylisocyanide are shown in Figure 6. As in the case of the oxidation of carbon monoxide, the oxidation of *n*-butylisocyanide by the equatorial Mo=O group occurs in the first step, which has an energy barrier of 2.4 kcal/mol. Unlike the oxidation of carbon monoxide, isomerization toward a thiocarbamate-like intermediate is barrierless and takes only one step. Direct release of the product from Int01''' was investigated. Since the energy keeps increasing even when $r(\text{Cu}-\text{C}(\text{O}=\text{C}=\text{N}))$ is about 4.1 Å and angle $\text{O}=\text{C}=\text{N}$ in the product is already close to linear, direct release of the product was considered as energetically implausible. Moreover, Int02b''' is very stable, which is 35.2 kcal/mol lower in energy than Int01'''.

Figure 6. Energy diagram (in kcal/mol) and transition-state geometries for the oxidation of *n*-butylisocyanide by Mo-Cu CODH. Key bond distances are shown in Å.

The energy profiles in Figure 3-5 show that the formation of a thiocarbonate intermediate is kinetically more favorable than the immediate release of CO₂; this is consistent with the reaction mechanisms presented by Hofmann *et al.* and Siegbahn *et al.* Nevertheless, we also observed that the energy barriers from the thiocarbonate intermediate (Int03b, Int03b', and Int03b'') to the final product (Int02a, Int02a', and Int02a'') are 24.0 kcal/mol, 21.8 kcal/mol, and 18.6 kcal/mol respectively. We, therefore, propose a mechanism consisting of the following three major steps (see Scheme 2):

- 1) The equatorial Mo=O group in CuSMo(=O)₂ reacts with carbon monoxide.
- 2) After two-step isomerization, a thiocarbonate intermediate is formed.
- 3) Finally, the thiocarbonate intermediate undergoes a reverse process of the isomerization and releases a free CO₂.

The rate-limiting step of this proposed mechanism is the process from the thiocarbonate intermediate to the final product.

Scheme 2. Our proposed mechanism for the Mo-Cu CODH catalyzed carbon monoxide oxidation.

Our mechanism is essentially consistent with the mechanisms proposed based on DFT studies, in the sense that the formation of CO₂-bound intermediate is involved. However, our QM/MM calculations suggest that a water molecule is not required for CO₂ release. For example, in Siegbahn *et al.*'s calculations, the energy barrier for the step from the thiocarbonate intermediate to the final product was too high (28.2 kcal/mol), and thus they proposed the pathway in Figure 1b as an alternative mechanism, which involves the binding of a water molecule to Mo. By contrast, the CO₂ release in our proposed mechanism is energetically plausible even without a water molecule. A major difference between the energy profiles obtained in the ONIOM and DFT calculations is that the thiocarbonate intermediate (Int03b) in our ONIOM calculations is 10.2 kcal/mol lower in energy than the first intermediate (Int01), whereas this energy difference in their study is 19.8 kcal/mol. As can be seen in Table 1, key distances in Int03b are shorter than the corresponding distances in the structure obtained by DFT. In particular, the Cu-S distance is 0.06 Å shorter and the C-S distance is 0.05 Å shorter in Int03b. These differences suggest that the active site is less relaxed in the protein environment, resulting in a smaller energy gap between Int01 and Int03b.

Table 1. Key distances in the thiocarbonate intermediate obtained by Siegbahn *et al.* and our optimized Int03b.

	DFT ¹	QM/MM
Cu-S (Å)	3.50	3.44
Cu-O (Å)	1.92	1.91

C-S (Å)	1.83	1.78
Mo-O (Å)	2.20	2.18
C-O ₁ (Å)	1.27	1.26
C-O ₂ (Å)	1.33	1.30

¹ Data from Ref. 7.

In the energy profiles obtained when Glu763 ionized and protonated, the energy barriers are somewhat different. Even when Glu763 is protonated, barriers are quite different if Glu763 has different conformations. For example, the energy barrier for CO₂ release is 13.8 kcal/mol when Glu763 is protonated and adopts Conf I while this barrier is 6.0 kcal/mol when protonated Glu763 has Conf II. When Glu763 is deprotonated, this barrier is 6.2 kcal/mol. The high barrier when Glu763 adopts Conf I could be rationalized by the hydrogen bond between Glu763 and the equatorial Mo=O group, which stabilizes the Int01. Another thing worth noting is the conformation of CO₂ in the transition state when it is released. When Glu763 is protonated, no matter Conf I or Conf II, CO₂ is somewhat perpendicular to the plane formed by S-Cu-S in the active site while it, to some extent, is parallel to this plane when Glu763 is ionized. This conformational difference results in a larger displacement of the water molecule included in QM region when Glu763 is protonated. In Int02a', due to this displacement, CO₂ moves up and rests above the active site while CO₂ in Int02a stays near the active site, perhaps due to the different conformation of Glu763. Relative stabilities of Int02a and Int02a' respect to their corresponding first intermediates (Int01 and Int01'), therefore, are quite different. (Figure 3 & 4) More importantly, the key energy barrier for the process from the thiocarbonate intermediate to the final product is higher when Glu763 is protonated, especially when it adopts Conf I. The high barrier (24.0 kcal/mol) in Conf I mainly comes from the release process of CO₂,

which has a barrier of 13.8 kcal/mol relative to Int01 while this barrier is about 6 kcal/mol in the other two cases. However, considering the barrier when protonated Glu763 adopts Conf II, the reaction is still plausible, and the mechanistic scenario is not altered. We therefore conclude that the ionization status of Glu763 affects the oxidation of carbon monoxide but the effect is insignificant. This result is consistent with the near pH-independence observed in the kinetic studies for Mo-Cu CODH.

There is a water cluster near the active site in MoCu-CODH and one water molecule in this cluster is included in the QM region in our ONIOM(QM:MM) calculations. We believe that this water cluster provides water molecules to regenerate the active species CuMo(=O)_2 after the release of CO_2 . More importantly, the coordination of a water molecule to Mo stabilizes the desired product very largely and the resulting species (Int02a_2, Int02a'_2, and Int02a''_2) is the most stable one among all the species involved in our proposed reaction mechanism.

The energy profile for *n*-butylisocyanide oxidation (Figure 6) indicates that the formation of Int02b''', an inhibited form of Mo-Cu CODH observed in the crystal structure, is barrierless. And this Int02b''' is 35.2 kcal/mol lower than the first intermediate Int01''' in energy. Therefore, once Int02b''' is formed, because of the high energy barrier from Int02b''' to Int01''', it is hard for Int02b''' to undergo a reverse process as in the reaction of carbon monoxide. Moreover, unlike carbon monoxide, *n*-butylisocyanide has a bulky *n*-butyl chain, and as can be seen in the crystal structure, this *n*-butyl chain is accommodated in a pocket near the active site, where the aforementioned water cluster rests. Therefore, there is no water molecule to assist the release of the final product even if could be produced. The absence of water molecules and stability of Int02b''' collectively explain why Int02b''' was trapped in the crystal structure of Mo-Cu CODH with *n*-butylisocyanide.

4.4 Conclusion

In summary, ONIOM(QM:MM) calculations were performed to investigate the oxidation of carbon monoxide and *n*-butylisocyanide catalyzed by Mo-Cu CODH. Based on the obtained energy profiles, we conclude that the reaction begins with the oxidation of the substrate by the equatorial Mo=O group in the active site, followed by two-step isomerization leading to a thiocarbonate-like intermediate. When carbon monoxide is the substrate, its thiocarbonate intermediate should undergo a reverse process of isomerization to release a free CO₂ molecule. However, when *n*-butylisocyanide is the substrate, its thiocarbamate-like intermediate is quite stable and thus unlikely to overcome the high energy barrier in the reverse process, resulting in enzyme inhibition. Our proposed mechanism also supports the experimentally observed near pH-independence of Mo-Cu CODH.

4.5 References

1. Meyer, O.; Schlegel, H. G. *Arch. Microbiol.* **1978**, *118*, 35-43.
2. Meyer, O.; Schlegel, H. G. *J. Bacteriol.* **1980**, *141*, 74-80.
3. Dobbek, H.; Gremer, L.; Meyer, O.; Huber, R. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8884-8889.
4. Morsdorf, G.; Frunzke, K.; Gadkari, D.; Meyer, O. *Biodegradation* **1992**, *3*, 61-82.
5. Dobbek, H.; Gremer, L.; Kiefersauer, R.; Huber, R.; Meyer, O. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15971-15976.
6. Hofmann, M.; Kassube, J. K.; Graf, T. *J. Biol. Inorg. Chem.* **2005**, *10*, 490-495.
7. Siegbahn, P. E. M.; Shestakov, A. F. *J. Comput. Chem.* **2005**, *26*, 888-898.

8. Zhang, B.; Hemann, C. F.; Hille, R. *J. Biol. Chem.* **2010**, 285, 12571-12578.
9. Stein, B. W.; Kirk, M. L. *Chem. Commun.* 2014, 50, 1104-1106.
10. Rokhsana, D.; Large, T. A. G.; Dienst, M. C.; Retegan, M.; Neese, F. *J. Biol. Inorg. Chem.* **2016**, 21, 491-499.
11. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucl. Acids Res.* **2004**, 32, W665-W667.
12. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. (2008), AMBER 10, University of California, San Francisco.
13. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, 2, 815-826.
14. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, 2, 327-350.
15. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.;

- Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Rev. D.01*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
16. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
 17. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.
 18. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.
 19. Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. *Ab Initio molecular orbital theory*. John Wiley & Sons, New York, NY, USA, 1986.
 20. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Kollman, J.; Wang, J. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
 21. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
 22. Weigned, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
 23. Weigned, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065.

Chapter 5 Mechanism of an Asymmetric Ring-Opening Reaction of Epoxide with Amine Catalyzed by a Metal-Organic Framework: Insights from Combined Quantum Mechanics and Molecular Mechanics Calculations

5.1 Introduction

Metal-organic frameworks (MOFs) feature highly extended network architectures and are now attracting widespread interest in the fields of chemistry and materials science. Gas separation and storage are among the most important applications of MOFs.¹⁻⁵ MOFs have also been shown to act as heterogeneous catalysts, and thus in recent years, applications of MOFs to catalysis have been increasingly explored.⁶⁻¹⁶ In general, MOFs, as heterogeneous catalysts, have advantages over homogeneous catalysts, in the sense that they permit easy separation, efficient recycling, etc. MOFs, as a new class of porous materials, have much greater surface areas than those of traditional porous materials such as zeolites, leading to higher capacity for gas storage and enhanced catalytic activity. Additionally, compared to zeolites, MOFs have a much wider range of pore sizes, allowing a larger scope of substrates from small molecules to larger molecules such as carbohydrates and proteins. Moreover, various metal ions or metal secondary building units (SBUs) and organic linkers offer great tunability for MOFs, which potentially allows desired properties to be induced, such as a certain structural topology and selectivity for particular chemical reactions. Indeed, MOFs have been shown to act as asymmetric heterogeneous catalysts, and the development of such catalysts has been an active area of research.¹⁷⁻¹⁹

For example, in 2008, Tanaka *et al.* synthesized a novel MOF, which we are concerned with herein, by treating chiral *R*-2,2'-dihydroxyl-1,1'-binaphthalene-5,5'-dicarboxylic acid (*R*-5,5'-H₂BDA) containing BINOL with Cu(NO₃)₂ (Figure 1a).²⁰

Each pair of Cu(II) ions is bridged by four ionized *R*-5,5'-H₂BDA—i.e., *R*-5,5'-BDAs—to form a dinuclear metal paddlewheel (PDW) unit. This Cu-based MOF (**1**) essentially had a two-dimensional structure and was shown to catalyze several asymmetric ring-opening reactions (AROs) of *meso*-epoxides with aniline-type amines. For example, with catalyst **1**, cyclohexene oxide reacted with aniline to afford an (*R,R*)- β -amino alcohol in 51% e.e. and 51% yield at room temperature under solvent-free conditions.

AROs of *meso*-epoxides with nucleophiles offer promising routes to desymmetrized compounds, and thus synthetic methods for AROs have been investigated extensively.²¹⁻²⁵ For example, Jacobsen and coworkers have pioneered the asymmetric catalysis of AROs based on chromium(III)-salen catalysts.^{22,23} There was also an attempt to immobilize chromium(III)-salen catalysts within a zeolite, to heterogenize the catalysts.²⁶

Among the various known AROs, those of *meso*-epoxides with amines are interesting because they afford β -amino alcohols that are important intermediates for the synthesis of fine chemicals and drugs. A number of catalysts for this purpose have been developed so far,²⁷⁻⁵⁷ and several of them have employed BINOL ligands to support a metal ion and to induce enantioselectivities. These BINOL-based catalysts are described essentially as BINOLates, because BINOL coordinates to a metal ion via its deprotonated hydroxyl groups. During a given ARO reaction, the metal ion near the hydroxyl groups is postulated to act as a Lewis acid center and thus coordinates to the epoxide oxygen. It is interesting to note that catalysis driven by hydrogen bonding has also been developed. Kureshy and coworkers reported organocatalytic ARO reactions in which hydrogen bonding interactions between a substrate and the organocatalyst seems to be key to the catalytic activity.⁴⁰

Importantly, despite its inclusion of BINOL linkers, catalyst **1** differs from the existing BINOL-metal type Lewis acid catalysts because it maintains neutral hydroxyl groups of BINOL (Figure 1a). Thus, there are two different acidic sites that may activate the epoxide. First, the copper centers of the PDW units may be used for the ARO reaction as Lewis acidic sites.²⁵ It may also be possible that the hydroxyl groups of BINOL are used instead. Nevertheless, simultaneous use of these acidic sites for catalysis would be difficult because they are distant from each other (Figure 1a). The A–B–A-type stacking mode makes it likely that the catalytic reaction takes place on the surface of the solid material of **1**.²⁰ Therefore, instead of using multiple acidic sites from different MOF layers to achieve the catalysis, one layer of the MOF may already act as the catalyst.

Figure 1. (a) MOF catalyst **1** examined in this study.²⁰ (b) Model of MOF **1** used for QM/MM calculations (3×3 layer). Cyan carbon atoms were kept fixed during geometry optimization. (c) QM region for QM/MM calculations. Carbon atoms of the substrates (cyclohexene oxide and aniline) are colored magenta.

Quantum chemistry enables quantum mechanical descriptions of molecules and thus constitutes a powerful tool for dissecting the mechanisms of chemical reactions of molecules that involve bond-forming and -breaking events. A practically feasible

approach that allows both quantum mechanical descriptions of the reaction site and realistic modeling of large MOF structures is a hybrid quantum mechanics and molecular mechanics (QM/MM) method.⁵⁸⁻⁶⁵ Many research groups applied QM/MM methods to large enzymatic systems and thereby gained valuable insights into their intricate reaction mechanisms. QM/MM methods should also be applicable to MOF systems,⁶⁶⁻⁶⁹ even though classical force fields, such as the universal force field (UFF)⁷⁰ and MM3,⁷¹⁻⁷³ are not always available for MOFs. Therefore, several groups have recently proposed modifying conventional force field parameters, so that structures of MOFs can be described by MM methods. For example, the Schmid group developed a scheme whereby a first-principles-based force field called MOF-FF is derived by use of a genetic algorithm.⁷⁴ Addicoat *et al.* reported what they call UFF4MOF parameters, which were developed based on UFF and can be applied to MOFs.⁷⁵ These parameters reproduce well the quantum mechanically determined structures of typical SBUs of MOFs, such as PDW and $\text{Zr}_4\text{O}(\text{CO}_2)_6$, thereby offering greater opportunities to execute not only MM simulations but also QM/MM calculations of MOF systems to gain quantum mechanical insights.⁷⁶

In this study, QM/MM calculations were performed on the reaction between cyclohexene oxide and aniline catalyzed by **1** to address the question of how this catalytic ARO reaction takes place. Another important goal of our study is to identify the factor that induces the enantioselectivity of the reaction.

5.2 Methodology

5.2.1 Model setup

A model of the Cu-based MOF (**1**) for QM/MM calculations, containing 9 unit cells (3×3 layer, Figure 1b), was built from the crystal structure determined by

Tanaka *et al.*²⁰ Peripheral, terminal 6-hydroxy-1-naphthoic acid groups were truncated, and hydrogen atoms were attached to their C5 atoms. Water molecules coordinating to Cu ions were removed because the substrate—namely, cyclohexene oxide—is likely to coordinate to the Cu site (a Lewis acid center) when the reaction occurs. The heavy atoms of truncated terminal 6-hydroxy-1-naphthoic acid groups were fixed at their original positions in the crystal structure (Figure 1b). The Merz-Kollman scheme was used to determine atomic point charges.⁷⁷ The atomic charges for the *R*-5,5'-BDA linker were determined by performing a DFT calculation on lithium 6-hydroxy-1-naphthoate. The atomic charge of the C5 atom in the *R*-5,5'-BDA linker was defined as the sum of the atomic charges of the C5 and H5 atoms in lithium 6-hydroxy-1-naphthoate. The atomic charges for the Cu-PDW unit were derived using Cu₂(O₂CH)₄ in the triplet spin state. All these calculations were carried out using Gaussian 09 software,⁷⁸ employing the B3LYP functional⁷⁹⁻⁸¹ and the aug-cc-pVDZ-PP basis set⁸²⁻⁸⁴ for Cu atoms and the aug-cc-pVDZ basis set⁸⁵ for the other atoms. We used van der Waals radii of 2.26 Å for a Cu ion and 2.55 Å for a Li ion.⁸⁶ The derived atomic charges for Cu-PDW agreed reasonably well with those reported in previous studies⁸⁷⁻⁸⁹.

5.2.2 QM/MM calculations

All QM/MM calculations were performed using ChemShell software⁹⁰ that in this study integrated QM calculations and MM calculations performed by TURBOMOLE⁹¹ and DL_POLY,⁹² respectively. Different from ONIOM used in previous chapters, the additive QM/MM scheme was utilized in these calculations. The QM region included cyclohexene oxide, aniline, and the Cu-PDW building unit that consists of four 6-hydroxy-1-naphthoate moieties and two Cu ions (Figure 1c). The

electrostatic embedding scheme,⁹³ in which point charges of MM atoms are included in the one-electron Hamiltonian of the QM calculations, was employed for QM/MM calculations. As for the treatment of the QM–MM boundary, hydrogen link atoms were introduced to saturate the valence of the QM boundary atoms, and the charge-shift scheme was used.⁹⁴ For geometry optimization calculations, the dispersion-corrected B3LYP (B3LYP-D3) functional^{79-81,95} was used in combination with the SV(P) basis set⁹⁶ for the QM part of QM/MM, with zero damping used for the D3 correction. In order to discern the dispersion effect on the reaction effectively, the B3LYP functional (without D3) was also employed for selected reactions. Single-point energy calculations on the optimized structures were performed using the TZVP basis set for the QM calculation in QM/MM.⁹⁷ Although the open-shell singlet state should be the ground state of the Cu-PDW unit (Table 1), we considered the almost equally stable triplet spin state to simplify the computational treatment. As for the MM part, UFF⁷⁰ and UFF4MOF⁷⁵ parameter sets were utilized with atomic charges derived with the method described in the Model setup section. The HDLC optimizer⁹⁸ was used to locate stationary points (minimum-energy and saddle points) on the potential energy surface (PES). The HDLC optimizer employs a microiterative scheme to locate transition states, in which a system is divided into two regions: a reaction core and its environment. The geometry optimization calculations for the reaction core and environment are executed alternately until convergence is reached. Numerical frequency calculations for the reaction core were performed to confirm that each optimized transition-state structure corresponds to a first-order saddle point. All reported relative energies are based on QM/MM energies without zero-point energy (ZPE) corrections.

Table 1. The energy differences between the triplet state and the open-shell singlet state in RCs. Relative energies shown in kcal/mol are with respect to the triplet state in each RC.

	<i>(R,R)</i> -1-RC	<i>(R,R)</i> -2-RC	<i>(S,S)</i> -1-RC	<i>(S,S)</i> -2-RC
triplet	0.0	0.0	0.0	0.0
open-shell singlet	-0.6	-0.6	-0.6	-0.6

5.3 Results and Discussion

5.3.1 Stereochemistry in the ring-opening reaction of cyclohexene oxide

A possible mechanism for the ring-opening reaction of cyclohexene oxide with aniline is displayed in Figure 2a. The reaction may be divided into two steps: an epoxide ring-opening step and a proton-transfer step. In this mechanism, the nitrogen atom of aniline attacks one of the two carbon atoms in the epoxide moiety of cyclohexene oxide, from the backside of a C–O bond, resulting in ring-opening of the epoxide. After the ring-opening process, there will be a proton transfer from the ammonium moiety to the alkoxide oxygen atom to yield the β -amino alcohol product. Although the cyclohexene oxide substrate is an achiral molecule, the ring-opening reaction possibly leads to the formation of two types of desymmetrized products (*(R,R)*- and *(S,S)*-isomers). As shown in Figure 2b, the attack of the nitrogen atom on the C1 atom gives the *(R,R)*-isomer, whereas the attack on the C2 atom leads to the formation of the *(S,S)*-isomer.

Figure 2. (a) Two possible steps in the reaction of cyclohexene oxide with aniline. (b) Formation of (*R,R*)- and (*S,S*)- β -amino alcohols.

The chair conformation of cyclohexane is the most stable one of the four possible conformations (half-chair, boat, distorted boat, and chair conformations). According to the trans-diaxial effect (the Fürst–Plattner rule), the conformation of cyclohexene oxide is an important factor that determines the regioselectivity of reactions between nucleophiles and cyclohexene oxide. The two possible conformations of cyclohexene oxide as **A** and **B** are schematically illustrated in Figure 3. The nitrogen atom of aniline is likely to attack the C1 atom in conformation **B** because the product will adopt the most stable chair conformation. In this conformation, the nitrogen atom of aniline is unlikely to attack the C2 atom because a less stable twist-boat conformation product is formed. However, in the case of **A**, the nitrogen atom favors the attack on the C2 over the C1 atom. On the basis of these

arguments, in the following calculations, we calculated energy profiles for the formation of the most stable chair conformation products; thus, **A** and **B** were used to study the pathways leading to the (*S,S*) and (*R,R*) isomers, respectively.

Figure 3. Influence of the ring conformation on the stereochemistry in the ring-opening reaction of cyclohexene oxide with aniline. The resultant cyclohexane ring prefers the chair conformation.

5.3.2 Epoxide ring-opening step

We first assumed that the epoxide of the substrate is activated by the copper center of the Cu-PDW unit. In the stage of reactant complex (RC), the oxygen atom of cyclohexene oxide coordinates to a Cu(II) ion within the MOF, which acts as a Lewis acid for catalysis. When aniline approaches the epoxide molecule on Cu(II) for the reaction, it can form a hydrogen bond with a carboxylate ligand of Cu-PDW. However, hydroxyl groups on the linkers are more than 8 Å away from the nearest Cu(II) ion in the crystal structure; therefore, aniline cannot form a hydrogen bond with any of the hydroxyl groups during the reaction. Figure 4 shows four possible geometries of RC, which were obtained from QM/MM geometry optimization calculations. As mentioned above, cyclohexene oxide should assume conformation **A** for the (*S,S*)-isomer

formation ((*S,S*)-1-RC and (*S,S*)-2-RC) and conformation **B** for the (*R,R*)-isomer formation ((*R,R*)-1-RC and (*R,R*)-2-RC). The carboxylate group of the Cu-PDW unit is bound to the C5 atom of a naphthol group in the linker. Therefore, the two sides—namely, the empty side and the crowded side—with respect to the carboxylate plane offer distinct environments, where the crowdedness on the latter side originates from the presence of a naphthol ring. In (*S,S*)-1-RC and (*R,R*)-2-RC, the phenyl ring of aniline comes to the empty side, whereas it stays on the crowded side in the case of (*R,R*)-1-RC. In all these three cases, aniline forms a hydrogen bond with the carboxylate oxygen of the linker and thus stays close to the epoxide moiety. By contrast, aniline cannot come close to the epoxide in (*S,S*)-2-RC, because the naphthol group blocks the approach of aniline. Nevertheless, (*S,S*)-2-RC is less stable than the other species, and thus this species is unlikely to play any important role in the reaction. Even if (*S,S*)-2-RC is formed, it should be converted to more stable (*S,S*)-1-RC, and then the ring-opening reaction on the (*S,S*) pathway may begin.

Figure 4. Four possible structures of RC. Cyclohexene oxide has conformation **A** in (*S,S*)-1-RC and (*S,S*)-2-RC and conformation **B** in (*R,R*)-1-RC and (*R,R*)-2-RC. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

Pathways starting from the two most stable RC states in Figure 4; i.e., (*S,S*)-1-RC and (*R,R*)-1-RC, were examined first. Figure 5 shows the energy profiles and the optimized geometries of transition states (TS1) and intermediates (Int1) for the ring-

opening step on the (*S,S*)-1 and (*R,R*)-1 pathways. As mentioned earlier, the nitrogen atom of aniline attacks the C1 and the C2 atom in the (*R,R*)-1 and the (*S,S*)-1 pathway, respectively (Figure 2). The attack of aniline on the epoxide moiety triggers the cleavage of the C–O bond on the other side. When going from RC to TS, the C2–O bond is lengthened by 0.43 Å from 1.44 (RC) to 1.87 Å (TS1) in the (*S,S*)-1 pathway. Likewise, the C1–O bond is elongated by 0.43 Å from 1.44 (RC) to 1.87 Å (TS1) in the (*R,R*)-1 pathway. The Cu(II) ion, which coordinates to the epoxide oxygen, serves as a Lewis acid and facilitates the ring-opening reaction. The ring-opening process leads to the formation of an alkoxide ion, and the strong binding of the alkoxide oxygen to the Cu(II) ion causes the dissociation of one of the Cu–O coordination bonds of Cu-PDW. As a result, the negative charge of the alkoxide ion is stabilized effectively by the copper center, which therefore stabilizes TS1 and Int1. In Figure 5, we can also see that in the stage of Int1, the cleavage of the C–O bond of the epoxide and the formation of a new C–N bond have been completed. The oxygen atom of the resultant alkoxide ion binds to the Cu(II) ion, leading to the formation of a distorted square-planar-like structure around the copper center. This geometric distortion should facilitate the subsequent proton transfer step. In fact, at Int1, the amine group of aniline still maintains the hydrogen bond with the carboxylate oxygen, but the hydrogen-bond distance at Int1 is shorter than at RC. The shorter hydrogen-bond distance observed for Int1 is partly because the nitrogen atom now has ammonium-like character, and also because the carboxyl oxygen has been detached from the copper center and thus acts as a more negatively charged hydrogen-bond acceptor.

The energies of the transition states for the two reaction pathways, (*S,S*)-1-TS1 and (*R,R*)-1-TS1, differ only by 0.5 kcal/mol, with the latter state being slightly more stable (Figure 5). Close inspection of the geometry of (*R,R*)-1-TS1 reveals that there

are two CH- π interactions between the phenyl ring of aniline and the naphthol ring, which may account for the somewhat higher stability of (*R,R*)-1-TS1 compared with (*S,S*)-1-TS1. The formation of CH- π interaction is possible in (*R,R*)-1-TS1 because the Cu-O bond cleavage allows the aniline to come close to the naphthol group.

Figure 5. Energy diagram and optimized structures of TS1 and Int1 for the (*S,S*)-1 and (*R,R*)-1 pathways. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

Another reaction pathway was also examined. As mentioned above, the reaction on the (*S,S*) pathway is not likely to proceed directly from (*S,S*)-2-RC. Therefore, we

examined the (*R,R*)-2 pathway, in which (*R,R*)-2-RC undergoes the ring-opening reaction. The reaction along the (*R,R*)-2 pathway is largely similar to that along the (*R,R*)-1 pathway, as can be seen in Figure 6. Thus, upon the attack of aniline on the epoxide, one of the Cu–O bonds is cleaved to enhance the Lewis acidity of the Cu(II) ion and to stabilize TS1 or Int1; the same trend has been observed for the (*S,S*)-1 and (*R,R*)-1 pathways. As with the case of (*R,R*)-1-TS1, CH– π interactions are observed between aniline and a naphthol group in (*R,R*)-2-TS1, and key C–H distances are 2.77 and 2.72 Å. As shown in Figure 5, the two C---H distances associated with the CH– π interactions in (*R,R*)-1-TS are 3.01 and 2.69 Å, suggesting that one of the two interactions is relatively weak here. By contrast, both of the two C---H distances are short in (*R,R*)-2-TS1 (Figure 6), which suggests that the stability of (*R,R*)-2-TS1 is influenced by CH– π interactions to a larger extent. Indeed, (*R,R*)-2-TS1 is more stable than (*R,R*)-1-TS1 by 0.8 kcal/mol; therefore, the (*R,R*)-2 pathway should be the most plausible pathway for formation of the (*R,R*) isomer.

Figure 6. Energy diagram and optimized structures of TS1 and Int1 for the (*R,R*)-2 pathway. Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored

magenta and light green, respectively.

We performed Atoms-in-Molecules (AIM) analysis⁹⁹ at the MP2/SV(P) level of theory to confirm the presence of CH- π interactions between aniline and the naphthol group at (*R,R*)-2-TS1. For the AIM analysis, the structures of the aniline and naphthol groups were extracted from the QM/MM-optimized (*R,R*)-2-TS1, and the carboxylate group was replaced with a hydrogen atom. The position of the appended hydrogen atom was optimized. As shown in Figure 7, we can see bond paths and bond critical points between aniline and naphthol, supporting our hypothesis that there are CH- π interactions in that region, which, largely via the dispersion effect, stabilize the transition state.

Figure 7. AIM analysis on the structure of (*R,R*)-2-TS1. The carboxylate group of a linker was replaced with a hydrogen atom. Bond paths and bond critical points are denoted by black lines and light green dots, respectively.

To further investigate the role of dispersion in determining the relative stability of transition states, we evaluated the relative energies of key species formed in the ring-opening step using the B3LYP functional with and without the dispersion correction for the QM calculations within QM/MM. It should be noted that for these two types of calculations, geometry optimization was done separately, to assess the

effect of dispersion on geometries as well. Table 2 summarizes the QM/MM-calculated relative energies of the optimized RC, TS1, and Int1 species for the four possible pathways.

Table 2. Relative QM/MM energies (in kcal/mol) of RC, TS1, and Int1 in the four possible pathways, which were evaluated by using the B3LYP-D3 or B3LYP functional for the QM part. Relative energies are measured from (*R,R*)-1-RC.

		RC	TS1	Int1
B3LYP-D3 (B3LYP/MM) [D3]	<i>(R,R)</i> -1	0.0 (0.0) [0.0]	17.4 (20.4) [-3.0]	-2.7 (1.3) [-4.0]
		2.3 (-1.2) [3.5]	16.6 (19.3) [-2.7]	-4.0 (0.3) [-4.3]
	<i>(S,S)</i> -1	0.9 (1.8) [-0.9]	17.9 (15.4) [2.5]	-2.6 (-1.3) [-1.3]
		3.6 (-7.4) [11.0]	-	-
	<i>(R,R)</i> -1	0.0	21.0	1.9
B3LYP	<i>(R,R)</i> -2	0.6	21.0	1.5
	<i>(S,S)</i> -1	-0.5	17.4	-0.1
	<i>(S,S)</i> -2	-5.1	-	-

With the B3LYP functional (without the D3 correction), (*S,S*)-2-RC is calculated as much more stable than the other RCs. However, as mentioned before, the ring-opening reaction is not likely to take place from (*S,S*)-2-RC because aniline cannot access the epoxide. The relative energy of (*S,S*)-1-TS1 measured from (*R,R*)-1-RC (17.4 kcal/mol) is almost unchanged from that in the B3LYP-D3 calculation (17.9 kcal/mol), indicating that the dispersion effect does not play a very important role in

the (*S,S*)-1 pathway. By contrast, the relative energies of (*R,R*)-1-TS1 and (*R,R*)-2-TS1 increase significantly by 3.6 kcal/mol (17.4 $\rightarrow$ 21.0 kcal/mol) and 4.4 kcal/mol (16.6 $\rightarrow$ 21.0 kcal/mol), respectively. These noticeable increases are attributable to the fact that, with B3LYP, significant portions of CH- π interactions, which could be described by B3LYP-D3, cannot be accounted for. Indeed, the CH- π distances between the aniline and the naphthol group are significantly longer—4.47 and 5.79 Å at (*R,R*)-1-TS1 and 3.06 and 3.52 Å at (*R,R*)-2-TS1—in the geometries optimized with the B3LYP functional (see Figure 8) than the distances obtained with the B3LYP-D3 functional. These results collectively suggest that the dispersion effect in CH- π interactions affects the stability and geometry of (*R,R*)-1-TS1 and (*R,R*)-2-TS1 considerably, thereby playing an important role in determining the enantioselectivity. The somewhat different increases in the relative energies of (*R,R*)-1-TS1 (3.6 kcal/mol) and (*R,R*)-2-TS1 (4.4 kcal/mol), observed when the dispersion correction was removed, also support our argument that (*R,R*)-2-TS1 is stabilized by CH- π interactions to a larger extent than in the case of (*R,R*)-1-TS1.

Figure 8. QM/MM-optimized structures of (*R,R*)-1-TS1 and (*R,R*)-1-TS1 obtained with the B3LYP functional. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

The above computational examinations are based on the assumptions that the copper site of Cu-PDW acts as a Lewis acid site for the ARO reaction. We now consider the case in which the hydroxyl groups of BINOL are used instead, to activate the epoxide substrate. Tanaka *et al.* reported that the yield was low when (*S*)-BINOL was used as the catalyst. This result alluded to the unimportant role of hydrogen bonds in the catalysis. Nevertheless, in the environment of the MOF, the hydroxyl groups of BINOL might acquire somewhat different roles. To assess the catalytic role of the hydroxyl groups in the BINOL moiety of MOF **1**, we examined the geometry and stability of the state in which aniline coordinates to the Cu(II) center while an OH group of a linker forms a hydrogen bond with the epoxide oxygen of cyclohexene oxide. For this purpose, an extended QM region was adopted so that one of the four ligands of Cu-PDW includes an entire BINOL moiety in the QM region. The most stable hydrogen-bonded RC species (RCext2), obtained by these calculations, was less stable than the Cu-bound RCext1, which is essentially the same as (*R,R*)-1-RC in Figure 4 but was reoptimized in the extended QM region (QMext1), by 1.5 kcal/mol (Figure 9). Thus, the hydrogen-bonded configuration (RCext2) is slightly less stable than RCext1 featuring the coordination of the epoxide to Cu. In addition, additional calculations on the BINOL-catalyzed ring-opening reaction showed that the barrier for the H-bond-assisted process is 22.7 kcal/mol with DFT (Figure 10) or 18.4 kcal/mol with QM/MM (Figure 11), which is somewhat higher than the barriers obtained for the reactions in which Cu(II) acts as the Lewis acid site (Figures 5 and 6). It should be noted that the QM/MM model used here included the BINOL moiety and the substrates in the QM region. From these results, it is more likely that the reaction proceeds at the copper center, although the possibility cannot be ruled out completely

that the reaction also takes place via the hydrogen bond assisted mechanism, given the small energy differences between the transition states in these mechanisms.

Figure 9. Several RC species optimized using the extended QM region; RCext1 in which epoxide coordinates to the Cu(II) ion; RCext2 and RCext3 in which aniline coordinates to the Cu(II) ion and epoxide forms a hydrogen bond with the BINOL moiety. Relative energies shown in kcal/mol are with respect to RCext1. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. QM atoms are shown in stick.

Figure 10. Energy diagram and optimized structures of key species for the hydrogen bond assisted ring-opening reaction (DFT). Relative energies shown in kcal/mol are with respect to RC1. Key distances are indicated in Å. B3LYP-D3/6-311+G**//B3LYP-D3/6-31G* was used.

Figure 11. Energy diagram and optimized structures of key species for the hydrogen bond assisted ring-opening reaction (QM/MM). Relative energies shown in kcal/mol are with respect to RC_{BINOL} . Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively. QM atoms are shown in stick.

5.3.3 Proton-transfer step

After the formation of an alkoxide intermediate (Int1), the alkoxide oxygen accepts a proton to yield the final product, β -amino alcohol. The most plausible proton donor is the ammonium nitrogen atom at Int1. In the previous section, we showed that the (*S,S*) and (*R,R*) isomers should be formed via the (*S,S*)-1 and (*R,R*)-2 pathways, respectively. Therefore, the proton-transfer step was examined for these pathways. For steric reasons, a proton cannot be transferred directly from the ammonium nitrogen atom to the distant alkoxide-oxygen atom. Rather, the alkoxide intermediates should be converted into β -amino alcohol via several minor proton migration steps. Overall,

the proton migration steps in the two pathways have net barriers that are not very high (<10 kcal/mol) compared with the barriers for the first ring-opening step. Therefore, the first step should be the rate-determining step of the reaction, and the proton transfer step should not have any influence on the enantioselectivity.

The mode of proton transfer appears to be slightly different in the (*S,S*)-1 and (*R,R*)-2 pathways. As shown in Figure 12, in the (*R,R*)-2 pathway, (*R,R*)-2-Int1 is converted into (*R,R*)-2-Int3 via (*R,R*)-2-Int2, while (*S,S*)-1-Int1 is directly converted to (*S,S*)-1-Int3 in the (*S,S*)-1 pathway. (*R,R*)-2-Int1 is converted into slightly less stable (*R,R*)-2-Int2 via a proton transfer from the ammonium to the carboxylate oxygen without a barrier. Although (*R,R*)-2-Int2 was obtained in the (*R,R*)-2 pathway, its counterpart, (*S,S*)-1-Int2, could not be obtained in the (*S,S*)-1 pathway. In the species on the (*R,R*)-2 pathway, there are CH- π interactions between aniline and a naphthol group, but not for the species on the (*S,S*)-1 pathway. Thus, the CH- π interactions might play some role in determining whether the Int2 species can be formed. After a rotation of the OH group, (*R,R*)-2-Int3 is formed. (*R,R*)-2-Int3 is less stable than (*R,R*)-2-Int1 by 8.2 kcal/mol, but this energy increase is not critical. The rotation of the hydroxyl group changes the position of the atoms in the aniline group significantly. At (*R,R*)-2-Int1, the relatively strong hydrogen bond between the ammonium moiety and a carboxylate oxygen atom restricted the movement of the aniline group. However, at (*R,R*)-Int3, the aniline can move more easily because there is no strong geometric restriction imposed by hydrogen bonding.

Figure 12. Energy diagram for the proton transfer step. Relative energies shown in kcal/mol are with respect to *(R,R)*-1-RC.

At *(R,R)*-2-Int3, the hydroxyl group, which contains the migrating proton, forms a hydrogen bond with a neighboring carboxylate group that is located opposite to the aniline. We found another possible intermediate, *(R,R)*-2-Int3-2, in which the hydroxyl group forms a hydrogen bond with another neighboring carboxylate group near the aniline (Figure 13). However, *(R,R)*-2-Int3-2 is higher in energy than *(R,R)*-2-Int3 by 1.6 kcal/mol. Thus up to this point, the aniline molecule is seen to undergo significant repositioning to facilitate the proton transfer event.

Figure 13. QM/MM-optimized structures of (*S,S*)-1-Int3-2 and (*R,R*)-2-Int3-2. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

In the final proton transfer step, the O---H distance is shortened from 3.21 Å at (*R,R*)-Int3 to 2.12 Å at (*R,R*)-TS3. During this process, the geometry of the cyclohexane is altered noticeably because the phenyl group of aniline is loosely fixed onto a naphthol group through CH- π interactions. It is found that overall, the proton transfer step is exothermic; (*R,R*)-2-PC is more stable than (*R,R*)-2-Int1 by 12.1 kcal/mol. As we have seen above, the Cu(II) center forms a distorted square-planar-like structure to stabilize the alkoxide ion in (*R,R*)-2-Int1. Interestingly, the product formation drives the movement of the detached carboxylate group back to the copper center, to restore the original Cu-PDW structure. Thus, the product formation also removes the geometry distortions around the copper center, thereby contributing to the stabilization of (*R,R*)-2-PC. At the product stage, we observe no CH- π interaction anymore between the phenyl group and the naphthol group, because the reformed Cu-O bond precludes aniline from coming close to the naphthol group.

In the (*S,S*)-isomer formation pathway, (*S,S*)-1-Int1 is converted directly into

(*S,S*)-1-Int3. As in the case of the (*R,R*)-2 pathway, we obtained another possible intermediate, (*S,S*)-1-Int3-2 (4.2 kcal/mol, see Figure 13), which is almost as stable as (*S,S*)-2-Int3 (4.0 kcal/mol). In this pathway, there are no CH- π interactions; thus, the aniline can move more flexibly to find a more stable conformation. As such, after the substrate repositioning, the alkoxide oxygen can come closer to the carboxyl hydrogen atom, which appears to make (*S,S*)-1-TS3 more stable than (*R,R*)-2-TS3. The structure of (*S,S*)-1-Int3 is similar to that of (*S,S*)-1-TS3, whereas there was a significant geometry change in going from (*R,R*)-2-Int3 to (*R,R*)-2-TS3. The energy difference between TS3 and Int3 is also smaller in the (*S,S*)-isomer formation pathway (1.6 kcal/mol in the (*R,R*)-2 pathway vs 0.2 kcal/mol in the (*S,S*)-1 pathway). The proton transfer step is also exothermic by 13.5 kcal/mol (Int1 $\rightarrow$ PC) in the (*S,S*)-1 pathway.

To gain a broader view of the reaction, the full energy diagram is shown in Figure 14. We can clearly see in the figure that the overall reaction is exothermic. Importantly, the first ring-opening step is the rate-determining step, and the relative energies of TS1 species in the two pathways determine the selectivity of the reaction. Therefore, the enantioselectivity should be improved, if the energy gap between the two ring-opening transition states is increased. One possible way of doing this may be to increase the magnitude of CH- π interactions via introduction of substituents to the linker or the substrate. This may be achieved, for example, by attaching an electron-donating substituent like NH₂ at the C7 position of a 6-hydroxy-1-naphthoate moiety of the linker. The activation energy for the proton transfer step (Int1 $\rightarrow$ TS3) is 9.8 kcal/mol in the (*R,R*)-2 pathway and 6.8 kcal/mol in the (*S,S*)-1 pathway. These barriers are lower than the barriers for the first step; thus, the second step is less important in the context of enantioselectivity.

Figure 14. Full energy diagram for the reaction. Relative energies shown in kcal/mol are with respect to (R,R)-1-RC. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

5.4 Conclusion

This QM/MM computational study provided detailed insights into the mechanism of the asymmetric ring-opening reaction of cyclohexene oxide with aniline catalyzed by a two-dimensional MOF that contains a Cu-PDW unit. The MOF has two different acidic sites that may possibly activate the epoxide moiety of the substrate, but our calculations suggest that the copper center of Cu-PDW should be preferentially used for this purpose. With this catalytic mode, in the first step of the reaction, aniline attacks one of the two carbon atoms of the epoxide moiety in cyclohexene oxide, and the fate of the reaction is determined depending on which carbon atom is attacked.

This ring-opening step is identified as the rate-determining step, and therefore, increasing the energy gap between the transition states for the two different pathways should improve the enantioselectivity. We showed that CH– π interactions between aniline and a naphthol group play an important role in inducing the enantioselectivity. The CH– π interactions stabilize the transition state only in the (*R,R*) pathway, thereby facilitating the formation of (*R,R*)- β -amino alcohol. By contrast, CH– π interactions are not formed in the (*S,S*) pathway. Another interesting finding is that one of the carboxylate groups is detached from the copper center upon formation of the alkoxide intermediate, which demonstrates rather flexible behavior of the Cu-PDW unit during catalysis. After the ring-opening step, the generated alkoxide intermediate is readily converted into the β -amino alcohol product via several minor proton migration steps, because these steps do not entail high energy costs.

5.5 References

1. Kitagawa, S.; Kitaura, R.; Noro, S. *Angew. Chem. Int. Ed.* **2004**, *43*, 2334-2375.
2. Kitagawa, S.; Matsuda, R. *Coord. Chem. Rev.* **2007**, *251*, 2490-2509.
3. Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402*, 276-279.
4. Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, *112*, 724-781.
5. Li, J.-R.; Sculley, J.; Zhou, H.-C. *Chem. Rev.* **2012**, *112*, 869-932.
6. Fujita, M.; Kwon, Y. J.; Washizu, S.; Ogura, K. *J. Am. Chem. Soc.* **1994**, *116*, 1151-1152.
7. Hasegawa, S.; Horike, S.; Matsuda, R.; Furukawa, S.; Mochizuki, K.; Kinoshita, Y.; Kitagawa, S. *J. Am. Chem. Soc.* **2007**, *129*, 2607-2614.
8. Horike, S.; Dincă, M.; Tamaki, K.; Long, J. R. *J. Am. Chem. Soc.* **2008**, *130*,

5854-5855.

9. Ferey, G. *Chem. Soc. Rev.* **2008**, *37*, 191-214.
10. Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. *Chem. Soc. Rev.* **2009**, *38*, 1450-1459.
11. Kuppler, R. J.; Timmons, D. J.; Fang, Q.-R.; Li, J.-R.; Makal, T. A.; Young, M. D.; Yuan, D.; Zhao, D.; Zhuang, W.; Zhou, H.-C. *Coord. Chem. Rev.* **2009**, *253*, 3042-3066.
12. Farrusseng, D.; Aguado, S.; Pinel, C. *Angew. Chem. Int. Ed.* **2009**, *48*, 7502-7513.
13. Corma, A.; García, H.; Llabrés i Xamena, F. X. *Chem. Rev.* **2010**, *110*, 4606-4655.
14. Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. *Science* **2013**, *341*, 974-986.
15. Liu, J.; Chen, L.; Cui, H.; Zhang, J.; Zhang, L.; Su, C.-Y. *Chem. Soc. Rev.* **2014**, *43*, 6011-6061.
16. Eddaoudi, M.; Sava, D. F.; Eubank, J. F.; Adil, K.; Guillerm, V. *Chem. Soc. Rev.* **2015**, *44*, 228-249.
17. Ma, L.; Abney, C.; Lin, W. *Chem. Soc. Rev.* **2009**, *38*, 1248-1256.
18. Liu, Y.; Xuan, W.; Cui, Y. *Adv. Mater.* **2010**, *22*, 4112-4135.
19. Yoon, M.; Srirambalaji, R.; Kim, K. *Chem. Rev.* **2012**, *112*, 1196-1231.
20. Tanaka, K.; Oda, S.; and Shiro, M. *Chem. Commun.* **2008**, 820-822.
21. Ager, D. J.; Prakash, I.; Schaad, D. R. *Chem. Rev.* **1996**, *96*, 835-876.
22. Jacobsen, E. N. *Acc. Chem. Res.* **2000**, *33*, 421-431.
23. Nielsen, L. P. C.; Jacobsen, E. N. Catalytic Asymmetric Epoxide Ring-Opening Chemistry. In *Aziridines and Epoxides in Organic Synthesis*; Yudin, A., Ed.;

Wiley: New York, 2006.

24. Hodgson, D. M.; Gibbs, A. R.; Lee, G. P. *Tetrahedron* **1996**, *52*, 14361-14640.
25. Pastor, I. M.; Yus, M. *Curr. Org. Chem.* **2005**, *9*, 1-29.
26. Gigante, B.; Corma, A.; García, H.; Sabater, M. J. *Catal. Lett.* **2000**, *68*, 113-119.
27. Yamashita, H. *Chem. Lett.* **1987**, 525-528.
28. Yamashita, H. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 1213-1220.
29. Fu, X. L.; Wu, S. H. *Synth. Commun.* **1997**, *27*, 1677-1683.
30. Hou, X.-L.; Wu, J.; Dai, L.-X.; Xia, L.-J.; Tang, M.-H. *Tetrahedron: Asymmetry* **1998**, *9*, 1747-1752.
31. Sagawa, S.; Abe, H.; Hase, Y.; Inaba, T. *J. Org. Chem.* **1999**, *64*, 4962-4965.
32. Sekine, A.; Ohshima, T.; Shibasaki, M. *Tetrahedron* **2002**, *58*, 75-82.
33. Carrée, F.; Gil, R.; Collin, J. *Tetrahedron Lett.* **2004**, *45*, 7749-7751.
34. Carrée, F.; Gil, R.; Collin, J. *Org. Lett.* **2005**, *7*, 1023-1026.
35. Martin, M.; Bezzenine-Lafollée, S.; Gil, R.; Collin, J. *Tetrahedron: Asymmetry* **2007**, *18*, 2598-2605.
36. Kureshy, R. I.; Singh, S.; Khan, N. H.; Abdi, S. H. R.; Suresh, E.; Jasra, R. V. *Eur. J. Org. Chem.* **2006**, 1303-1309.
37. Kureshy, R. I.; Prathap, K. J.; Agrawal, S.; Khan, N. H.; Abdi, S. H. R.; Jasra, R. V. *Eur. J. Org. Chem.* **2008**, 3118-3128.
38. Kureshy, R. I.; Prathap, K. J.; Kumar, M.; Bera, P. K.; Khan, N. H.; Abdi, S. H. R.; Bajaj, H. C. *Tetrahedron* **2011**, *67*, 8300-8307.
39. Tak, R.; Kumar, M.; Kureshy, R. I.; Choudhary, M. K.; Khan, N. H.; Abdi, S. H. R.; Bajaj, H. C. *RCS Adv.* **2016**, *6*, 7693-7700.
40. Kumar, M.; Kureshy, R. I.; Saravanan, S.; Verma, S.; Jakhar, A.; Khan, N. H.; Abdi, S. H. R.; Bajaj, H. C. *Org. Lett.* **2014**, *16*, 2798-2801.

41. Bartoli, G.; Bosco, M.; Carlone, A.; Locatelli, M.; Massaccesi, M.; Melchiorre, P.; Sambri, L. *Org. Lett.* **2004**, *6*, 2173-2176.
42. Bao, H.; Zhou, J.; Wang, Z.; Guo, Y.; You, T.; Ding, K. *J. Am. Chem. Soc.* **2008**, *130*, 10116-10127.
43. Bao, H.; Wu, J.; Li, H.; Wang, Z.; You, T.; Ding, K. *Eur. J. Org. Chem.* **2010**, 6722-6726.
44. Bao, H.; Wang, Z.; You, T.; Ding, K. *Chin. J. Chem.* **2013**, *31*, 67-71.
45. Arai, K.; Salter, M. M.; Yamashita, Y.; Kobayashi, S. *Angew. Chem. Int. Ed.* **2007**, *46*, 955-957.
46. Arai, K.; Lucarini, S.; Salter, M. M.; Ohta, K.; Yamashita, Y.; Kobayashi, S. *J. Am. Chem. Soc.* **2007**, *129*, 8103-8111.
47. Azoulay, S.; Manabe, K.; Kobayashi, S. *Org. Lett.* **2005**, *7*, 4593-4595.
48. Ogawa, C.; Azoulay, S.; Kobayashi, S. *Heterocycles* **2005**, *66*, 201-206.
49. Ogawa, C.; Wang, N.; Boudou, M.; Azoulay, S.; Manabe, K.; Kobayashi, S. *Heterocycles* **2007**, *72*, 589-598.
50. Kokubo, M.; Naito, T.; Kobayashi, S. *Tetrahedron* **2010**, *66*, 1111-1118.
51. Kamal, A.; Ramu, R.; Azhar, M. A.; Khanna, G. B. R. *Tetrahedron Lett.* **2005**, *46*, 2675-2678.
52. Sun, J.; Dai, Z.; Yang, M.; Pan, X.; Zhu, C. *Synthesis* **2008**, 2100-2104.
53. Schneider, C.; Sreekanth, A. R.; Mai, E. *Angew. Chem. Int. Ed.* **2004**, *43*, 5691-5694.
54. Mai, E.; Schneider, C. *Synlett* **2007**, 2136-2138.
55. Gao, B.; Wen, Y.; Yang, Z.; Huang, X.; Liu, X.; Feng, X. *Adv. Synth. Catal.* **2008**, *350*, 385-390.
56. Bonollo, S.; Fringuelli, F.; Pizzo, F.; Vaccaro, L. *Synlett* **2008**, 1574-1578.

57. Plancq, B.; Ollevier, T. *Chem. Commun.* **2012**, 48, 3806-3808.
58. Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, 103, 227-249.
59. Field, M. J.; Bash, P. A.; Karplus, M. *J. Comp. Chem.* **1990**, 11, 700-733.
60. Gao, J.; Xia, X. *Science* **1992**, 258, 631-635.
61. Tuñón, I.; M. Martins-Costa, T. C.; Millot, C.; Ruiz-López, M. F.; Rivail, J. L. *J. Comp. Chem.* **1996**, 17, 19-29.
62. Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, 48, 1198-1229.
63. Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, 110, 949-1017.
64. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. *WIREs Comput. Mol. Sci.* **2012**, 2, 327-350.
65. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. *Chem. Rev.* **2015**, 115, 5678-5796.
66. Choomwattana, S.; Maihom, T.; Khongpracha, P.; Probst, M.; Limtrakul, J. *J. Phys. Chem. C* **2008**, 112, 10855-10861.
67. Oxford, G. A. E.; Snurr, R. Q.; Broadbelt, L. J. *Ind. Eng. Chem. Res.* **2010**, 49, 10965-10973.
68. Zheng, M.; Liu, Y.; Wang, C.; Liu, S.; Lin, W. *Chem. Sci.* **2012**, 3, 2623-2627.
69. Hirao, H.; Ng, W.; Moeljadi, A. M. P.; Bureekaew, S. *ACS Catal.* **2015**, 5, 3287-3291.
70. Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, 114, 10024-10035.
71. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, 111, 8551-8566.
72. Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, 111, 8566-8575.

73. Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576-8582.
74. Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. *Phys. Status Solidi B* **2013**, *250*, 1128-1141.
75. Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. *J. Chem. Theory Comput.* **2014**, *10*, 880-891.
76. Fang, Z.; Dürholt, J. P.; Kauer, M.; Zhang, W.; Lochenie, C.; Jee, B.; Albada, B.; Metzler-Nolte, N.; Pöppel, A.; Weber, B.; Muhler, M.; Wang, Y.; Schmid, R.; Fischer, R. A. *J. Am. Chem. Soc.* **2014**, *136*, 9627-9636.
77. Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431-439.
78. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. *J. Gaussian 09*, Gaussian, Inc., Wallingford, CT, USA, 2009.
79. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
80. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**,

37, 785-789.

81. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.
82. Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H. *J. Chem. Phys.* **1999**, *110*, 7667-7676.
83. Figgen, D.; Rauhut, G.; Dolg, M.; Stoll, H. *Chem. Phys.* **2005**, *311*, 227-240.
84. Peterson, K. A.; Puzzarini, C. *Theor. Chem. Acc.* **2005**, *114*, 283-296.
85. Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
86. Allinger, N. L.; Zhou, X.; Bergsma, J. *J. Mol. Struct. (THEOCHEM)* **1994**, *312*, 69-83.
87. Yang, Q.-Y.; Zhong, C.-L. *J. Phys. Chem. B* **2006**, *110*, 17776-17783.
88. Liu, J.; Rankin, R. B.; Johnson, J. K. *Mol. Simul.* **2009**, *35*, 60-69.
89. Tafipolsky, M.; Amirjalayer, S.; Schmid, R. *J. Phys. Chem. C* **2010**, *114*, 14402-14409.
90. Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendric, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1-28.
91. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165-253.
92. Smith, W.; Forester, T. R. *J. Mol. Graph.*, **1996**, *14*, 136-141.
93. Bakowies, D.; Thiel, W. *J. Phys. Chem.*, **1996**, *100*, 10580-10594.
94. Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.*, **1997**, *106*, 79-92.
95. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

96. Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571-2577.
97. Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829-5835.
98. Billeter, S. R.; Turner, J. A.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177-2186.
99. AIMAll (Version 13.05.06), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2013 (aim.tkgristmill.com).

Appendix

Appendix A Chapter 2

Table S1. ONIOM energy data for the first step of the conversion catalyzed by CYP19A1.

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	E(M06/B2,EE) [hartree]	$\Delta(E(B3LYP/B2,$ EE)+ZPE) [hartree]	$\Delta(E(M06/B2,$ EE)+ZPE) [hartree]
RCd	-2493.081640	7.304793	-3633.333252	-3631.829751	0.0	0.0
RCq	-2493.081283	7.304847	-3633.333065	-3631.828794	0.2	0.6
TS1d	-2493.041577	7.296767	-3633.303263	-3631.803762	13.8	16.3
TS1q	-2493.037934	7.296720	-3633.299301	-3631.799307	16.2	19.1
Int1d	-2493.053061	7.300320	-3633.317060	-3631.824082	7.4	3.6
Int1q	-2493.052560	7.300357	-3633.316500	-3631.823584	7.7	3.9
TS2q	-2493.050136	7.299784	-3633.315196	-3631.824377	8.2	3.4
Int2d	-2493.142913	7.308350	-3633.408261	-3631.918121	-44.8	-55.5
Int2q	-2493.151763	7.307741	-3633.414965	-3631.930277	-49.4	-63.1

Table S2. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data for the first step of the conversion catalyzed by CYP19A1.

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RCd	-3633.333252	-3614.836318	-18.496934
RCq	-3633.333065	-3614.836187	-18.496879
TS1d	-3633.303263	-3614.804010	-18.499253
TS1q	-3633.299301	-3614.798898	-18.500403
Int1d	-3633.317060	-3614.819764	-18.497296
Int1q	-3633.316500	-3614.819408	-18.497093
TS2q	-3633.315196	-3614.818662	-18.496534
Int2d	-3633.408261	-3614.911004	-18.497258
Int2q	-3633.414965	-3614.924406	-18.490559

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Table S3. ONIOM energy data for the second step of the conversion catalyzed by CYP19A1 following path I and II.

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	E(M06/B2,EE) [hartree]	Δ (E(B3LYP/B2, EE)+ZPE) [hartree]	Δ (E(M06/B2, EE)+ZPE) [hartree]
RCI	-2568.286993	7.311376	-3708.560047	-3707.02662	5.3	6.7
RCII	-2568.292576	7.311376	-3708.568431	-3707.03731	0.0	0.0
TS01I	-2568.25865	7.30575	-3708.536818	-3707.00662	16.3	15.7
TS01II	-2568.259670	7.304400	-3708.539595	-3707.01054	13.7	12.4
Int01I	-2568.28583	7.30867	-3708.56684	-3707.04428	-0.7	-6.1
Int01II	-2568.261300	7.306802	-3708.542739	-3707.01695	13.3	9.9
TS02II	-2568.253463	7.302720	-3708.536122	-3707.01017	14.8	11.6
Int02I	-2568.36374	7.314029	-3708.64004	-3707.11976	-43.3	-50.1
Int02II	-2568.347486	7.308686	-3708.633579	-3707.10959	-42.6	-47.0

Table S4. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data for the second step of the conversion catalyzed by CYP19A1 following path I and II.

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RCI	-3708.560047	-3690.073792	-18.486255
RCII	-3708.568431	-3690.087474	-18.480958
TS01I	-3708.536818	-3690.051685	-18.485133
TS01II	-3708.539595	-3690.060128	-18.479466
Int01I	-3708.566836	-3690.081286	-18.485550
Int01II	-3708.542739	-3690.064973	-18.477766
TS02II	-3708.536122	-3690.060613	-18.475510
Int02I	-3708.640039	-3690.161094	-18.478945
Int02II	-3708.633579	-3690.161163	-18.472416

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Table S5. ONIOM energy data for the second step of the conversion catalyzed by CYP19A1 following path III.

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	E(M06/B2,EE) [hartree]	Δ (E(B3LYP/B2, EE)+ZPE) [hartree]	Δ (E(M06/B2, EE)+ZPE) [hartree]
RCIII	-2644.455420	7.327342	-3784.765212	-3783.202976	0.0	0.0
TS01III	-2644.418382	7.323028	-3784.726988	-3783.170372	21.3	17.8
Int01III	-2644.421798	7.324122	-3784.728219	-3783.170462	21.2	18.4
TS02III	-2644.410598	7.319106	-3784.718341	-3783.156297	24.2	24.1
Int02III	-2644.526724	7.324249	-3784.843178	-3783.291112	-50.9	-57.2

Table S6. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data for the second step of the conversion catalyzed by CYP19A1 following path III.

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RCIII	-3784.765212	-3766.170918	-18.594294
TS01III	-3784.726988	-3766.140178	-18.586810
Int01III	-3784.728219	-3766.147345	-18.580874
TS02III	-3784.718802	-3766.132132	-18.586670
Int02III	-3784.718341	-3766.130869	-18.587472

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Table S7. ONIOM energy data for the second step of the conversion catalyzed by CYP19A1 following path III with two water molecules connecting Asp309 and Thr310.

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	E(M06/B2,EE) [hartree]	$\Delta(E(B3LYP/B2, EE)+ZPE)$ [hartree]	$\Delta(E(M06/B2, EE)+ZPE)$ [hartree]
RCIII'	-2797.275404	7.383854	-3937.658122	-3936.037869	0.0	0.0
TS01III'	-2797.240311	7.378898	-3937.625257	-3936.001056	17.5	20.0
Int01III'	-2797.244070	7.379752	-3937.628302	-3936.001648	16.1	20.2
TS02III'_a	-2797.242488	7.375770	-3937.627188	-3936.001853	14.3	17.5
TS02III'_b	-2797.242705	7.375809	-3937.627372	-3936.128171	14.2	-58.3
Int02III'_a	-2797.351302	7.381179	-3937.744947	-3936.001672	-56.2	17.7
Int02III'_b	-2797.249211	7.381356	-3937.632184	-3936.005185	14.7	18.9

Table S8. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data for the second step of the conversion catalyzed by CYP19A1 following path III with two water molecules connecting Asp309 and Thr310.

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RCIII'	-3937.658122	-3919.095646	-18.562476
TS01III'	-3937.625257	-3919.066968	-18.558289
Int01III'	-3937.628302	-3919.066972	-18.561331
TS02III'_a	-3937.627188	-3919.065436	-18.561752
TS02III'_b	-3937.627372	-3919.066118	-18.561254
Int02III'_a	-3937.744947	-3919.189621	-18.555326
Int02III'_b	-3937.632184	-3919.075246	-18.556939

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Table S9. ONIOM energy data for the third step of the conversion catalyzed by CYP19A1.

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	E(M06/B2,EE) [hartree]	Δ (E(B3LYP/B2, EE)+ZPE) [hartree]	Δ (E(M06/B2, EE)+ZPE) [hartree]
RCa	-2643.527107	7.316239	-3783.834404	-3782.279464	0.3	0.0
RCb	-2643.524894	7.315304	-3783.833983	-3782.274574	0.0	2.5
TS1a	-2643.499894	7.309552	-3783.811168	-3782.255616	10.7	10.8
TS1b	-2643.494788	7.310040	-3783.812363	-3782.258603	10.3	9.2
TS1c	-2643.484184	7.308940	-3783.800066	-3782.249680	17.3	14.1
Int1a	-2643.550840	7.313297	-3783.863752	-3782.311046	-19.9	-21.7
Int1a'	-2643.557559	7.311312	-3783.872340	-3782.313655	-26.6	-24.5
Int1c	-2643.509539	7.313235	-3783.828144	-3782.282561	2.4	-3.8
TS2a	-2643.536532	7.308305	-3783.858693	-3782.298725	-19.9	-17.1
TS2c	-2643.508810	7.311533	-3783.824993	-3782.280393	3.3	-3.5
Int2a	-2643.596477	7.313547	-3783.917085	-3782.363084	-53.3	-54.2
Int2b	-2643.618635	7.314506	-3783.939497	-3782.391682	-66.7	-71.5
Int2c	-2643.587331	7.317470	-3783.909095	-3782.364618	-45.8	-52.7
RCd	-2567.018844	7.285461	-3707.290623	-3705.764145	0.0	0.0
TS1d	-2566.995065	7.279396	-3707.275050	-3705.750130	6.0	5.0
TS1e	-2566.969736	7.277655	-3707.250191	-3705.728462	20.5	17.5
Int1d	-2567.015804	7.284050	-3707.297381	-3705.780902	-5.1	-11.4
Int1e	-2566.997271	7.282106	-3707.280467	-3705.763507	4.3	-1.7
Int2d	-2567.124876	7.288769	-3707.401007	-3705.887367	-67.2	-75.2

Table S10. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data for the third step of the conversion catalyzed by CYP19A1.

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RCa	-3783.834404	-3765.339264	-18.495139
RCb	-3783.833983	-3765.335664	-18.498319
TS1a	-3783.811168	-3765.316215	-18.494953
TS1b	-3783.812363	-3765.311374	-18.500990
TS1c	-3783.800066	-3765.303918	-18.496148
Int1a	-3783.863752	-3765.380148	-18.483604
Int1a'	-3783.872340	-3765.388212	-18.484129
Int1c	-3783.828144	-3765.333121	-18.495023
TS2a	-3783.858693	-3765.375864	-18.482829
TS2c	-3783.824993	-3765.329349	-18.495644
Int2a	-3783.917085	-3765.430192	-18.486894
Int2b	-3783.939497	-3765.442691	-18.496806
Int2c	-3783.909095	-3765.410972	-18.498123
RCd	-3707.290623	-3688.879270	-18.411353
TS1d	-3707.275050	-3688.871890	-18.403161
TS1e	-3707.250191	-3688.837534	-18.412657
Int1d	-3707.297381	-3688.890284	-18.407097
Int1e	-3707.280467	-3688.869303	-18.411164
Int2d	-3707.401007	-3688.997516	-18.403492

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Keto-enol tautomerization

(a)

(b)

Figure S1. Energy profiles (in kcal/mol) for keto-enol tautomerization in (a) the 1-water case and (b) the 2-water case, obtained at the B3LYP(SCRf(solvent=water))/6-311+G(d,p)//B3LYP/6-31G* level with zero-point energy corrections.

Appendix B Chapter 3

Table S1. Raw energy data

(a) B3LYP/B1 in pure DFT calculations

	ΔE [kcal/mol]	$\Delta(E+ZPE)$ [kcal/mol]	E [hatree]	ZPE [hatree]	E+ZPE [hatree]
² 1bB	0.0	0.0	-1563.515672	0.346144	-1563.169528
⁴ 1bB	1.1	1.0	-1563.513907	0.345986	-1563.167922
² TS_ooc	16.3	16.7	-1563.489671	0.346743	-1563.142928
⁴ TS_ooc	31.2	30.4	-1563.465905	0.344779	-1563.121126
² Int_ooc	5.3	7.0	-1563.507171	0.348753	-1563.158418
⁴ Int_ooc	23.4	24.0	-1563.478424	0.347148	-1563.131276
² TS_o-o	18.2	18.7	-1563.486731	0.347079	-1563.139652
² Int_o-o	13.8	13.5	-1563.493649	0.345671	-1563.147979
² TS_cyc	17.6	17.2	-1563.487613	0.345459	-1563.142154
² Int_cyc	6.1	6.5	-1563.506006	0.346847	-1563.159159
² TS_dis	18.6	17.3	-1563.485985	0.344024	-1563.141962
² Int_dis	-26.1	-27.4	-1563.557288	0.344153	-1563.213135

(b) B3LYP(SCRF)/B2//B3LYP/B1 in pure DFT calculations

	ΔE [kcal/mol]	$\Delta(E+ZPE)$ [kcal/mol]	E [hatree]	ZPE [hatree]	E+ZPE [hatree]
² 1bB	0.0	0.0	-2703.717931	0.346144	-2703.371787
⁴ 1bB	0.3	0.2	-2703.717465	0.345986	-2703.371479
² TS_ooc	14.1	14.5	-2703.695448	0.346743	-2703.348705
⁴ TS_ooc	30.9	30.0	-2703.668717	0.344779	-2703.323938
² Int_ooc	3.4	5.1	-2703.712434	0.348753	-2703.363681
⁴ Int_ooc	21.5	22.1	-2703.683726	0.347148	-2703.336578
² TS_o-o	13.1	13.7	-2703.697095	0.347079	-2703.350016
² Int_o-o	8.9	8.6	-2703.703716	0.345671	-2703.358045
² TS_cyc	10.1	9.7	-2703.701767	0.345459	-2703.356308
² Int_cyc	2.5	2.9	-2703.713967	0.346847	-2703.367119
² TS_dis	16.1	14.8	-2703.692242	0.344024	-2703.348219
² Int_dis	-29.5	-30.8	-2703.764986	0.344153	-2703.420833

(c) ONIOM energy data

	E(B3LYP/B1,ME) [hatree]	ZPE(B3LYP/B1,ME) [hatree]	E(B3LYP/B2,EE) [hatree]	$\Delta(E(B3LYP/B2,EE)+ZPE)$ [kcal/mol]
RC	-1576.750723	13.905959	-2717.382689	0.0
TS_ooc	-1576.715543	13.905448	-2717.352477	18.6
TS_ooc'	-1576.720916	13.907531	-2717.353465	19.3
Int_ooc	-1576.744590	13.907501	-2717.377364	4.3
Int_ooc'	-1576.737416	13.909172	-2717.368877	10.7
TS_o-o	-1576.726339	13.905821	-2717.367692	9.3
TS_o-o'	-1576.716908	13.907240	-2717.357712	16.5
Int_o-o	-1576.728434	13.905185	-2717.370844	6.9
Int_o-o'	-1576.732788	13.901320	-2717.384145	-3.8
Int_cyc	-1576.774709	13.905081	-2717.409548	-17.4
Int_cyc'	-1576.755700	13.903179	-2717.392450	-7.9
TS_dis	-1576.765280	13.902873	-2717.398398	-11.8
TS_dis'	-1576.738503	13.900743	-2717.380514	-1.9
Int_dis	-1576.840552	13.905540	-2717.444077	-38.8
Int_dis'	-1576.817960	13.904538	-2717.413968	-20.5

Table S2. Mulliken group spin populations**(a) B3LYP/B1 in pure DFT calculations**

	Fe	Op	Od	Por	O(Por)	His
² 1bB	0.28	0.80	0.90	-0.76	-0.21	-0.01
⁴ 1bB	0.34	0.79	0.89	0.78	0.22	0.00
² 1bA	0.00	0.31	0.69	0.00	0.00	0.00
² 1bA'	2.21	-0.31	-0.69	-0.22	-0.02	0.04
⁴ 1bA	2.20	0.32	0.69	-0.23	-0.02	0.04
⁴ 1bA'	3.86	-0.32	-0.69	0.11	-0.01	0.04
⁶ 1bA	3.77	0.33	0.69	0.18	-0.01	0.04
² TS _{_ooc}	0.41	0.45	0.36	-0.17	-0.05	0.00
⁴ TS _{_ooc}	0.88	0.51	0.35	1.11	0.15	-0.01
² Int _{_ooc}	0.87	0.17	0.02	-0.06	0.00	-0.01
⁴ Int _{_ooc}	0.88	0.18	0.05	1.79	0.10	0.00
² TS _{_o-o}	1.17	0.48	-0.54	-0.10	0.00	-0.01
² Int _{_o-o}	1.15	0.92	-0.69	-0.23	-0.13	-0.01
² TS _{_cyc}	1.16	0.90	-0.21	-0.85	0.01	-0.01
² Int _{_cyc}	1.18	0.86	-0.02	-1.00	-0.01	-0.01
² TS _{_dis}	1.18	0.86	0.00	-0.96	-0.08	-0.01
² Int _{_dis}	1.18	0.87	-0.06	-0.98	0.00	-0.01

(b) B3LYP(SCRf)/B2//B3LYP/B1 in pure DFT calculations

	Fe	Op	Od	Por	O(Por)	His
² 1bB	0.46	0.71	0.80	-0.79	-0.19	0.00
⁴ 1bB	0.52	0.67	0.77	0.85	0.19	-0.01
² 1bA	0.04	0.34	0.66	-0.03	0.00	0.00
² 1bA'	2.32	-0.33	-0.66	-0.36	-0.04	0.07
⁴ 1bA	2.33	0.33	0.66	-0.36	-0.03	0.07
⁴ 1bA'	3.82	-0.34	-0.66	0.14	-0.01	0.05
⁶ 1bA	3.80	0.34	0.66	0.17	-0.02	0.04
² TS _{_ooc}	0.61	0.33	0.22	-0.12	-0.03	0.00
⁴ TS _{_ooc}	0.77	0.39	0.23	1.50	0.13	-0.01
² Int _{_ooc}	0.88	0.16	0.01	-0.05	0.00	0.00
⁴ Int _{_ooc}	0.82	0.15	0.03	1.89	0.10	-0.01
² TS _{_o-o}	1.26	0.38	-0.48	-0.17	0.00	0.00
² Int _{_o-o}	1.21	0.88	-0.65	-0.29	-0.14	-0.01
² TS _{_cyc}	1.18	0.87	-0.15	-0.91	0.01	0.00
² Int _{_cyc}	1.22	0.83	-0.02	-1.02	-0.01	0.00
² TS _{_dis}	1.22	0.83	0.00	-0.97	-0.08	0.00
² Int _{_dis}	1.21	0.85	-0.06	-1.00	0.00	0.00

Table S3. Mulliken group charges**(a) B3LYP/B1 in pure DFT calculations**

	Fe	Op	Od	Por	Opor	His
² 1bB	0.62	-0.08	-0.05	-0.25	-0.46	0.23
⁴ 1bB	0.63	-0.09	-0.06	-0.25	-0.46	0.23
² 1bA	0.57	-0.15	-0.16	-0.08	-0.42	0.25
² 1bA'	0.66	-0.15	-0.16	-0.11	-0.42	0.19
⁴ 1bA	0.66	-0.15	-0.16	-0.11	-0.42	0.19
⁴ 1bA'	0.74	-0.15	-0.16	-0.22	-0.42	0.22
⁶ 1bA	0.69	-0.15	-0.16	-0.17	-0.41	0.21
² TS _{ooc}	0.61	-0.18	-0.20	-0.05	-0.44	0.26
⁴ TS _{ooc}	0.70	-0.17	-0.17	-0.21	-0.44	0.30
² Int _{ooc}	0.70	-0.29	-0.29	0.05	-0.44	0.26
⁴ Int _{ooc}	0.70	-0.29	-0.29	0.06	-0.45	0.26
² TS _{o-o}	0.74	-0.32	-0.34	0.12	-0.44	0.24
² Int _{o-o}	0.72	-0.35	-0.34	0.16	-0.39	0.20
² TS _{cyc}	0.72	-0.35	-0.50	0.35	-0.42	0.21
² Int _{cyc}	0.73	-0.37	-0.48	0.29	-0.37	0.20
² TS _{dis}	0.69	-0.37	-0.46	0.24	-0.31	0.20
² Int _{dis}	0.68	-0.36	-0.48	0.14	-0.17	0.21

(b) B3LYP(SCRf)/B2//B3LYP/B1 in pure DFT calculations

	Fe	Op	Od	Por	O(Por)	His
² 1bB	-1.98	-0.12	0.04	2.60	-0.36	-0.17
⁴ 1bB	-1.23	-0.19	0.03	2.03	-0.36	-0.28
² 1bA	-2.20	0.15	-0.17	1.86	-0.27	0.62
² 1bA'	-1.70	0.15	-0.17	1.59	-0.27	0.40
⁴ 1bA	-1.70	0.15	-0.17	1.60	-0.28	0.40
⁴ 1bA'	-1.79	0.14	-0.16	1.45	-0.25	0.62
⁶ 1bA	-1.62	0.15	-0.16	1.43	-0.27	0.47
² TS _{ooc}	-0.63	-0.35	0.00	1.46	-0.32	-0.16
⁴ TS _{ooc}	-0.89	-0.25	0.10	1.71	-0.31	-0.36
² Int _{ooc}	-1.08	-0.26	0.00	1.90	-0.29	-0.27
⁴ Int _{ooc}	-1.01	-0.25	0.01	1.96	-0.30	-0.40
² TS _{o-o}	-1.28	-0.28	-0.13	2.25	-0.29	-0.27
² Int _{o-o}	-0.04	-0.55	-0.07	1.32	-0.18	-0.49
² TS _{cyc}	0.01	-0.60	-0.20	1.34	-0.24	-0.29
² Int _{cyc}	0.10	-0.69	0.01	1.18	-0.22	-0.38
² TS _{dis}	0.02	-0.67	0.07	1.09	-0.19	-0.32
² Int _{dis}	-0.11	-0.62	-0.04	0.89	0.06	-0.18

Table S4. CO binding to ferrous peroxy radical

(a) B3LYP/B1

	E(B1) [au]	ZPE [au]	$\Delta E(B1)$ [kcal/mol]	$\Delta(E(B1)+ZPE)$ [kcal/mol]
^1CO	-113.306911	0.005036		
$^41bA'$	-1563.490003	0.344123		
$^41bA' + ^1\text{CO}$	-1676.796914	0.349159		
$^2[1bA\text{-CO}]$	-1676.817168	0.355678	0.6	1.5
$^2[\text{Int_o-o}\cdots\text{CO}]$	-1676.818090	0.354280	0.0	0.0

(b) B3LYP/B2-SCRF

	E(B2,solv) [au]	ZPE(B1) [au]	$\Delta E(B2,solv)$ [kcal/mol]	$\Delta(E(B2,solv)+ZPE)$ [kcal/mol]
^1CO	-113.349406	0.005036		
$^41bA'$	-2703.689105	0.344123		
$^41bA' + ^1\text{CO}$	-2817.038511	0.349159		
$^2[1bA\text{-CO}]$	-2817.057379	0.355678	2.6	3.5
$^2[\text{Int_o-o}\cdots\text{CO}]$	-2817.061486	0.354280	0.0	0.0

(c) Binding energy

	E [kcal/mol]	E+ZPE [kcal/mol]
B3LYP/B1	-11.8	-7.7
B3LYP(SCRF)/B2//B3LYP/B1	-12.7	-8.6

Appendix C Chapter 4

Table S1. Raw energy data

	E(B3LYP/B1,ME) [hartree]	ZPE(B3LYP/B1, ME) [hartree]	E(B3LYP/B2, EE) [hartree]	$\Delta(E(B3LYP/B2,$ EE)+ZPE) [hartree]
RC	-4242.418828	18.985450	-5686.979003	16.5
RC'	-4242.434323	18.983989	-5687.003849	0.0
TS01	-4242.415835	18.983956	-5686.976381	17.2
TS01'	-4242.430759	18.983788	-5686.995621	5.0
Int01	-4242.435445	18.986690	-5686.992118	9.1
Int01'	-4242.443247	18.985387	-5686.998335	4.3
TS02a	-4242.407493	18.987519	-5686.970890	22.9
TS02a '	-4242.425379	18.986467	-5686.989935	10.3
Int02a	-4242.414470	18.987141	-5686.987246	12.4
Int02a'	-4242.457465	18.988883	-5687.038808	-18.9
Int02a_2	-4242.471163	18.988341	-5687.019929	-7.4
Int02a'_2	-4242.472275	18.989075	-5687.040308	-19.7
TS02b	-4242.422094	18.986755	-5686.980263	16.5
TS02b'	-4242.434120	18.986021	-5686.994231	7.3
Int02b	-4242.433663	18.988581	-5686.995914	7.9
Int02b'	-4242.448041	18.988499	-5687.011056	-1.7
TS03b	-4242.429794	18.987996	-5686.994471	8.4
TS03b'	-4242.445793	18.987611	-5687.010588	-2.0
Int03b	-4242.444349	18.988732	-5687.010340	-1.1
Int03b'	-4242.460790	18.988298	-5687.026506	-11.5
RC''	-4241.604990	18.975497	-5686.200573	0.0
TS01''	-4241.600805	18.975596	-5686.191488	5.8
Int01''	-4241.622604	18.976996	-5686.206016	-2.5
TS02a ''	-4241.598916	18.975503	-5686.194632	3.7
Int02a''	-4241.601304	18.976119	-5686.201894	-0.4
Int02a''_2	-4241.660216	18.981953	-5686.263186	-35.2
TS02b''	-4241.612575	18.978025	-5686.203005	0.1
Int02b''	-4241.626606	18.979664	-5686.216980	-7.7
TS03b''	-4241.618629	18.978723	-5686.213101	-5.8
Int03b''	-4241.632591	18.980115	-5686.228890	-14.9
RC'''	-4303.446977	19.091715	-5748.002612	0.0
TS01'''	-4303.434936	19.089395	-5747.996505	2.4
Int01'''	-4303.459247	19.093526	-5748.014555	-6.4
TS02b'''	-4303.451429	19.090755	-5748.012861	-7.0
Int02b'''	-4303.503985	19.092560	-5748.069815	-41.6

Table S2. QM-MM energy components (ONIOM(B3LYP/B2:AMBER)-EE) data

	E(ONIOM) [hartree]	E(QM) [hartree]	E(QM-MM) [hartree]
RC	-5686.979003	-5639.226572	-47.752431
RC'	-5687.003849	-5639.236930	-47.766920
TS01	-5686.976381	-5639.229503	-47.746878
TS01'	-5686.995621	-5639.226587	-47.769035
Int01	-5686.992118	-5639.235852	-47.756266
Int01'	-5686.998335	-5639.236468	-47.761868
TS02a	-5686.970890	-5639.206847	-47.764043
TS02a '	-5686.989935	-5639.212181	-47.777755
Int02a	-5686.987246	-5639.235188	-47.752057
Int02a'	-5687.038808	-5639.257853	-47.780955
Int02a_2	-5687.019929	-5639.195707	-47.824222
Int02a'_2	-5687.040308	-5639.248443	-47.791865
TS02b	-5686.980263	-5639.234822	-47.745441
TS02b'	-5686.994231	-5639.235321	-47.758910
Int02b	-5686.995914	-5639.254629	-47.741285
Int02b'	-5687.011056	-5639.251865	-47.759191
TS03b	-5686.994471	-5639.253036	-47.741435
TS03b'	-5687.010588	-5639.255765	-47.754823
Int03b	-5687.010340	-5639.270267	-47.740072
Int03b'	-5687.026506	-5639.272958	-47.753548
RC''	-5686.200573	-5638.563028	-47.637545
TS01''	-5686.191488	-5638.553255	-47.638233
Int01''	-5686.206016	-5638.578657	-47.627359
TS02a ''	-5686.194632	-5638.559131	-47.635501
Int02a''	-5686.201894	-5638.567350	-47.634544
Int02a''_2	-5686.263186	-5638.615626	-47.647560
TS02b''	-5686.203005	-5638.563480	-47.639526
Int02b''	-5686.216980	-5638.578581	-47.638399
TS03b''	-5686.213101	-5638.576663	-47.636438
Int03b''	-5686.228890	-5638.597814	-47.631075
RC'''	-5748.002612	-5700.135154	-47.867458
TS01'''	-5747.996505	-5700.134589	-47.861916
Int01'''	-5748.014555	-5700.144492	-47.870062
TS02b'''	-5748.012861	-5700.158976	-47.853885
Int02b'''	-5748.069815	-5700.218838	-47.850978

* E(QM-MM) was defined as E(ONIOM) – E(QM).

Appendix D Chapter 5

Table S1 QM/MM and QM energies of key species in four possible pathways.

Relative energies shown in kcal/mol are with respect to (*R,R*)-1-RC

	RC	TS1	Int1	Int2	Int3	Int3-2	TS3	PC
<i>(R,R)</i> -1								
QM/MM	0.0	17.4	-2.7	-	-	-	-	-
QM	0.0	17.0	-4.5	-	-	-	-	-
<i>(S,S)</i> -1								
QM/MM	0.9	17.9	-2.6	-	4.0	4.2	4.2	-16.1
QM	1.8	18.8	-2.0	-	5.2	4.7	5.7	-14.9
<i>(R,R)</i> -2								
QM/MM	2.3	16.6	-4.0	-2.1	4.2	5.8	5.8	-16.1
QM	1.2	15.6	-5.7	-4.1	2.2	4.4	3.4	-16.8
<i>(S,S)</i> -2								
QM/MM	3.6	-	-	-	-	-	-	-
QM	4.4	-	-	-	-	-	-	-

Figure S1. Optimized structure of lithium 6-hydroxy-1-naphthoate.

Figure S2. Optimized structure of $\text{Cu}_2(\text{O}_2\text{CH})_4$.

Figure S3. Assigned atomic charges for the Cu-paddlewheel MOF.

Figure S4. QM/MM-optimized structures of (R,R) -1-TS1 and (R,R) -1-TS1 obtained with the B3LYP functional. Key distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

Figure S5. QM/MM-optimized structures of species in the proton transfer step. Interatomic distances are indicated in Å. Carbon atoms in the substrates and the MOF are colored magenta and light green, respectively.

List of Publications

1. Hirao, H.; Xu, K.; Chuanprasit, P.; Moeljadi, A. M. P.; Morokuma, K. Key Concepts and Applications of ONIOM Methods. In *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*; Tunon, I., Moliner, V., Royal Society of Chemistry: Cambridge, 2016. (Chapter 1)
2. Xu, K.; Wang, Y.; Hirao, H. Mechanistic Scenario of the Second Step in CYP19A1-Catalyzed Estrogen Formation Derived from Multiscale QM/MM Calculations. *in preparation*. (Chapter 2)
3. Xu, K.; Wang, Y.; Hirao, H. Estrogen Formation via H-Abstraction from the O-H bond of *gem*-Diol by Compound I in the Reaction of CYP19A1: Mechanistic Scenario Derived from Multiscale QM/MM Calculations. *ACS Catal.* **2015**, *5*, 4175-4179. (Chapter 2)
4. Xu, K.; Chuanprasit, P.; Hirao, H. Ferric-Superoxide Intermediate in the Verdoheme Formation Step of Heme Oxygenase. *in preparation*. (Chapter 3)
5. Xu, K.; Hirao, H. The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase: A QM/MM Study. *Submitted*. (Chapter 4)
6. Doitomi, K.; Xu, K.; Hirao, H. Mechanism of an Asymmetric Ring-Opening Reaction of Epoxide with Amine Catalyzed by a Metal-Organic Framework: Insights from Combined Quantum Mechanics and Molecular Mechanics Calculations. *Dalton Trans.* **2017**, *46*, 3470-3481. (Chapter 5)

Other Publications

1. Hirao, H.; Thellamurege, N. M.; Chuanprasit, P.; Xu, K. *Int. J. Mol. Sci.* **2013**, *14*, 24692-24705.
2. Wu, X.; See, J. W. T.; Xu, K.; Hirao, H.; Roger, J.; Hierso, J.-C.; Zhou, J. S. *Angew. Chem. Int. Ed.* **2014**, *53*, 13573-13577.
3. Jia, Y.-X.; Li, B.-B.; Li, Y.; Pullarkat, S. A.; Xu, K.; Hirao, H.; Leung, P.-H. *Organometallics* **2014**, *33*, 6053-6058.
4. Jia, Y.-X.; Yang, X.-Y.; Tay, W. S.; Li, Y.; Pullarkat, S. A.; Xu, K.; Hirao, H.; Leung, P.-H. *Dalton Trans.* **2016**, *45*, 2095-2101.
5. Wu, C.; Yue, G.; Nielsen, C. D.-T.; Xu, K.; Hirao, H.; Zhou, J. S. *J. Am. Chem. Soc.* **2016**, *138*, 742-745.
6. Kee, J. W.; Ng, Y. Y.; Kulkarni, S. A.; Muduli, S. K.; Xu, K.; Ganguly, R.; Lu, Y.; Hirao, H.; Soo, H. S. *Inorg. Chem. Front.* **2016**, *3*, 651-662.
7. Shi, Y. X.; Xu, K.; Clegg, J. K.; Ganguly, R.; Hirao, H.; Frišćić, T.; García, F. *Angew. Chem. Int. Ed.* **2016**, *55*, 12736-12740.
8. Amaniampong, P. N.; Karam, A.; Trinh, Q. T.; Xu, K.; Hirao, H.; Jérôme, F.; Chatel, G. *Sci. Rep.* **2017**, *7*, 40650.
9. Ong, D. Y.; Tejo, C.; Xu, K.; Hirao, H.; Chiba, S. *Angew. Chem. Int. Ed.* **2017**, *56*, 1840-1844.
10. Wu, J.; Xu, K.; Hirao, H.; Yoshikai, N. *Chem. Eur. J.* **2017**, *23*, 1521-1525.

Presentations

1. Xu, K.; Hirao, H. *Computational Studies of the Metabolism of Fluoxetine by Cytochrome P450 Enzyme*, 41st International Conference on Coordination Chemistry (ICCC-41), 21st-25th July, 2014 (poster).
2. Xu, K.; Wang, Y.; Hirao, H. *Estrogen Formation via H-Abstraction from the O-H bond of gem-Diol by Compound I in the Reaction of CYP19A1: Mechanistic Scenario Derived from Multiscale QM/MM Calculations*, The Chemical Bonds at the 21st Century (CB-2015), 14th-18th June, 2015 (poster).
3. Xu, K.; Wang, Y.; Hirao, H. *Estrogen Formation via H-Abstraction from the O-H bond of gem-Diol by Compound I in the Reaction of CYP19A1: Mechanistic Scenario Derived from Multiscale QM/MM Calculations*, MGMS Conference: Exploring Mechanisms in Biology: Theory and Experiment (CB-2015), 25th-27th Nov., 2015 (poster).
4. Xu, K.; Hirao, H. *The Mechanism of Mo-Cu CO Dehydrogenase*, Asian Network for Natural & Unnatural Materials (ANNUM), 8th-10th June, 2016 (poster).
5. Xu, K.; Hirao, H. *The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase: A QM/MM Computational Study*, The 14th International Symposium for Chinese Organic Chemists (ISCOC) & The 11th International Symposium for Chinese Inorganic Chemists (ISCIC), 8th-10th Dec., 2016 (poster).
6. Xu, K.; Hirao, H. *The Reaction Mechanism of Mo-Cu Carbon Monoxide Dehydrogenase: A QM/MM Computational Study*, 9th Singapore International Chemistry Conference (SICC9), 11th-14th Dec., 2016 (poster, **Energy & Environmental Science Poster Prize**).