2014

Computational Studies into Iron-Based Biocatalysis

Rami Gherib
University of Windsor

Computational Studies into Iron-Based Biocatalysis

By

Rami Gherib

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Chemistry and Biochemistry

in Partial Fulfillment of the Requirements for

the Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

Computational Studies into Iron-Based Biocatalysis

by

Rami Gherib

APPROVED BY:

_____

S. Goodwin
School of Computer Science

_____

J. Rawson
Department of Chemistry & Biochemistry

_____

J. Gauld, Advisor
Department of Chemistry & Biochemistry

May 16, 2014

Declaration of Co-Authorship

I. Co-Authorship Declaration

I hereby declare that this thesis material incorporates material that is a result of joint research as follows:

Chapter 3 was done in collaboration with Daniel Simard and Vanessa Susevski under the supervision of Prof. James W. Gauld.

Chapter 4 was done in collaboration with Dr. Eric A. C. Bushnell and Anthony J. Deschamps under the supervision Prof. James W. Gauld.

Chapter 5 was done in collaboration with Dr. Eric A. C. Bushnell and Anthony J. Deschamps under the supervision Prof. James W. Gauld.

Chapter 6 was done in collaboration with Dr. WenJuan Huang under the supervision Prof. James W. Gauld.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above material in my thesis.

I certify that, with the above qualifications, this thesis, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publications

This thesis included five original papers that have been previously published for publications in peer-reviewed journals as follows:

	Citation	Status
	Gherib, R.; Dokainish, H. M.; Gauld, J. W. *Int. J. Mol. Sci.* **2013**, *15*, 401.	Published
Chapter 3	Gherib, R.; Simard, D. J.; Susevski, V.; Gauld, J. W. *In preparation*	To be submitted
Chapter 4	Gherib, R.; Bushnell. A. C.; Deschamps, A. J.; Gauld, J. W. *In preparation*	To be submitted
Chapter 5	Gherib, R.; Bushnell. A. C.; Deschamps, A. J.; Gauld, J. W. *In preparation*	To be submitted
Chapter 6	Gherib, R.; Huang, W.-J.; Gauld, J. W. *In preparation*	To be submitted

I certify that I have obtained written permission from the copyright owners to include the above-published materials in my thesis. I certify that the above material describes work completed while registered as a graduate student at the University of Windsor.

I declare that to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violates any proprietary right and that the ideas techniques, quotations or any other material, are fully acknowledged in accordance with the standard practices. Furthermore, to the extend that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owners to include such material in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies Office, and that this thesis has not been submitted for a higher degree to any other university or institutions.

Abstract

Iron-containing enzymes are ubiquitous in nature and play several key roles in living organisms. Unfortunately, like many other types of enzymes, their catalytic mechanisms are often elusive. The work presented herein investigates computationally the catalytic processes of three iron-containing enzymes: glycerophosphosdiesterase from *Enterobacter aerogenes* (GpdQ), coral allene oxide synthase (cAOS) and S-ribosylhomocysteinase (LuxS). While some were investigated primarily using density functional theory (DFT), others used quantum mechanics/molecular mechanics (QM/MM) and molecular dynamics (MD) simulations in a complementary fashion.

Dedication

I dedicate this work to my mother Monia Ayadi, my father Faouzi Ghrib, my brother Amine Gherib and my girlfriend WenJuan Huang.

Acknowledgements

Academic research is not an easy thing to do because it mostly involves failing. For every successful result, there is usually behind it a very long line of failed attempts. It can often be demotivating and highly stressful. However my overall experience during these past 2 years has been extremely pleasant and I owe it mostly to the people around me. I want to take this opportunity to acknowledge them and thank them.

Firstly, I would like to thank my supervisor Prof. James Gauld. During my graduate studies he has always made himself available to help me resolve research-related issues. Problems that I would spend months trying to solve would be instantaneously resolved as soon as I'd describe them to him. He has been an excellent supervisor. I am very grateful for having been his student.

I would also like to thank Prof. Jeremy Rawson and Prof. Scott Goodwin for taking their time in reading my thesis.

I am excessively grateful for all past and present group members of the Gauld group. I would like to thank Dr. Eric Bushnell for all of his teachings about computational and quantum chemistry he has given me over the years. I would like to thank Grant Fortowsky for all the jokes and the humor he brought to the group. I would like to thank Bogdan Ion for his superhuman-like-tolerance of all the crazy things Eric, Grant and I did in the office. I would like to thank Daniel Simard for his tolerance of the crazy things Eric, Grant and I used to do with Bogdan as well as allowing me to steal his fruits. I would like to thank Hisham Dokainish for his frequent, sudden and comical outbursts during many afternoons. I would like to thank Mohamed Aboelnga for often teasing Hisham and only Hisham. I would like to thank Phil De Luna for his swag. I would also like to thank the undergraduate students with whom I worked with; Anthony Deschamps and Vanessa Susevski for helping me generate important results in a small amount of

time. I would also like to extend my thanks to Erum Kazim, Gyllian Porteous, Peter Boateng and Omid Beiraghi for their sunny personalities.

I would like to thank my family for their unconditional love and support. I am grateful to my mother, Monia Ayadi, for her tenderness, her affection and for reminding me to eat once in a while. I would like to thank my father, Faouzi Ghrib, for his devotion, for always looking out after my future and for all the conversations we had in the car to and from school. I would like to thank my brother, Amine Gherib, for his friendship, for his jokes, for changing my mind off research when I got home. From the bottom of my heart, I would to thank my wonderful girlfriend WenJuan Huang for all of the support she has given me in these past years. She is by far the best thing I have acquired from joining the Gauld group.

Finally I would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), SHARCNET and Compute Calcul Canada for additional computational resources.

Table of Content

Chapter 3

The Different Catalytic Rate of Fe(II)/M(II)-GpdQ Explained with DFT

Chapter 4

Modeling Compound I in Coral Allene Oxide Synthase: A Challenge for Density Functional Theory

Chapter 5

Compound I Reactivity in Coral Allene Oxide Synthase: A QM/MM Study

Chapter 6

Insights from MD and QM/MM into the Biocatalysis of LuxS

Chapter 7

Conclusions

List of Tables

List of Figures

List of Schemes

List of Appendices (on CD)

Table A1.Cartesian coordinates of optimized structure of $N/W_T/O_B$, $P/OT/W_B$, $N/W_T/W_B$, $N/W_T/O_B$, $N/O_T/O_B$, $P/W_T/W_B$ and $P/W_T/O_B$.

Table A2.Cartesian coordinates of optimized structures of $RC–W_T/W_B$, $RC–W_T/O_B$, $RC–O_T/O_B$, $RC–W_B$ and $RC–O_B$.

Table A3.Cartesian coordinates of optimized structure of species along the mechanistic pathway of Fe(II)/M(II)-GpdQ with neutral His81.

Table A4. Cartesian coordinates of optimized structure of species along the mechanistic pathway of Fe(II)/M(II)-GpdQ with positively charged His81.

Table B1. Cartesian coordinates of Cpd I/cAOS optimized with 6-31G, 6-31G(d), 6-31G(d,p), 6-31+G, 6-31+G(d), 6-31+G(d,p), BS1, B2 and BS3 (see text).

Table B2. Cartesian coordinates of Cpd/cAOS optimized with ONIOM(**DFT_i**/6-31G(d,p):AMBER96) where **DFT_i**=BLYP, B3LYP*, B3LYP, M06, M06-L.

Table C1. Cartesian coordinates of optimized structures of species along the hydroxylation catalytic mechanism of arachidonic acid by Cpd I/cAOS.

Table C2. Cartesian coordinates of optimized structures of species along the epoxydation catalytic mechanism of arachidonic acid by Cpd I/cAOS.

Table D1. LuxS annealing protocol.

Table D2. Cartesian coordinates of optimized structures at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory. Single point energies were obtained at ONIOM(6-311+G(2df,p):AMBER96)–ME level of theory.

List of Abbreviations and Symbols

8R, 9S-EET	8R,9S-epoxyeicosatetranoic acid
8R-HPETE	8-hydroperoxyeicosatetreaenoic acid
8R-HETE	8-hydroxyperoxyeicosatetreaenoic acid
AI-2	Autoinducer-2
AO	Atomic orbital
bpNPP	Bisparanitrophenol
BO	Born-Oppenheimer
BS	Basis set
cAOS	Coral allene oxide synthase
Cpd 0	Compound 0
Cpd I	Compound I
Cpd II	Compound II
DFT	Density functional theory
DMP	Dimethylphosphate
DPD	Dihydroxy-2,3-pentadione
GGA	Generalized gradient approximation
GpdQ	Glycerophosphodiesterase
HF	Hartree-Fock
HK	Hohenberg-Kohn
HOMO	Highest occupied molecular orbital
HSAB	Hard soft acid base
IEFPCM	Integral equation formalism polarization continuum method

LCAO	Linear combination of atomic orbital
LDA	Local density approximation
LUMO	Lowest occupied molecular orbital
LuxS	S-ribosylhomocysteinase
MD	Molecular dynamics
MM	Molecular mechanics
MO	Molecular orbital
MOE	Molecular operating environment
MSR	Multi-state reactivity
ONIOM	Our Own N-Integrated Molecular Orbital Molecular Mechanics
OP	Organophosphate
PC	Product complex
PES	Potential energy surface
PhIO	Iodosylbenzene
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
QS-2	Quorum sensing-2
RC	Reactant complex
RMSD	Root mean square deviation
SCF	Self-consistent field
SRH	S-ribosylhomocysteine
TDSE	Time-dependent Schrodinger equation
TISE	Time-independent Schrodinger equation

Chapter 1

Introduction

Gold is for the mistress — silver for the maid —
Copper for the crafsman cunning at his trade.
"Good!" said the Baron, sitting in his hall,
"But Iron — Cold Iron — is master of them all."

Cold Iron
Rudyard Kipling

1.1 Biological Iron

After aluminum, iron is the second most abundant metal in the earth's crust.[1] In its earliest stages, life was mainly devoid of oxygen and employed Fe^{II}.[2] However, the emergence of O_2 in the later stages of life enabled the oxidation of Fe^{II} to Fe^{III}; an essential process for many life forms.[2] Nearly all organisms require iron for survival.[2,3] The latter performs a variety of biological processes including for example oxygen transport,[4] reduction of ribonucleotides,[5] electron transfers,[6] oxygen activation,[7] peroxide decomposition[8] and phosphoester bond hydrolysis.[9] Nature relies on iron due to its chemical versatility and abundance. Iron can acquire many oxidation states (from -2 to +6), coordination geometries (octahedral, trigonal bipyramidal and tetrahedral) and spin states.[10]

1.2 Catalysis

Catalysts are compounds that increase the rate of chemical reactions without being consumed.[11] There exist different types of catalysts including organometallic compounds,[12] nanoparticles,[13] zeolites[14] and enzymes.[15] From a practical point-of-view, catalysts are very important industrially since approximately 80% of all chemical and pharmaceutical products are synthesized with their aid.[1] Notably, organometallic compounds are prominent catalysts used in a variety of applications. Unfortunately, many of them use precious metals such as palladium, rhodium and iridium, which can often be costly and toxic.[1,16] The use of iron-based organometallic catalysts is advantageous due to its low cost and availability.[1] However, iron-based catalysis is currently limited partly due to the lack of insights concerning their mechanism.[17]

1.3 The Efficiency of Enzymes

Enzymes are biological catalysts and critical for life-sustaining processes. They can often enhance the rate of chemical reactions by $\approx 10^{15}$ fold.[18] For example, the non-

enzymatic hydrolysis of urea at 20.8°C occurs at a rate of 3×10^{-10} s^{-1}. But when catalyzed by urease at the same temperature, this reaction occurs at a rate of 3×10^{4} s^{-1}.[15] Others, like cysteinyl tRNA synthetase are highly specific and are capable of discriminating serine from cysteine by a factor of 10^{8}.[19] Understandably, their catalytic power and their incredible specificity allow them to serve as models for synthetic catalysis.[17] The insights acquired from studying their mechanisms, can be used for practical purposes including designing novel synthetic catalysts.

1.4 Iron-Based Enzymes

Iron-containing enzymes are ubiquitous in nature. Many have a catalytic prowess that remains unrivalled by their synthetic counterparts. Indeed, some like cytochrome P450 use O_2 to catalyze C–H and C=C oxidation with high regiospecificity.[7,20] Others like Purple Acid Phosphatases and Glycerophosphodiesterase (GpdQ) are binuclear metalloenzymes and catalyze the hydrolysis of phosphate esters.[21] Certain enzymes can catalyze the production of unstable chemical species' like coral allene oxide synthase (cAOS),[22-24] which converts fatty acid hydroperoxide to allene oxide. While some require redox active cofactors, eg. Nitric Oxide Synthase,[25] others including S-ribosylhomocysteinase (LuxS) do not.[26,27] Some iron-containing enzymes, like catalase, are one of the most efficient and can decompose 10^{6} H_2O_2 molecules per second.[28]

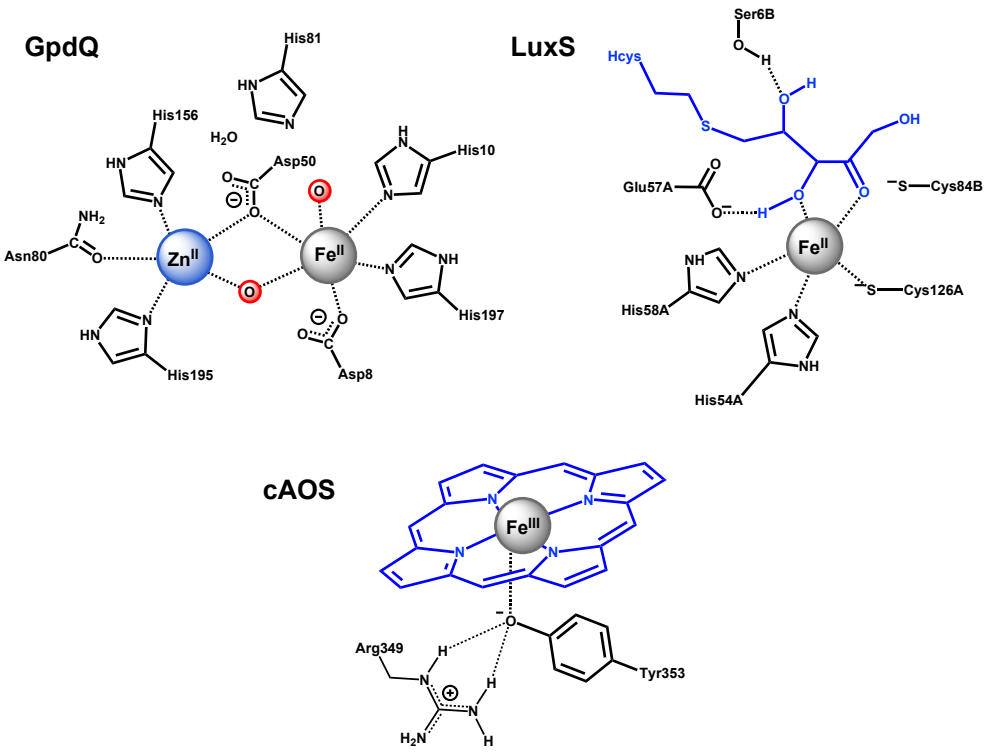


Figure 1.1. Schematic representation of selected Fe–containing enzymatic active sites.

1.5 Computational Enzymology

Many enzymes increase the rate of reactions by providing alternative pathways involving lower energy barriers. Hence, to understand the origin of the catalytic efficiency of enzymes, it is often important to characterize the alternative pathway it provides. This involves elucidating the nature of the intermediates and transitions states formed during the enzymatic reaction. Unfortunately many of these species are unstable, short-lived or not easily characterized by experimental methodologies.

One of the strengths of computational chemical methods, such as density functional theory (DFT)[29] and quantum mechanics/molecular mechanics (QM/MM),[30] is their ability to model and characterize intermediates and transition states formed during chemical reactions. Also, their ability to compute relative free energies allows them to evaluate the kinetic and thermodynamic properties of reactions. Other methods, including molecular dynamics (MD) simulation,[31] can also model the structural dynamics of enzymes and

describe the interactions between active sites and substrates. It is no wonder why in recent years computational enzymology has drastically enhanced our understanding of biocatalysis.[32,33]

1.6 Computational Investigations into Iron-Based Biocatalysis

One of the greatest challenges of modern-day biochemistry and biophysics is to determine the origin of enzyme's incredible catalytic power.[33,34] Despite many years of research in the topic, our understanding of biocatalysis at the molecular level remains incomplete. Many theories on enzymes' rate-enhancing abilities have been advanced.[32] Indeed, while some are quantum-mechanical in nature (e.g. orbital steering[35] and quantum tunneling[36]) others lean towards macromolecular levels (e.g. enzyme conformational dynamics).[37] Consequently, current studies into biocatalysis often employ a multiscale approach.[33,38]

It is the objective of the work presented herein to elucidate some of the catalytic strategies employed by iron-containing enzymes and the role of iron in these strategies. While some studies use primarily a DFT-cluster approach, others synergistically employ MD simulations and QM/MM methods. The mechanisms of iron heme, mononuclear non-heme and binuclear non-heme enzymes are investigated. The culmination of the results and analyses presented herein contribute in providing a systematic rationalization on how iron-enzymes reduce activation barriers in biocatalysis.

1.7 References

(1) Enthaler, S.; Junge, K.; Beller, M. *Angew. Chem. Int. Ed.* **2008**, *47*, 3317.

(2) Pierre, J. L.; Fontecave, M.; Crichton, R. R. *Biometals* **2002**, *15*, 341.

(3) Beard, J. L.; Dawson, H.; Piñero, D. J. *Nutr. Rev.* **1996**, *54*, 295.

(4) Lehninger, A.; Nelson, D. L.; Cox, M. M. *Lehninger's Principles of Biochemistry*; 4 ed.; New York, New York: WH Freeman and Company, 2005.

(5) Reichard, P. *Science* **1993**, *260*, 1773.

(6) Breuer, M.; Rosso, K. M.; Blumberger, J. *Proc. Natl. Acad. Sci.* **2014**, *111*, 611.

(7) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279.

(8) Halliwell, B.; Gutteridge, J. *Biochem. J.* **1984**, *219*, 1.

(9) Vogel, A.; Spener, F.; Krebs, B. *Encyclopedia of Inorganic and Bioinorganic Chemistry* **2001**.

(10) Plietker, B. *Iron Catalysis in Organic Chemistry: Reactions and Applications*; 1 ed.; Wiley-VCH: Weinheim, 2008.

(11) Coontz, R.; Fahrenkamp-Uppenbrink, J.; Szuromi, P. *Science* **2003**, *299*, 1683.

(12) Michalak, A.; Ziegler, T. *Organometallics* **1999**, *18*, 3998.

(13) Bell, A. T. *Science* **2003**, *299*, 1688.

(14) Karge, H. G.; Weitkamp, J. *Zeolites as Catalysts, Sorbents and Detergent Builders: Applications and Innovations*; Elsevier, 1989.

(15) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications: Toronto, 1987.

(16) Bullock, R. M. *Angew. Chem. Int. Ed.* **2007**, *46*, 7360.

(17) Fürstner, A. *Angew. Chem. Int. Ed.* **2009**, *48*, 1364.

(18) Lewis, C. A.; Wolfenden, R. *Proc. Nat. Acad. Sci.* **2008**, *105*, 17328.

(19) Zhang, C.-M.; Christian, T.; Newberry, K. J.; Perona, J. J.; Hou, Y.-M. *J. Mol. Biol.* **2003**, *327*, 911.

(20) Guengerich, F. P. *J. Biol. Chem.* **2013**, *288*, 17063.

(21) Schenk, G.; Mitić, N.; Hanson, G. R.; Comba, P. *Coord. Chem. Rev.* **2013**, *257*, 473.

(22) Bushnell, E. A. C.; Gherib, R.; Gauld, J. W. *J. Phys. Chem. B* **2013**, *117*, 6701.

(23) Oldham, M. L.; Brash, A. R.; Newcomer, M. E. *Proc. Natl. Acad. Sci.* **2005**, *102*, 297.

(24) Brash, A. R. *Phytochemistry* **2009**, *70*, 1522.

(25) Tejero, J.; Stuehr, D. *IUBMB Life* **2013**, *65*, 358.

(26) Pei, D.; Zhu, J. *Curr. Opin. Chem. Biol.* **2004**, *8*, 492.

(27) Huang, W.; Gherib, R.; Gauld, J. W. *J. Phys. Chem. B* **2012**, *116*, 8916.

(28) Vainshtein, B. K.; Melik-Adamyan, W. R.; Barynin, V. V.; Vagin, A. A.; Grebenko, A. I. *Nature* **1981**.

(29) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651.

(30) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(31) Karplus, M.; McCammon, J. A. *Nat. Struc. Mol. Biol.* **2002**, *9*, 646.

(32) Llano, J.; Gauld, J. W. Mechanistics of Enzyme Catalysis: From Small to Large Active−Site Models, In Quantum Biochemistry:Electronic Structure and Biological Activity; Matta, C. F., Ed.; Wiler-VCH: Weinheim, **2010**; Vol. 2, p 920.

(33) Gherib, R.; Dokainish, H. M.; Gauld, J. W. *Int. J. Mol. Sci.* **2013**, *15*, 401.

(34) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.

(35) Storm, D. R.; Koshland, D. E. *Proc. Natl. Acad. Sci.* **1970**, *66*, 445.

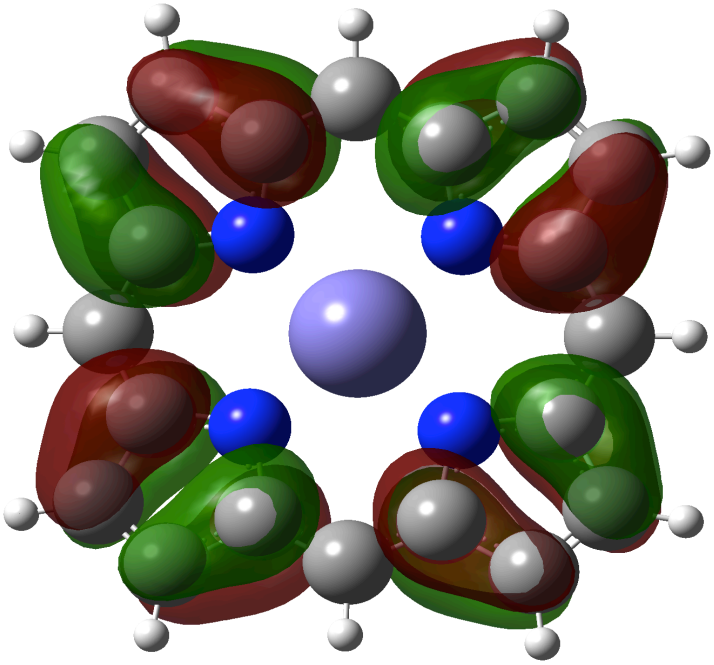
(36) Masgrau, L.; Roujeinikova, A.; Johannissen, L. O.; Hothi, P.; Basran, J.; Ranaghan, K. E.; Mulholland, A. J.; Sutcliffe, M. J.; Scrutton, N. S.; Leys, D. *Science* **2006**, *312*, 237.

(37) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422.

(38) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41.

Chapter 2

Theoretical Methods



2.1 Introduction

The last several decades have made computational chemistry a required part of chemical research.[1] Today theoretical models are to be considered with the same weight as spectroscopic methods.[2] The foundations of computational chemistry are rooted in quantum chemistry. Many of the theoretical methods used in this thesis, such as density functional theory, are quantum mechanical in nature. Some, like molecular mechanics, although being essentially empirical have been designed to reproduce quantum mechanical methods. Hence this chapter will provide a brief summary of the methods and their underlying concepts used throughout this work.

2.2 General Principles of Quantum Chemistry

2.2.1 The Time-Dependent Schrödinger Equation

Quantum mechanics (QM) provides mathematical methodologies allowing for the properties of systems to be calculated. Many conventional *ab initio* methods involve the mathematical manipulation of a system's wavefunction. The wavefunction inherently contains all information concerning the system in question.

For many systems, the wavefunction as well as their properties, change throughout time. This is the case for atoms/molecules subjected to electromagnetic radiation. The time-dependent Schrödinger equation (TDSE) describes how wavefunctions change throughout time and allows future wavefunctions to be calculated starting from an initial wavefunction (see Equation 2.1).

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \widehat{H}\Psi(r,t) \qquad \textbf{(2.1)}$$

The term $\widehat{H}$ is the Hamiltonian operator (see Equation 2.2). For any given system, the eigenvalues of $\widehat{H}$ correspond to the possible values of the system's total energy (E). It is composed of a kinetic energy term, $\frac{-h^2}{2m}\nabla^2$ and a time-dependent potential energy term $V(r,t)$:

$$\widehat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(r,t) \qquad \textbf{(2.2)}$$

In cases where the potential energy does not change in time, the TDSE simplifies to the time-independent Schrödinger (TISE) equation (see Equation 2.3) where the potential energy term is solely spatially dependent (see equation 2.4):

$$\widehat{H}\Psi(r) = E\Psi(r) \quad \textbf{(2.3)}$$

$$\widehat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(r) \quad \textbf{(2.4)}$$

Solving the TISE can be seen as a partial differential problem or as an eigenvalue problem where the total energy of the system corresponds to the eigenvalue of $\widehat{H}$ and $\Psi(r)$ corresponds to its eigenfunction.

2.2.2 The Born-Oppenheimer Approximation

For molecules, the field free non-relativistic Hamiltonian operator takes the following general form (see equation 2.5):

$$\widehat{H} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{\hbar^2}{2m_N}\sum_N \nabla_N^2 + \sum_{i>j}\frac{1}{r_{ij}} + \sum_{A>B}\frac{Z_A Z_B}{R_{AB}} - \sum_{i,A}\frac{Z_A}{r_i A} \quad \textbf{(2.5)}$$

The total energy of a molecule can be expressed as a sum of other energetic contributions:

- Kinetic energy of electrons (first sum in equation 2.5)
- Kinetic energy of nuclei (second sum in equation 2.5)
- Coulombic repulsion between electrons (third sum in equation 2.5)
- Coulombic repulsion between nuclei (fourth sum in equation 2.5)
- Coulombic attraction between electrons and nuclei (last sum in equation 2.5)

The Born-Oppenheimer (BO) approximation posits that due to their significantly higher mass, nuclei travel much slower than electrons. Using a classical analogy; during their cycle of electronic motion, the change in nuclear position is negligible. Hence to a reasonable approximation, the Hamiltonian operator (equation 2.5) can be simplified by omitting the nuclear kinetic energy terms and fixing internuclear distances. The latter

allows the energy due to nuclear repulsions to be treated as a constant rather than a part of the Hamiltonian. The BO approximation yields the electronic Hamiltonian operator (see equation 2.6):

$$\widehat{H}_E = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_{i>j}\frac{1}{r_{ij}} - \sum_{i,A}\frac{Z_A}{r_iA} \quad \textbf{(2.6)}$$

2.2.3 Atomic and Molecular Orbitals

The second term in the electronic Hamiltonian (see equation 2.6) makes the electronic Schrodinger equation unsolvable for multielectronic systems. To a reasonable approximation we can assume that each electron moves independently of one another. This allows each electron to be associated to its own one-electron orbital. An atomic orbital (AO) is the wavefunction of an electron in an atom (ϕ_i). A molecular orbital (MO) is the wavefunction of an electron in a molecule. The linear combination of atomic orbital approximation asserts that MOs (ψ) can be expressed as a linear combination of atomic orbitals (LCAO) (see equation 2.7) where the terms c_i are expansion coefficients and correspond to the contribution each AO has on the MO (see equation 2.7).

$$\psi(r) = \sum_{i=1} c_i\phi_i(r) \quad \textbf{(2.7)}$$

2.2.4 The Slater Determinant

Because of their half-integral spin angular momenta, the total wavefunction of multi-electronic systems requires antisymmetric wavefunctions. These obey the requirement that, upon interchanging the spatial and spin coordinates of two particles, their sign changes (see eq.

$$\Psi(r_1, r_2, \dots, r_N) = -\Psi(r_2, r_1, \dots, r_N) \quad \textbf{(2.8)}$$

The antisymmetry requirement is also called the Pauli principle. In non-relativistic QM, the Pauli principle is treated as an additional postulate. It has the important consequence of completely preventing electrons form having identical spatial and spin coordinates.

This consequence, also called the Pauli exclusion principle, forces electrons of same spin to keep apart from one another, which ultimately reduces the total energy of the system.

The Slater determinant satisfies the requirement of antisymmetry. The components of a Slater determinant are spin-orbitals, which are the product of a MO and a one-electron spin function. For a system with N electrons, the Slater determinant is as follows (see equation 2.9), where the $\frac{1}{\sqrt{N!}}$ term allows the Slater determinant be normalized.

$$\Psi(r_1, r_2, \ldots, r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \cdots & \psi_N(1)\alpha(1) \\ \psi_1(2)\alpha(2) & \psi_1(2)\beta(2) & \cdots & \psi_N(2)\alpha(2) \\ \vdots & \ddots & \ddots & \vdots \\ \psi_1(N)\alpha(N) & \psi_1(N)\beta(N) & \cdots & \psi_N(N)\alpha(N) \end{vmatrix} \quad \textbf{(2.9)}$$

The α and β terms correspond to the eigenfunctions of the total spin angular-momentum operator. They are also eigenfunctions of the z-axis spin angular momentum with eigenvalues of $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ respectively.

2.2.5 The Variational Theorem

In the framework of time-independent wavefunction based QM, the variational theorem asserts that if ϕ is any well-behaved wavefunction that obeys the same boundary of the exact wavefunction and where E_0 is the lowest-energy eigenvalue of the corresponding Hamiltonian operator then:

$$E_0 \leq E_\phi = \frac{\int \phi^* \hat{H} \phi \, d\tau}{\int \phi^* \phi \, d\tau} \quad \textbf{(2.10)}$$

It should be noted that the notation for the lowest-energy eigenvalue (i.e. E_0) in equation 2.10 has been used repeatedly in numerous quantum chemistry textbooks.[3-6]

2.3 Hartree-Fock Method

The Hartree-Fock (HF) method is the corner stone of conventional *ab initio* quantum chemical calculations. It assumes that electrons are moving in a smeared electrostatic potential created by the average instantaneous interactions between electrons.

Furthermore, based on the central-field approximation, it assumes that the effective potential acting on an electron can be approximated by a function of radii only, while averaging over the angles. The wavefunctions used in HF theory are Slater determinants whose MO components are LCAOs. The expansion coefficients (see section 2.2.3) are optimized to minimize the variational integral.

From this process emerge the Roothaan-Hall equations, which can be written in matrix form (see equation 2.11). It possesses a Fock matrix operator (F), a coefficient matrix (C), an overlap matrix (S) and the orbital energies matrix (ε). The components of the C and ε matrices correspond not only to the eigenvectors and eigenvalues of the Fock matrix but also to the optimized expansion coefficients and corresponding orbital energies.

$$FC = SC\varepsilon \tag{2.11}$$

The components of F are themselves functionals of AOs. Therefore, prior to deriving the component of C and ε, an initial set of expansion coefficients is required to generate the Fock matrix operator. The alternating process of deriving new expansion coefficients before inserting them into the Fock matrix operator is repeated until the variational energy ceases to change. At such point, the calculation is said to have reach self-consistency. This process is called the self-consistent field (SCF) cycle.

2.4 Density Functional Theory

2.4.1 The Electron Density

The electron density describes the probability of finding any one of the electrons within a certain volume at a specific distance from the molecular nuclear framework. This quantity can be derived from the modulus squared of the wavefunction and integrating over N-1 electron coordinates (see equation 2.12). It should be noted that unlike the wavefunction, the electron density always has three variables independent from the size of the system.[7]

$$\rho(r) = N \int \cdots \int |\Psi(r_1, r_2, \cdots r_N)|^2 dr_2 \cdots dr_N \tag{2.12}$$

In density functional theory (DFT), the properties of atoms and molecules are calculated from their corresponding electron densities. In application, the electron density of the overall system is the sum of the electron density of spin up electrons and spin down electrons. These are expressed as the sum of alpha and beta MOs respectively (see equation 2.13).[8]

$$\rho(r) = \rho_\alpha(r) + \rho_\beta(r) = \sum_{i\epsilon\alpha} \psi_i^2\,(r) + \sum_{i\epsilon\beta} \psi_i^2(r) \qquad \textbf{(2.13)}$$

DFT not only has rigorous theoretical foundations but also tends to provide results that can rival post-HF methods in terms of accuracy at the fraction of their computational costs. For such reasons, it is currently the workhorse of most computational chemical studies.[9]

2.4.2 Hohenberg-Kohn Theorems

The Hohenberg-Kohn (HK) theorems use some of the basic foundations of quantum mechanics to validate DFT. The first HK theorem demonstrates that the eigenfunctions (i.e. wavefunctions) of two Hamiltonian operators with differing external potentials can never produce the same electron densities.[10] In time-independent adiabatic DFT, the external potential is generated from the nuclear charges and their spatial positions. Hence, from the first HK theorem, it stems that for any given electron density, there exists a single nuclear framework whose wavefunctions correspond to it. The ramification of the first HK theorem is that any ground state properties of a molecule is a functional of the ground state electron density. The second HK theorem is analogous to the variational theorem. It states that any trial electron density (ρ_t) will give an energy value (E_t) higher than the true ground state energy (E_0) (see equation 2.14).[10]

$$E_0[\rho_0] \leq E_t[\rho_t] \qquad \textbf{(2.14)}$$

2.4.3 Local Density Approximation

In DFT, the properties of atoms and molecules are calculated from functionals of the electron density. One particularly important functional is the one that calculates the energy of atoms and molecules. Unfortunately the exact mathematical expression of such functionals is unknown and current DFT methods rely on approximative functionals. The local density approximation (LDA) assumes that the electron density can be treated as a uniform electron gas at a given point. LDA functionals depend only on the electron density. The Thomas-Fermi,[11,12] Dirac[13] and Vosko-Wilk-Nusair[14] functionals use the LDA. They calculate respectively the kinetic, exchange and correlation energies. The Dirac exchange energy functional is provided underneath (see equation 2.15):[13]

$$E_X^D = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int \rho_\alpha^{4/3}(r)dr \qquad \textbf{(2.15)}$$

2.4.4 Generalized Gradient Approximation

Generalized gradient approximations (GGA) DFT methods employ functionals that depend not only the electron density but also on their gradients. Compared to LDA functionals, which perform poorly for systems with nonhomogeneous electron densities, such as atoms and molecules, many GGA functional perform relatively well.[5] Indeed, the LDA underestimates the exchange energy by 10%.[15] Many GGA functionals calculate more accurately exchange and correlation energies contributions, which overall provide different but cost-effective ways of considering electron correlation compared to wavefunction-based methodologies.[8]

Arguably, one of the most popular GGA functionals at the present time is the Becke 88 exchange functional.[15] It has the capacity to reproduce the same exchange energies as those of the exact Fock exchange. It remains highly potent and widely used in computational chemical studies.[5,8,16] In its earliest forms, Becke 88 adopted the following format (see equation 2.16):

$$E_x^B = E_x^D - 0.0042 \int {\rho_\alpha}^{4/3}(r) \frac{\left(\frac{|\nabla\rho_\alpha(r)|}{{\rho_\alpha}^{4/3}(r)}\right)^2}{1+0.024\left(\frac{|\nabla\rho_\alpha(r)|}{{\rho_\alpha}^{4/3}(r)}\right) sinh^{-1}\left(\frac{|\nabla\rho_\alpha(r)|}{{\rho_\alpha}^{4/3}(r)}\right)} dr \qquad \textbf{(2.16)}$$

2.4.5 Hybrid Density Functional Theories and B3LYP

Hybrid DFT functionals contain the exact exchange from HF theory. The exact exchange component uses Kohn-Sham orbitals instead of the electron density and corresponds to a system of non-interacting electrons. Arguably, the most popular functional today is the hybrid B3LYP developed by Becke[17] and later modified by Stephens *et al*.[18] The B3LYP functional currently employed in the Gaussian 09 suite of programs is expressed as follows (see equation 2.17):[19]

$$E_{XC}^{B3LYP} = 0.2E_X^{HF} + 0.8E_X^D + 0.72E_X^B + 0.19E_C^{VWN} + 0.81E_C^{LYP} \qquad \textbf{(2.17)}$$

Due to its surprisingly good performance, B3LYP is widely used today for many applications including open-shell transition-metal chemistry.[20]

2.5 Basic Machinery of MO/DFT Calculations

2.5.1 Slater and Gaussian Basis Sets

Atomic orbitals in molecular calculations are expressed in term of linear combination of basis functions (see equation 2.18). These can correspond to any known mathematical functions that can efficiently be manipulated to model AOs, MOs and total electronic wavefunctions. During calculations, their expansion coefficients are optimized as to minimize the total energy of the system's wavefunction.

$$\phi = \sum_{i=1} c_i \chi_i \qquad \textbf{(2.18)}$$

A basis set is a set of basis functions. Amongst the wide variety of existing basis functions are Slater and Gaussian-types basis functions. The former are quasi-near ideal descriptions of AOs. They decrease exponentially far from the nuclei and obey Kato's cusp condition[21] where their electron density has a cusp at the nuclear position. On the

other hand, Gaussian-type basis functions do not adequately model AO near nuclear positions and fall relatively fast far from the nuclei (see 2. 2.1).[22]

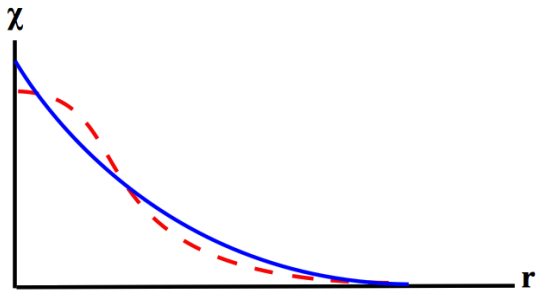


Figure 2.1: Schematic representation of Slater (blue/solid) and Gaussian (red/dashed) type basis functions.

Although Gaussian-type basis functions do not accurately model AO, they are preferred in practice because they allow for efficient calculation of molecular integrals.[22] Today, most computational software employ Gaussian-type basis sets, although it should be noted that some like Amsterdam Density Functional (ADF) use Slater-type basis sets.[23]

2.5.2 Gaussian Split Valence Basis Sets

Split valence basis set divide valence orbitals into two parts: an inner shell and an outer shell. Since chemical bonds involve primarily valence orbitals, the motivation behind split valence basis sets is that it describes more accurately MO's, which in turn describes more accurately wavefunctions (if we are using *ab initio* methods) and electron densities (if we are using DFT).

To understand the composition of split valence basis set, we will consider the example of 3-21G. This basis set represents atomic orbitals of core electrons by a single basis function expressed as a linear combination of three Gaussian primitives. The atomic orbitals of valence electrons are represented by two Gaussian basis functions corresponding to inner and outer shells. The basis function corresponding to the inner

shell is composed of two basis primitives. The outer shell basis function is composed of a single basis primitive. It should be noted that during the SCF cycle, the coefficient underlining the contribution of Gaussian primitive inside individual basis functions do not change. Only those of overall basis functions are optimized as to minimize the energy of the system.

2.5.3 Polarization and Diffuse Functions

Polarization functions allow electron orbital and distributions to be polarized by inclusion of functions of higher angular momentum. For hydrogen and helium atoms, this corresponds to the incorporation of *p*-functions. For second-row atoms, this involves the incorporation of *d*-functions. They are routinely used in molecular calculations and permit orbitals to distort from their original atomic symmetry. 6-31G(d,p) is a common example of a polarized split-valence basis set.[22]

Diffuse functions allow for a more accurate description of orbitals in regions far from nuclei, which emerge in anions and lone-pairs. Diffuse functions make the value of atomic orbitals fall very slowly, as electrons move away from their nuclei. Diffuse functions are denoted by a (+) sign.[22]

2.5.4 Effective Core Potentials

Starting with the third row on the periodic table, the number of electron becomes large enough to slow down calculations considerably. For many of these systems, the cost comes primarily from the large number of core electrons, which are not always chemically relevant. Furthermore, additional complications arise due to the emergence of relativistic effects in elements in the lower half part of the periodic table. Effective core potentials (ECP) are often used to treat this issue. This approach modifies the Fock operator to only treat valence electrons explicitly. It also involves the use of basis sets that

are have been specially parameterized for the use of ECPs and implicitly consider orbital contractions produced from relativistic effects.[22] LANL2DZ is a basis set used in ECP's.

2.5.5 Electron Correlation

Electron correlation refers to the interaction between a pair of electrons. Different computational methods often differ in the way they consider and model electron correlation. In HF, the repulsion between electrons is treated with every electron moving in a smeared-out average electrostatic field created by all other electrons. However this does not depict reality since in fact, each electron moves under the repulsion of individual electrons rather than average electron clouds. HF tends to underestimate the electron-electron repulsion and increase the total electronic energy. In the context of wavefunction-based methods, alternative approaches exist whose treatment of electron correlation are more accurate than that of HF. However, these *post-Hartree Fock methods* can sometimes involve multiple Slater determinants. Due to the resulting computational costs, they cannot typically be used routinely for large systems.

2.6 Molecular Mechanics

Molecular mechanics (MM) is based on the model of a molecule as a set of balls held together by springs. The former corresponds to individual atoms and the latter to chemical bonds.[24] The general approach to MM is to mathematically express the potential energy of a molecule as a function of its resistance towards bond stretching (E_S), bending (E_b), torsion (E_T) and non-bonding interactions through the use of an empirically derived forcefield (see equation 2.19). Prototypical MM forcefields will often express bond stretching and bending in terms of Hooke's law (see equation 2.20). Torsion potentials often rely on combinations of sine and cosine functions (see equation 2.21).

$$E_{Total} = \sum E_S + \sum E_B + \sum E_T + \sum E_V + \sum E_E \quad \textbf{(2.19)}$$

$$E_{S/B} = k_{S/B}\left(d - d_{eq}\right)^2 \quad \textbf{(2.20)}$$

$$E_T = k_0 + \sum_{r=1} k_T[1 + cos(r\theta)] \quad \textbf{(2.21)}$$

The non-bonding interactions are typically dissected into two contributions: Van der Waals (E_V) and electrostatic interactions (E_E). The former are usually evaluated through Lennard-Jones potentials (see equation 2.23) and the latter through Coulomb's law (see equation 2.24) and parameterized partial charges.

$$E_V = k_V \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \quad \textbf{(2.22)}$$

$$E_E = \frac{q_i q_j}{r_{ij}} \quad \textbf{(2.23)}$$

It should be noted that different forcefields exist, each often specially parameterized for a particular group/type of molecules. The Assisted Model Building with Energy Refinement[25] (AMBER) is a forcefield particularly adept in modeling biopolymers such as carbohydrates, proteins and nucleic acids.

One of the strengths of MM methods is their speed and relatively low computational costs. Adequately parameterized forcefields are often very accurate and can rival *ab initio* and DFT calculations. One of their weaknesses is their inability to model electronic properties. Typical MM methods cannot model transition states and the parameterization of certain systems including those involving transition metals remains challenging.

2.7 Quantum Mechanics/Molecular Mechanics

Quantum mechanics/molecular mechanics (QM/MM) is a hybrid method that combines the ability of quantum chemical methods to model electronic properties with the low computational costs of molecular mechanics. It consists in dividing the systems into two layers: one whose energy is obtained through QM theory and the other obtained from MM. The QM layer models the reactive parts of the system and considers the dynamics of electrons in bond formation/cleavage. The MM layer considers surroundings and its steric and electrostatic effects on the QM layer.[26]

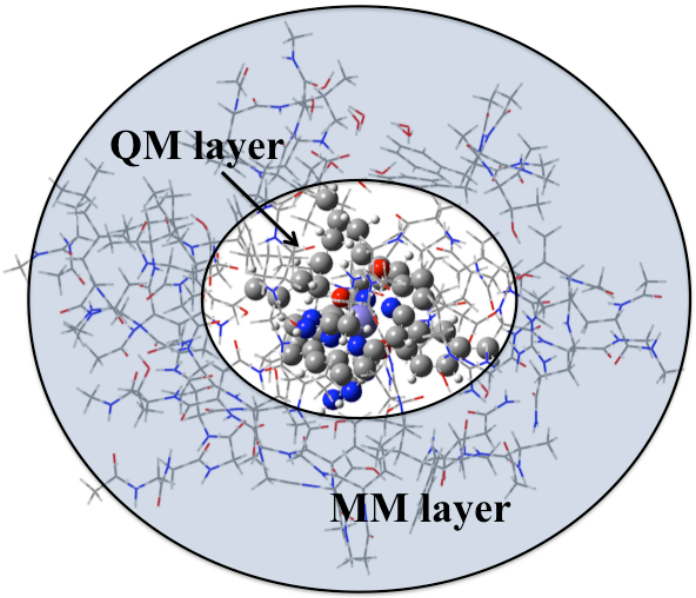


Figure 2.2 Representation of a QM/MM model. The shaded region corresponds the MM layer and the center corresponds to the QM region.

In computational enzymology, QM/MM enhances the ability to model the non-homogenous nature of active sites. DFT is typically employed in modeling the QM layer while a forcefield parameterized for proteins is used for the MM region. The energy of a system calculated through QM/MM can be partitioned in two ways: through an additive or a subtractive scheme. Our Own N-Integrated Molecular Orbital Molecular Mechanics (ONIOM) is a subtractive scheme that calculates the total energy as follows (see equation 2.25). The work presented herein uses the ONIOM scheme as implemented in the Gaussian suite of programs.[19]

$$E_{Tot}^{ONIOM} = E_{Tot}^{MM} - E_{RE}^{MM} + E_{RE}^{QM} \quad \textbf{(2.24)}$$

2.8 Classical Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are used to model the trajectory of atoms and simulate their time-dependent behavior. They can model conformational changes in proteins and sample their configurations. Classical MD simulations involve the

calculations of forces applied on each atom and velocities derived from kinetic theory. The forces are calculated from the MM forcefield and correspond to the negative of the potential energy gradient with respect to the atomic spatial coordinates. Starting from a set of initial atomic positions, the atomic trajectories are calculated by integrating Newton's law of motion.[24]

2.9 Potential Energy Surfaces

The BO approximation allows the ground state energy of a molecule to be expressed as a function of the nuclear positions. Potential energy surfaces (PES) correspond to multidimensional surfaces relating energy to nuclear positions. In chemical reactions, the chemically relevant species are the reactant, transition states, intermediates and products. Reactants, intermediates and products correspond to minima on the PES where $\frac{\partial E}{\partial q_i} = 0$ and $\frac{\partial^2 E}{\partial {q_i}^2} > 0$ for all nuclear coordinates (q_i). Transition states correspond to first order saddle points on the PES where $\frac{\partial E}{\partial q_i} = 0$ for all nuclear coordinates, $\frac{\partial^2 E}{\partial {q_i}^2} < 0$ for one nuclear coordinate and $\frac{\partial^2 E}{\partial {q_i}^2} > 0$ for all other nuclear coordinates.[22,24]

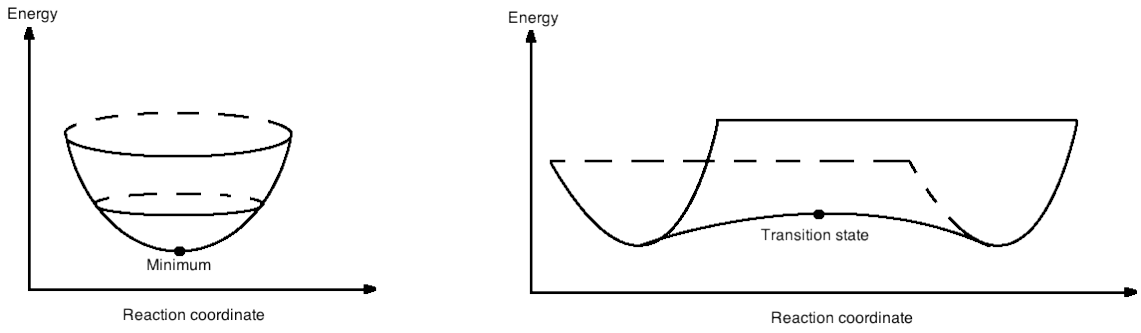


Figure 2.3 Schematic representation of a minimum/reactant/intermediate/product (left and saddle point/transition state (right)

2.10 References

(1) Rode, B. M.; Hofer, T. S.; Kugler, M. D. *The Basics of Theroretical and Computational Chemistry*; Wiley-VCH: Weinheim, 2007.

(2) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. *Chem. Rev.* **2014**.

(3) McQuarrie, D. A. *Quantum Chemistry*; University Science Books: Sausalito, 1983.

(4) Szabo, A.; Ostlund, N. S. *Mondern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, 1989.

(5) Koch, W.; Holthausen, M. C.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; 2 ed.; Wiley-Vch Weinheim, 2001.

(6) Schatz, G. C.; Tatner, M. A. *Quantum Chemistry in Chemistry*; Dover Publications, Inc.: Toronto, 2002.

(7) Levine, I. N. *Quantum chemistry*; Prentice Hall Upper Saddle River, NJ, 2000.

(8) Gill, P. M. W. *J. Chem. Phys.* **1994**, *100*, 5066.

(9) Salahub, D. R.; De La Lande, A.; Goursot, A.; Zhang, R.; Zhang, Y. *Recent Progress in Density Functional Methodology for Biomolecular Modeling*; Springer: Heidelberg, 2013.

(10) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(11) Thomas, L. H. *Proc. Camb. Phil. Soc.* **1927**, *23*, 542.

(12) Fermi, E. *Rend. Accad. Lincei.* **1927**, *6*, 602.

(13) Dirac, P. A. M. *Proc. Camb. Phil. Soc.* **1930**, *26*, 376.

(14) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(15) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(16) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Chem. Rev.* **2011**, *112*, 289.

(17) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(18) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(19) 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., J. A. ; 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. P., 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, Inc., CT: Wallingford 2009.

(20) Sameera, W. M. C.; Maseras, F. *WIREs Comp. Mol. Sci.* **2012**, *2*, 375.

(21) Kato, T. *On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics*; Wiley Subscription Services, Inc., A Wiley Company, 1957; Vol. 10.

(22) Lewars, E. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Springer: Boston, 2010.

(23) Te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comp. Chem.* **2001**, *22*, 931.

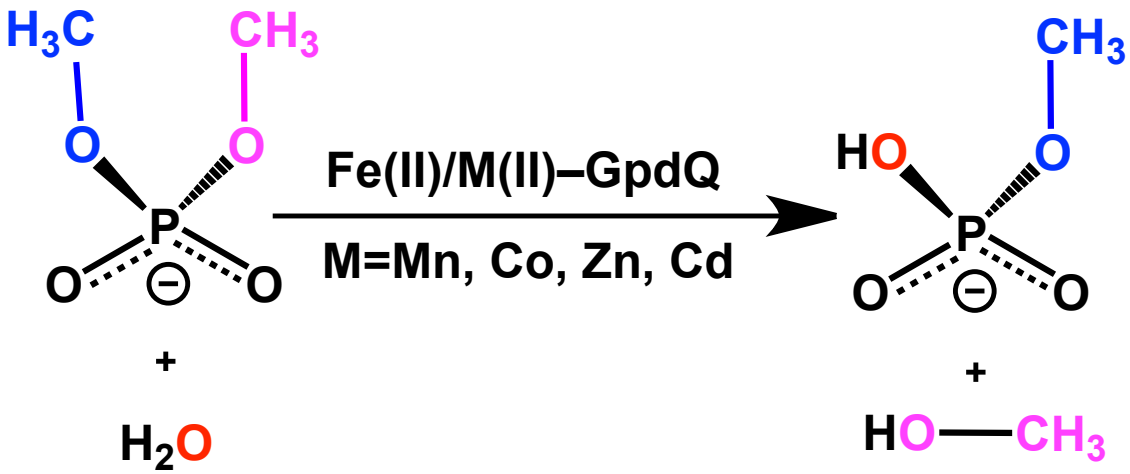
(24) Andrew, R. L. *Molecular Modelling: Principles and Applications*; 3 ed.; Pearson Education: Essex, 2001.

(25) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham Iii, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(26) van der Kamp, M. W.; Mulholland, A. J. *Biochemistry* **2013**, *52*, 2708.

Chapter 3

The Different Catalytic Rates of Fe(II)/M(II)-GpdQ Explained with DFT



3.1 Introduction

The glycerophosphodiesterase from *Enterobacter aerogenes* (GpdQ) is a binuclear metalloenzyme whose ability to hydrolyse P–O bonds in a wide range of phosphoesters has recently attracted attention.[1,2] Indeed, while the specificity of most phosphoesterases centers around a particular class of phosphoesters, GpdQ can hydrolyse mono, di and triesters.[3,4] Furthermore it can breakdown some organophosphate pesticides and EA 2192, a toxic product generated from the degradation of organophosphate (OP) nerve agents.[5] Consequently, it can potentially be used as a enzymatic bioremediator.[6,7]

Existing GpdQ crystal structures indicate the presence of two distinct metal binding sites (see Figure 3.1).[2] The first, labeled as the α site, contains Asp8, His10, Asp50 and His197 and the second, labeled as the β site, contains Asp50, Asn80, His156 and His195. Metal ion affinity studies have suggested that metal ions coordinated to the α site are more tightly bound then those occupying the β site.[1,2,8] Furthermore, it has been shown that the metal ion affinity of the β site increases in the presence of a substrate.[9]

The metal ion composition of GpdQ in biological conditions is elusive.[10] However it has been shown that enzymatic activity can manifest from different divalent metal ions such as Mn(II), Co(II), Zn(II) and Cd(II).[1,11] A recent study by Daumann *et al*.[10] have investigated the catalytic activity of GpdQ with different heterodinuclear metal ion compositions. Specfically, by generating different Fe(II)/M(II)-GpdQ compositions (with M(II) occupying the β site), it has been shown that GpdQ acquires differing catalytic rates of approximately 0.21 s^{-1} for Fe(II)/Zn(II), 0.68 s^{-1} for Fe(II)Co(II), 2.15 s^{-1} for Fe(II)/Mn(II) and 2.35 s^{-1} for Fe(II)/Cd(II) for the hydrolysis of bisparanitrophenol (bpNPP). It remains uncertain as to why these differing metal ion compositions would cause significant deviations in catalytic rates.

The catalyic mechanism of GpdQ remains elusive[12] partly due to the ambigious nature of its protonation state. Although crystal structures indicate the presence of bridging and

terminal oxygens,[1,2] it remains uncertain as to whether these correspond to hydroxide or water ligands. Importantly, uncovering their nature would allow one to discern whether these could act as nucleophiles. Furthermore, the protonation of His81, which occupies the second coordination sphere remains unknown. The latter could have significant implications where substrate coordination could be affected or by being directly involved in the mechanism.

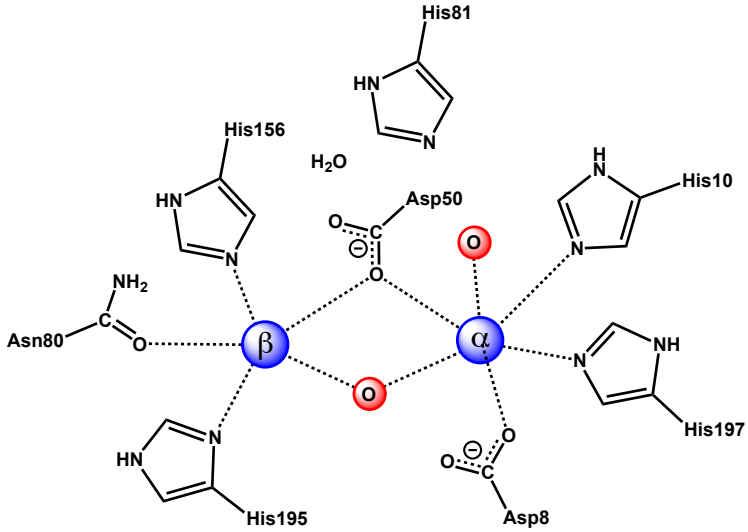
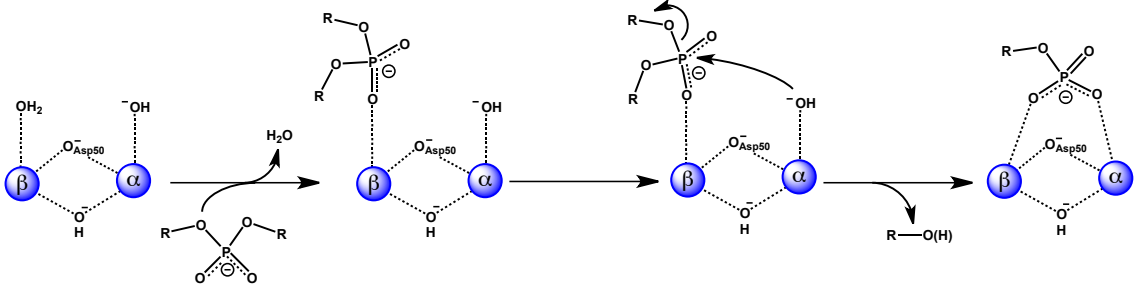


Figure 3.1. Schematic representation of the GpdQ active site.[1,2]

From structural and spectroscopic investigations, Hadler *et al*.[2,9,13,14] have suggested a mechanism whereby the terminal bound hydroxide/water molecule attacks the substrate's phosphorus atom (see Scheme 3.1). The bridging OH^- has been proposed to activate the terminal nucleophile inducing a nucleophilic attack resulting in the release of an alcohol product.



Scheme 3.1. Proposed hydrolytic mechanism of GpdQ by Hadler *et al*.[9,13,14]

This study attempts to provide, through density functional theory (DFT), detailed insights into the catalytic mechanism of GpdQ. We initially investigated the protonation state of the pre-reactive complex (PRC), containing two Co(II) metals ions occupying the α site and the β site but lacking a substrate. The optimized models were then compared to the corresponding crystal structure (PDB: 3D03) and allowed to identify the nucleophile. Secondly, via molecular docking and DFT, we modeled several possible reactant complexes with differing protonation states and substrate coordinations. The calculations of free energies allowed the most thermodynamically favoured reactant complexesto be determined. Thereon, we modeled different reaction pathways for the hydrolysis of dimethylphosphate (DMP) by GpdQ. The hydrolysis of DMP by Fe(II)/M(II)-GpdQ (where M=Zn, Co, Mn and Cd) has been modeled. The trends in computational activation energies are compared with the trends in experimental catalytic rates. A tentative rational relying on hard and soft Lewis acid and base (HSAB) theory and the properties of the HOMO in the reactant complexes has been provided.

3.2 Computational Methods

All computational models were generated starting from one of the existing crystal structure of GpdQ (PDB: 3D03).[2] Because it does not contain a substrate bound in its active site, molecular docking was performed using the Molecular Operating Environment (MOE)[15] to model Michaelis complexes. Dimethylphosphate was modeled and docked in the active site using the London dG scoring function in conjunction with AMBER12EHT force field. The resulting model was thereafter truncated to include the metal ions' first coordination sphere and the side chain of His81. The C_α of all residues were fixed to maintain the integrity of the active site. DFT calculations were performed using the Gaussian 09 suite of programs.[16]

Geometry optimization and frequency calculations were done at the B3LYP/6-31G(d) level of theory for Mn, Co and Zn containing complexes and B3LYP/LANL2DZ(Cd);6-31G(d)(H,C,N,O,P,Fe) for Cd containing ones. Single point calculations were performed at the B3LYP/6-311+G(2d,2p) level of theory for Mn, Co and Zn containing complexes. The latter has been used in previous computational studies pertaining to phosphotriesterases[17] and phosphatases.[18] Models containing Cd atoms used the aug-cc-pVTZ-PP on Cd centers and 6-311+G(2d,2p) for all other atoms.

To model the average contribution of the protein surrounding the active site, all single point calculations were conducted using the IEFPCM method with a dielectric constant of 4.0. Relative Gibbs free energy corrections were obtained via B3LYP/6-31G(d) for Mn, Co and Zn containing models and B3LYP/LANL2DZ(Cd);6-31G(d)(H,C,N,O,P,Fe) for models with Cd. Preliminary calculations on the Michaelis complexes were performed in order to determine the most stable spin-states for all Fe(II)/M(II) combinations. The lowest energies were obtained for cases where both metals were in the high-spin configurations. It should be noted that all attempts to model antiferromagnetic spin-states failed and either resulted in repetitive SCF convergence failures or highly unstable models. To our knowledge, there exists no experimental evidence that neither supports nor refutes the emergence of antiferromagnetism between heteronuclear metals centers in GpdQ.

3.3 Results and Discussion

3.3.1 Protonation State of the Active Site

The study presented herein began with an assessment of the protonation state of the active site of Co(II)/Co(II)-GpdQ prior to substrate binding. Specifically, our objective for this part was to answer the following questions:

- Does the bridging oxygen belong to a water or hydroxide?
- Does the terminal oxygen belong to a water or hydroxide?

- Is His81 neutral or protonated charged?

To answer these questions, all possibilities have been modeled using DFT-clusters starting from the crystal structure. In total, eight protonation states have been considered. For each neutral His81 (hereon denoted as 'N') and His81–H^+ (hereon denoted as 'P') containing scenarios, the following protonation states have been modeled:

- Terminal H_2O/Bridging H_2O (hereby denoted as W_T/W_B)
- Terminal H_2O/Bridging OH^- (hereby denoted as W_T/O_B)
- Terminal OH^-/Bridging H_2O (hereby denoted as O_T/W_B)
- Terminal OH^-/Bridging OH^- (hereby denoted as O_T/O_B)

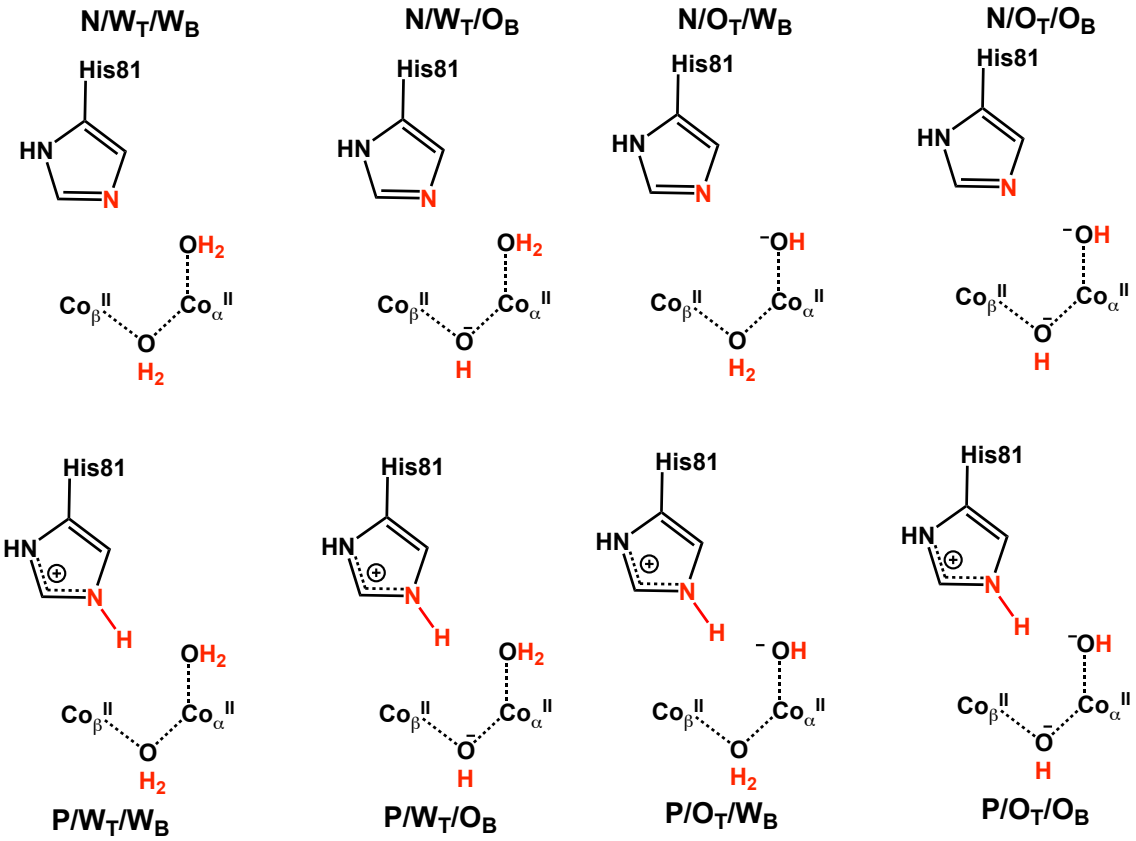
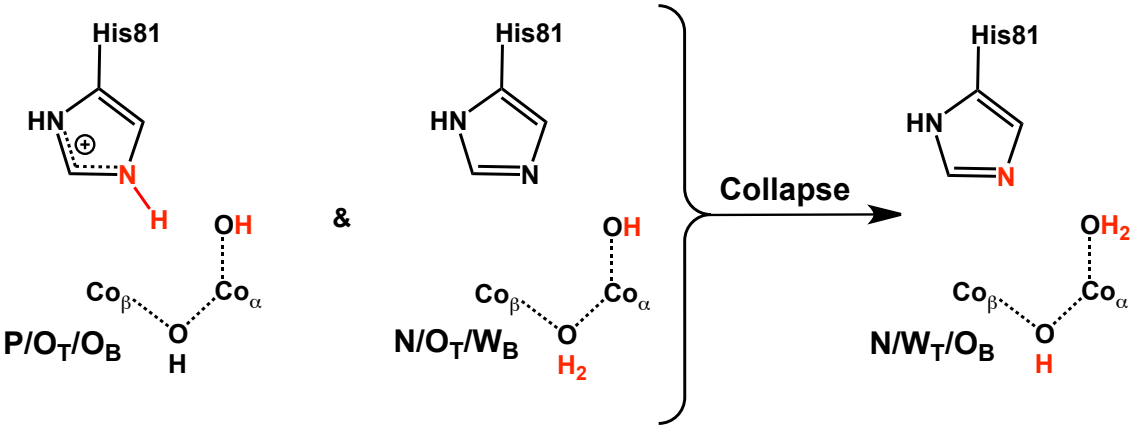


Figure 3.2. Schematic representations of all considered models and protonation states of the GpdQ active site.

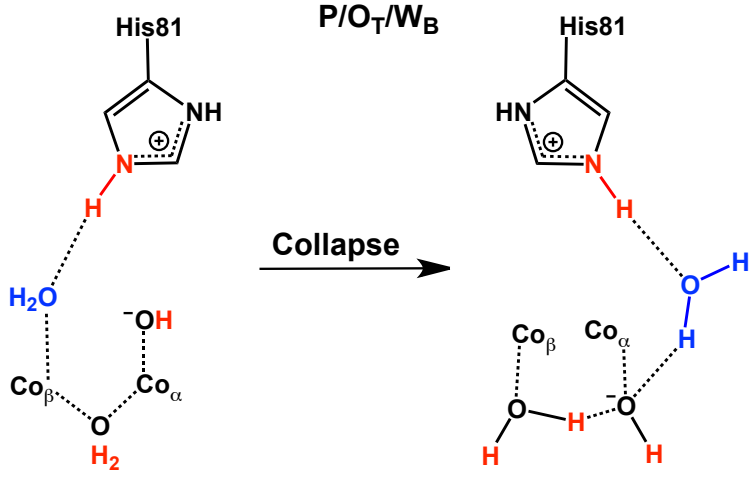
The optimized structures were obtained for all eight possibilities. In each case, all interatomic distances between heavy atoms were computed. The distances of the DFT models were then compared to those of the crystal structure by computing the root mean square deviation (RMSD). The RMSD of each DFT model quantifies the deviation that exists structurally between the optimized DFT model and the crystal structure. It is assumed that the DFT model with the smallest RMSD has the protonation state which corresponds to that of the crystal structure and the pre-reactive complex.

Two of the considered models, $P/O_T/O_B$ and $N/O_T/W_B$, collapsed to $N/W_T/O_B$, i.e. both having a neutral His81, a terminal water and bridging hydroxyl group . In the former case of $P/O_T/O_B$, during the minimization, the proton on His81 (marked in red in Scheme 3.2) transferred from His81 to the terminal hydroxide. During the optimization of latter, $N/O_T/W_B$, one of the protons on the bridging water dislocated to the terminal hydroxide. These calculations indicate that the presence of a terminal hydroxide is unstable and unlikely to correspond to the crystal structure. At the current level of theory (B3LYP/LANL2DZ(Co);6-31G(d)(H,C,N,O)), the Co occupying the α site is not a strong enough Lewis acid to stabilize the formation of a terminal hydroxide. Ultimately, these computational results suggest that the terminal oxygen may not be a hydroxide ligand for $Co_\alpha^{II}/ Co_\beta^{II}$ metal ion composition



Scheme 3.2. Schematic representations of the optimizations of $P/O_T/O_B$ and $N/O_T/W_B$. At the (B3LYP/LANL2DZ(Co);6-31G(d)(H,C,N,O) level of theory, both initial structures break down and form $N/W_T/O_B$.

Our structural analysis did not consider $P/O_T/W_B$ since it optimized to a conformation with two bridging components (see Scheme 3.3). Furthermore, during the minimization, the water originally in the β site translocated to the α site. The nature of such models does not correlate with that of the crystal structure and hence provides grounds to suggest that the pre-reactive complex does not contain a protonated His81, a terminal hydroxide and a bridging water.



Scheme 3.3. Schematic representations of the optimization of $P/O_T/W_B$ at the (B3LYP/LANL2DZ(Co);6-31G(d)(H,C,N,O) level of theory.

Our analysis of interatomic distances and RMSD's were done on three scales. Initially a small scale involving only seven interatomic interactions (see Table 3.1) was considered. Several key interactions in each DFT model were measured and then compared to those in the crystal structures by subsequently calculating the RMSD's. All six monomers of the GpdQ were considered and the specified interatomic distances were measured for each. Thereon, the average of these distances were calculated and represented in Table 3.1 as 'average crystal structure (Avg Xtal Struc.)'.

The model with the lowest RMSD (0.49 Å) corresponds to $N/W_T/O_B$ where His81 is neutral and Co_α is coordinated to a terminal water and a bridging hydroxide. The models with the second lowest RMSD (0.56 and 0.58 Å) correspond to $N/O_T/O_B$ and $P/W_T/O_B$ both have a bridging hydroxide but differ in the protonation states of His81 and terminal oxygen. The model $N/W_T/W_B$, which contains a bridging water, has a moderately larger RMSD (0.80 Å). $P/W_T/W_B$ corresponds to a fully protonated active site. Its corresponding RMSD is excessively large (3.85 Å) and indicates that the pre-reactive/crystal structure does not involve a protonated His81 with bridging and terminal waters. Since the three models with the lowest RMSDs contained bridging hydroxide ligands, the current DFT results suggest that the bridging ligand is most likely a hydroxide.

Table 3.1. Selected interatomic distances in DFT models and averaged crystal structure (PDB: 3D03). The bridging and terminal oxygen atoms are denoted as 'B' and 'T' respectively. The interaction betweenthe nitrogen atoms on the imidazole moiety of His81 (marked in red Figure 3.2) and the active site water associated with the β site has also been considered.

Interactions (Å)	Co_{α}–Co_{β}	Co_{α}–B	Co_{β}–B	Co_{α}–T	Co_{β}–T	B–T	N_{His81}–O_W	RMSD (Å)
Avg Xtal Struc.	3.69	2.41	2.60	2.83	4.38	2.52	3.45	**N/A**
$N/W_T/W_B$	3.79	2.15	3.86	4.14	5.28	2.70	2.87	**0.80**
$N/W_T/O_B$	3.27	2.06	2.03	2.34	4.07	2.99	4.13	**0.49**
$N/O_T/O_B$	3.29	1.99	2.02	1.94	3.97	3.11	3.02	**0.56**
$P/W_T/W_B$	3.68	2.15	3.83	8.57	10.47	8.17	2.82	**3.85**
$P/W_T/O_B$	3.54	2.02	2.09	3.92	3.93	2.68	2.73	**0.58**

The second (medium) scale considered involved the two Co(II) metals, the atoms with direct coordination, the oxygen atom on the β site water and the nitrogen atom on His81. At this scale all interatomic distances have been measured and RMSD computed (see Table 3.2). The third (large) scale involved all heavy atoms in the models as well as all interatomic distances. For both analyses, the DFT models have been ranked in terms of their RMSD, indicative of their divergence away from the conformation of the crystal structure.

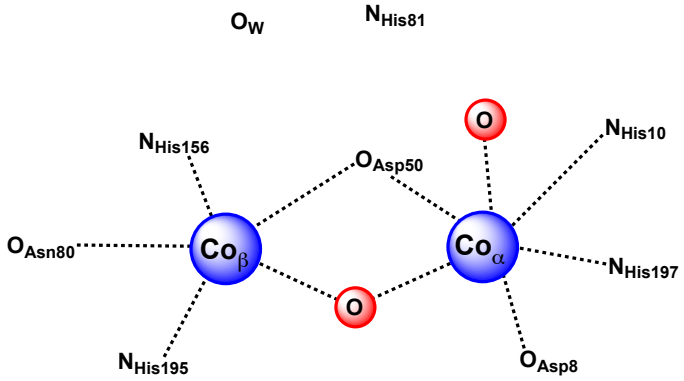


Figure 3.3. Atoms considered in the second (medium-sized) scale. Only the positions of the β site water (O_W), the nitrogen atom on His81 and the atoms in the first coordination sphere have been considered. Charges were excluded from this Figure to enhance its clarity.

The three highest ranking DFT models (with the lowest average RMSD) all have a bridging hydroxide ligand. At both scales, the model corresponding to a neutral His81, terminal water and bridging hydroxide deviates the least from the crystal structure. Similarly to the small scale analysis, a fully protonated active site gives rise to significant structural changes and an optimized structure that differs significantly from that of the crystal structure.

Table 3.2. RMSD of the DFT models at the medium (Med) and large-scale (Lrg). The average between both scale have been computed as well as the corresponding ranking.

Models	RMSD (Å)			Rank*		
	Med-Scale	Lrg-Scale	Avg.	Med-Scale	Lrg-Scale	Avg.
$N/W_T/W_B$	1.97	0.82	1.40	4	2	4
$N/W_T/O_B$	0.76	0.52	0.64	1	1	1
$N/O_T/O_B$	0.99	0.86	0.93	2	3	2
$P/W_T/W_B$	2.36	1.63	2.00	5	5	5
$P/W_T/O_B$	1.00	0.89	0.95	3	4	3

*A rank of 1 corresponds to the optimized structure that best matched the crystal structure.

The DFT results presented herein strongly suggest that the pre-reactive complex as modeled in the crystal structure contains a neutral His81 side-chain, a terminal water ligand and a bridging hydroxide. The set of possibilities involving a terminal hydroxide adjacent to potential proton donor (either a positively charged His81 or a bridging water) can be excluded as these collapsed to terminal water containing species.

3.3.2 Substrate Coordination and Displacement of the Terminal Water

In addition to investigating the pre-reactive complex, we have also examined the reactant complex. Specifically, we considered the proposal by Hadler *et al*.,[2,9,13,14] which suggests that the terminal ligand may act as a nucleophile by either directly attacking the substrate or by being activated by the bridging ligand before attacking. We studied the possibility that a DMP substrate may displace either a terminal water or terminal hydroxide ligated to the α site and by doing so bind in a bidentate fashion to both metals. In the following calculations, we have used the Fe(II)/Zn(II) metal ion composition with

the former occupying the α site and the latter occupying the β site. It should also be noted that the models considered herein did not involve the side chain of His81.

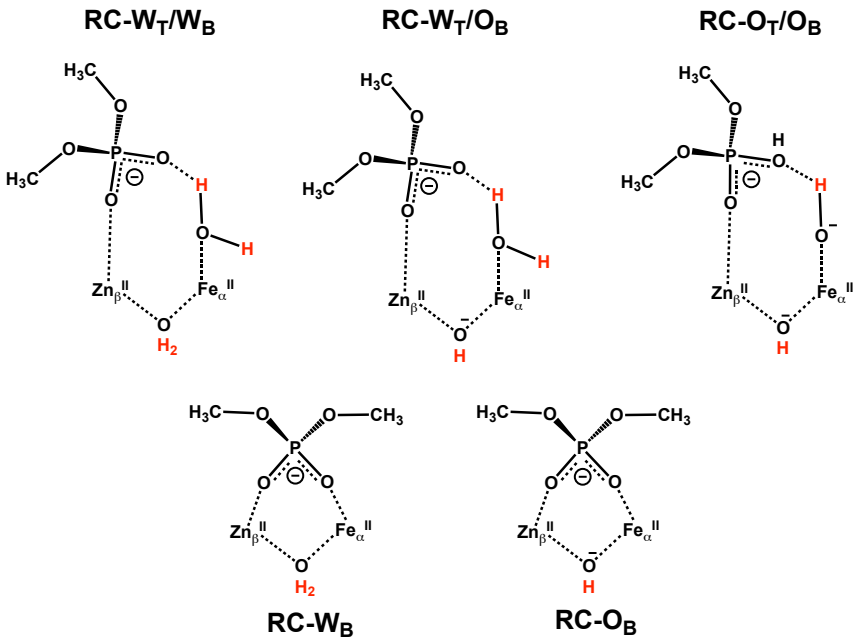
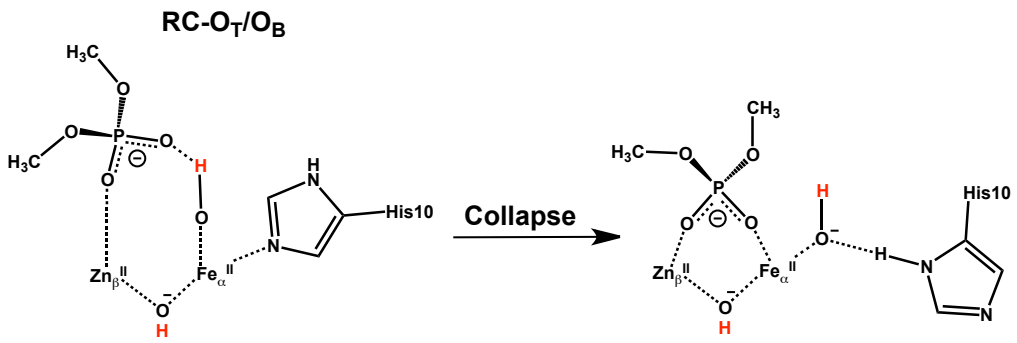


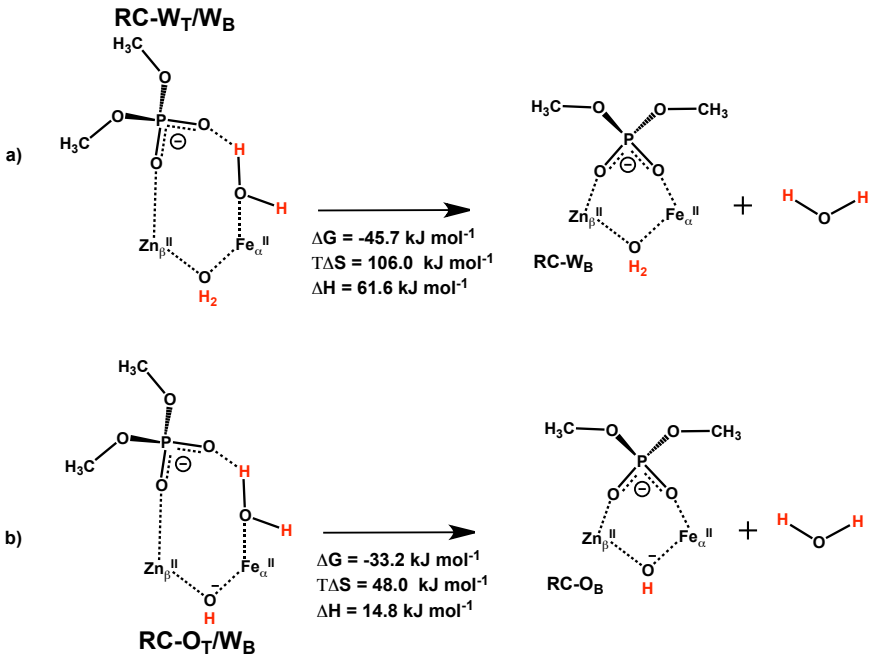
Figure 3.4. Schematic representations of all considered reactant models in the active site of GpdQ.

During the minimization of RC–O_T/O_B, DMP displaced the hydroxide terminal ligand thereby adopting a bidentate coordination. The hydroxide, in turn, displaced the side chain of His10 and maintained its ligation to Fe(II). The optimized conformation of RC–O_T/O_B indicates the unlikelihood of the terminal hydroxide to act as a nucleophile. Indeed, in the considered case it appears more likely for the bridging hydroxide to act as nucleophile as it is closer to the phosphorus atom than the terminal hydroxide by 0.68 Å. In addition, from the optimized conformation, it appears that nucleophilic attack by the terminal hydroxide would be hindered by the adjacent phosphate oxygen coordinated to the Fe(II).



Scheme 3.4. Schematic representations of the optimizations of RC-O_T/O_B. At the (B3LYP/LANL2DZ(Fe,Zn);6-31G(d)(H,C,N,O) level of theory, it breaks down to form a bidentate coordination and the hydroxide displacing His10.

Gibbs free energy differences were computed to determine if the displacement of a terminal water ligand by DMP was thermodynamically favorable. For the cases where the bridging ligand is a water and a hydroxide, such displacement was computed to be favorable as indicated by Gibbs free energy differences of -45.7 and -33.2 kJ mol^{-1} respectively. In both cases, the displacement is driven by the increased entropy acquired by the water after leaving its coordination site. Although these processes are endothermic, at 298 K the entropic contributions outweigh the enthalpic costs. Presently, our DFT results suggest that the replacement of a terminal water by DMP in the α site for systems involving a bridging water or hydroxide is favorable.



Scheme 3.5. Schematic representations of the displacement of a terminally water by DMP and corresponding thermodynamical properties in a) water and b) hydroxide bridging ligand complexes obtained at the B3LYP/6-311+(2d,2p)//B3LYP/LANL2DZ(Fe,Zn);6-31G(d)(H,C,N,O)) level of theory

3.3.3 Mechanistic study of Fe(II)/M)(II)–GpdQ Catalysis

Previous calculations, performed herein, suggest the bridging ligand to be a hydroxide and the terminal ligand to be displaced by the DMP substrate. These conclusions provide a starting point for the mechanistic study of the hydrolysis of DMP by Fe(II)/M(II)–GpdQ. We considered four different metals ion compositions: M=Mn, Co, Zn and Cd and looked at both cases where His81 was neutral and protonated.

We began by neutralizing His81 in the Fe(II)/Zn(II) model. However during its minimization, its imidazole moiety rotated almost completely away from the substrate and the rest of the active site. The His81 side chain in the resulting minimized model had no significant interactions with the active site components (almost none with the substrate). This conformation is believed to occur from repulsion between the imidazole

and the negatively charged substrate. It most likely occurs due to a lack of steric hindrance that would normally be present from the rest of the enzyme. In any case, due to the lack of interactions between the neutral His81 and the rest of the active site, the former was removed from the rest of the model.

Many binuclear hydrolases have been suggested to utilize bridging hydroxides as nucleophiles without the necessary requirement of the catalytic involvement of surrounding residues.[12,19] Hence, we investigated the direct attack of the hydroxide on the phosphorus center. For all metal ion compositions, it was found that the nucleophilic attack occurred through an S_N2 type mechanism with a concerted proton transfer from the bridging hydroxide to the methoxide leaving-group (see Figure 3.4). The geometries between the different metal ion compositions are similar with the exception of slight elongation in Cd(II) containing systems (see table 3.3). The reactant complexes all involved DMP in bidentate metal ion coordination. The direct bridging hydroxide attack involved 4-membered ring transition states and high energy barriers of approximately 201.0 kJ mol^{-1} for all metal compositions. All reactions were computed to be thermodynamically favorable.

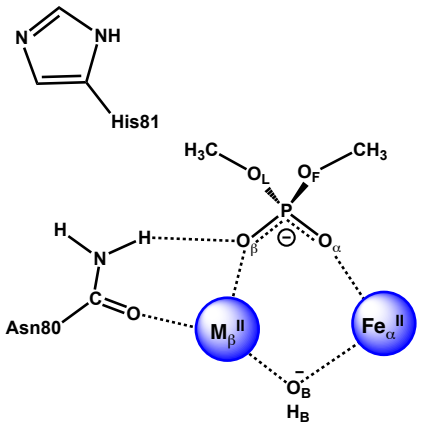


Figure 3.5. Schematic representations of neutral His81 containing reactant models and designation of atoms.

The computed energy barriers (see Figure 3.6) suggest that the mechanism of GpdQ involve a neutral His81 and direct attack by a bridging hydroxide. As our current results suggest, such a mechanism scheme too slowly for biological purposes. Furthermore the trends in energy barriers do not correspond to the experimental catalytic rates. Assuming the mechanism of GpdQ follows Arrhenius behavior (with quantum effects such as proton tunneling being negligible), experimental catalytic rates suggest that Cd(II)–GpdQ has the lowest energy barrier and Zn(II)–GpdQ the highest one. However computed energy barriers suggest Co(II)–GpdQ to hydrolyze DMP the fastest and Mn(II)–GpdQ the slowest and these contradict experimental results. Furthermore the energy barriers between different metal ion compositions are very close to one another and go in opposition to the experimental rates, which suggest significant differences.

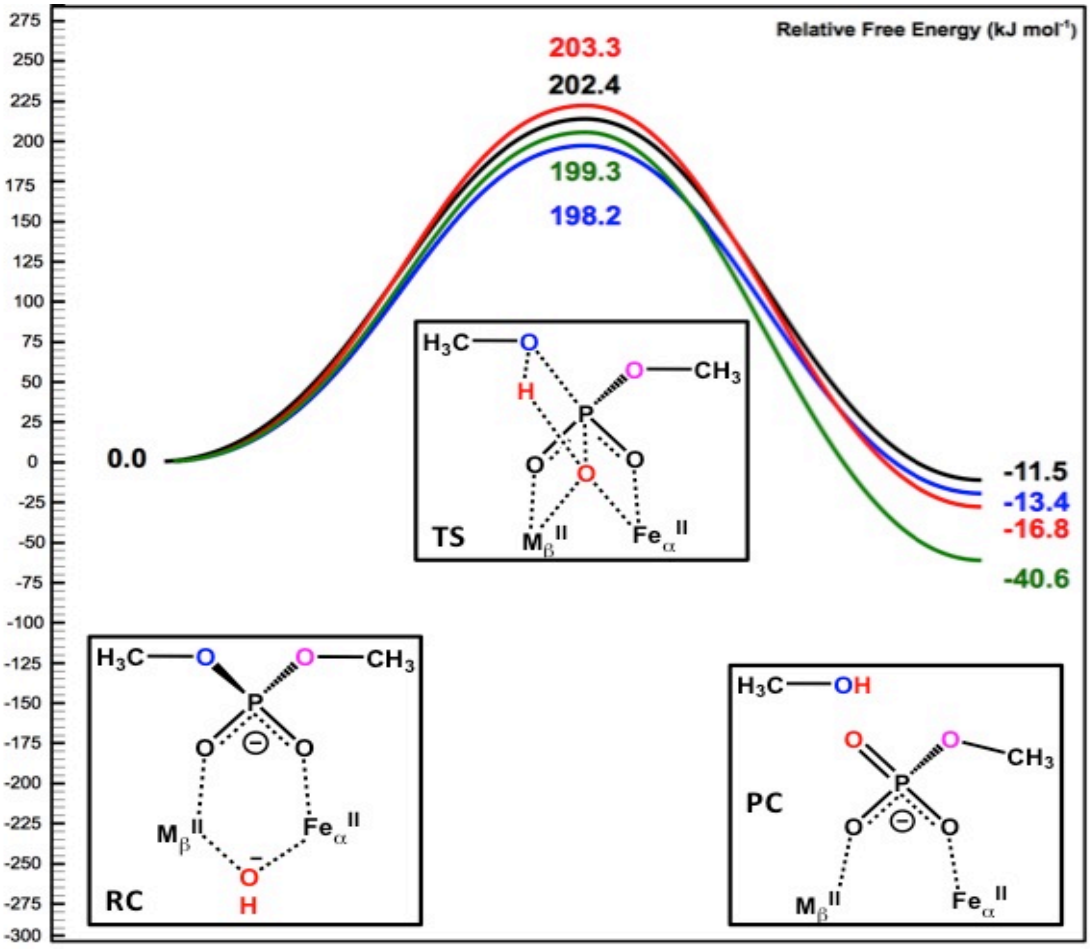


Figure 3.6. Schematic free energy surface of the DMP hydrolysis by Fe(II)/M(II)–GpdQ with the bridging hydroxide ligand acting as the nucleophile and neutral His81. Color key: black/M=Zn, red/M=Mn, blue/M=Co, green/M=Cd.

Table 3.3. Selected atomic distances (Å) of optimized species along the reaction pathway of DMP hydrolysis with neutral His81 model.

	RC				TS				PC			
	Mn	**Co**	**Zn**	**Cd**	**Mn**	**Co**	**Zn**	**Cd**	**Mn**	**Co**	**Zn**	**Cd**
Fe–M	3.26	3.22	3.23	3.34	3.50	3.44	3.49	3.58	3.62	3.69	3.63	3.79
Fe–O_B	2.03	2.05	2.02	2.01	2.26	2.21	2.25	2.27	4.01	3.99	4.01	4.10
M–O_B	2.09	2.02	2.04	2.22	2.30	2.21	2.30	2.43	4.29	4.21	4.28	4.33
O_B–P	3.18	3.16	3.16	3.24	1.75	1.75	1.74	1.75	1.52	1.52	1.52	1.53
H_B–O_F	2.82	2.77	2.78	2.67	1.37	1.31	1.38	1.41	1.00	1.00	1.00	1.00
H_B–O_B	0.97	0.97	0.97	0.97	1.08	1.12	1.08	1.06	1.66	1.66	1.66	1.68
Fe–O_α	2.36	2.31	2.35	2.43	2.19	2.19	2.18	2.18	1.93	1.93	1.93	1.94
M–O_β	2.27	2.19	2.23	2.38	2.17	2.21	2.11	2.29	2.00	1.95	1.98	2.18
P–O_F	1.64	1.64	1.64	1.64	2.27	2.23	2.27	2.29	3.71	3.67	3.72	4.10

We investigated the notion that the mechanism of GpdQ requires a protonated His81 acting as a catalytic acid. It should be noted that a very similar mechanistic pathway has been previously investigated computationally for Fe(III)–Zn(II) purple acid phosphatases.[18] At the B3LYP/6-311+(2d,2p)//B3LYP/6-31G(d)+ΔG level of theory, this mechanism proceeds via a two-step reaction. For Mn(II), Co(II) and Zn(II) containing complexes, the inclusion of a protonated His81 altered the coordination of the DMP substrate to the Fe(II) in the α site from bidentate to monodentate. However the Cd(II) containing reactant complex maintained its bidentate coordination. For the latter, the different substrate coordination occurred due to the greater size of the Cd(II), which pushes the residues occupying the β site away from itself and allows greater accessibility from the substrate.

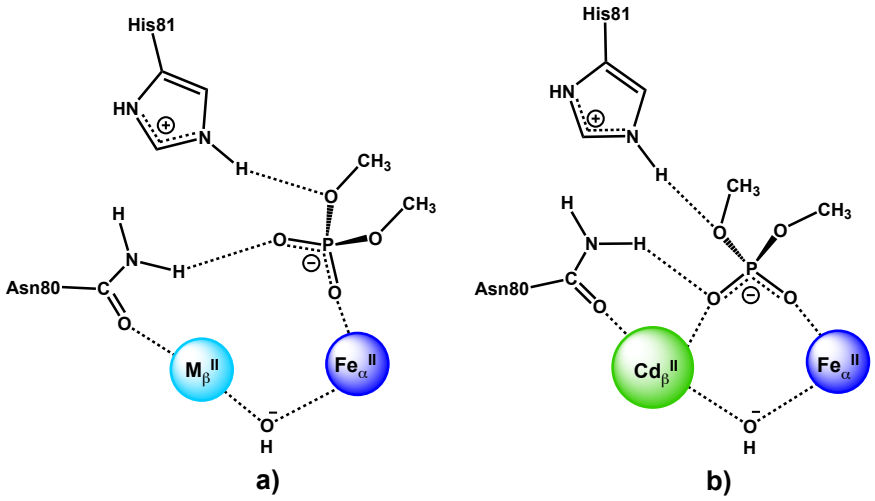


Figure 3.7. Schematic representations of protonated His81 reactant models with a) Mn(II), Co(II), Zn(II) and b) Cd(II) and designation of atoms.

The initial step involved the nucleophilic attack of the hydroxide bridging ligand and the formation a 5-coordinate phosphorus centered complex. The barriers related to this initial step for Zn, Co, Mn, and Cd complexes were 71.3, 65.5, 64.9 and 48.9 kJ mol^{-1} respectively. Relative to each corresponding reactant complexes, the resulting intermediate was raised by 86.6, 70.8, 69.4 and 50.0 kJ mol^{-1} for Zn, Co, Mn and Cd containing complexes respectively. In each intermediate, a O_B–P bond is partially formed with lengths of 2.00, 2.06, 1.94 and 1.99 Å for Mn, Co, Zn and Cd complexes respectively. It should be noted that these geometries differs from those predicted in the case of Fe(III)–Zn(II) purple acid phosphatases where the O_B–P bond was 1.85 Å.[18]

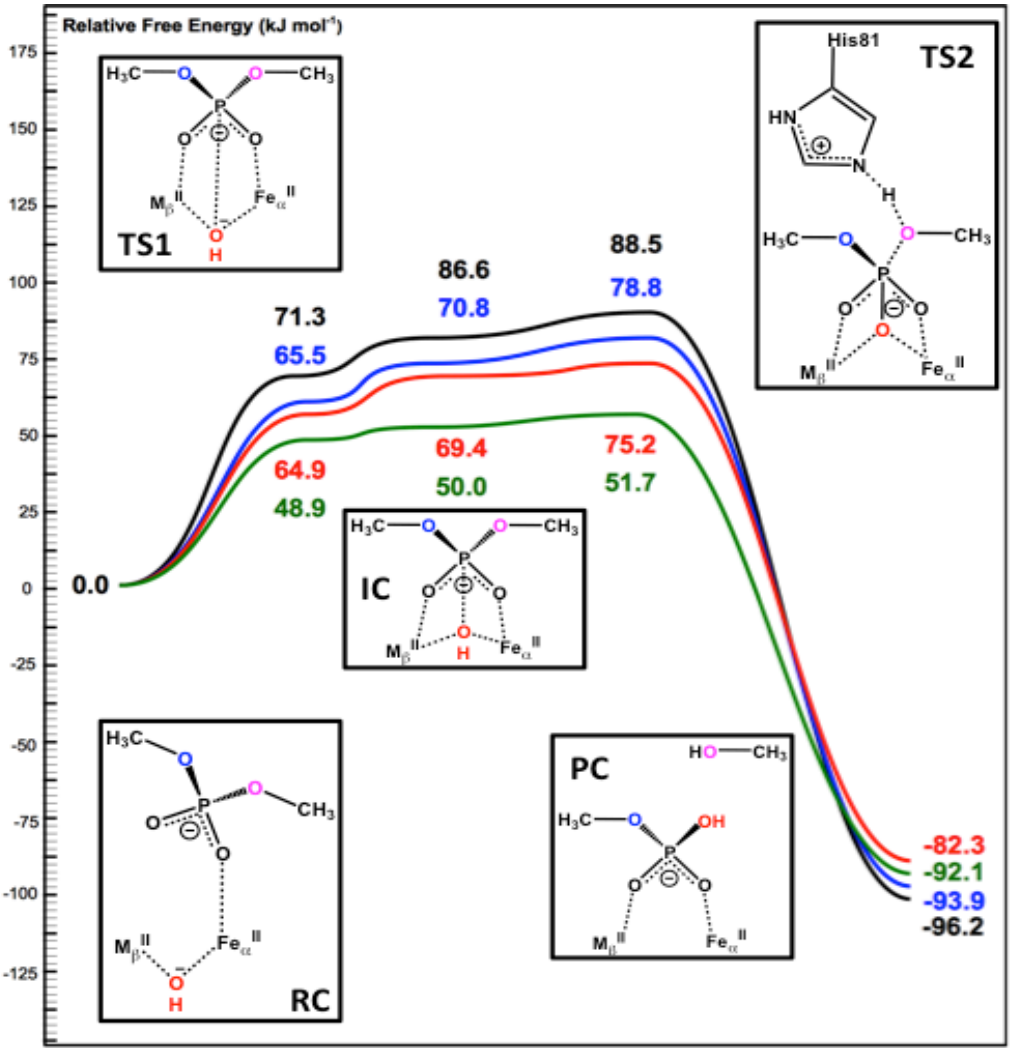


Figure 3.8. Schematic free energy surface of the DMP hydrolysis by Fe(II)/M(II)–GpdQ with the bridging hydroxide ligand acting as the nucleophile and an acidic positively charged His81. Color key: black/M=Zn, red/M=Mn, blue/M=Co, green/M=Cd. It should be noted that while the coordination of the RC is monodentate, such coordination only applies to Mn, Co and Zn containing complexes.

The second mechanistic step involves the highest energy barrier and is rate-determining. This step involves three key molecular events that occur concertedly. While the bond between P–O_L breaks allowing the methoxide to act as a leaving group, His81 transfers its H^+ to O_L, thereby stabilizing it and forming methanol. Furthermore, the P–O_L bond

cleavage occurs with the shortening of the O_B–P bond from 2.00, 2.06, 1.94 and 1.99 Å for Mn, Co, Zn, Cd IC's to approximately 1.58 Å in all optimized PC.

The trends obtained from the computated energy barriers fit very well with those of the catalytic rates obtained experimentally. The catalytic rate of the hydrolysis of bpNPP of Zn, Co, Mn and Cd containing complexes are 0.21, 0.68, 2.15 and 2.35 s^{-1} respectively generating a trend of Zn<Co<Mn<Cd in terms of rates. The computational energy barriers for the rate determining step in the hydrolysis of DMP are 88.5, 78.8, 75.2 and 51.7 kJ mol^{-1} establishing the following trend in terms of energy barriers: Zn>Co>Mn>Cd. The computational trend fits with the experimental results. This correlation between computational and experimental results strongly suggest that the mechanism of GpdQ involves the nucleophilic attack of a hydroxide bridging ligand and the protonation of the substrate by His81.

Table 3.4. Selected atomic distances (Å) of optimized species along the reaction pathway of DMP hydrolysis with protonated His81 Mn(II) and Co(II) containing models.

	RC		**TS1**		**IC**		**TS2**		**PC**	
	Mn	**Co**	**Mn**	**Co**	**Mn**	**Co**	**Mn**	**Co**	**Mn**	**Co**
Fe–M	3.37	3.32	3.48	3.42	3.50	3.48	3.58	3.56	3.87	3.82
Fe–O_B	2.05	2.04	2.18	2.18	2.19	2.23	2.26	2.25	3.73	3.74
M–O_B	2.01	1.94	2.20	2.11	2.22	2.13	2.35	2.30	3.85	3.86
O_B–P	3.25	3.33	2.05	2.07	2.00	2.06	1.85	1.84	1.58	1.58
Fe–O_α	2.21	2.22	2.13	2.12	2.13	2.11	2.13	2.12	2.01	2.01
M–O_β	3.82	3.87	2.28	2.19	2.26	2.18	2.23	2.10	2.05	1.97
P–O_L	1.68	1.69	1.78	1.76	1.79	1.76	1.93	1.93	4.85	4.84

Table 3.5. Selected atomic distances (Å) of optimized species along the reaction pathway of DMP hydrolysis with protonated His81 Zn(II) and Cd(II) conatining models.

	RC		TS1		IC		TS2		PC	
	Zn	**Cd**	**Zn**	**Cd**	**Zn**	**Cd**	**Zn**	**Cd**	**Zn**	**Cd**
Fe–M	3.29	3.43	3.44	3.51	3.50	3.55	3.59	3.60	3.82	3.90
Fe–O_B	2.04	2.11	2.17	2.16	2.21	2.19	2.25	2.26	2.01	3.62
M–O_B	1.93	2.20	2.10	2.29	2.19	2.33	2.37	2.42	1.97	3.82
O_B–P	3.31	3.17	2.09	2.10	1.94	1.99	1.82	1.86	1.58	1.59
Fe–O_α	2.19	2.12	2.13	2.17	2.11	2.14	2.11	2.14	2.01	2.02
M–O_β	4.04	2.48	2.20	2.40	2.15	2.38	2.12	2.37	1.97	2.19
P–O_L	1.69	1.67	1.77	1.77	1.80	1.80	1.93	1.92	4.83	5.47

3.3.4 The Origin of GpdQ's Catalytic Power

While previous computational results strongly correlate with experimental results and tentatively shed light on the mechanistic nature of GpdQ, they do not directly nor specifically establish the underlying sources of the different catalytic rates. In the current section, our objective is to answer the following questions:

- Why is the catalytic rate/energy barrier of Fe(II)/Cd(II)–GpdQ so high/low compared to other considered metal ion compositions?
- How do the electronic properties of the different Fe(II)/M(II)–GpdQ systems vary?
- Can these properties explain why Fe(II)/M(II)–GpdQ have different catalytic rates?

It is relatively well known that enzymes lower free energy barriers by increasing and decreasing the free energies of reactant and product complexes respectively. Hence we looked at the possibility that the reactivity of the reactants and rate-determining transition states (TS2) differ between different Fe(II)/M(II)–GpdQ systems.

With minor exceptions in the M–O_β and M–Fe bonds, the geometries of the TS2 complexes in different Fe(II)/M(II)–GpdQ systems are very similar. The conformations and geometries of the bipyramidal transition state substrates are consistent in different metal ion compositions. These similarities suggest that transition state stabilization is not the cause of the different of Fe(II)/M(II)–GpdQ catalytic rates.

The geometries of reactant complexes, on the other hand, differed significantly in two key aspects: the M–O_β interaction (see Figure 3.5 and associated text) and the $\angle O_\beta$–P–O_α angle. Indeed the length of the former for Mn, Co, Zn and Cd complexes are 2.01, 1.94, 1.93 and 2.20 Å respectively. The $\angle O_\beta$–P–O_α angles are 122.9°, 123.1°, 123.2° and 118.8° for Mn, Co, Zn and Cd complexes respectively. This means that the $\angle O_\beta$–P–O_α angle for the Cd–RC is roughly 4.3° smaller than those of Mn, Co and Zn which resemble one another. This deviation results from the bidentate coordination of the substrate, which compresses the $\angle O_\beta$–P–O_α angle. More importantly by computing the relative energies of the DMP substrates in their bound active site conformations, this $\angle O_\beta$–P–O_α compression destabilizes DMP by 20.8 kJ mol^{-1} compared to DMP bound to Mn, Co and Zn containing GpdQ. This reactant destabilization corresponds to 88.5 and 56.5% of the $\Delta\Delta G^{\ddagger}$ between Cd/Mn and Cd/Zn mechanisms respectively.

Although reactant destabilization may constitute a significant portion of the enhanced catalytic rate/lowering of the energy barrier, it does not make up entirely its totality. We therefore attempted to investigate the electronic properties of the reactant active sites by looking at their HOMO. Although the HOMO was delocalized over many atomic centers and had an antibonding nature, a significant proportion was located around the bridging hydroxide moiety. The size of that portion is large in the case of the Cd, medium-sized in Mn and Co cases and small in the Zn containing system (see figure 3.5). Furthermore, HOMO-LUMO energy differences for Zn, Co, Mn and Cd systems are -4.52, -4.41, -3.99 and -4.04 eV respectively, consistent with the notion that the HOMO in Cd and Mn

systems is destabilized. The notion that in Cd and Mn-GpdQ, the electrons occupying the HOMO are more polarizable, more distant from the nuclei and more loosely held to the hydroxide moiety than corresponding Co and Zn-GpdQ is consistent with the sizes and HOMO-LUMO energy differences.

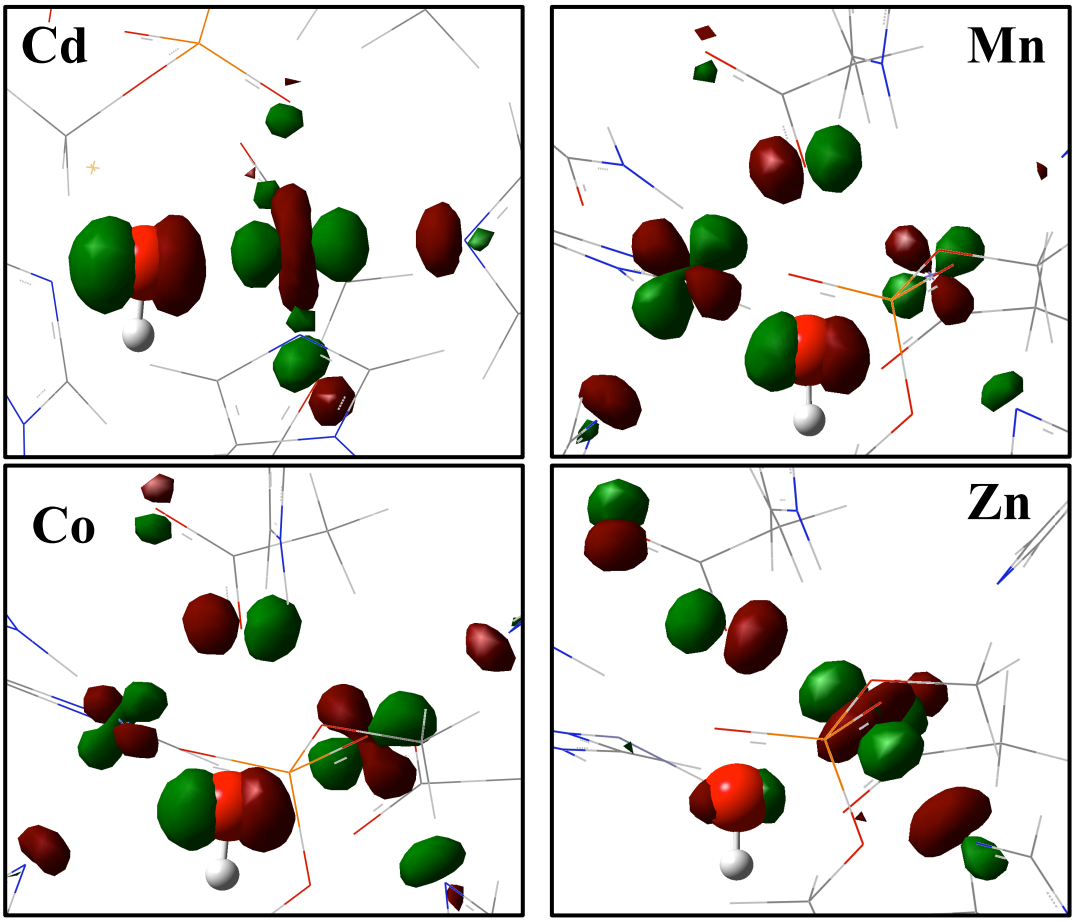


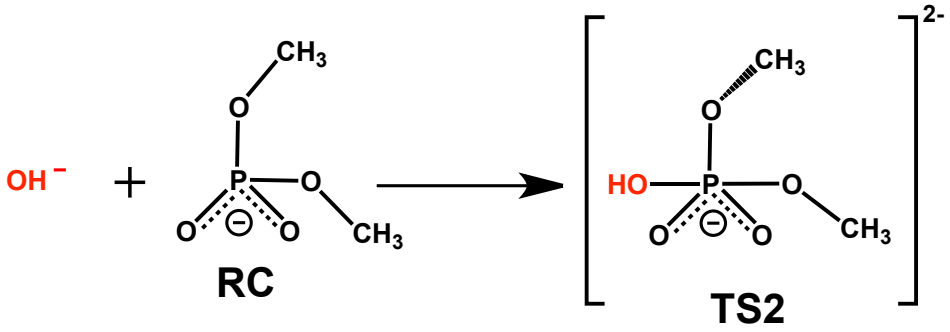
Figure 3.9. Selected portion of HOMO (isovalue=0.07) of Cd (up/left), Mn (up/right), Co (down/left) and Zn (down/right) containing RC's obtained at B3LYP/aug-cc-pVTZ-PP(Cd),6-311+G(2d,2p)(H,C,N,O,Fe) for Cd and B3LYP/6-311+G(2d,2p) for Mn, Co and Zn containing systems.

Hard-soft acid-base (HSAB) theory[20] identifies Lewis acid and bases as either being hard or soft. Hard Lewis acid/base are characterized as being hard to polarize, having

small size and high oxidation states. Examples of hard Lewis bases include F^-, Cl^-, CO_3^{2-} and OH^-. Soft acid/bases are characterized as being easy to polarize, having larger sizes and lower oxidation states. Examples of soft Lewis acids include Cu^+, Ag^+, Pd^{2+} and Pt^{2+}. The central notion in HSAB theory is that "hard-likes-hard and soft-likes-soft." That is, the formation of complexes involving hard acid–bases or soft acid–bases is favored thermodynamically. The hardness (η) of electrophiles (Lewis acids) and nucleophiles (Lewis bases) can be approximated using the following equation:[21]

$$\eta = \frac{E_{LUMO} - E_{HOMO}}{2} \tag{3.1}$$

The mechanistic steps extending from RC to TS2 can be seen as the formation of a phosphate-based bipyramidal complex from a Lewis acid (a distorted DMP) and Lewis base (OH^-) (see scheme 3.6). While it is known that OH^- is a hard nucleophile, we computed the hardness of the distorted DMP to be 2.24 eV using equation (3.1), which is similar in value to that of 4-hydroxy-2-nonenal (a soft electrophile[22] with η=2.55 eV[23]). This suggests that the formation of the phosphate-centered bipyramidal complex is unstable partly because it is a soft acid/hard base complex.



Scheme 3.6. Schematic representation of the formation of a phosphate based bipyramidal complex with hydroxide Lewis base and DMP as Lewis acid.

As seen previously, it appears that the hardness of the hydroxide bridging ligand varies between different Fe(II)/M(II)–GpdQ metal ion compositions. Specifically, in Cd and Mn-GpdQ the hydroxide is a softer nucleophile than in Co and Zn-GpdQ. Hence by using HSAB theory, it is plausible that the phosphate bipyramidal complex in TS2 is more stable for Cd and Mn complexes than for Co and Zn ones due the ability of the former two to "soften" the Lewis basicity of the nucleophile allowing for an enhanced association to the distorted DMP, a soft electrophile. A more stable TS2 bipyramidal complex would translate into an enhanced catalytic rate. Indeed, such a rationale is consistent with experimental rates of 2.35 and 2.15, 0.68 and 0.21 s^{-1} for Cd, Mn, Co and Zn,[10] which like our DFT results show Cd and Mn to be in a class of their own in terms of catalytic rates distinguished from Co and Zn.

3.4 Conclusions

The mechanism of glycerophosphodiesterase of *Enterobacter aerogenes* (GpdQ) has been investigated through the use of DFT. Many insights that highly correlate with experimental trends have been acquired and provide an atomistic scale description of the molecular events and driving forces behind the biocatalysis of GpdQ. The key results and analyses are provided as follow:

- An assessment of the protonation state of the active site of GpdQ involving a statistical comparision between DFT models and crystal structure suggests the bridging ligand to be a hydroxide, the terminal ligand to be a water and His81 to adopt a neutral state.
- A small assessment of substrate binding coupled with free energy calculations indicate that dimethylphosphate (DMP) can spontaneously displace both hydroxide and water terminal ligands allowing DMP to coordinate to the metal in the α site. In

all cases, optimized DFT models indicate that a bridging hydroxide is a more favorable nucleophile compared to the terminal ligand.

- Two mechanistic scenario have been investigated: the direct nucleophilic attack of DMP by a bridging hydroxide involving a neutral His81 and an acidic His81 aided mechanism that also involved a nucleophilic attack by a bridging hydroxide. The free energy barriers calculated for the former were excessively high and did not concur with experimental results. The latter provide reasonable energy barrier that agreed qualitatively with experimental trends. Coupled with the analysis from the active site protonation state DFT assessment, this two-step mechanism suggests that substrate binding increases the pKa of His81, allowing the latter to play a key mechanistic role.
- The source of Fe(II)/Cd(II)–GpdQ's enhanced catalytic ability is due partly to its ability to structurally distort the substrate in the reactant complex. Fe(II)/Cd(II)–GpdQ destabilizes the reactant as a catalytic strategy which accounts for 56.5% of its enhanced rate in comparision to Fe(II)/Zn(II)–GpdQ. An investigation of the electronic state of Fe(II)/M(II)–GpdQ system suggests that Cd and Mn change the reactivity of the bridging hydroxide nucleophile by making it a softer Lewis base. This ultimately allows for the formation of a more stable activated phosphate-centered bipyramidal complex during the rate-determing steps of the mechanism. This enhanced stability lowers the Gibbs free energy barrier and increases the reaction rate as correlated with experimental results.

3.5 References

(1) Jackson, C. J.; Carr, P. D.; Liu, J.-W.; Watt, S. J.; Beck, J. L.; Ollis, D. L. *J. Mol. Biol.* **2007**, *367*, 1047.

(2) Hadler, K. S.; Tanifum, E. A.; Yip, S. H.-C.; Mitic, N.; Guddat, L. W.; Jackson, C. J.; Gahan, L. R.; Nguyen, K.; Carr, P. D.; Ollis, D. L. *J. Am. Chem. Soc.* **2008**, *130*, 14129.

(3) McLoughlin, S. Y.; Jackson, C.; Liu, J.-W.; Ollis, D. L. *Appl. Environ. Microbiol.* **2004**, *70*, 404.

(4) Gerlt, J. A.; Whitman, G. J. *J. Biol. Chem.* **1975**, *250*, 5053.

(5) Ghanem, E.; Li, Y.; Xu, C.; Raushel, F. M. *Biochemistry* **2007**, *46*, 9032.

(6) Ely, F.; Foo, J. L.; Jackson, C. J.; Gahan, L. R.; Ollis, D. L.; Schenk, G. *Curr. Top. Biochem. Res.* **2008**, *2008*, 63.

(7) Ghanem, E.; Raushel, F. M. *Toxicol. Appl. Pharm.* **2005**, *207*, 459.

(8) Jackson, C. J.; Hadler, K. S.; Carr, P. D.; Oakley, A. J.; Yip, S.; Schenk, G.; Ollis, D. L. *Acta Crystallograph. Sect. F Struct. Biol. and Cryst. Commun.* **2008**, *64*, 681.

(9) Hadler, K. S.; Mitic, N.; Ely, F.; Hanson, G. R.; Gahan, L. R.; Larrabee, J. A.; Ollis, D. L.; Schenk, G. *J. Am. Chem. Soc.* **2009**, *131*, 11900.

(10) Daumann, L. J.; McCarthy, B. Y.; Hadler, K. S.; Murray, T. P.; Gahan, L. R.; Larrabee, J. A.; Ollis, D. L.; Schenk, G. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 425.

(11) Mirams, R. E.; Smith, S. J.; Hadler, K. S.; Ollis, D. L.; Schenk, G.; Gahan, L. R. *J. Biol. Inorg. Chem.* **2008**, *13*, 1065.

(12) Schenk, G.; Mitić, N.; Gahan, L. R.; Ollis, D. L.; McGeary, R. P.; Guddat, L. W. *Acc. Chem. Res.* **2012**, *45*, 1593.

(13) Hadler, K. S.; Gahan, L. R.; Ollis, D. L.; Schenk, G. *J. Inorg. Biochem.* **2010**, *104*, 211.

(14) Hadler, K. S.; Mitic, N.; Yip, S. H.-C.; Gahan, L. R.; Ollis, D. L.; Schenk, G.; Larrabee, J. A. *Inorg. Chem.* **2010**, *49*, 2727.

(15) Inc., C. C. G. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.

(16) 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., J. A. ; 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. P., 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, Inc., CT,: Wallingford 2009.

(17) Chen, S.-L.; Fang, W.-H.; Himo, F. *J. Phys. Chem. B* **2007**, *111*, 1253.

(18) Alberto, M. E.; Marino, T.; Ramos, M. J.; Russo, N. *J. Chem. Theory Comput.* **2010**, *6*, 2424.

(19) Mitic, N.; Smith, S. J.; Neves, A.; Guddat, L. W.; Gahan, L. R.; Schenk, G. *Chem. Rev.* **2006**, *106*, 3338.

(20) Pearson, R. G. *J. Am. Chem. Soc.* **1963**, *85*, 3533.

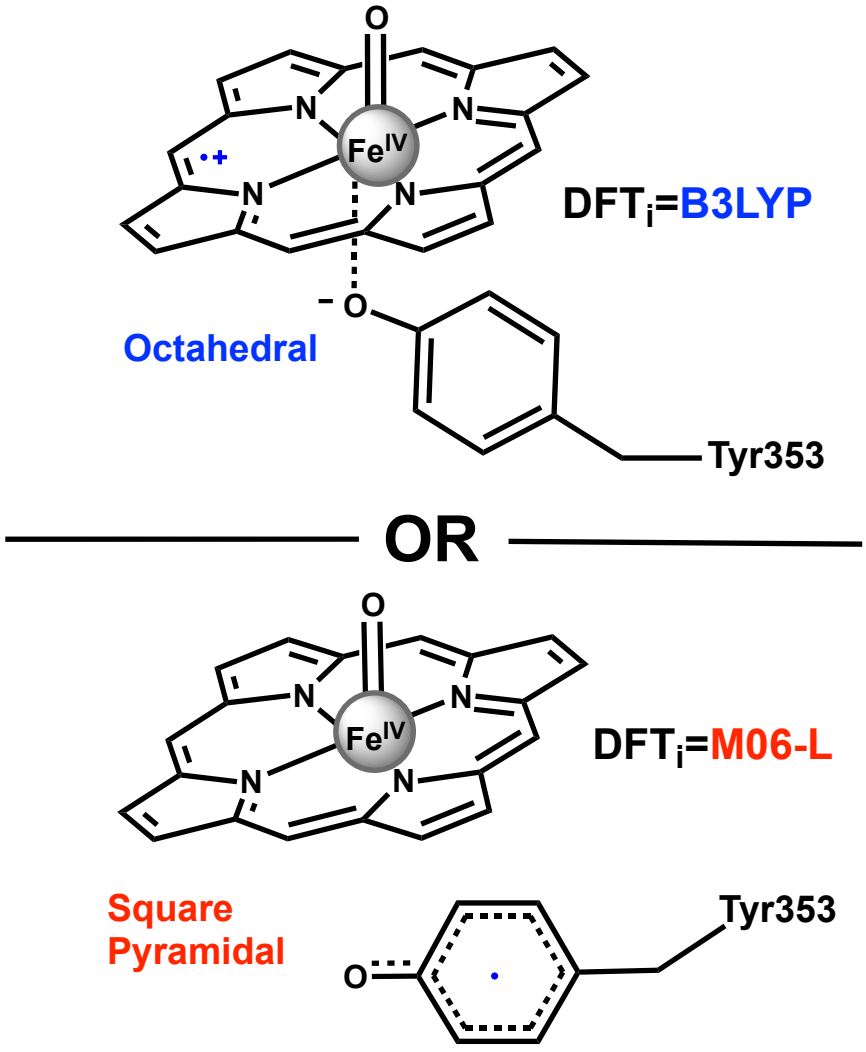
(21) Pearson, R. G. *Inorg. Chem.* **1988**, *27*, 734.

(22) LoPachin, R. M.; Geohagen, B. C.; Gavin, T. *Toxicol. Sci.* **2009**, *107*, 171.

(23) LoPachin, R. M.; Gavin, T.; DeCaprio, A.; Barber, D. S. *Chem. Res. Toxicol.* **2011**, *25*, 239.

Chapter 4

Modeling Compound I in Coral Allene Oxide Synthase: A Challenge for DFT



4.1 Introduction

Many iron-heme enzymes such as peroxidases, catalases, P450s and nitric oxide synthases, use high-valent iron oxo intermediates in their mechanisms.[1] Of particular importance amongst such species is compound I (Cpd I); which has been suggested to be one of the strongest oxidizing agents found in nature.[2] As a result, the properties and reactivity of Cpd I has been the subject of numerous experimental and computational investigations.[1-5] In general, it has been shown that although some properties may alter from one enzyme to another, others remain consistent. For instance, in most cases Cpd I has low lying near-degenerate doublet and quartet spin-states.[2,6,7] In P450s, for example, the doublet spin-state has two electrons with parallel spin located on the Fe(IV)=O moiety and a single anti-ferromagnetically coupled electron residing on the porphyrin ring ($Por^{\bullet+}$). In contrast, in the quartet spin-state they are ferromagnetically coupled (Figure 4.1).

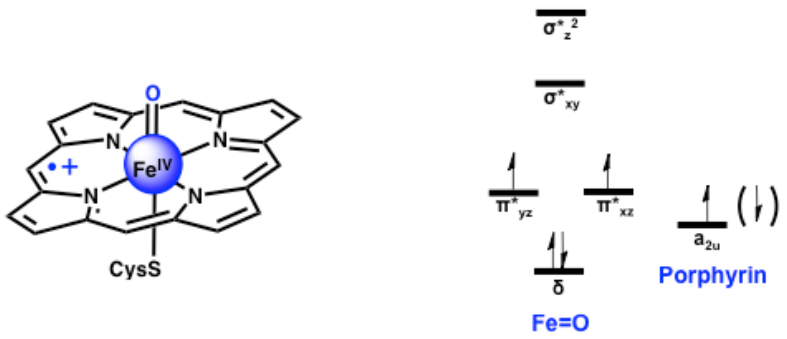


Figure 4.1. Schematic representations of (left) Cpd I and (right) electronic configurations of its near degenerate lowest-lying doublet (in parentheses) quartet and doublet states.

In addition to its natural formation, Cpd I can be generated biosynthetically using reactive oxidative agents such as H_2O_2, m-CPBA,[8] peroxyacetic acid,[9] and iodosylbenzene (PhIO).[10] This approach has been used for instance to investigate the electronic properties of Cpd I in catalases.[11] EPR spectroscopy and rapid-quenching

freeze methods have suggested that while Cpd I complexes in some catalases (e.g. bovine liver catalase (BLC)) involve a radical on the Iron's axial tyrosinate ligand,[11] others (e.g. *Proteus mirbilis* (PCM)[12] and *Micrococcus luteus* (MLC)[13]) involve a porphyrin radical. In the case of PCM an X-Ray crystal structure of a putative Cpd I-like intermediate gave an Fe—O_{Tyr337} (Tyr337 being the axial ligand) bond length of approximately 2.12 Å.[14] Meanwhile, DFT studies using a chemical cluster model of a catalase-type active site have suggested that Cpd I has Fe—O_{Tyr337} bond distances of 2.30 and 2.17 Å for doublet and quartet respectively, with an electron radical delocalized between the porphyrin moiety and the tyrosinate axial Fe-ligand.[15]

Coral allene oxide synthase (cAOS) is a heme-enzyme with an 11% amino acid sequence identity to catalase.[16] Like catalase, its active site contains an axial tyrosinate ligand and distal His, Asn and Ser ligands.[17,18] However, unlike catalase it cannot perform the decomposition of H_2O_2 into H_2O and O_2.[18] Remarkably, it instead catalyzes the formation of an allene epoxide from a fatty acid hydroperoxide without the need for the reactive Cpd I intermediate.[19,20] However, a number of studies have induced the formation of Cpd I in cAOS and subsequently investigated its electronic structure and reactivity. That is, its formation has been induced in an active site that appears suited for neither its formation nor reactivity. For example, in a recent study Boeglin et al.[19] used PhIO to generate Cpd I in cAOS and examined the subsequent stereospecific hydroxylation and epoxidation of arachidonic acid.[19] Meanwhile, based on the results of their experimental EPR spectroscopy study, Wu et al.[21] suggested that Cpd I in cAOS (Cpd I/cAOS) may in fact have tyrosyl radical-like characteristics. However, despite such detailed studies, the properties and reactivity of Cpd I/cAOS, and how they differ from or are similar to those in catalase and P450s, remains unclear.

Computational approaches have been shown to be suited to examination of such species.[1,15,22]. DFT methods, often in a QM-chemical cluster or increasingly a QM/MM approach, have become the methodology of choice for modeling the electronic and

structural properties of Cpd I.[2,6] Amongst the existing extensive variety of exchange-correlation functionals; B3LYP is generally preferred for P450-related studies[23] and is currently the workhorse of most DFT studies.[24] However, several studies have suggested that B3LYP may have inherent limitations in predicting the structural and energetic properties of metal-containing systems.[25-27] For instance, it has been suggested that reducing the amount of Hartree-Fock (HF) contribution to the exchange from 20% to 15% (B3LYP*) may better reproduce experimental data.[27] Thus, in recent years there has been efforts to develop functionals that may provide greater accuracy and reliability for such systems.[28] Indeed, some studies have suggested that meta- and hybrid-meta GGA's (e.g. M06 and M06-L) may be suitable.[29]

The study presented herein is an assessment of the ability of a variety of pure-, hybrid, meta- and meta-hybrid-DFT functionals (i.e., BLYP, B3LYP*, B3LYP, M06, M06-L), within an ONIOM(QM/MM) framework, to model Cpd I within cAOS, in the presence of arachidonic acid. In addition, using a B3LYP/MM framework, the effect of basis set choice upon key structural parameters such as the Fe(IV)···O_{Tyr} distance and Cpd I/cAOS···substrate interactions, and the electronic configuration of Cpd I/cAOS are examined. Such a study provides insights into the effect of DFT/MM method choice when modeling high-valent heme-iron oxo species, and serves as an important step towards reliable mechanistic studies of Cpd I reactivity in such enzymes as cAOS.

4.2 Computational Method

4.2.1 Molecular Docking

Several X-Ray crystal structures are now available[30] of cAOS and relevant complexes that may be used as templates to model Cpd I within cAOS. The starting point of this study was an X-Ray crystal structure of a holoenzyme of coral allene oxide synthase 8R-lipoxygenase (PDB code: 3DY5).[30] As it does not contain substrate (arachidonic acid) bound in the active site, molecular docking was performed using the London dG scoring

function and the AMBER99 force field within the Molecular Operating Environment (MOE)[31] software package. It is noted that as AMBER99 does not possess the parameters to model Cpd I, the substrate analogue 8R-HPETE was docked instead of arachidonic acid, with an oxygen atom bound to Fe(III). The top one hundred scoring Michaelis complexes were subsequently minimized with AMBER99 until their potential energies' root-mean-square gradients fell under 0.05 kcal au^{-1}. The resulting minimized complex with the highest binding free energy, as calculated via the London dG scoring function in conjunction with AMBER99, was then used to obtain a suitable chemical model for the subsequent QM/MM studies (see below).

4.2.2 QM/MM Model

The above MM optimized structure was modified to contain Cpd I and arachidonic acid. Specifically, 8R-HPETE was mutated in silico to arachidonic acid while an oxo ligand was added in the axial position to the iron center. The QM region contained all reactive parts of the polyunsaturated substrate, the heme moiety (porphyrin side chains were inserted in the MM layer), and side-chains of Thr66, His67, Arg349 and Tyr353. It should be noted that the roles of Thr66 and His67 have been previously investigated experimentally and computationally and shown to be catalytically relevant. The C_α centers of all residues in the MM layer were held fixed at their AMBER96 minimized positions (see above) so as to ensure the overall integrity of the model.

4.2.3 QM/MM Calculations

All QM/MM calculations were performed using a two-layer ONIOM[32] scheme with mechanical embedding as implemented in the Gaussian 09[33] suite of programs. The QM region, that in which the reaction occurs, was modeled with either the BLYP, B3LYP*,[27] B3LYP,[34] M06[29] or M06-L[29,35] functionals. The MM region was modeled using the AMBER96[36] force field. Geometry optimizations involved the use of different basis sets

(see subsequent text for details). For all calculations, unless otherwise noted, the LANL2DZ basis set has been used on the Fe center. BS1, BS2 and BS3 denote the use of 6-31G, 6-31G(d) and 6-31G(d,p) respectively on C, H, N, O and LANL2DZ(f) on Fe, i.e., an *f*-type polarization function has been included on Fe. For all reported spin-densities and relative energies herein, unless otherwise noted, the 6-311+G(2df,p) basis set has been used. It should be noted that throughout this study the doublet, quartet and sextet spin states of each species has been considered.

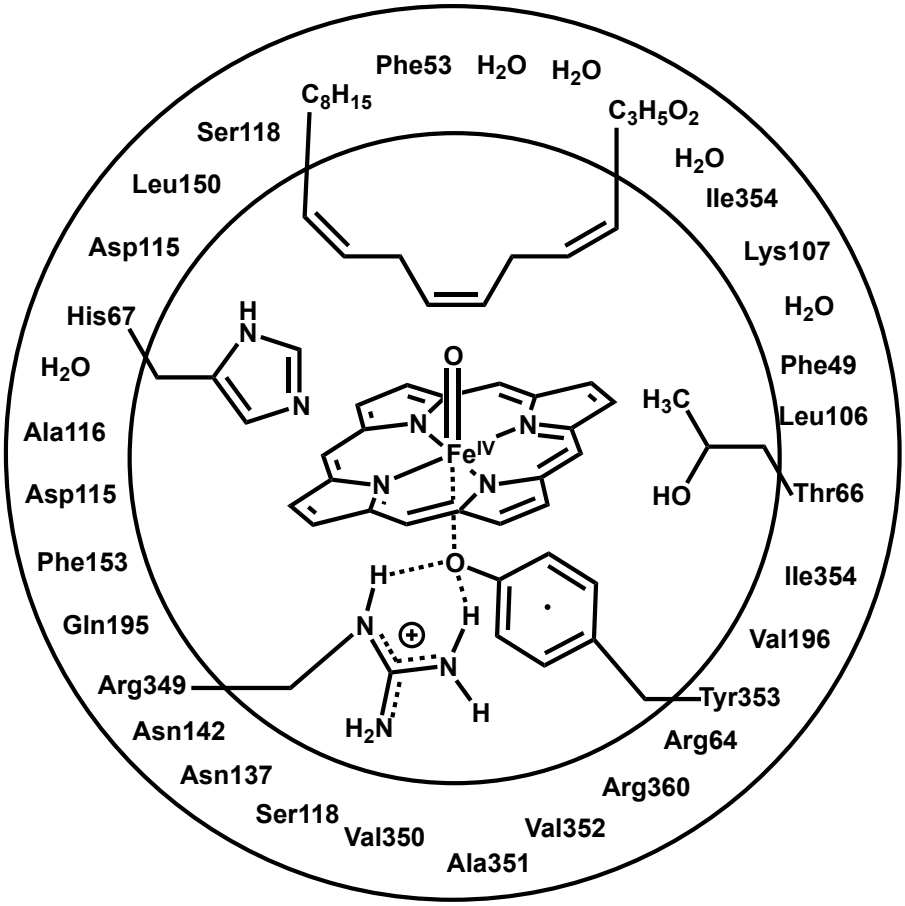


Figure 4.2. Schematic representation of the QM/MM active site model for Cpd I/cAOS complex. The components in the inner circle occupy the QM layer while those in the outer circle occupy the MM layer.

4.3 Results and Discussion

4.3.1 Basis Set Assessment

At present B3LYP is arguably the most common DFT method of choice for computational studies on Cpd I and related systems, and has been widely applied previously with success.[2,6,15] Hence, using the B3LYP method, we investigated the effects of basis set choice on the optimized geometry of Cpd I/cAOS.

Initially, the optimized structure of Cpd I/cAOS was obtained using the 6-31G basis set on all atoms except iron, for which the ECP basis set LANL2DZ was used. It is noted that this level of theory has been widely used in, for example, previous P450 studies.[2,6] The resulting optimized structures with select distances are shown in Figure 4.3. At this level of theory Tyr353 is ligated via its side-chain oxygen to the Fe center with Fe···O_{Tyr353} distances of 2.28, 2.23 and 2.12 Å for the doublet, quartet, and sextet states, respectively. Meanwhile, the Fe–O_{Distal} distance is consistently approximately 1.64 Å, in agreement with previous DFT-based studies on Cpd I in P450.[2,6] In addition, the double bond (C1=C2) of the arachidonic acid substrate to be oxidized is situated near the Fe=O moiety with r(FeO···C_1) and r(FeO···C_2) distances ranging from 3.72 to 3.84 Å, respectively. Furthermore, short and strong hydrogen bond interactions between the guanidinium of Arg349 and the phenolate oxygen of Tyr353 are observed, with $_{Arg349}NH_2$···O_{Tyr353} distances of 1.82 to 1.97 Å (Figure 4.3).

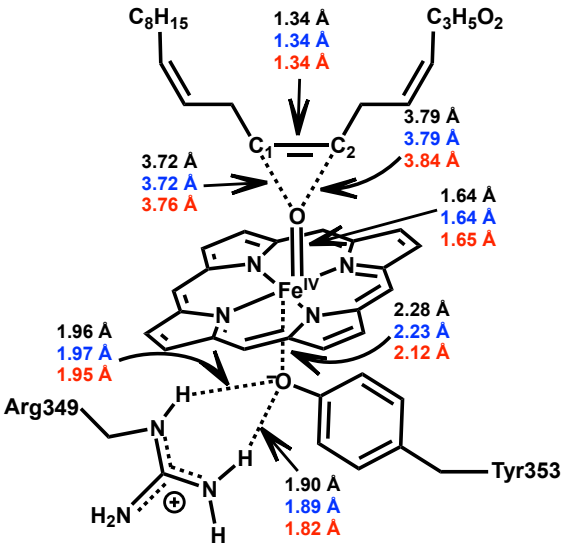


Figure 4.3. Schematic representation of the structure of Cpd I/cAOS optimized at the ONIOM(B3LYP/6-31G:AMBER96) level of theory with selected bond distances (angstroms). Color code: black (singlet), blue (quartet) and red (sextet).

The largest change in the optimized geometry obtained for Cpd I/cAOS was observed upon including *d*-type polarization functions on all first-row heavy atoms, i.e., 6-31G to 6-31G(d), as can be seen in Table 4.1. Most significantly, it results in cleavage of the Fe···O_{Tyr353} ligation interaction which has now increased by at least 0.76 Å for all three spin-states. Consequently, at the B3LYP/6-31G(d) level of theory Cpd I/cAOS is calculated to be a square pyramidal rather than octahedral complex, as commonly seen in most P450 systems.[2] It is noted that the inclusion of *d*-functions has only quite minor or negligible affects on other interactions.

Table 4.1. Optimized Fe–O_{Tyr} distances (Angstroms) obtained at the ONIOM(B3LYP/**BS_i**:AMBER96) level of theory.

Basis Set (BS$_i$)	**Spin State**	**Fe···O$_{Tyr}$ (Å)**
6-31G[a]	Doublet	2.28
	Quartet	2.23
	Sextet	2.12
6-31G(d)[a]	Doublet	3.06
	Quartet	3.08
	Sextet	3.51
6-31G(d,p)[a]	Doublet	3.11
	Quartet	3.13
	Sextet	3.53
6-31+G[a]	Doublet	2.28
	Quartet	2.24
	Sextet	2.11
6-31+G(d)[a]	Doublet	3.11
	Quartet	3.12
	Sextet	3.50
BS1[b]	Doublet	2.28
	Quartet	2.24
	Sextet	2.12
BS2[b]	Doublet	3.04
	Quartet	3.08
	Sextet	3.51
BS3[b]	Doublet	3.13
	Quartet	3.13
	Sextet	3.52

[a] with LANL2DZ on the Fe center. [b] with LANL2DZ(f) on the Fe center.

Despite the O_{Tyr353} possibly having some anionic character, the inclusion of diffuse functions to the 6-31G basis set (i.e., increasing it to 6-31+G) has only quite minor affects on the Fe···O_{Tyr353} distance. In contrast, addition of such functions to the 6-31G(d) basis set (i.e., increasing it to 6-31+G(d)) changes the Fe···O_{Tyr353} length by 0.04 – 0.05 Å for the doublet and quartet states, but only 0.01 Å for the sextet state (see Table 4.1). The inclusion of *f*-type functions on the Iron center resulted in only quite minor changes in the

optimized geometry (see Table 4.1). For instance, optimized geometries obtained using the BS1, BS2 and BS3 basis sets give Fe···O_{Tyr353} distances within 0.02 Å of those obtained using the corresponding basis set without such functions (see Table 4.1).

To further investigate the effects of Fe–O_{Tyr353} ligation on the electronic properties of Cpd I/cAOS, and given the above observed basis set dependence effects, single point calculations were performed at the ONIOM(B3LYP/(6-311+G(2df,p):AMBER96)// ONIOM(B3LYP/**6-31G**:AMBER96) and ONIOM(B3LYP/(6-311+G(2df,p):AMBER96)// ONIOM(B3LYP/**6-31G(d,p)**:AMBER96) levels of theories. The results obtained are shown in Table 4.2. Conformational energy differences are also given and are the differences in energy (for the same spin-state) between single point calculations performed at the same level of theory, but with geometries optimized using ONIOM(B3LYP/6-31G:AMBER96) and ONIOM(B3LYP/6-31G(d,p):AMBER96). That is, they are indicative of the effects of ligation of Tyr353 to Fe(IV) and delocalization of the electron radical.

For both choices of optimized geometry the spin density values of Fe and O_{Distal} are, in general, consistent with the notion that the latter has radical character. There are, however, some notable geometry dependent differences observed. For example, for geometries obtained at the ONIOM(B3LYP/6-31G:AMBER96) level, the spin densities of Tyr353 and porphyrin indicate that ligation of O_{Tyr353} to Fe causes delocalization of an unpaired electron (originating from the electron hole in the porphyrin moiety) between the porphyrin ring and Tyr353. However, in the case of geometries obtained using ONIOM(B3LYP/6-31G(d,p):AMBER96), which notably has no ligation between Fe and O_{Tyr353}, the spin densities values of Ty353 and porphyrin posits that the electron radical is located on Tyr353 instead of the porphyrin ring as seen in conventional P450 systems[2] and Cpd I/catalase.[15] An exception to the general trends, for both selected geometry optimization levels, is seen in the spin densities of the sextet states. For structures obtained using ONIOM(B3LYP/6-31G(d,p):AMBER96), the sextet manifests itself with

Fe adopting a high spin configuration. In contrast, with structures obtained using ONIOM(B3LYP/6-31G:AMBER96) the Fe has a low spin configuration with a triplet-state porphyrin ring.

Table 4.2. Selected spin densities, relative spin splitting and conformational energies[a] obtained at the ONIOM(B3LYP/(6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/**BS$_i$**:AMBER96) level of theory

Basis Set (BS_i)	Spin State	Fe	O_{Distal}	Tyr353	Porphyrin	Spin Splitting ΔE (kJ mol^{-1})	Conformational ΔE (kJ mol^{-1})[a]
6-31G	Doublet	1.32	0.87	-0.49	-0.68	0.0	0.0
	Quartet	1.25	0.89	0.46	0.37	0.9	0.0
	Sextet	1.27	0.89	0.26	2.53	206.8	0.0
6-31G(d,p)	Doublet	1.34	0.80	-0.96	-0.18	0.0	-32.5
	Quartet	1.32	0.80	0.96	-0.09	0.0	-33.4
	Sextet	3.53	0.49	0.98	0.14	41.6	-197.6

[a] E[ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G:AMBER96)] – E[ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/**BS$_i$**:AMBER96)]

The spin densities in Table 4.2 for Cpd I/cAOS obtained using geometries optimized at the ONIOM(B3LYP/6-31G:AMBER96) level do not appear to correlate with the experimental EPR results of Wu et al.[21] The former suggests the delocalization of an electron radical between Tyr353 and the porphyrin moiety, while the latter instead suggests the formation of a tyrosyl radical (i.e. without delocalization). On the other hand, the use of ONIOM(B3LYP/6-31G(d,p):AMBER96) optimized geometries does generate spin densities that bear closer correlation with experiment; i.e. both indicate the presence of a tyrosyl radical. However, it should be noted that Wu *et al*.,[21] based on results of their experimental mutagenesis and EPR studies, suggested that a tyrosyl (Tyr193) somewhat removed from the active site was the site of the radical in Cpd I/cAOS. In contrast, the present results suggest that the radical species in question may instead be the iron's axially ligated tyrosyl (Tyr353).

Regardless of their electronic configurations the use of either optimized geometry suggests that the lowest lying doublet and quartet spin states are degenerate. However

what is remarkable is the destabilization of the sextet in the octahedral iron (i.e., $\mathbf{BS_i}$ = 6-31G) compared to the square pyramidal ($\mathbf{BS_i}$ = 6-31G(d,p)) case. Indeed, in the former complex, the sextet is destabilized significantly by 206.8 kJ mol^{-1}, whereas in the latter this destabilization is reduced to 41.6 kJ mol^{-1}! As can be seen in Table 4.2, the conformational relative energy differences indicate that the Fe–O_{Tyr353} coordination destabilizes Cpd I/cAOS by 32.5, 33.4 and 197.6 kJ mol^{-1} for doublet, quartet and sextet spin states respectively. In the case of the sextet, its high energy may originate from its adoption of a triplet porphyrin moiety.

4.3.2 Axial Ligand Environment

The environment surrounding the axial thiolate ligand in many hemeproteins is reasonably conserved.[37,38] In P450cam the thiolate's anionic sulfur is surrounded by the backbone amide groups of Gln360, Gly359 and Leu358. These have been suggested to stabilize Fe-S_{Cys357} coordination and the sulfur's negative charge.[38,39] This suggests that the hydrogen-bonding environment surrounding the cysteine may also be catalytically relevant. Hence, the effect of hydrogen-bonding environment surrounding Ty353 in Cpd I/cAOS, was considered. To do so, the QM/MM model was modified by moving the guanidinium of Arg349 from the QM to the MM layer. For all spin states, at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level the Fe–O_{Distal} distance shortens to 2.17 Å. It should be noted that these results also appear consistent with a previous computational study by Wang *et al*.[22] on Cpd I within catalase.[40] In particular, using a B3LYP/LACVP- chemical cluster model approach that included Tyr337 (modeled as tyrosinate) and Arg333 (modeled as a guanidinium), they obtained Fe-S_{Cys357} distances of 2.117 and 2.135 Å for the doublet and quartet states respectively.[22] Thus, the present results further demonstrate the importance using basis sets with polarization in conjunction with a proper hydrogen-bonding environment when modeling high-valent oxo-iron species. Furthermore, they show that inclusion of the proximal ligands' hydrogen-bonding

environment can influence the structural properties of Cpd I in cAOS and the coordination environment of the Fe center.

4.3.3 DFT Functional Assessment

For the computational study of systems such as iron-containing metalloenzymes, B3LYP has long been the method of choice.[2,6,15] While in many instances it has been shown to give results in good agreement with experiment, in other cases it has lead to significant differences.[41,42] This should come as no surprise since B3LYP was not parameterized with metal-containing reference molecules.[33,34] In addition, it suffers from a number of limitations including the inability to accurately model non-covalent interactions (i.e. Van der Waals) and π-π stacking.[28,43]

The modeling of Cpd I/cAOS presents itself as a challenge for DFT methods since it requires an exchange-correlation functional that not only performs well for metal complexes but also for medium-range interactions. Indeed the functional needs to accurately model the coordination of the iron center and the dispersion interactions between the polyunsaturated hydrocarbon substrate, porphyrin ring, and tyrosinate functional group. Furthermore, Cpd I is an open-shell system hence the DFT method must adequately predict orbital occupations.

Thus, the ability of a range of functionals (BLYP, B3LYP*, B3LYP, M06 and M06-L), within an ONIOM framework, to reliably predict the structure and properties of Cpd I/cAOS were examined. It should be noted that while M06 has been shown to possess good accuracy in general for transition metals and medium-range correlation interactions, M06-L has been shown to be particularly reliable in modeling transition metal complexes.[29] The results obtained are shown in Table 4.3.

The optimized geometries of Cpd I/cAOS obtained at the ONIOM(DFT_i/6-31G(d,p):AMBER96) level (DFT_i = BLYP, B3LYP, B3LYP*, M06, M06-L) revealed dramatic deviations from those obtained using B3LYP, as shown in Table 4.3. Indeed,

while the use of B3LYP predicted every spin state to be pentacoordinated, each of the above four alternate functionals predict the doublet and quartet complexes to be hexacoordinated. For the sextet, as for B3LYP, B3LYP* and M06 predict the complex to be square pyramidal with Fe–O_{Tyr353} distances of 3.53, 3.49, and 3.28 Å respectively. Although these three functionals show qualitative correspondence in coordination, it should be stressed that the standard deviation of predicted Fe–O_{Tyr353} distances of pentacoordinated complexes is 0.13 Å indicating pronounced quantitative disparities between functionals. M06-L gives Fe–O_{Tyr353} distances for the doublet and quartet complexes that are 0.21-0.23 Å shorter than obtained using M06. For the sextet, however, M06-L gives a distance that is 0.74 Å less than that observed with M06. Notably, BLYP and M06-L consistently suggest the Fe to be octahedral for all spin-states.

Table 4.3. Select optimized distances (Angstroms) obtained at the ONIOM(**DFT_i**/6-31G(d,p):AMBER96) level of theory (**DFT_i** = BLYP, B3LYP*, B3LYP, M06, M06-L).

DFT_i	**Spin State**	**Fe–O_{Tyr} (Å)**	**O_{Distal}–C1 (Å)**	**O_{Distal} –C2 (Å)**
BLYP	Doublet	2.32	3.60	3.64
	Quartet	2.42	3.73	3.82
	Sextet	2.47	3.73	3.78
B3LYP*	Doublet	2.40	3.70	3.76
	Quartet	2.38	3.70	3.75
	Sextet	3.49	3.75	3.68
B3LYP	Doublet	3.11	3.74	3.76
	Quartet	3.13	3.74	3.76
	Sextet	3.53	3.76	3.70
M06	Doublet	2.47	3.37	3.37
	Quartet	2.48	3.37	3.31
	Sextet	3.28	3.38	3.28
M06-L	Doublet	2.24	3.21	3.08
	Quartet	2.27	3.29	3.21
	Sextet	2.54	3.30	3.20

In addition to the shortening of the Fe–O_{Tyr353} interaction, M06 and M06-L cause a significant decrease in the separation between O_{Distal} and the substrate's alkene functional group (C1=C2) at which reaction is to occur, i.e., $r(O_{Distal}\cdots C1)$ and $r(O_{Distal}\cdots C2)$. For

BLYP, B3LYP and B3LYP* the shortest $O_{Distal}\cdots C1$ and $O_{Distal}\cdots C2$ distances are 3.60, 3.70 and 3.78 Å respectively, whereas for M06 and M06-L the *longest* $O_{Distal}\cdots C1$ and $O_{Distal}\cdots C2$ distances are 3.38 and 3.37 Å respectively. These differences may in part be due to the fact that M06 and M06-L were parameterized to describe non-covalent interactions such as those likely present between the substrate and heme moiety. It was noted that this "pulling" of the substrate closer to the oxidant by M06 and M06-L results in the energies of the protein surroundings of Cpd I/cAOS were on average 57.6 kJ mol^{-1} (excluding the sextet state of M06-L) lower than for the models optimized using BLYP, B3LYP* and B3LYP. These results suggest that the choice of functional to be used in a QM/MM framework may not only affect the high-layer but also that of the low-layer.

Spin densities of and relative energy differences between the states were then obtained at the ONIOM(DFT_i/(6-311+G(2df,p):AMBER96)//ONIOM(DFT_i/6-31G(d,p): AMBER96) level of theory and are given in Table 4.4. With the exception of BLYP, which finds the doublet to be 3.5 kJ mol^{-1} lower in energy than the quartet, all functionals predict the doublet and quartet spin states to be essentially degenerate; lying within 0.7 kJ mol^{-1}. Of the functionals considered herein, B3LYP gave the smallest energy difference between the doublet and quartet states at 0.0 kJ mol^{-1}. This reflects perhaps in part the fact that pure density functional methods (e.g., BLYP) tend to favor low spin states.[44] BLYP also significantly destabilizes the sextet, which now lies 121.6 kJ mol^{-1} above the doublet. It is noted that while B3LYP* and B3LYP predict different coordination numbers for the Fe center they give similar relative energies for the sextet state. In particular, B3LYP predicts that the sextet lies 41.6 kJ mol^{-1} higher in energy than the doublet, whereas B3LYP* predicts a difference of 54.6 kJ mol^{-1}. This is a result of B3LYP having 5% more HF contribution than B3LYP* and thus, favoring to a greater extent the sextet spin-state.

As noted above, both M06 and M06-L predict the doublet and quartet states to be near-degenerate; the former favoring the doublet by 0.5 kJ mol^{-1} while the latter favors the

quartet by 0.7 kJ mol^{-1}. However, a significant disparity occurs for the spin splitting energies involving the sextets. Similar to BLYP, B3LYP and B3LYP*, M06-L predicts the sextet to be higher in energy than the doublet, though it gives a small difference of just 22.7 kJ mol^{-1}. In contrast, M06, unlike all other functionals considered, predicts the sextet to lie *lower* in energy than the doublet by 13.3 kJ mol^{-1}. This disparity may in part arise from their different optimized geometries. In particular, the overstabilization of the sextet by M06 may be due to the elongation of the Fe–O_{Tyr353} interaction (cf. Table 4.3). It is noted that computational studies on Cpd I do not always consider its sextet state as it is typically decidedly higher in energy than the doublet and quartet states.[7,45,46]

Table 4.4. Selected spin densities and relative energies obtained at the ONIOM(**DFT_i**/(6-311+G(2df,p):AMBER96)//ONIOM(**DFT_i**/6-31G(d,p):AMBER96) level of theory

DFT_i	Spin State	Fe	O_{Distal}	Tyr353	Porphyrin	ΔE (kJ mol^{-1})
BLYP	Doublet	1.20	0.71	-0.48	-0.34	0.0
	Quartet	1.45	0.84	0.52	0.11	3.5
	Sextet	3.12	0.68	0.54	0.53	121.6
B3LYP*	Doublet	1.33	0.84	-0.64	-0.52	0.0
	Quartet	1.29	0.85	0.64	0.20	0.1
	Sextet	3.30	0.55	0.94	0.18	54.7
B3LYP	Doublet	1.34	0.80	-0.96	-0.18	0.0
	Quartet	1.34	0.80	0.96	-0.09	0.0
	Sextet	3.39	0.49	0.98	0.14	41.6
M06	Doublet	1.19	0.89	-0.77	-0.27	0.0
	Quartet	1.05	0.93	0.86	0.08	0.5
	Sextet	3.25	0.45	0.96	0.21	-13.3
M06-L	Doublet	1.18	0.82	-0.50	-0.35	0.0
	Quartet	1.23	0.95	0.54	0.21	-0.7
	Sextet	3.20	0.55	0.75	0.40	22.7

As seen previously (cf. Table 4.2), the spin densities of Cpd I/cAOS are highly dependent on the whether the complex is penta- or hexacoordinated, which is in turn dependent on the level of theory used. Namely, using the B3LYP method, in pentacoordinated complexes the electron radical is delocalized between the porphyrin

ring and the Tyr353 side chain. In contrast, in hexacoordinated complexes using the same DFT method the radical is localized on the side of chain of Ty353 while the porphyrin is now closed-shell. All of the other functionals considered herein also predict that the doublet and quartet states prefer to be hexacoordinated. Thus, we also examined whether they also predict the electron hole to be delocalized between Tyr353 and the porphyrin moiety or formation of an open-shell Tyr353 adjacent to a closed-shell porphyrin.

As can be seen from Table 4.4, despite having optimized structures that differ substantially (cf. Table 4.3), M06 agrees with B3LYP in predicting the localization of the electron on Tyr353. In contrast, the use of the DFT methods BLYP, B3LYP* and M06-L to describe the high-layer predicts the electron to be delocalized across Tyr353 and the porphyrin. In particular, M06 predicts the Fe=O moiety to contain two parallel-spin electrons delocalized between the iron and oxo centers for the doublet and quartet states. Additionally, the measured spin-densities on the tyrosinate are -0.77 and 0.86 for the doublet and quartet spin states, respectively. Their magnitudes being close to unity is consistent with the notion that the electron is mostly located on Tyr353 and antiparallel in the doublet and parallel in the quartet. For the sextet state, although it is favoured by M06 while B3LYP destabilizes it, both methods predict similar spin densities. Thus, although B3LYP and M06 predict different optimized geometries as well as at times conflicting spin spitting energies for Cpd I/cAOS, they qualitatively give the same electronic configuration.

From Tables 4.3 and 4.4 one can see how the proximity of porphyrin moiety and Tyr353 side-chain enhances the delocalization of the unpaired electron. This notion is consistent with the geometries and corresponding spin-densities calculated using BLYP, B3LYP, B3LYP*, and M06-L. Indeed, the latter three not only predict the coordination of O_{Tyr353} to Fe, with the exception of the sextet B3LYP*, but also delocalization of the radical between the porphyrin and tyrosinate. Although the following discussion mostly considers M06-L, the same arguments apply to BLYP and B3LYP*. For M06-L, the spin

densities of the Tyr353 side-chain are -0.50, 0.54 and 0.75 for the doublet, quartet and sextet respectively, while those on the porphyrin moiety are -0.35, 0.21 and 0.4 (Table 4.4). Their sums are -0.85, 0.75 and 1.15 respectively and their magnitudes are close to unity. This is consistent with the concept that in the doublet the Tyr353 and porphyrin have an unpaired electron antiparallel to those centered on low-spin Fe=O, while in the quartet it is parallel. In contrast, the sextet has a high-spin Fe center with a ferromagnetically coupled delocalized electron.

B3LYP* appears to be unique in that it gives the doublet and quartet to be octahedral and the sextet as square pyramidal while undergoing a change in atomic occupations between conformational changes. In particular, for the doublet and quartet states the spin densities indicate that the electron is delocalized between the porphyrin and Tyr354 whereas in the sextet the electron is localized on Tyr353 (Table 4.4). This is illustrated by, for example, the fact that the magnitudes of the Tyr353 centered spin densities are 0.64 in both the doublet and quartet. However, as the Fe–O_{Tyr353} distance increases by 1.11 Å on going from the quartet to the sextet, this spin density increases to 0.94 indicating a fully localized unpaired electron centered on Tyr353.

The electronic structures of Cpd I/cAOS predicted by B3LYP and M06 appear to be in reasonable agreement with experimental EPR results,[21] while those from BLYP, B3LYP* and M06-L do not. Since B3LYP destabilizes the sextet spin state, similar to most other functionals considered herein, the present results suggest that B3LYP may provide a more consistent model of Cpd I/cAOS than M06.

In order to gain further insights into the effect of DFT method choice on the predicted properties of Cpd I/cAOS, single-point calculations were performed at the ONIOM(DFT_i/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory. The differences in energy from those obtained at the ONIOM(DFT_i/6-311+G(2df,p):AMBER96)//ONIOM (DFT_i/6-31G(d,p):AMBER96) level give insight into whether at larger basis sets the functional favors the pentacoordinated iron and longer

substrate···heme distances predicted at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory. The results are summarized in Table 4.5.

Table 4.5. Difference in energies (ΔE; kJ mol^{-1}) between models obtained at the ONIOM(**DFT$_i$**/6-311+G(2df,p):AMBER96//ONIOM(B3LYP/6-31G(d,p):AMBER96) and ONIOM(**DFT$_i$**/ 6-311+G(2df,p):AMBER96//ONIOM(**DFT$_i$**/6-31G(d,p):AMBER96) levels of theory.

DFT_i	**Spin State**	**ΔE (kJ mol^{-1})**
BLYP	Doublet	7.1
	Quartet	5.9
	Sextet	30.9
B3LYP*	Doublet	-7.6
	Quartet	-7.8
	Sextet	-2.9
M06	Doublet	91.0
	Quartet	90.9
	Sextet	89.6
M06-L	Doublet	94.4
	Quartet	94.7
	Sextet	46.5

As can be seen, all methods except B3LYP* show a destabilization of the doublet, quartet and sextet states upon adopting the ONIOM(B3LYP/6-31G(d,p):AMBER96) optimized geometry. That is, they all disfavor a pentacoordinated Fe center and/or long O_{Distal}···C1/C2 distances. The smallest degree of destabilization is observed for BLYP with the doublet, quartet and sextet to calculated to now lie higher in energy by 7.1, 5.9 and 30.9 kJ mol^{-1} respectively. It is noted that both methods gave similarly long optimized O_{Distal}···C1/C2 distances, thus these energy differences are likely due in most part to the change in coordination environment of the Fe (cf. Table 4.3). In the case of M06 and M06-L, the destabilization observed upon adopting the B3LYP-obtained geometry is more significant. In the case of M06 destabilization amounts to 91.0, 90.9

and 89.6 kJ mol^{-1} for the doublet, quartet and sextet respectively. Similarly, for M06-L the destabilization is 94.4, 94.7, and 46.5 kJ mol^{-1} respectively.

As mentioned above, B3LYP* is the exception amongst the functionals considered herein. Specifically, the use of B3LYP-optimized geometries Cpd I/cAOS stabilizes the doublet, quartet and sextet states by 7.6, 7.8 and 2.9 kJ mol^{-1} respectively. In the case of the sextet the stabilization is minor as the optimized geometries obtained at the ONIOM(DFT_i/6-31G(d,p):AMBER96) level for each method are quite similar. Regardless, the results suggest that for B3LYP* as the basis set size is increased, a five- rather than the six-coordinate Fe-center is preferred. By extension it also suggests that when modeling Cpd I/cAOS the basis set requirements for B3LYP* may differ from those of B3LYP.

4.4 Conclusions

Through the use of multiple density functional methods (BLYP, B3LYP*, B3LYP, M06 and M06-L) employed within a ONIOM QM/MM approach, the structural and electronic properties of Cpd I in cAOS has been investigated. In addition, using B3LYP, the effect of high-layer basis set choice for optimizations has been assessed as well as the hydrogen-bonding environments effect on the coordination of the Fe center.

Use of the pure-DFT method BLYP results in prediction of a hexacoordinated Fe complex but the lowest energy doublet and quartet spin-states are not degenerate. Specifically, the quartet and the sextet are destabilized relative to the doublet by 3.5 and 121.6 kJ mol^{-1} respectively. Moreover, it predicts the delocalization of the radical between Tyr353 and the porphyrin moiety to occur for all three spin-states.

In contrast, the use of B3LYP results in formation of a pentacoordinated Fe complex, with a now significantly lengthened Fe$\cdots O_{Tyr353}$ (its proximal ligand) distance. However, the lowest energy doublet and quartet spin-states are predicted to be degenerate while the

sextet is destabilized by 41.6 kJ mol^{-1}. Furthermore, it predicts the electron hole to be localized on the side chain of Ty353 rather than the porphyrin as typically seen in P450s.

Reduction of the % Hartree-Fock contribution to the exact exchange by use of the B3LYP* method again results in the prediction of a hexacoordinated Fe complex for the doublet and quartet spin-states, but a pentacoordinated Fe for the sextet. The doublet and quartet are calculated to be degenerate with the sextet lying higher in energy by 54.7 kJ mol^{-1}. For the doublet and quartet states the electron radical is delocalized between Tyr353 and the porphyrin, but localized on Ty353 in the sextet.

Structurally and electronically, M06 behaves similarly to B3LYP*. Importantly, however, and unlike all other functionals assessed herein it instead overstabilizes the sextet by 13.3 kJ mol^{-1}.

In contrast, M06-L consistently predicts Cpd I/cAOS to be a hexacoordinated Fe complex. The quartet is slightly energetically favored with respect to the doublet by 0.7 kJ mol^{-1} while destabilizing the sextet by 22.7 kJ mol^{-1}. It predicts the electron hole to be delocalized across Tyr353 and the porphyrin ring.

This current study shows that the properties predicted for Cpd I within cAOS can differ significantly with the choice of functional. Furthermore, it suggests that of the functionals assessed, only B3LYP was able to (i) reproduce the electronic states consistent with predicted by experimental EPR studies where Cpd I/cAOS involves a localized tyrosyl radical species and (ii) simultaneously give spin-splitting energies that are consistent with other functionals considered herein.

4.5 References

(1) Hersleth, H.-P.; Ryde, U.; Rydberg, P.; Görbitz, C. H.; Andersson, K. K. *J. Inorg. Biochem.* **2006**, *100*, 460.

(2) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279.

(3) Rittle, J.; Green, M. T. *Science* **2010**, *330*, 933.

(4) Hohenberger, J.; Ray, K.; Meyer, K. *Nat. Commun.* **2012**, *3*, 720.

(5) Shaik, S.; de Visser, S. P.; Kumar, D. *J. Biol. Inorg. Chem.* **2004**, *9*, 661.

(6) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2009**, *110*, 949.

(7) Ogliaro, F.; Harris, N.; Cohen, S.; Filatov, M.; de Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.* **2000**, *122*, 8977.

(8) Egawa, T.; Shimada, H.; Ishimura, Y. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 1464.

(9) Schünemann, V.; Jung, C.; Trautwein, A. X.; Mandon, D.; Weiss, R. *FEBS Lett.* **2000**, *479*, 149.

(10) Groves, J. T.; Nemo, T. E.; Myers, R. S. *J. Am. Chem. Soc.* **1979**, *101*, 1032.

(11) Ivancich, A.; Jouve, H. M.; Gaillard, J. *J. Am. Chem. Soc.* **1996**, *118*, 12852.

(12) Ivancich, A.; Jouve, H. M.; Sartor, B.; Gaillard, J. *Biochemistry* **1997**, *36*, 9356.

(13) Benecky, M. J.; Frew, J. E.; Scowen, N.; Jones, P.; Hoffman, B. M. *Biochemistry* **1993**, *32*, 11929.

(14) Andreoletti, P.; Pernoud, A.; Sainz, G.; Gouet, P.; Jouve, H. M. *Acta Crystallogr. Sect. D* **2003**, *59*, 2163.

(15) Rydberg, P.; Sigfridsson, E.; Ryde, U. *J. Biol. Inorg. Chem.* **2004**, *9*, 203.

(16) Koljak, R.; Boutaud, O.; Shieh, B.-H.; Samel, N.; Brash, A. R. *Science* **1997**, *277*, 1994.

(17) Oldham, M. L.; Brash, A. R.; Newcomer, M. E. *Proc. Natl. Acad. Sci.* **2005**, *102*, 297.

(18) Tosha, T.; Uchida, T.; Brash, A. R.; Kitagawa, T. *J. Biol. Chem.* **2006**, *281*, 12610.

(19) Boeglin, W. E.; Brash, A. R. *J. Biol. Chem.* **2012**, *287*, 24139.

(20) Bushnell, E. A. C.; Gherib, R.; Gauld, J. W. *J. Phys. Chem. B* **2013**, *117*, 6701.

(21) Wu, F.; Katsir, L. J.; Seavy, M.; Gaffney, B. J. *Biochemistry* **2003**, *42*, 6871.

(22) Wang, R.; de Visser, S. P. *J. Inorg. Biochem.* **2007**, *101*, 1464.

(23) Ghosh, A.; Taylor, P. R. *Curr. Opin. Chem. Biol.* **2003**, *7*, 113.

(24) Himo, F.; Siegbahn, P. E. M. *Chem. Rev.* **2003**, *103*, 2421.

(25) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 4388.

(26) Harvey, J. N. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203.

(27) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48.

(28) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Chem. Rev.* **2011**, *112*, 289.

(29) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(30) Gilbert, N. C.; Niebuhr, M.; Tsuruta, H.; Bordelon, T.; Ridderbusch, O.; Dassey, A.; Brash, A. R.; Bartlett, S. G.; Newcomer, M. E. *Biochemistry* **2008**, *47*, 10665.

(31) Molecular Operating Environment; 2012.10 ed.; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2012.

(32) Morokuma, K. *Bull. Korean Chem. Soc.* **2003**, *24*, 797.

(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., J. A. ; 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. P., 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, Inc., CT: Wallingford 2009.

(34) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(35) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(36) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comp. Chem.* **2004**, *25*, 1157.

(37) Poulos, T. L. *J. Biol. Inorg. Chem.* **1996**, *1*, 356.

(38) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *J. Mol. Biol.* **1987**, *195*, 687.

(39) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism, and Biochemistry*; 3 ed.; Kluwer Academic/Plenum Publishers: New York, 2005.

(40) Andreoletti, P.; Pernoud, A.; Sainz, G.; Gouet, P.; Jouve, H. M. *Acta Crystallogr. Sect. D* **2004**, *60*, 617.

(41) Hughes, T. F.; Friesner, R. A. *J. Chem. Theory Comput.* **2010**, *7*, 19.

(42) Wodrich, M. D.; Corminboeuf, C. m.; Schleyer, P. v. R. *Org. Lett.* **2006**, *8*, 3631.

(43) Johnson, E. R.; DiLabio, G. A. *Chem. Phys. Lett.* **2006**, *419*, 333.

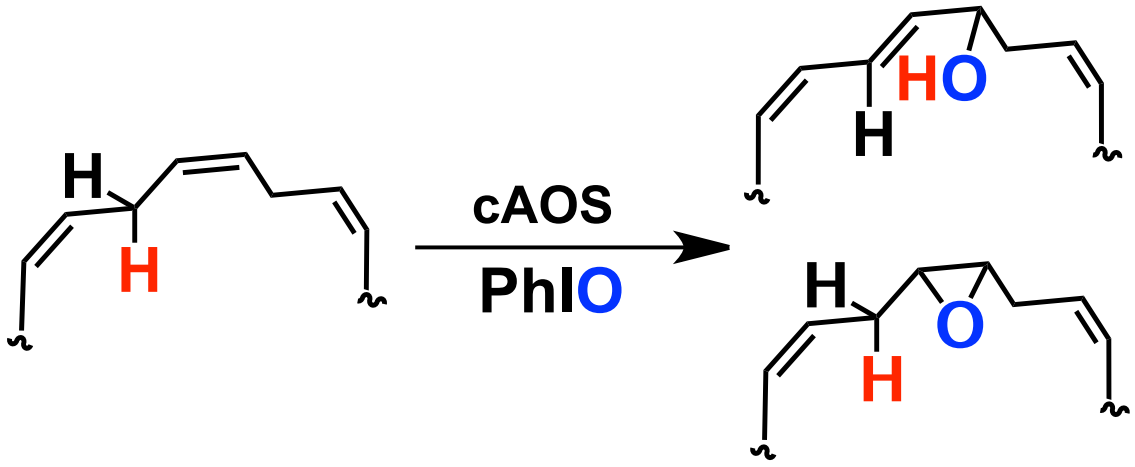
(44) Poli, R.; Harvey, J. N. *Chem. Soc. Rev.* **2003**, *32*, 1.

(45) Ogliaro, F.; de Visser, S. P.; Groves, J. T.; Shaik, S. *Angew. Chem.* **2001**, *113*, 2958.

(46) Schöneboom, J. C.; Neese, F.; Thiel, W. *J. Am. Chem. Soc.* **2005**, *127*, 5840.

Chapter 5

Compound I Reactivity in Coral Allene Oxide Synthase: A QM/MM Study



5.1 Introduction

Cytochrome P450s are ubiquitous in nature being found in a wide variety of biosystems including microbes, fungi, plants, animals and humans.[1] They metabolize 75% of all known parmaceuticals and are involved in hormone biosynthesis.[2,3] Their extensive involvement in drug metabolism has in part made them targets of biomedical and pharmaceutical studies.[4] In addition to their biological relevance, the chemistry involved in their biocatalysis constitutues a very active area of research.[5] Although they possess a wide variety of chemical reactivities, e.g., epoxidation and hydroxylation, arguably they are best known for their ability to insert a single oxygen into organic substrates through C–H activation (e.g., their monooxygenase abilities).[6]

The catalytic mechanism of P450s has been the subject of much discussion and debate.[7] Central to understanding their mechanism is the characterization of their generally accepted key intermediate Compound I (Cpd I).[8] The latter formally contains a porphyrin radical cation ($Por^{\bullet+}$) and a Fe(IV)=O center (Figure 5.1).[9] Notably, it has been suggested to be one of the most powerful oxidative agents of biochemical systems.[5] Unfortunately, experimentally Cpd I is an elusive species and does not accumulate in the mechanism.[10] However, computational investigations have produced key insights into its reactivity and electronic ground state.[5,8] For example, it has been established the doublet and quartet spin-states lie close in energy. Furthermore, mechanistic studies on P450 have suggested that Cpd I can utilize multistate reactivity (MSR); its nearly-degenerate doublet and quartet spin-states contributing to reactivity by exploiting differing reaction barriers and/or mechanisms.[11,12]

The overall mechanism of oxidation by Cpd I typically involves the abstraction of two electrons from the substrate to the $Por^{\bullet+}$(Fe(IV)=O) moiety.[5] More specifically, the conventional C–H hydroxylation mechanism involves an initial proton-coupled electron transfer (PCET) from the substrate to generate compound II (Cpd II). The latter contains a closed shell or radical cation porphyrin and a Fe(IV) or Fe(III) center respectively, ligated

to a distal hydroxyl moiety (Figure 5.1).[13] This is followed by a rebound mechanism thereby forming a ferryl alcohol (Fe(IV)-OH) complex. Typically, the initial PCET activating the C–H bond constitutes the rate-determining step with doublet and quartet transition states lying close in energy. It is noted that mechanistically the doublet and quartet potential energy surfaces differ for the rebound step: for the doublet it often occurs without a barrier while the quartet involves a barrier.[7]

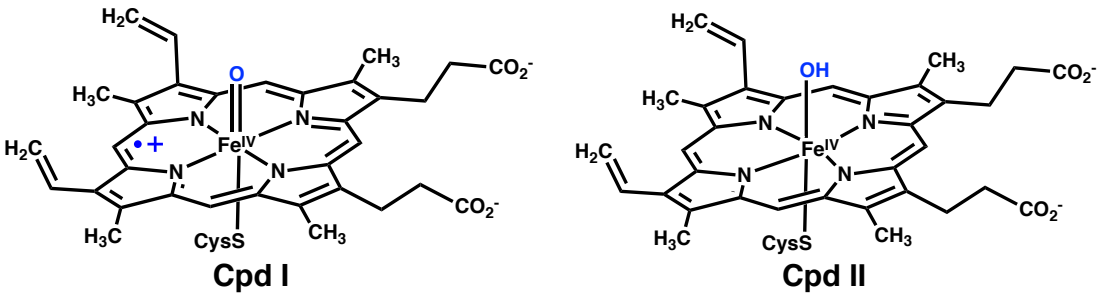


Figure 5.1. Schematic representation of Cpd I and Cpd II in P450.

Alkene hydroxylation and epoxidation are amongst the many chemical transformations P450s can catalyse,[14] and have been the subject of numerous experimental[3,15-18] and computational studies.[19-22] Like their typical alkane C–H activation adjoint, they are also thought to involve MSR. For example, it has been sugggested that Cpd I can react with an alkene substrate at a carbon center of its double bond, ultimately yielding a ferric-epoxide product.[20] The rate-determining step of alkene epoxidation, abstraction of an electron from the substrate onto the heme moiety, may occur via the doublet or quartet spin-states with approximately similar energy barriers. However, similar to the rebound step in C–H hydroxylation, a subsequent ring closure to give the epoxide is barrierless for the doublet spin-state but not the nearly degenerate quartet.

The catalytic role of the axial ligand in hemoproteins is elusive.[23,24] In the case of P450s, studies suggest that the thiolate axial ligand influences redox potentials, stabilizes Cpd I, mediates oxygen insertion in organic substrates and impacts chemoselectivity between epoxidation and hydroxylation.[25] A study by Ogliaro *et al*.[25] indicates that the

thiolate may help prevent the reduction of Cpd I precursors and faciliate protonation of Cpd 0 and O–O lysis, thereby promoting its conversion to Cpd I. It provides evidence for the notion that the thiolate may serve to gate the catalytic cycle. However, some Cpd I mimetic studies have suggested that the thiolate's presence may not be catalytically essential[26-28] while others have suggested significant axial ligand effects on reactivity.[29-31]

Coral allene oxide synthase (cAOS) is a heme-containing enzyme that catalyzes the production of unstable epoxides from fatty acid hydroperoxide.[32,33] Notably, it shares some resemblances with P450s and catalases as well as some differences. For instance, structurally it resembles catalases in that both have a tyrosinate axial ligand and both catalyze peroxide bond cleavage.[33] Furthermore, like both catalases and P450s, the biocatalysis of cAOS involves the formation of Cpd II.[34,35] However despite their structural and mechanistic similarities, cAOS does not have catalase activity or the ability to innately form Cpd I from either peroxide (as in catalases) or molecular oxygen (as in P450s). Indeed, unlike P450s and catalases the formation of Cpd II bypasses that of Cpd I and, in its wild-type endogenous state, cAOS cannot act as a monooxygenase.[36] The outlined properties of cAOS, catalases and P450s are consistent with the notion that the ability of hemoproteins to catalyze epoxidation and hydroxylation reactions requires cysteine as a proximal ligand.

In recent work, Boeglin *et al*.[36] investigated the formation of Cpd I in cAOS by using iodosylbenzene (PhIO) as an oxygen donor. Remarkably, they were able to demonstrate that in the presence of PhIO cAOS can act as a stereospecific monooxygenase, catalyzing the hydroxylation and epoxidation of arachidonic acid and forming 8R-HETE and 8R,9S-EET respectively. These results suggest that P450-like reactivity does not in fact require a thiolate proximal ligand. Furthermore, they invite one to consider P450-like reactivity within a system not inherently built for monooxygenation.

The present study examines the properties of Cpd I in cAOS (Cpd 1/cAOS) as well as its ability to perform hydroxylation and epoxidation reactions, via the complementary use

of docking, molecular dynamics (MD) and ONIOM(QM/MM) computational methods. In particular, the kinetic and thermodynamics properties of Cpd I/cAOS reactivity are examined and compared to results obtained from previous computational studies on Cpd I/P450. The results provide insights into not only the role of the axial ligand in P450 but also how the active site of cAOS adapts to form Cpd I and reacts towards an unsaturated fatty acid.

5.2 Computational Methods

The Molecular Operating Environment (MOE)[37] software package was used to perform all docking and molecular dynamics (MD) simulations, while all ONIOM(QM/MM) calculations were performed using the Gaussian 09[38] suite of programs. It is noted that we have previously successfully used the methodology described herein in a study of the catalytic mechanism of wild-type cAOS.[39]

5.2.1 Molecular Docking

A crystal structure of AOS 8R-lipoxygenase (PDB: 3DY5) from *P. homomalla* was used as a starting point. Counterions, crystallographic waters and the C-terminal lipoxygenase domain were removed while hydrogen atom coordinates were added using the MOE default method. The substrate, initially modeled as 8R-hydroperoxyeicosatetraneoic acid (8R-HPETE), was docked at proximity from the heme-cofactor in the active site. Binding free energies were approximated via the London dG scoring function in conjunction with the AMBER99 force field and the top one hundred scoring complexes were determined. These were subsequently optimized using the AMBER99 force field until the root-mean-square gradient of the potential energy fell below 0.05 kcal au^{-1}. The binding free energies were then recalculated with the London dG scoring function and the top scoring Michaelis complex was chosen for subsequent

QM/MM calculations. It should be noted that in all docked complexes the peroxy moiety was ligated to the Fe center.

5.2.2 QM/MM Model and Calculations

An active site model was extracted from the top scoring Michaelis complex. The peroxy moiety was deleted and the substrate was manually mutated from 8R-HPETE to arachidonic acid. An oxygen atom was inserted to ligate to the Fe center thereby generating an iron oxo moiety. The QM/MM active site model (see Figure 5.2) contained in total 1168 atoms and the calculations were performed using the ONIOM scheme.[40] The QM region (93 atoms) contained the reactive part of the substrate, the iron center, the porphin moiety (side-chains were placed in the MM layer) and the side-chains of Thr66, His67, Arg349 and Tyr353. Thr66 and His67 have been shown experimentally and computational to be involved in the biocatalysis of cAOS.[33,39] Geometry optimizations and frequency calculations (to obtain the Gibbs corrections and validate the nature of stationary point) were performed using the hybrid B3LYP functional[41] combined with LANL2DZ on iron and 6-31G(d,p) basis set on all other atom (hereby designated as BS1). Single point calculations were obtained at the ONIOM (B3LYP/BS2:AMBER96//B3LYP/BS1:AMBER96)-ME level of theory (where BS2 corresponds to 6-311+G(2df,p) basis set). The C_α located beyond the extremities of the QM layer were fixed as to maintain the integrity of the system during optimization.

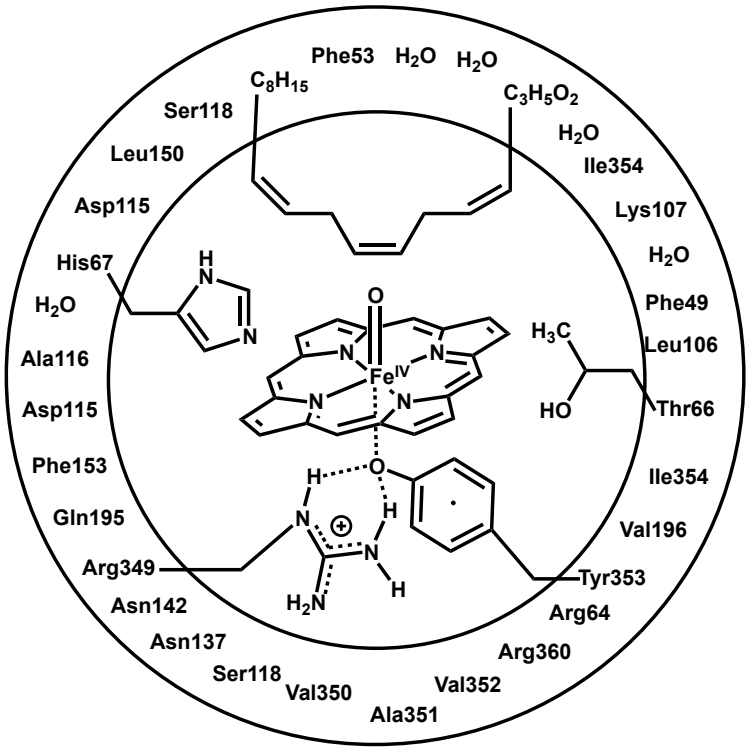


Figure 5.2. Schematic representation of the QM/MM active site model for Cpd I/cAOS complex. The components in the inner circle occupy the QM layer while those in the outer circle occupy the MM layer.

It should be noted that this study was performed in conjunction with a DFT/MM functional assessment, which described thoroughly the structural and electronic properties of Cpd I/cAOS predicted by BLYP, B3LYP*, B3LYP, M06 and M06-L. However, it should be noted that the B3LYP functional has been successfully used to model many heme-catalyzed reactions.[5,8] Thus, for the present investigation B3LYP has been used to model the hydroxylation and epoxidation of arachidonic acid by Cpd I/cAOS.

5.3 Results and Discussion

5.3.1 Cpd I/cAOS Hydroxylation of Arachidonic Acid

At the B3LYP/BS1 level of theory, Cpd I/cAOS adopts a pentacoordinated geometry centering on the iron (see Figure 5.4). Specifically, the tyrosinate ligand does not coordinate to the iron center and positions itself directly onto the guanidino group of

Arg349. The oxo-ligand coordinates to iron center with a bond distance of 1.60 Å for both ^{2}RC and ^{4}RC and 1.62 Å for ^{6}RC. The substrate is not adequately positioned in the active site to be directly deprotonated by the oxo-ligand. The corresponding hydrogen (shown red in Figure 5.4) is not oriented towards the oxo-ligand and lies far from it at distances of 3.43 for ^{2}RC and ^{4}RC and 3.50 Å for ^{6}RC. Instead, the hydrogen points to the imidazole group of His67 and form relatively long interactions (2.49 Å for all spin states). It should be noted that a previous study in our group on the endogenous catalysis of cAOS, affirms the role of His67 as a catalytic base where it deprotonates the hydrogen peroxide group on 8(R)-HPETE.[39]

The free energy surface of the reaction is provided in Figure 5.3. It shows that ^{2}RC and ^{4}RC are essentially degenerate with most of the energy difference arising from the Gibbs corrections. ^{6}RC lies 31.0 kJ mol^{-1} higher in energy than ^{2}RC, indicating its exclusion from the initial oxidative steps. As seen in Table 5.1, the spin densities reveal that the Fe=O moiety in ^{2}RC and ^{4}RC contains two unpaired electrons. These spin-densities are consistent with the generally accepted nature of Cpd I/cAOS, which contains a low-spin Fe^{IV} center. Typically Cpd I/P450 contains a porphyrin radical cation. However, Cpd I/cAOS differs in that the porphyrin is closed-shell and the electron radical is localized on the tyrosinate for all spin states investigated. As seen in Table 5.1, ^{2}RC and ^{4}RC differ in that the electron radical on the tyrosinate is anti-parallel with respect to those localized on the Fe=O moiety. The spin-densities of the Fe=O, porphyrin and tyrosinate moiety in ^{6}RC are 3.88, 0.14 and 0.98 respectively, indicating a high-spin Fe^{IV} with an electron radical localized on the tyrosinate.

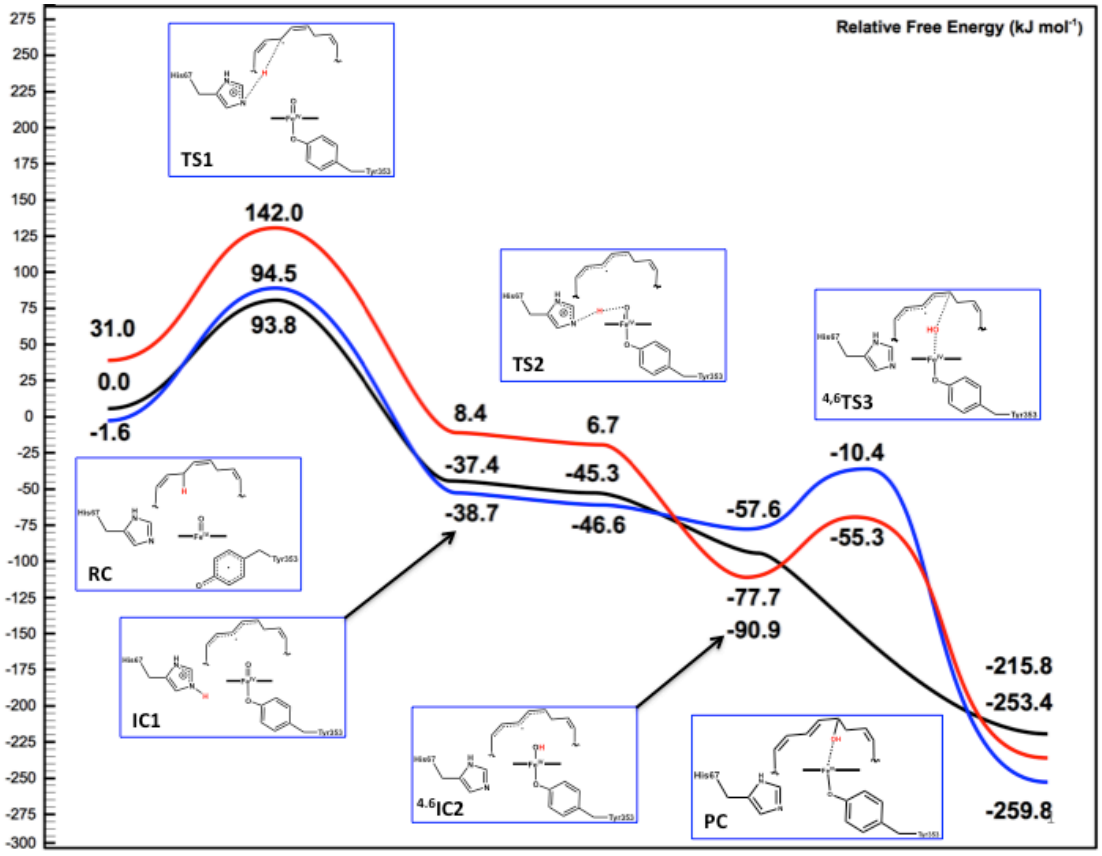


Figure 5.3. Schematic free energy surface of the hydroxylation of arachidonic acid by Cpd I/cAOS. Color code: doublet (black), quartet (blue) and sextet (red) spin-states.

Due to the large distance separating the oxo-ligand and the substrate, we considered the notion that His67 may act as a proton shuttle between the two moieties. The initial step, in the hydroxylation of arachidonic acid by Cpd I/cAOS is a PCET where a H^+ is transferred from the substrate to His67 while an electron is transferred from the substrate to the tyrosinate radical. The geometries of the corresponding transition states indicate an elongation of the $C_{10}\cdots H$ distance from 1.10 to 1.35 Å and a shortening of the $N_{His67}\cdots H$ interaction from 2.49 to 1.44 Å approximately for all spin states (see Figure 5.4). Upon being deprotonated, the substrate reduces the tyrosinate thereby generating a closed shell tyrosinate, as can be seen from corresponding spin densities (-0.01, 0.10 and 0.15 for $^2TS1_H$, $^4TS1_H$ and $^6TS1_H$ respectively), and an open shell substrate with a delocalized electron radical, (consistent with spin densities of -0.68, 0.70 and 0.70 for $^2TS1_H$, $^4TS1_H$

and $^{6}TS1_{H}$ respectively). It should also be noted that in the case of $^{2}TS1_{H}$, that electron spin on the substrate radical is antiparallel to that iron-oxo species. Upon reduction, the tyrosinate coordinates to the iron center by shortening the $Fe\cdots O_{Tyr353}$ by 1.02, 1.00 and 1.43 Å for doublet, quartet and sextet respectively (see Figure 5.4).

This initial step involves multistate reactivity and includes doublet and quartet species. The energy barriers for doublet and quartet pathways are 93.8 and 94.5 kJ mol^{-1} respectively. The barrier for the sextet is high (142 kJ mol^{-1}) indicating that its involvement in this initial step is unlikely. When comparing these results to a QM/MM study done by Cohen *et al*.[42] on the oxidation of cyclohexene by Cpd I/P450, several notable differences emerge. Specifically, the latter calculated the average energy barrier of the hydroxylation of cyclohexane for doublet and quartet spin states to be 66.6 and 70.3 kJ mol^{-1} respectively. This corresponds to a significant decrease of 27.2 and 24.2 kJ mol^{-1} the energy barriers for doublet and quartet spin states with respect to Cpd I/cAOS hydroxylation. Considering that every increment of 5.8 kJ mol^{-1} corresponds to ten-fold reduction in reaction rate, the results herein imply that Cpd I/cAOS performs hydroxylation roughly 27 000 times slower than Cpd I/P450.

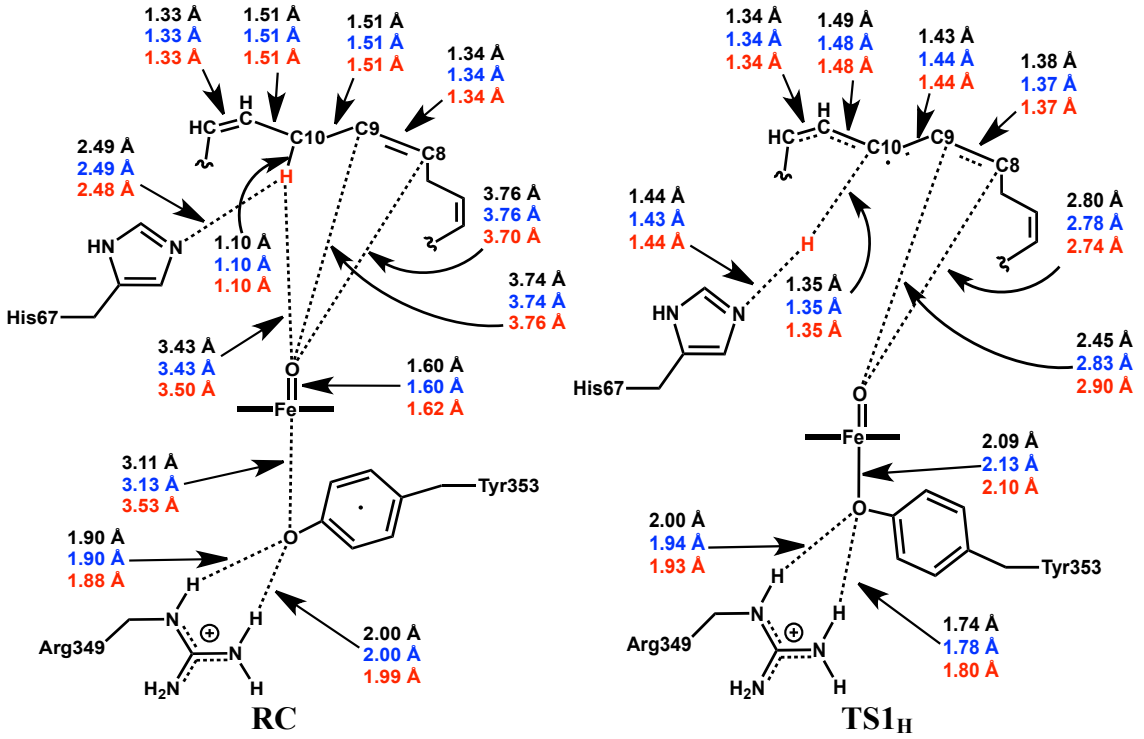


Figure 5.4. Schematic representations of optimized geometries of **RC** (i.e. Cpd I/cAOS) and **TS1$_H$** with selected bond lengths (Å). Color code: doublet (black), quartet (blue) and sextet (red) spin-states

With complete deprotonation of the substrate His67 shifts its orientation in the active site now forming an H-bond with the oxygen of Fe=O. As seen in Figure 5.5, His67 in IC1$_H$ forms a strong H-bond with the oxo-ligand with a N_{His67}–H···O_{Fe} distance of 1.66 Å for ^{2}IC1$_H$ and 1.67 Å for ^{4}IC1$_H$ and ^{6}IC1$_H$. As a results, the protonated His67 is adequately positioned to transfer its H^+ to the Fe=O moiety and fulfill its role as proton shuttle. The electron radical is delocalized over the substrate as seen in Figure 5.5, with the presence of partial double bonds on the substrate and as seen in Table 5.1 with the spin densities on the substrate being -0.98, 1.03 and 0.98 for ^{2}IC1$_H$, ^{4}IC1$_H$ and ^{6}IC1$_H$ respectively. The relative free energies of ^{2}IC1$_H$, ^{4}IC1$_H$ and ^{6}IC1$_H$ are -37.4, -38.7 and 8.4 kJ mol^{-1} respectively. Similarly to the RCs, the doublet and quartet are essentially degenerate while the sextet lies significantly higher in energy.

The formation of $IC1_H$ is quickly followed by the H^+ transfer from His67 to the oxo-ligand. During this step the N_{His67}–H···O_{Fe} distance decreases to 1.33 Å for $^2TS2_H$ and $^4TS2_H$ and to 1.43 Å for $^6TS2_H$. On the other hand, The N_{His67}···H distance increases to 1.19 Å for $^2TS2_H$ and $^4TS2_H$ and to 1.12 Å for $^6TS2_H$. These structural changes occur in conjunction with the lengthening of the Fe–O bond from 1.55 to 1.67 Å approximately. It is noted that at the B3LYP/BS1 level of theory, barriers for all spin states were obtained. However, upon increasing the size of the basis set, in the course of calculating single point energies, the energy barrier separating $IC1_H$ and $IC2_H$ vanished with $^2TS2_H$, $^4TS2_H$ and $^6TS2_H$ being 7.9, 7.9 and 1.7 kJ mol^{-1} respectively lower in energy than the corresponding $IC1_H$. This ensures not only that the H^+ transfer from His67 to the oxo-ligand is a barrierless process but also that His67 acts as a very fast proton shuttle between the substrate and the oxo-ligand.

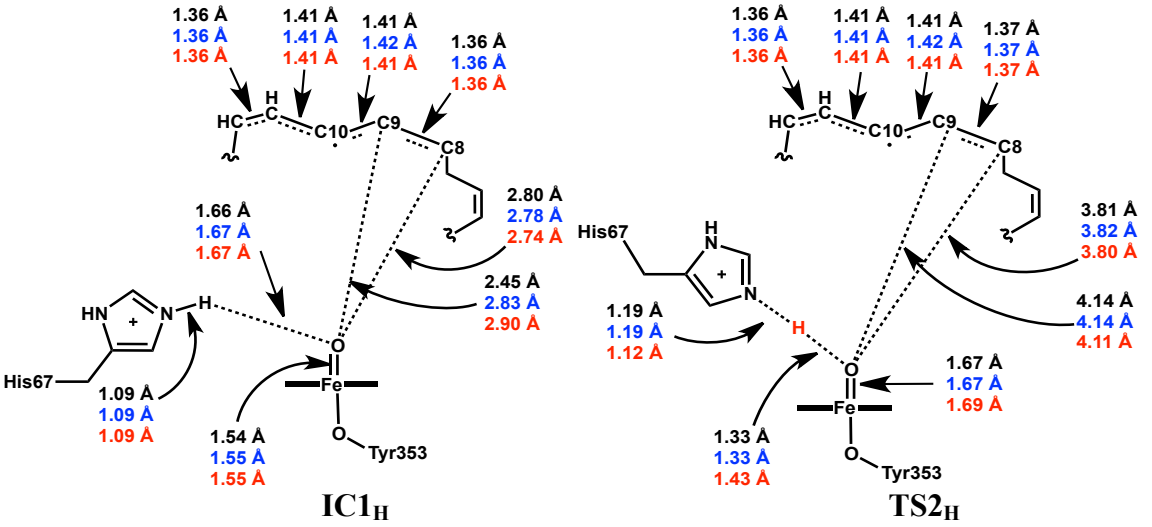


Figure 5.5. Schematic representations of optimized geometries of **$IC1_H$** and **$TS2_H$** with selected bond lengths (Å). Color code: doublet (black), quartet (blue) and sextet (red) spin-states.

Cpd II/cAOS (i.e. $IC2_H$) is formed following the reduction of tyrosinate by the substrate and the His67 mediated H^+ transfer from the substrate to the oxo-ligand. The relative free

energies of $^{2}IC2_{H,}$ $^{4}IC2_{H}$ and $^{6}IC2_{H}$ are -77.7, -57.6 and -90.0 kJ mol^{-1} respectively (Figure 5.3). These values indicate the potential involvement of the sextet in the last mechanistic step. The hydroxyl-ligand is located closer to $C8_{Substrate}$ than to $C9_{Substrate}$ by 0.32, 0.31 and 0.29 Å for $^{2}IC2_{H,}$ $^{4}IC2_{H}$ and $^{6}IC2_{H}$ respectively which aids in the formation of an alcohol at the $C8_{Substrate}$ site. This demonstrates, at the molecular level, how the active of cAOS adapts to the formation of Cpd II and induces hydroxylation, the latter not being part of its intrinsic nature.

The step that follows is the rebound step, whose controversial nature has been studied thoroughly.[7,43,44] As is the case, in the commonly accepted mechanism of P450s, the low and high-spin surfaces of Cpd II/cAOS react differently. Specifically, the doublet reacts barrierlessly, while the quartet and sextet involve energy barriers of 47.2 and 35.6 kJ mol^{-1} with respect to $^{4}IC2_{H}$ and $^{6}IC2_{H}$ respectively. To further verify the absence of an energy barrier, the PES of the doublet was scanned, by gradually shortening the Fe^{IV}—OH···$C8_{Substrate}$ distance from 3.83 to 1.46 Å. The scan reveals a negligible barrier of 0.7 kJ mol^{-1}, with respect to $^{2}IC2_{H}$ at the ONIOM(B3LYP/LANL2DZ(Fe), 6-31G(d,p)(C,H,N,O):AMBER99) level of theory, when the Fe^{IV}—OH···$C8_{Substrate}$ distance is 3.6 Å. Frequency calculations at this point were performed and did not reveal the presence of imaginary frequencies.

What is perhaps surprising is the low energy barrier of the sextet spin surface during this rebound process. The latter, is not routinely considered in the mechanism of P450s, due to its high instability in comparison to doublet and quartet species.[45] However, in the case of Cpd I/cAOS, the results herein suggest its involvement in the rebound step. With respect to $^{4}IC2_{H}$ and $^{6}IC2_{H}$, the free energies of $^{4}TS3_{H}$ and $^{6}TS_{H}$ are 47.2 and 22.4 kJ mol^{-1} respectively. The root of the enhanced stability of $^{6}TS_{H}$ compared to $^{4}TS_{H}$ can be attributed to its closer interaction between the Fe^{IV}—OH and the substrate. As seen in Figure 5.6, the Fe^{IV}—OH···$C8_{Substrate}$ distances for $^{4}TS_{H}$ and $^{6}TS_{H}$ are 2.81 and 2.55 Å and the Fe^{IV}—OH···$C9_{Substrate}$ distances for $^{4}TS_{H}$ and $^{6}TS_{H}$ are 3.42 and 3.19 Å. This

implies that the substrate is on average 0.25 Å closer to the Fe^{IV}—OH moiety in $^{6}TS_{H}$ compared to $^{4}TS_{H}$.

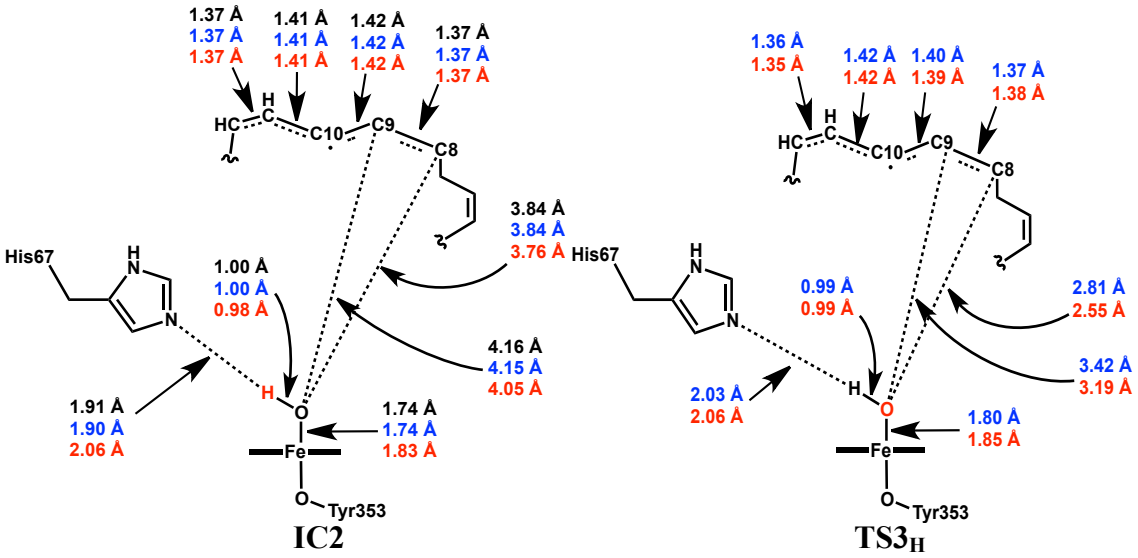


Figure 5.6. Schematic representations of optimized geometries of $IC2_{H}$ (Cpd II/cAOS) and $TS3_{H}$ with selected bond lengths (Å). Color code: doublet (black), quartet (blue) and sextet (red) spin-states.

The formation the PC_{H} results in an 8R-HETE bound to the iron center via the hydroxyl group. The $Fe\cdots O_{Susbtrate}$ distance increases along with the multiplicity of the system; 2.20, 2.66 and 2.86 Å for $^{2}PC_{H}$, $^{4}PC_{H}$ and $^{6}PC_{H}$ respectively. Upon the complete formation of 8R-HETE, the substrate becomes closed-shell and the unpaired electrons become localized on the iron center. The charge and spin states of the iron center can be inferred from Table 5.1, which suggest the occurrence of Fe^{III} with low, intermediate and high spin configurations for $^{2}PC_{H}$, $^{4}PC_{H}$ and $^{6}PC_{H}$ respectively. It is noted that the spin densities in the case of the sextet is subject to delocalization over the porphyrin, tyrosinate and iron moieties. The relative free energies $^{2}PC_{H}$, $^{4}PC_{H}$ and $^{6}PC_{H}$ are -215.8, -259.8 and -253.4 kJ mol^{-1} respectively indicating that the formation of the quartet is most favorable.

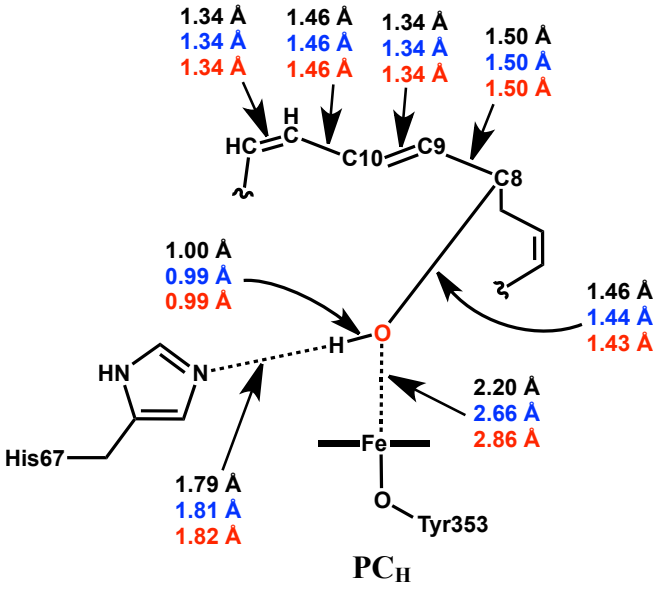


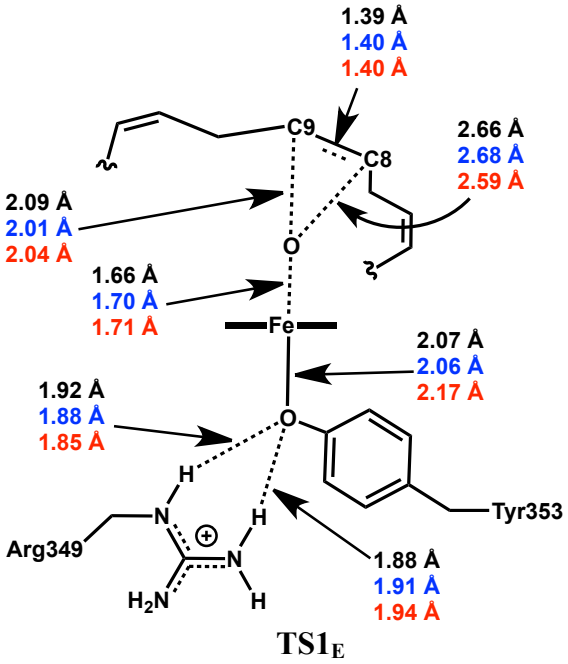
Figure 5.7. Schematic representations of optimized geometries of PC_H with selected bond lengths (Å). Color code: doublet (black), quartet (blue) and sextet (red) spin-states

Considering the relatively low energy of $^6IC2_H$, the barrierless collapse of $^2IC2_H$ and the preferential formation of ^4PC_H, the results herein support the notion that the rebound mechanism in Cpd I/cAOS occurs through double spin-crossover involving three different spin-states. IC2 initially manifests as a sextet ($^6IC2_H$) before undergoing a spin-crossover where it becomes temporarily a doublet ($^2TS3_H$) before undergoing a second spin-crossover where it collapses to a quartet product (^4PC_H).

5.3.2 Cpd I/cAOS Epoxidation of Arachidonic Acid

The first step of the epoxidation of arachidonic acid by Cpd I/cAOS is the attack of $C9_{Substrate}$ by the oxygen of Fe^{IV}=O. While the Fe=O···$C9_{Substrate}$ distance in ^2RC, ^4RC and ^6RC is 3.74, 3.74 and 3.76 Å respectively, in $^2TS1_E$, $^4TS1_E$ and $^6TS1_E$, it shortens significantly to 2.09, 2.01 and 2.04 Å. Concomitantly the Fe=O bond elongates slightly by 0.06, 0.10 and 0.09 Å in $^2TS1_E$, $^4TS1_E$ and $^6TS1_E$ respectively. Moreover, the C8=C9 double bond in the substrate also lengthens as seen by its increases of 0.05, 0.06 and 0.06

Å (Figure 5.8). Similar to $TS1_H$, the Fe···O-Tyr353 was found to shorten considerably with the metal center adopting a 6-coordinate geometry for all spin states.



$TS1_E$

Figure 5.8. Schematic representations of optimized geometries of $TS1_E$ with selected bond lengths (Å). Color code: doublet (black), quartet (blue) and sextet (red) spin-states

The energy barriers for doublet, quartet and sextet spin-states are 50.6, 67.3 and 117.4 kJ mol^{-1} respectively (Figure 5.9). These barriers differ significantly from the P450 scenario. In particular, in the QM/MM investigation by Cohen et al.[42] on the epoxidation of cyclohexene by P450 barriers of 96.2 and 97.1 kJ mol^{-1} were obtained for the doublet and quartet spin-states, respectively. Not only are the epoxidation barriers of Cpd I/cAOS lower by 45.6 and 29.8 kJ mol^{-1} but they also favor significantly the doublet species compared to Cpd I/P450. Again, with the present model and using the notion that increments of 5.8 kJ mol^{-1} increase the reaction rate by ten-fold, these results suggest that Cpd I/cAOS performs epoxidation roughly $7.3x10^8$ times faster than Cpd I/P450.

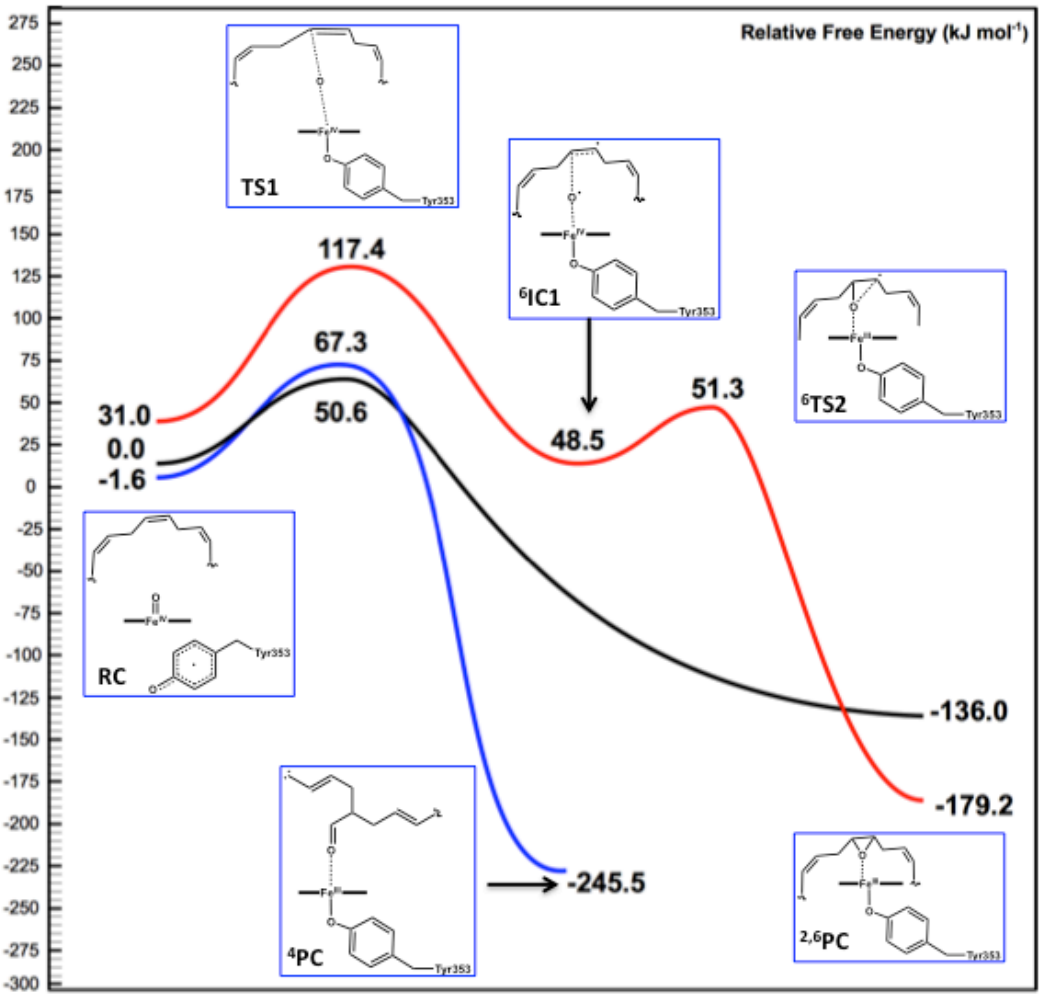


Figure 5.9. Schematic free energy surface of the epoxidation of arachidonic acid by Cpd I/cAOS. Color code: doublet (black), quartet (blue) and sextet (red) spin-states

During this initial step in the epoxidation, the substrate reduces the tyrosinate (as seen from the spin densities in Table 5.1 and 5.2 and while comparing the spin densities of the substrate for RC with those for $TS1_E$). For RC the substrate spin densities are 0.00 for all spin states. While for $^2TS1_E$, $^4TS1_E$ and $^6TS1_E$ spin densities of -0.50, 0.62 and 0.58 were respectively obtained thus indicating the partial transfer of an electron from the substrate.

The epoxidation at the doublet spin-surface occurs via a single step mechanism. Every attempt in optimizing $^2IC1_E$ resulted in its collapse to ^2PC_E thereby indicating the absence

of a transition state between $^{2}IC1_{E}$ and $^{2}PC_{E}$ at the B3LYP/BS1 level of theory. This notion is in agreement with previous computational Cpd I/P450 studies.[5]

Attempts to optimize $^{4}IC1_{E}$ resulted in the barrierless formation of a stable aldehyde byproduct (see Figure 5.10) coordinated to an intermediate spin ferric center (see Table 5.2). The formation of this complex (hereby denoted as $^{4}PC_{E}$) was very thermodynamically favorable with a relative free energy of -245.5 kJ mol^{-1} with respect to ^{2}RC. Although such a byproduct was not reported in the corresponding experimental study by Boeglin et al.,[36] experimental[46-49] and theoretical studies[21,22] have shown that double bond activation in P450s and other iron porphyrin-based catalysts coupled with iodosylbenzene can often be obstructed by the formation of aldehydes. Theoretical investigations[22] have suggested that these byproducts originate from quartet spin state surfaces and coincide with the results presented herein.

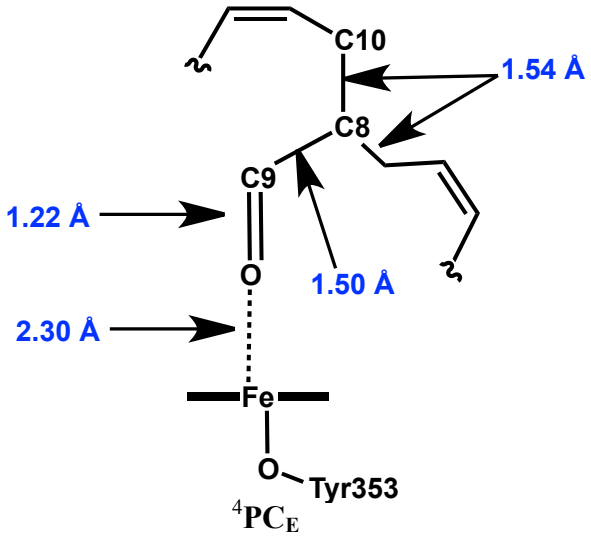


Figure 5.10. Schematic representations of optimized geometries of $^{4}PC_{E}$ with selected bond lengths (Å).

The optimization of $^{6}IC1_{E}$ resulted in the formation a biradical substrate coordinated to an intermediate spin ferric center (see Table 5.2). The two lone-pair electrons are localized on C8 and $O_{Substrate}$ (as seen in Table 5.2, from their respective spin density of

1.11 and 0.83). Starting from $^{6}IC1_{E}$, the formation of the epoxide product (8R,9S-EET) along the sextet spin surface, involves a barrier of 2.8 kJ mol^{-1}. The spin densities of $^{6}TS2_{E}$ reveal that closure of the epoxide ring forms a closed shell substrate. $^{6}TS2_{H}$ contains a high spin Fe^{III} with an electron delocalized over the porphyrin moiety, Tyr353 side-chain and $O_{substrate}$. During the optimization of $^{6}TS2_{E}$, the side chain of Thr66 rotated and formed a hydrogen bond with $O_{Substrate}$. Specifically, the Thr66—OH···$O_{Substrate}$ distance shortens from 4.68 Å in $^{6}IC1_{E}$ to 2.28 Å in $^{6}TS2_{E}$. It should be noted that a previous theoretical study done in our group suggests the formation of a hydrogen bond between Thr66 and a substrate bound oxygen radical in the endogenous reaction mechanism of cAOS.[39] These results herein demonstrate yet again how the active site of cAOS adapts to P450-like reactivity.

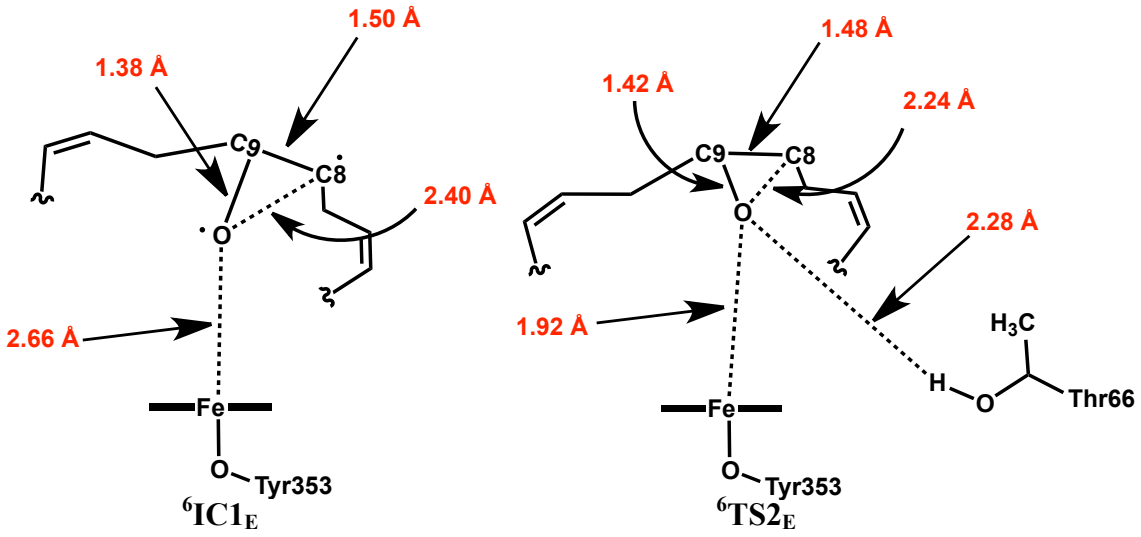


Figure 5.11. Schematic representations of optimized geometries of $^{6}IC1_{E}$ and $^{6}TS2_{E}$with selected bond lengths (Å).

The geometries of $^{2}PC_{E}$ and $^{6}PC_{E}$ have one prominent difference. Specifically, the Fe^{III}···$O_{Substrate}$ distance was found to be 2.33 and 3.63 Å, respectively. The relative free energies of $^{2}PC_{E}$, and $^{6}PC_{E}$ are -136.0 and -179.2 kJ mol^{-1} respectively and indicate that

the epoxide product complex will predominantly possess a high spin state Fe(III) and is not coordinated to 8R,9S-EET.

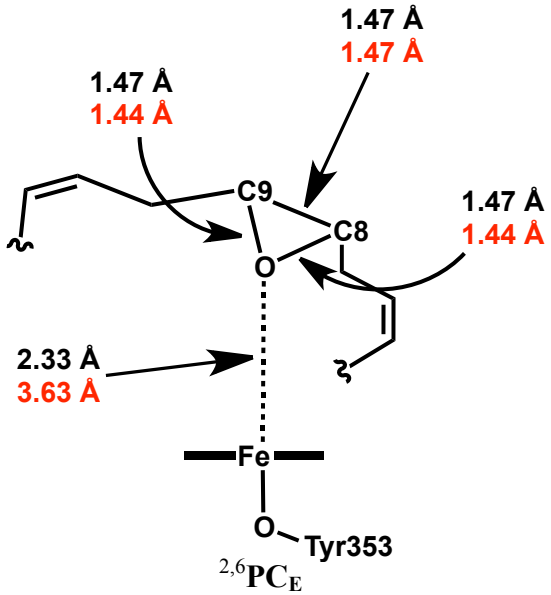


Figure 5.12. Schematic representations of optimized geometries of $^{2,6}PC_E$ with selected bond lengths (Å). Key distances for the doublet and sextet geometries are given in black and red, respectively.

Table 5.1: Selected spin densities obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96//B3LYP/LANL2DZ(Fe),6-31(d,p)(C,H,N,O):AMBER96) level of theory for the hydroxylation pathway.

Compound	Spin State	Fe	O_{Distal}	Tyr353	Porphyrin	Substrate
RC	Doublet	1.34	0.80	-0.96	-0.18	0.00
	Quartet	1.32	0.90	0.96	-0.09	0.00
	Sextet	3.39	0.49	0.98	0.14	0.00
$TS1_H$	Doublet	1.41	0.58	-0.01	-0.19	-0.68
	Quartet	1.38	0.80	0.10	0.13	0.70
	Sextet	3.26	0.60	0.05	0.24	0.70
$IC1_H$	Doublet	1.57	0.62	0.02	-0.22	-0.98
	Quartet	1.55	0.62	0.02	-0.19	1.03
	Sextet	3.36	0.49	-0.07	0.22	0.98
$TS2_H$	Doublet	1.66	0.54	0.02	-0.23	-0.98
	Quartet	1.63	0.54	0.02	-0.20	0.41
	Sextet	3.45	0.46	-0.13	0.21	0.98
$IC2_H$	Doublet	1.94	0.26	0.04	-0.25	-0.98
	Quartet	1.92	0.26	0.03	-0.22	0.98
	Sextet	4.15	0.42	0.84	0.43	-0.85
$TS3_H$	Quartet	2.23	0.15	-0.05	-0.23	0.92
	Sextet	2.64	0.43	0.75	0.15	1.03
PC_H	Singlet	1.09	0.00	0.00	-0.09	0.01
	Quartet	2.93	0.01	0.23	-0.19	0.02
	Sextet	4.24	0.01	0.30	0.44	0.01

Table 5.2: Selected spin densities obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96//B3LYP/LANL2DZ(Fe),6-31(d,p)(C,H,N,O):AMBER96) level of theory for the epoxidation pathway.

Compound	Spin State	Fe	O_{Distal}	Tyr353	Porphyrin	Substrate
$TS1_E$	Doublet	1.62	0.14	-0.01	-0.20	-0.50
	Quartet	1.67	0.73	0.00	-0.15	0.62
	Sextet	3.59	0.54	-0.02	0.24	0.58
$IC1_E$	Sextet	2.92	0.82	0.24	-0.11	1.11
$TS2_E$	Sextet	4.24	0.16	0.18	0.35	0.04
PC_E	Doublet	1.16	0.00	-0.05	-0.11	0.00
	Quartet	3.02	0.02	0.13	-0.23	0.06
	Sextet	4.20	0.00	0.29	0.51	0.00

5.5 Conclusions

In this current study, the catalytic mechanism of cAOS monooxygenation has been investigated using a QM/MM methodology with the assumption that compound I (Cpd I) is the active oxidative species. We have compared the corresponding thermodynamic and kinetic properties with the oxidation by P450 of cyclohexene. Our results not only help examine the function of the axial ligand in Cpd I reactivity but also provide atomistic details into the way the active site of cAOS adapts to the emergence of Cpd I and mediates its reactivity. Key results and analyses are summarized hereafter:

- At the ONIOM(B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d,p):AMBER96)-ME level theory, Cpd I in the active site of cAOS is pentacoordinated with an axial Tyr353 side-chain radical.

- The initial step in the hydroxylation of arachidonic acid is rate limiting and involves a proton coupled electron transfer (PCET) mediated by His67. The energy barriers of this step are 93.8 and 94.5 kJ mol^{-1} for doublet and quartet species. Compared to P450, where the initial oxidative barrier associated to the oxidation of cyclohexane corresponds to 66.6 and 70.3 kJ mol^{-1} respectively, our results suggests that the tyrosyl axial ligand decreases significantly the rate of reaction of Cpd I hydroxylation.
- The initial step in the epoxidation of arachidonic acid is rate limiting and involves a direct association of the oxo-ligand to the substrate. The initial barriers are 50.6 and 67.3 kJ mol^{-1} for doublet and quartet species respectively. Compared to the P450 analogue whose energy barriers for doublet and quartet spin-states are 96.2 and 97.1 kJ mol^{-1} respectively, our current results suggest that the tyrosyl axial ligand increases dramatically the rate of reaction of Cpd I epoxidation.
- Similarly to P450, for both hydroxylation and epoxidation, doublet and quartet spin states use differing mechanistic pathways, where the former is concerted and the latter can either be step-wise or result the formation of a stable aldehyde byproduct.

5.5 References

(1) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. *Chem. Rev.* **1996**, *96*, 2841.

(2) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism, and Biochemistry*; 3 ed.; Kluwer Academic/Plenum Publishers: New York, 2005.

(3) Rittle, J.; Green, M. T. *Science* **2010**, *330*, 933.

(4) Guengerich, F. P. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1.

(5) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279.

(6) Ortiz de Montellano, P. R. *Chem. Rev.* **2009**, *110*, 932.

(7) Shaik, S.; Cohen, S.; de Visser, S. P.; Sharma, P. K.; Kumar, D.; Kozuch, S.; Ogliaro, F.; Danovich, D. *Eur. J. Inorg. Chem.* **2004**, *2004*, 207.

(8) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2009**, *110*, 949.

(9) Chu, K.; Vojtchovský, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921.

(10) Davydov, R.; Makris, T. M.; Kofman, V.; Werst, D. E.; Sligar, S. G.; Hoffman, B. M. *J. Am. Chem. Soc.* **2001**, *123*, 1403.

(11) Shaik, S.; de Visser, S. P.; Kumar, D. *J. Biol. Inorg. Chem.* **2004**, *9*, 661.

(12) Hirao, H.; Kumar, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2005**, *127*, 13007.

(13) Guengerich, F. P. *J. Biochem. Mol. Toxicol.* **2007**, *21*, 163.

(14) Woggon, W.-D. In *Bioorganic Chemistry: Models and Applications*; Springer: New York, 1997, p 39.

(15) Hessenauer-Ilicheva, N.; Franke, A.; Meyer, D.; Woggon, W.-D.; van Eldik, R. *J. Am. Chem. Soc.* **2007**, *129*, 12473.

(16) Farinas, E. T.; Alcalde, M.; Arnold, F. *Tetrahedron* **2004**, *60*, 525.

(17) Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. *Proc. Natl. Acad. Sci.* **1998**, *95*, 3555.

(18) Alcalde, M.; Farinas, E. T.; Arnold, F. H. *J. Biomol. Screen.* **2004**, *9*, 141.

(19) de Visser, S. P.; Ogliaro, F.; Harris, N.; Shaik, S. *J. Am. Chem. Soc.* **2001**, *123*, 3037.

(20) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 11809.

(21) Kumar, D.; de Visser, S. P.; Shaik, S. *Chem. Eur. J.* **2005**, *11*, 2825.

(22) de Visser, S. P.; Kumar, D.; Shaik, S. *J. Inorg. Biochem.* **2004**, *98*, 1183.

(23) Kumar, D.; Sastry, G. N.; de Visser, S. P. *J. Phys. Chem. B* **2011**, *116*, 718.

(24) Nam, W.; Ryu, Y. O.; Song, W. J. *J. Biol. Inorg. Chem.* **2004**, *9*, 654.

(25) Ogliaro, F.; de Visser, S. P.; Shaik, S. *J. Inorg. Biochem.* **2002**, *91*, 554.

(26) Groves, J. T.; Quinn, R. *J. Am. Chem. Soc.* **1985**, *107*, 5790.

(27) Dawson, J. H.; Sono, M. *Chem. Rev.* **1987**, *87*, 1255.

(28) Groves, J. T.; Watanabe, Y. *J. Am. Chem. Soc.* **1986**, *108*, 507.

(29) Kumar, D.; Sastry, G. N.; de Visser, S. P. *J. Phys. Chem. B* **2011**, *116*, 718.

(30) de Visser, S. P.; Tahsini, L.; Nam, W. *Chem. Eur. J.* **2009**, *15*, 5577.

(31) Wang, R.; de Visser, S. P. *J. Inorg. Biochem.* **2007**, *101*, 1464.

(32) Tijet, N.; Brash, A. R. *Prostaglandins Other Lipid Mediat.* **2002**, *68*, 423.

(33) Oldham, M. L.; Brash, A. R.; Newcomer, M. E. *Proc. Natl. Acad. Sci.* **2005**, *102*, 297.

(34) Brash, A. R. *Phytochemistry* **2009**, *70*, 1522.

(35) Cho, K.-B.; Lai, W.; Hamberg, M.; Raman, C. S.; Shaik, S. *Arch. Biochem. Biophys.* **2011**, *507*, 14.

(36) Boeglin, W. E.; Brash, A. R. *J. Biol. Chem.* **2012**, *287*, 24139.

(37) Molecular Operating Environment: 2012.10 ed.; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2012.

(38) 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., J. A. ; 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. P., 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, Inc., CT,: Wallingford 2009.

(39) Bushnell, E. A. C.; Gherib, R.; Gauld, J. W. *J. Phys. Chem. B* **2013**, *117*, 6701.

(40) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.

(41) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(42) Cohen, S.; Kozuch, S.; Hazan, C.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 11028.

(43) Filatov, M.; Harris, N.; Shaik, S. *Angew. Chem. Int. Ed.* **1999**, *38*, 3510.

(44) Ogliaro, F.; Harris, N.; Cohen, S.; Filatov, M.; de Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.* **2000**, *122*, 8977.

(45) Yakushi, K.; Amao, Y.; Okura, I.; Miyashita, T.; Yoshizawa, K.; Shiota, Y.; Kagawa, Y.; Yao, H.; Momozawa, O.; Hamatani, T. *Bull. Chem. Soc. Jpn.* **2000**, *73*, 2643.

(46) Groves, J. T.; Myers, R. S. *J. Am. Chem. Soc.* **1983**, *105*, 5791.

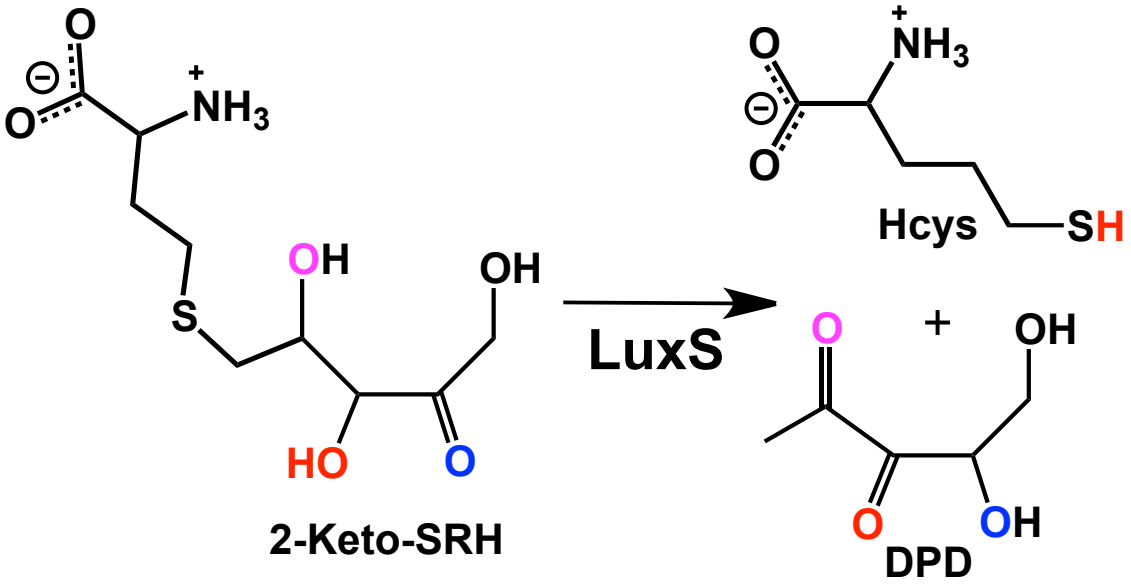
(47) Guengerich, F. P.; Macdonald, T. L. *Acc. Chem. Res.* **1984**, *17*, 9.

(48) Mansuy, D.; Leclaire, J.; Fontecave, M.; Dansette, P. *Tetrahedron* **1984**, *40*, 2847.

(49) Collman, J. P.; Kodadek, T.; Brauman, J. I. *J. Am. Chem. Soc.* **1986**, *108*, 2588.

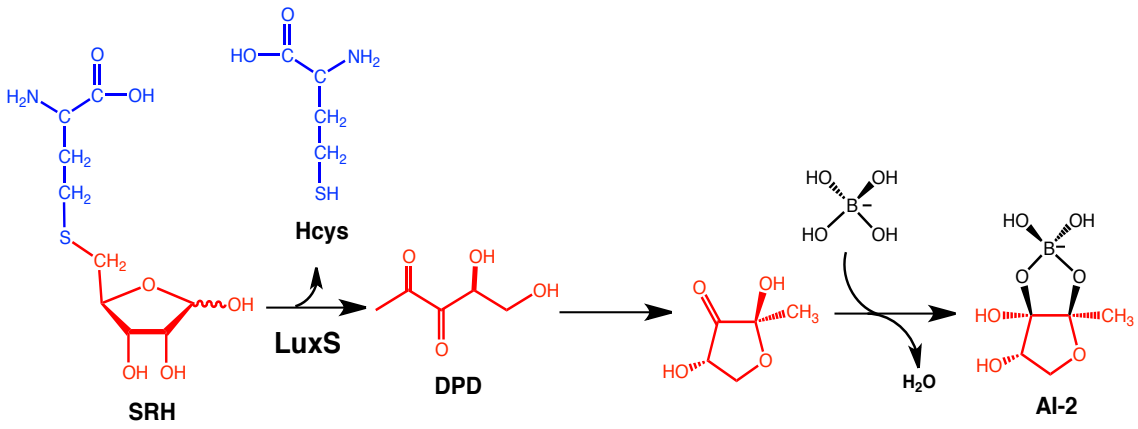
Chapter 6

Insights from MD and QM/MM into the Biocatalysis of LuxS



6.1 Introduction

From a chemical, biological and pharmacological perspective, S-ribosylhomocysteinase (LuxS) is an intriguing enzyme. From the chemist's viewpoint, this biocatalyst uses a Fe(II) cofactor[1] to mediate an internal redox reaction and catalyzes the cleavage of a stable thioether without redox-active cofactors through the conversion of S-ribosylhomocysteine (SRH) to 4,5-dihydroxy-2,3-pentanedione (DPD) and homocysteine (Hcys). From the biologist's standpoint, LuxS participates in the production of AI-2, a crucial signaling molecule in type-2 quorum sensing (QS-2).[2] AI-2 is formed from the spontaneous cyclization of DPD[3] and its complexation with borate[4]. Because of its involvement in interspecies communication, AI-2 has been proposed to act as a universal signaling molecule used by a wide variety of bacterial species including human pathogens[5]. From the pharmacologist's outlook, QS-2 and (by extension) LuxS mediate a variety of physiological processes that are involved in the spread of microbial infections such as biofilm formation,[6],motility,[7] sporulation[8] and production of toxins.[9] It is believed that AI-2/LuxS inhibitors are capable of serving as novel broad-spectrum antibiotics and playing a significant role in the emergence of antibiotic-resistant bacteria.[10]



Scheme 6.1. DPD as a biosynthetic precursor of AI-2

The structure of LuxS has been studied thoroughly and several X-ray crystal structures have been obtained from varying microbial species.[11-13] LuxS is a homodimer with active sites at the interface of its monomers. Its active sites contain a divalent metal cofactor, which has been subsequently identified as a Fe(II).[1] X-ray crystal structures have shown that, in the absence of the substrate, Fe(II) has a tetrahedral coordination to His54, His58, Cys126 and a water molecule. An X-ray crystal structure[14] of a C84A mutated, cobalt substituted LuxS (Co-LuxS) with a presumptive 2-keto mechanistic intermediate has indicated that the metal coordinates to the substrate through its O2 carbonyl oxygen. Furthermore, while the O3H hydroxyl oxygen was at proximity from the cobalt, it was not fully oriented towards it. Thus, it has been suggested[14] that during the formation of the Michaelis complex, the substrate replaces the metal-bound water while retaining Fe(II) centered tetrahedral geometry.

The functions of the active residues have been subject to many investigations. Specifically, genomic analyses have shown that Ser6, His11, Arg39, His54, Glu57, His58, Cys84, Tyr89 and Cys126 are conserved possibly to maintain the functionality of the enzyme.[1,13] Site-directed mutagenesis studies have indicated that Cys84 and Glu57 may be catalytically critical.[1] Cys84Ala and Glu57Ala mutations rendered the enzyme either inactive or with a negligible activity towards SRH. Spectroscopic analyses have determined the ionization state of the various active site components. Specifically, these studies have suggested that Cys84 and Glu57 in wild-type LuxS are found in their anionic forms while Arg39 is positively charged possibly stabilizing the thiolate.[15] Hence Cys84 and Glu57 have been presented as catalytic acid/bases.[1,2,10,15-17]

Based on many experimental findings, a biocatalytic mechanism has been proposed.[1,2,9,10,14,16,17] ^{13}C NMR studies have demonstrated the possible involvement of 2- and 3-ketone mechanistic intermediates.[17] By coupling mutagenesis and ^{13}C–NMR, it has been shown that Cys84 may be intimately involved in the formation of a 3-keto intermediate while Glu57 plays a critical role in the conversion of the 3-keto intermediate

to DPD and Hcys. Importantly, these studies suggest that the conversion of the 2-keto to the 3-keto intermediate is rate limiting.[17] The incorporation of deuterium into C1, C2 and C5 atoms in DPD (see Figure 6.1) when the catalysis of LuxS occurred in the presence of D_2O, indicated the involvement of proton transfers during catalysis.[18] Furthermore, by using deuterium labeled substrates, the stereochemical course of proton transfers was considered and suggested that the proton transfers occur rapidly but become partially rate limiting during the conversion of the 2-keto to 3-keto intermediate.[18] Importantly, this study correlates with preceding ^{13}C-NMR studies.

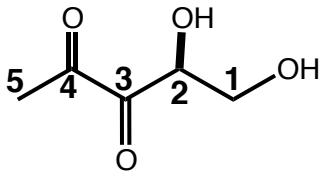


Figure 6.1. Carbon numbering in deuterium insertion in DPD

With regard to the metal, a spectroscopic study conducted on Co-LuxS in the presence of SRH showed that during catalysis the absorption spectra underwent dramatic and time-dependent changes.[1] This suggests that the first metal coordination sphere alters during catalysis perhaps giving the metal a catalytic role. Based on all of these experimental findings, an enzymatic mechanism has been proposed.[1-3,10,15-17]

The proposed mechanism can be divided in three stages. Initially, the ring opens and a carbonyl functional group is formed on C1. The corresponding double bond then moves from C1 to C2 thereby forming the 2-keto intermediate with Cys86 and Glu57 acting as acid/base catalytic residues. Subsequently, the carbonyl group moves from C2 to C3 thereby forming the 3-keto intermediate, again with the aid of Cys84 and Glu57 as acids/bases. Finally, the last step is a β-elimination, which involves mostly Glu57, results in the cleavage of the thioether bond.[10]

The first stage of the mechanism has been previously studied computationally by Huang *et al*.[19] Specifically, substrate binding processes were investigated through docking and MD simulations. The mechanism starting from SRH and leading to a 2-keto intermediate was subsequently studied using QM/MM. Docking and MD analyses showed that SRH can bind to the active site in its linear form and cyclic α- and β- forms. Furthermore, QM/MM calculations suggested the involvement of a water cofactor in the ring-opening process. This active site water, consistently positioned underneath the ribose ring, can shuttle a proton from $O2_{Substrate}$, ligated to Fe(II), to the ring oxygen. This process has been shown to occur concertedly with the deprotonation of $H_{C2-Substrate}$ by Cys84. Hence, QM/MM calculations suggest that the 2-keto intermediate results from a water-mediated ring opening mechanism. Overall, computational investigations show that LuxS can not only bind to all forms of SRH (cyclic and linear) but also catalyze subsequent reactions thereby forming a common mechanistic intermediate; the 2-keto SRH.

Notwithstanding all studies performed on LuxS, a substantial amount of its enzymatic mechanism remains elusive. Specifically, although the rate-determining step has been suggested to be the conversion of a putative 2-keto to a 3-keto intermediate, the precise mechanistic steps remain elusive. In addition, recent X-ray structures have indicated that perhaps Glu57 may be too distant to relay protons between different oxygen atoms in the substrate.[14,16] Certain studies[14,16] have therefore considered more complicated proton shuttling process involving His11 and Tyr89, which occupy different locations in the active site. The role of Glu57 and its involvement in the second and third stage also needs detailed atomistic consideration. In this present study, complementary computational methods have been used synergistically to study the latter two stages of the LuxS mechanism. QM/MM and MD simulations were used to answer previous questions and examine the details at the atomistic level of the conversion of a 2-keto intermediate into DPD resulting in the thioether bond cleavage.

6.2 Computational Method

6.2.1 Molecular Dynamic (MD) Simulations.

MD simulations were performed using the NAMD[20] set of algorithms and were set up using the Molecular Operating Environment (MOE) software package.[21] The models were generated from an X-ray crystal structure (PDB ID: 1YCL). Since the crystal structure did not correspond to the enzyme in its endogenous state, initial steps were undertaken to model it so. Specifically, Co(II) was modified to Fe(II) and the C84A mutation was manually reversed (i.e. a thiolate functional group was added to Ala84B). Substrates of interest were modeled using the presumed 2-keto intermediate (initially present in the crystal structure) as template.

Hydrogen atoms were inserted using the Protonate 3D application as installed in MOE. Subsequently, the entire model was solvated without boundary conditions. The two nitrogen and the sulfur atoms belonging to His58A, His54A and Cys126A respectively ligated to Fe(II) were fixed in their initial crystal structure positions to maintain the integrity of the first Fe(II) coordination sphere. Substrate atoms were not fixed. The final model was minimized using the AMBER12EHT force field. Each MD simulation underwent a 0.5 ns annealing run followed by 5.0 ns production run at 300 K using the NAMD set of algorithms in conjunction with AMBER12EHT force field.

6.2.2 QM/MM Calculations

All QM/MM calculations were done using the Gaussian 09 suite of programs[22] using the ONIOM[23] mechanical embedding (ME) formalism. The QM layer was modeled via density functional theory with the B3LYP[24] functional. Optimized structures and corresponding harmonic vibrational frequencies were obtained at the B3LYP/6-31G(d,p) level of theory for the QM layer and the AMBER96 force field[25] for the MM layer. Free energies were calculated using single point calculations performed at the

ONIOM(B3LYP/6-311+G(2df,p)//AMBER96)-ME level of theory in conjunction with Gibbs free energy corrections.

The QM/MM model (see Figure 6.2) was taken from the X-ray crystal structure of a C84A mutated Co-LuxS with an alleged SRH 2-keto intermediate (PDB: 1YCL)[14]. The QM layer included the substrate's 5-S-ethyl-5-thio-D-ribofuranose moiety, Fe(II), an active site water and the R-groups of Ser6B, Arg39B, Cys85B, Glu57A His54A, His 58B, Cys126A. Residues in the MM layer were included in their entirely as well as several H_2O molecules that interacted electrostatically with the QM layer. The integrity of the model was maintained by fixing a minimum number of atoms in their X-ray crystal positions.

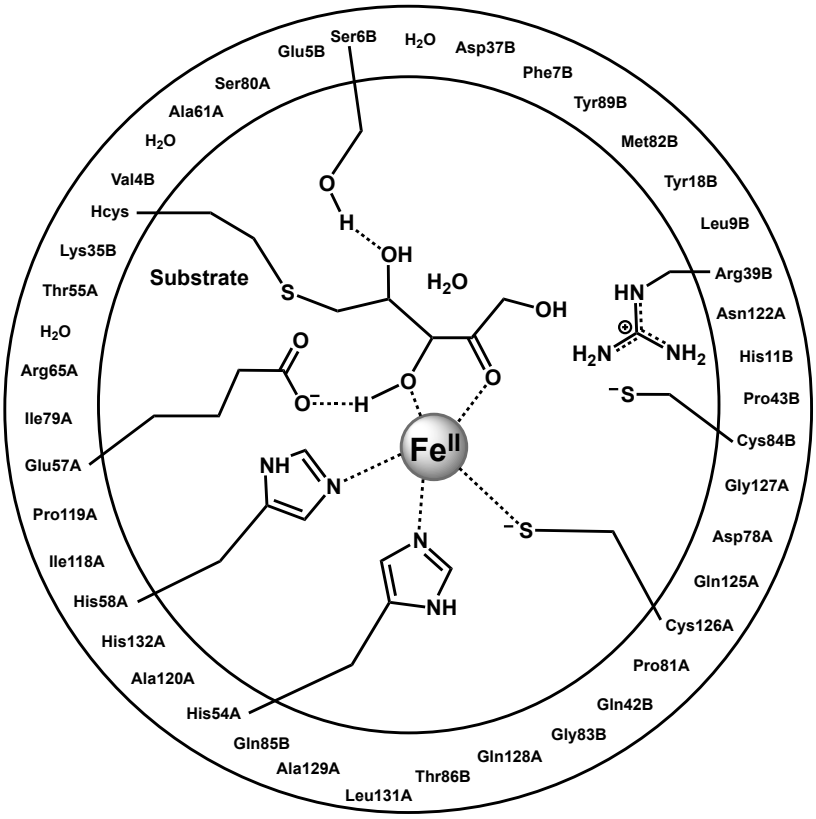


Figure 6.2. Schematic representation of the ONIOM QM/MM active site model with the putative 2-keto-SRH intermediate bound. The inner circle contains the moieties in the QM layer and the outer circle indicates those in the MM layer.

6.3 Results and Discussion

6.3.1 Deprotonation of $C3_{Substrate}$ by Cys84.

We began by modeling the 2-keto intermediate via QM/MM. The optimized structure (see Figure 6.3) contains an $O^{-}{}_{Glu57}\cdots H_{O3-Substrate}$ distance of 1.52 Å, indicating the presence of a strong hydrogen bond between the $H_{O3-Substrate}$ on the substrate and Glu57A. Furthermore the thiolate on Cys84B forms a moderate interaction with both the substrate and Arg39B where the $H_{C2-RC}\cdots{}^{-}S_{Cys84B}$ distance is 2.55 Å and the $H_{N-Arg39}\cdots S_{Cys84B}$ distance is 2.22 Å (not shown).

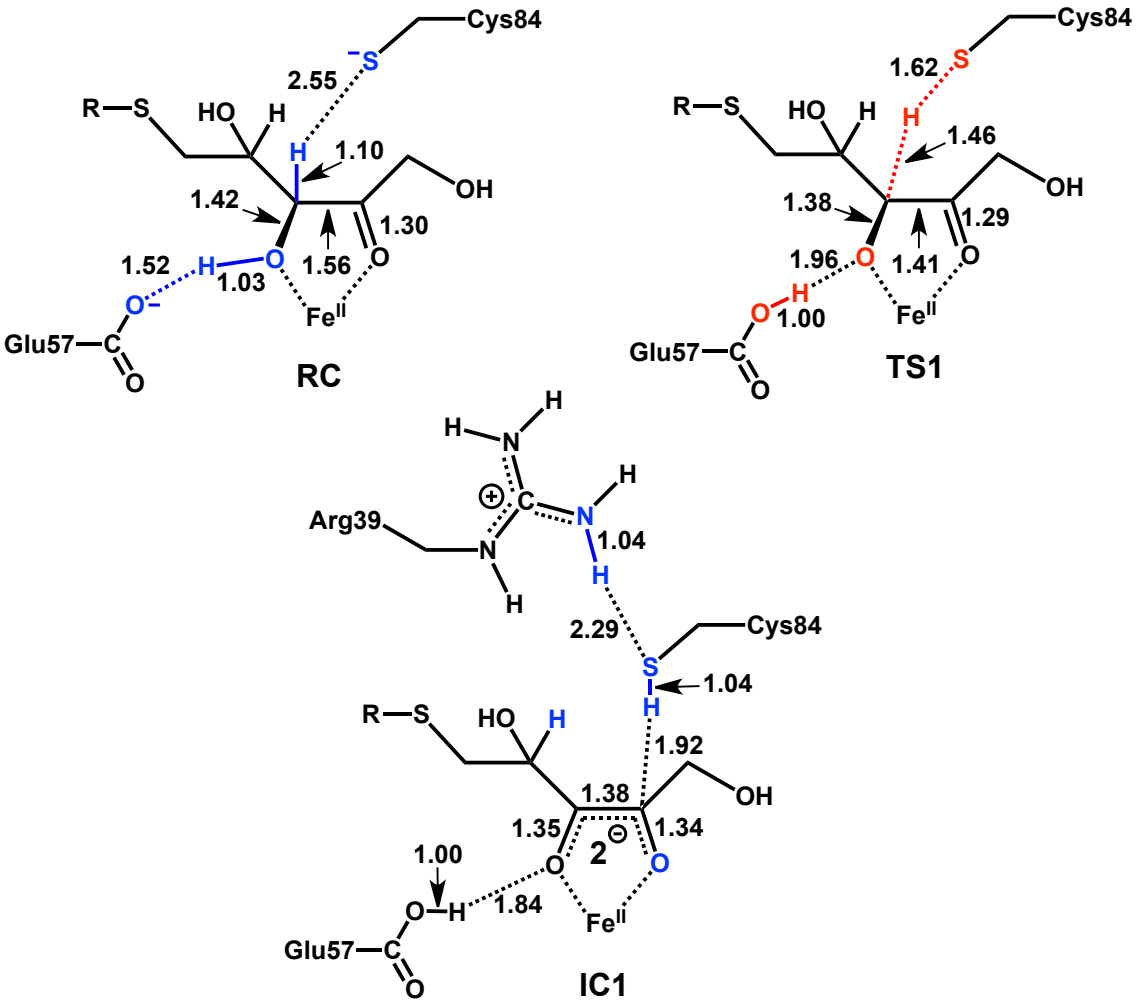


Figure 6.3. Schematic representation of optimized structures with selected bond length in Angstrom (Å) of RC, TS1 and IC1.

From thereon, we proceeded by modeling the first experimentally proposed step after the formation of the 2-keto intermediate, where Cys84B acts as a base and deprotonates $C3_{Substrate}$. The optimization of the corresponding transition state reveals that the $H_{C3–Substrate}$ transfer from $C3_{Substrate}$ to Cys84B occurs concertedly with the transfer of $H_{O3–Substrate}$ from $O3_{Substrate}$ to Glu57A. This step occurs with an energy barrier of 78.1 kJ mol^{-1} (see Figure 10) and generates a mechanistic intermediate (IC1) with a delocalized double bond extending between $O3_{Substrate}$, $C3_{Substrate}$, $C2_{Substrate}$ and $O2_{Substrate}$ (see Figure 6.3). The resulting substrate intermediate is a dianion that is stabilized by the delocalized double bond and the Fe(II) cofactor. Its formation increases the free energy of the system by 48.9 kJ mol^{-1} relative to RC. The optimized structure of IC1 indicates an interaction between $H_{S–Cys84B}$ and $C2_{Substrate}$ whereby the $H_{S–Cys84B}\cdots C2_{Substrate}$ distance is 1.92 Å in length and the $\angle S_{Cys84B}—H_{SCys84B}—C2_{Substrate}$ angle being 155.9°.

6.3.2 MD Simulations

The protonation of $O2_{Substrate}$ and the generation of the 3-keto intermediate starting from the 2-keto intermediate has been an elusive subject matters in experimental studies.[14,16] While some studies suggest Glu57B transfering a proton from $O3_{Substrate}$ to $O2_{Ssubstrate}$, others indicate that Glu57B may be too far from $O2_{Ssubstrate}$ and advance that His11B, Glu57A, Tyr89B, Cys84B as well as solvent molecules may form a proton transfer relay system, capable of transferring $H_{O3–Substrate}$ to $O2_{Substrate}$ generating the 3-keto intermediate. The atomistic details of latter propositions have not been thoroughly investigated.

The optimized structure of IC1 shows that it is unlikely that there is an interaction between $H_{O–Glu57A}$ and $O2_{Substrate}$. The distance between $H_{O–Glu57A}$ and $O2_{Substrate}$ is 4.24 Å and Glu57 remains firmly hydrogen bonded to the substrate via $H_{O–Glu57A}$ as seen in the $H_{O–Glu57A}\cdots O2_{Substrate}$ distance of 1.84 Å. Assuming that hydrogen bonding is a pre-requisite for proton transfer, the current IC1 QM/MM model suggests that Glu57A is unable to protonate $O2_{Substrate}$.

As one of the limitations of QM/MM is its inability to simulate moderate conformational changes between active site components, we performed a MD simulation to inquire into the possibility of either Glu57A or the substrate undergoing structural changes that may generate an interaction between $H_{O–Glu57A}$ and $O2_{Substrate}$. The system modeled for the MD simulation mimicked the QM/MM model. Figure 6.4 is a graphical analysis of the MD simulations done on IC1. It illustrates the changes in $H_{O–Glu57A}\cdots O3_{Substrate}$ and $H_{O–Glu57A}\cdots O2_{Substrate}$ distances throughout time. Specifically, it shows that the $H_{O–Glu57A}\cdots O2_{Substrate}$ distance remains consistently above 4 Å while the $H_{O–Glu57A}\cdots O3_{Substrate}$ distance remains on average underneath 2 Å. Indeed, upon closer analysis, the average $H_{O–Glu57}\cdots O2_{Substrate}$ distance was calculated to be 4.4 ± 0.2 Å while that of $H_{O–Glu57}\cdots O3_{Substrate}$ distance was 1.7 ± 0.2 Å approximately. These results suggest that, under the timescale considered, there are no driving forces bringing Glu57A closer to $O2_{Substrate}$ and that neither Glu57A nor the substrate possess enough flexibility to induce the hydrogen bonding required for proton transfer as suggested in the experimentally proposed mechanism.

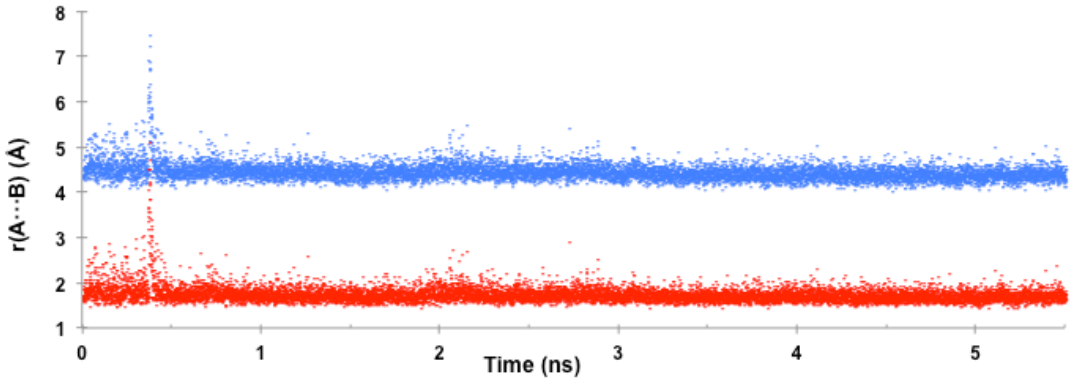


Figure 6.4. Plot of $H_{O–Glu57A}\cdots O2_{Substrate}$ (red) and $H_{O–Glu57A}\cdots O3_{Substrate}$ (blue) distances (Å) during MD simulation of IC1.

The final steps of the mechanism are proposed to involve Glu57A acting as a catalytic base by abstracting a proton from $C4_{Substrate}$ to subsequently transfering it to $S_{Substrate}$.

Previous QM/MM models and MD results, point to this step being unfeasible. Indeed, this step can only occur if Glu57A is anionic (which has been suggested of occurring as a result of transferring its acquired proton from $O3_{Substrate}$ to $O2_{Substrate}$) and hydrogen bonded to $C4_{Substrate}$. As our previous results show, Glu57A cannot protonate $O2_{Substrate}$ and hence cannot acquire an anionic form. Furthermore as subsequent MD results indicate, even if Glu57A were to acquire an anionic form with the putative 3-keto intermediate bound to the active site, it would still not be able to hydrogen bond to $C4_{Substrate}$.

Figure 6.5 (green scatter plot) plots the $O_{Glu57A}\cdots C4_{Substrate}$ distance with respect to simulation time during MD simulation. This simulation was conducted to determine if Glu57A or the substrate have enough flexibility to induce a hydrogen bond between O_{Glu57A} and $C4_{Substrate}$. As shown hereafter, the initial $O_{Glu57A}\cdots C4_{Substrate}$ distance is 4.47 Å. Subsequently the interaction shortens moderately, but not enough to promote hydrogen bonding between O_{Glu57A} and $C4_{Substrate}$. In general, the average $O_{Glu57A}\cdots C4_{Substrate}$ distance over the course of the simulation was 4.4 ± 0.3 Å. Hence, current MD results suggest that Glu57A cannot act as a catalytic base and abstract a proton from $C4_{Substrate}$.

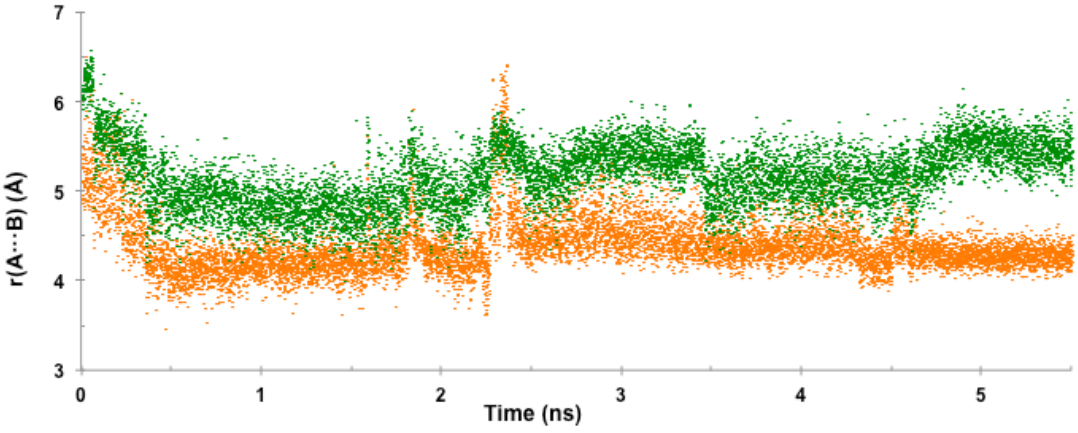


Figure 6.5. Plot of $O_{Glu57A}\cdots H_{C4–Substrate}$ (green) and $S_{Cys84B}\cdots H_{C4–Substrate}$ (orange) distances (Å) during the MD simulation of putative 3-keto intermediate bound to LuxS active site with anionic Glu57A.

Experimental studies have reviewed the proposed mechanism and suggested that Cys84B, rather than Glu57A, may be the catalytic base deprotonating $C4_{Substrate}$.[14,16] This proposal has also been investigated using MD simulations, by measuring the average $S_{Cys84B}\cdots H_{C4-Substrate}$ distance throughout time. The corresponding plot (see the orange scatter plot in Figure 6.5) shows that the initial distance is 5.6 Å and that the distance fluctuates around 5 Å approximately. Specifically, the average $S_{Cys84B}\cdots H_{C4-Substrate}$ distance is 5.2 ± 0.3 Å suggesting that S_{Cys84B} may be too far from $H_{C4-Substrate}$ to abstract it. Ultimately, current MD simulations results suggest that neither Glu57A nor Cys84B are adequately positioned with respect to the substrate to abstract $H_{C4-Substrate}$.

6.3.3 Water Mediated Transfer $H_{C4-Substrate}$ to $O2_{Substrate}$

From hereon, we considered and modeled alternative steps that could lead to a putative 3-keto intermediate. Specifically we considered the protonation of $C2_{Substrate}$ by Cys84B followed by a water-mediated proton transfer from $C4_{Substrate}$ to $O2_{Substrate}$. The transfer of $H_{S-Cys84B}$ from Cys84B to $C2_{Substrate}$ possessed a free energy barrier of 56.4 kJ mol^{-1} and generated a substrate $O2_{Substrate}$-oxyanion (as shown in IC2 in Figure 6.6). During this step, the $Fe(II)\cdots O2_{Substrate}$ distance shortened by 0.04 Å indicating an slight increase in coordination stabilizing the $O2_{Substrate}$-oxyanion in the intermediate (IC2) which had an relative Gibbs energy of -7.3 kJ mol^{-1} with respect to RC (see Figure 6.10). Optimization of both TS2 and IC2, indicated that the proton transfer from $H_{S-Cys84B}$ to $C2_{Substrate}$ occurred concertedly with the transfer of $H_{N-Arg39B}$ from Arg38B to the thiolate moiety on Cys84B (see TS2 in Figure 6.6).

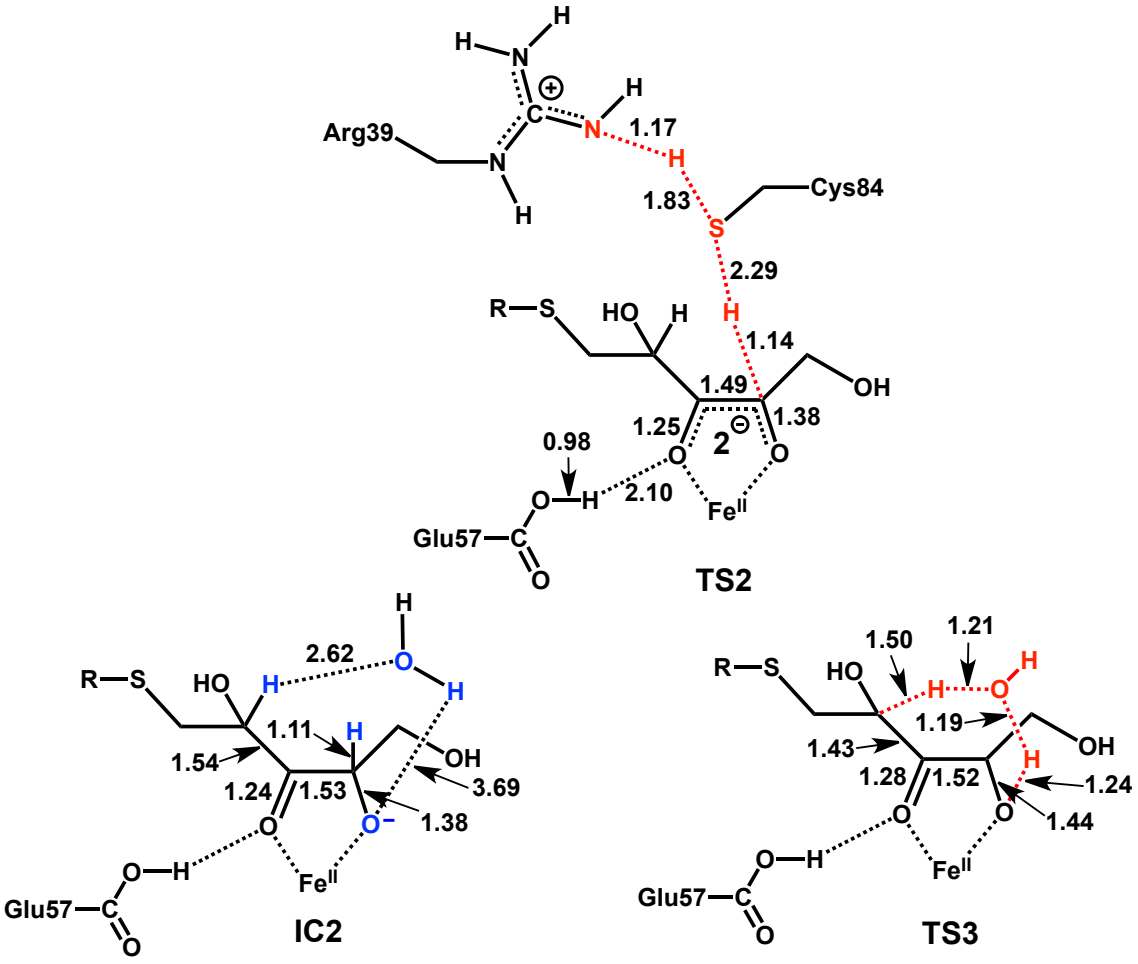


Figure 6.6. Schematic representation of optimized structures with selected bond length in angstrom (Å) of TS2, IC2 and TS3.

It should be noted that the protonation of $C2_{Substrate}$ by Cys84B was initially investigated with a QM/MM model with Arg38B in the MM layer. With this model, all attempts to optimize IC2 failed as IC2 consistently collapsed back to IC1 by barrierlessly transferring $H_{C2-Substrate}$ from $C2_{Substrate}$ back to S_{Cys84B}. With the addition of the guanidino group of Arg38B to the QM layer followed by a barrierless protonation of S_{Cys84B} by Arg39B, it was feasible to generate IC2. Although it is unclear whether Arg39B does indeed protonate Cys84B in vivo or if it is a result of running QM/MM calculations in vacuo, these results and outcomes suggest that Arg39B stabilizes the thiolate group on Cys84B. These

computational results coincide with previous experimental studies that have postulated a similar role for Arg39B.[15]

We have attempted to model the protonation of $O2_{Substrate}$. Specifically we investigated the possibility of an active site water shuttling a proton from $C4_{Substrate}$ to $O2_{substrate}$. The free energy barrier of this distant proton transfer is 119.4 kJ mol^{-1} and yields a 3-enolate (instead of the presumptive 3-keto) intermediate that lied 16.2 kJ mol^{-1} above RC. This high energy barrier constitutes the rate-determining step of this current investigation and coincides with ^{13}C-NMR studies, which have indicated that the rate-determining step occurred during the shift of the carbonyl group from $C2_{Substrate}$ to $C3_{Substrate}$.[17] It is also consistent with experimental deuterium labeling studies.[18] Indeed, experimental studies have shown that deuterium is not present in the final product (DPD) and may have been lost to the solvent. Our computational results provide a mechanistic rational whereby the deuterium bound to $C4_{Substrate}$ is transferred to the active site water and hence effectively lost to the solvent.

6.3.4 Thioether Bond Cleavage by Glu57

From hereon, as QM/MM can only provide a static representation of the model, we performed an MD simulation of IC2 to investigate the interaction between the substrate and Glu57A. Figure 6.4 plots the change throughout time of the $S_{Substrate}\cdots H_{O–Glu57A}$ distance in IC2. As seen in Figure 6.4, once Glu57A is protonated it has the capacity to hydrogen bond to $S_{Substrate}$. Indeed, the initial $S_{Substrate}\cdots H_{O–Glu57A}$ distance is 4.62 Å. But over the course of the simulation it acquires an average distance of 2.5 ± 0.8 Å. These MD results suggest that a protonated Glu57A spontaneously reorients in the active to hydrogen bond with $S_{Substrate}$.

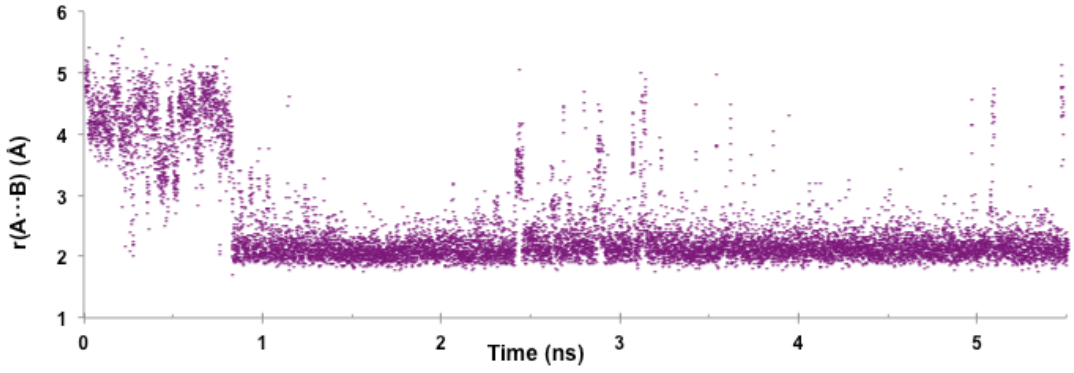


Figure 6.7. Plot of $S_{Substrate}\cdots H_{O-Glu57A}$ distance (Å) during MD simulation of IC2

In light of these MD results, we subsequently modeled, through QM/MM, a proton transfer from Glu57A to $S_{Substrate}$ in two steps. A first step (IC3→IC4) consisted of a dihedral bond rotation in Glu57A allowing $H_{O-Glu57A}$ to be more conveniently oriented towards $S_{Substrate}$. The second step (IC4→IC5) consisted of the $H_{O-Glu57A}$ transfer to $S_{Substrate}$ resulting in the thioether bond cleavage and in the formation of the product.

The dihedral bond rotation occurs with a negligible energy barrier of 38.3 kJ mol^{-1} with respect to RC. It results in an intermediate (IC4) that is 29.9 kJ mol^{-1} more stable than RC. Initially in IC3, the ∠O–C–O–H dihedral angle is 30.0° and increases to 134.0° in IC4. Over the course of the dihedral angle rotation, Glu57A remains hydrogen bonded to $O3_{Substrate}$; $r(H_{O-Glu57A}\cdots O3_{Ssubstrate})$ in TS4 is 1.86 Å.

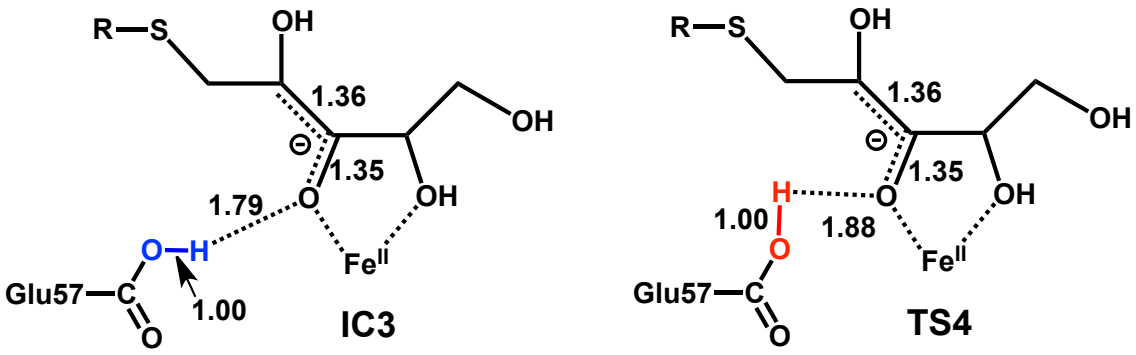


Figure 6.8. Schematic representation of optimized structures with selected bond length in angstrom (Å) of IC3 and TS4

The final step of the mechanism consists of $H_{O\text{–}Glu57A}$ transfer to $S_{Substrate}$ resulting in the thioether bond cleavage and in the formation of the product. The product complex (PC) lied 36.7 kJ mol^{-1} higher in free energy than RC (see Figure 6.10). Throughout the entirety of our QM/MM study (including the one by Huang et al.[19]), both $O2_{Substrate}$ and $O3_{Substrate}$ remained consistently coordinated to Fe(II) generating a stable Fe(II) 5-coordination complex. However, upon modeling PC, the coordination of $O2_{Substrate}$ to Fe(II) breaks and both components cease to interact. Indeed, while in IC4 the $O2_{Substrate}$···Fe(II) distance is 2.29 Å, in PC this distance increases significantly to 4.15 Å thereby producing a tetrahedral Fe(II) center whereby Fe(II) is coordinated only to $O3_{Substrate}$ by 2.00 Å. This suggests that upon cleaving the thioether bond, the DPD moiety's interaction with Fe(II) decreases, perhaps facilitating its departure from the active site. Upon transferring a proton to $S_{Substrate}$, the functional group on Glu57 becomes anionic thereby repulsing $O3_{Substrate}$, which has a partial negative charge. The thioether bond cleavage enhances the substrate's flexibility allowing it to move away from the $^{-}O_2C_{Glu57}$ functional group and towards Fe(II).

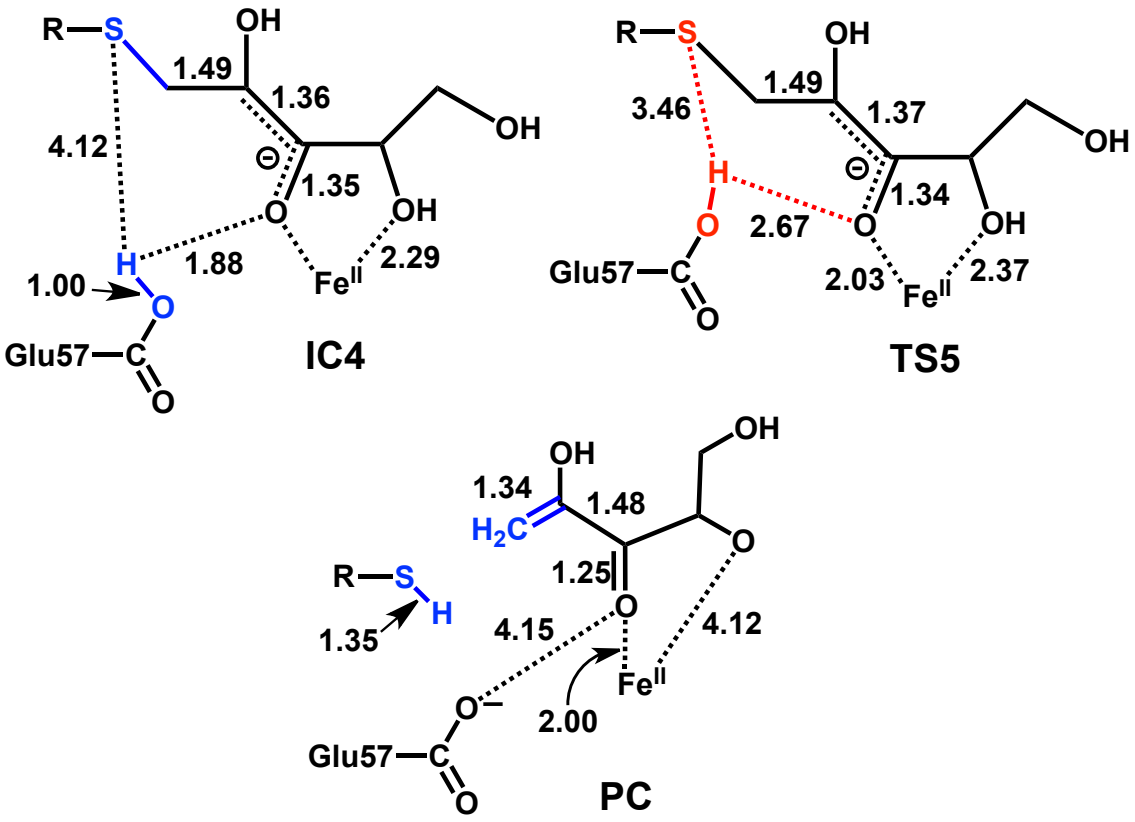


Figure 6.9. Schematic representation of optimized structures with selected bond length in angstrom (Å) of IC4, TS5 and PC.

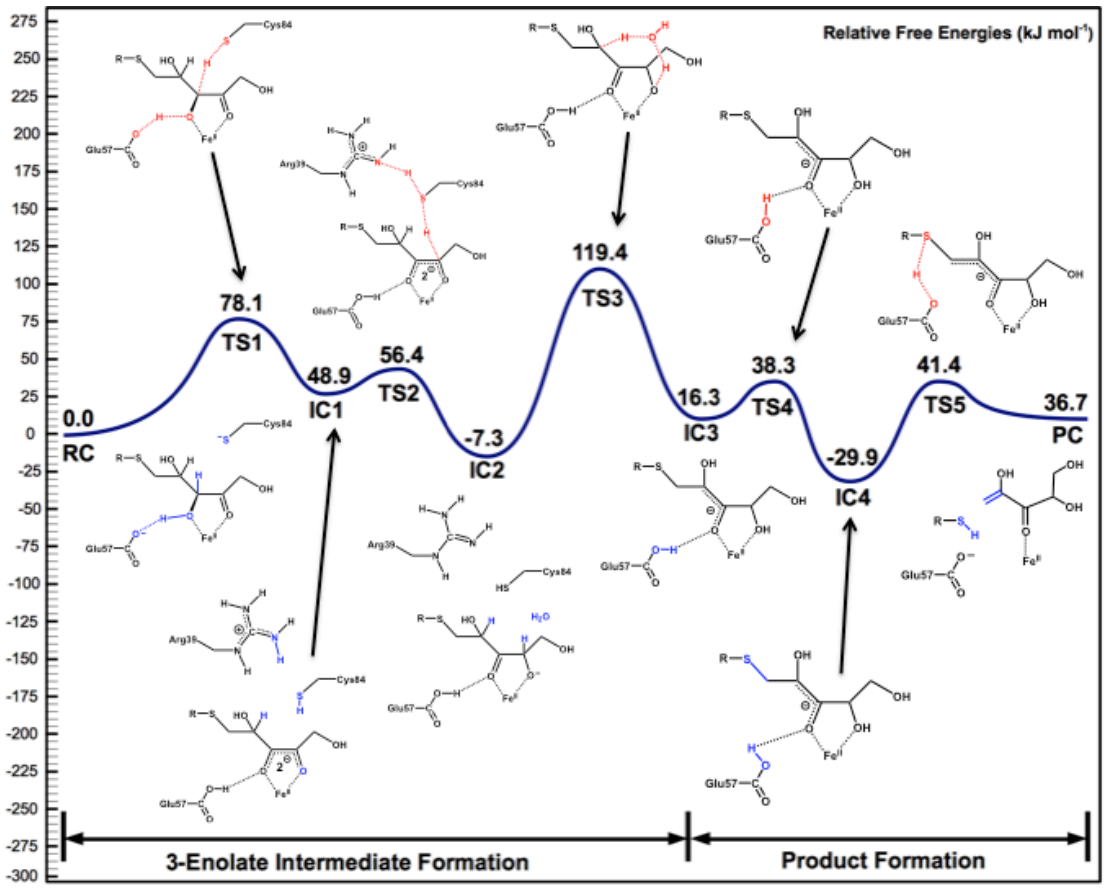


Figure 6.10. Schematic free energy surface of the conversion of putative 2-keto intermediate into DPD.

6.3.5 The Mechanistic Role of Fe(II)

Previous experimental studies[1] have suggested that Fe(II) is a Lewis acid and mediates the internal redox reaction. To provide further insights, we performed an additional set of single point calculations where Fe(II) was removed from the optimized structures. The resulting relative free energies suggest that Fe(II) dramatically lowers the relative free energies in the initial two steps thereby allowing the active site of LuxS to readily deprotonate the substrate and stabilize its emergent negative charges. Specifically in the first step of (RC→IC1), Fe(II) reduces the free energy barrier from 164.5 kJ mol^{-1} to 78.1 kJ mol^{-1} and stabilizes IC1 by 91.5 kJ mol^{-1} with respect to RC.

Intriguingly, Fe(II) slows down dramatically the subsequent steps in the mechanism and specifically the rate of the distant H^+ transfer during the rate-determining step (IC2→IC3) by increasing its free energy barrier by a considerable 108.6 kJ mol^{-1}. The relative free energy barrier of this step without Fe(II) is 10.8 kJ mol^{-1} and suggests that upon protonating $C2_{Substrate}$ and generating IC2 with its anionic $O2_{Substrate}$, the substrate intrinsically favors a water mediated H^+ transfer from $C4_{Substrate}$ to $O2_{Substrate}$. Hence, although Fe(II) considerably increases the rate by which Glu57A and Cys84B deprotonate the substrate, it decreases the rate of subsequent H^+ transfers leading to the cleavage of the thioether bond. However, it should be noted that the rate-enhancement provided by Fe(II) in LuxS outweighs its rate-reduction.

6.4 Conclusions

QM/MM and MD simulations results presented herein support an alternate mechanistic pathway for the conversion of a putative 2-keto intermediate to a DPD tautomer. This catalytic mechanism involves the initial concerted deprotonation of $C3_{Substrate}$ and $O3_{Substrate}$ by Cys84B and Glu57A respectively. Subsequently, Cys84B shuttles the H^+ to $C2_{Substrate}$ thereby generating an oxyanion on $O2_{Substrate}$. An active site water cofactor then acts as a long distance proton shuttle by transferring H^+ from $C4_{Substrate}$ to $O2_{Substrate}$ generating a stable 3-enolate intermediate. This step occurs with a free energy barrier of 119.4 kJ mol^{-1} and is the rate-determining step of this mechanism. Importantly this correlates with experimental results, which have suggested that the rate-determining step involves the conversion of a presumptive 2-keto to a 3-keto intermediate and that a deuterium bound to $C4_{Substrate}$ may be lost to the solvent during the mechanism. A fast dihedral bond rotation of ∠O–C–O–H in the side chain of Glu57A allows $H_{O-Glu57A}$ to be oriented towards $S_{Substrate}$. Glu57A then shuttles a H^+ from $O3_{Substrate}$ to $S_{Substrate}$ before the full cleavage of the thioether bond. The thioether bond cleavage decreases the interaction between the DPD moiety and Fe(II) as reflected by the $O2_{Substrate}$···Fe(II) distance, which elongates

significantly generating a tetrahedral Fe(II). Subsequent QM/MM calculations reveal that Fe(II) may stabilize the highly anionic species occurring in the first two mechanistic steps and increasing the rate by which Cys84B and Glu57A deprotonate the substrate. However this comes at the cost of increasing subsequent free energy barriers.

6.5 References

(1) Zhu, J.; Dizin, E.; Hu, X.; Wavreille, A.-S.; Park, J.; Pei, D. *Biochemistry* **2003**, *42*, 4717.

(2) Xavier, K. B.; Bassler, B. L. *Curr. Opin. Microbiol.* **2003**, *6*, 191.

(3) Schauder, S.; Shokat, K.; Surette, M. G.; Bassler, B. L. *Mol. Microbiol.* **2001**, *41*, 463.

(4) Chen, X.; Schauder, S.; Potier, N.; Van Dorsselaer, A.; Pelczer, I.; Bassler, B. L.; Hughson, F. M. *Nature* **2002**, *415*, 545.

(5) Federle, M. J.; Bassler, B. L. *J. Clin. Invest.* **2003**, *112*, 1291.

(6) Hammer, B. K.; Bassler, B. L. *Mol. Microbiol.* **2003**, *50*, 101.

(7) Tian, Y.; Wang, Q.; Liu, Q.; Ma, Y.; Cao, X.; Guan, L.; Zhang, Y. *Biosci. Biotechnol. Biochem.* **2008**, *72*, 1063.

(8) Bassler, B. L.; Losick, R. *Cell* **2006**, *125*, 237.

(9) Ohtani, K.; Hayashi, H.; Shimizu, T. *Mol. Microbiol.* **2002**, *44*, 171.

(10) Pei, D.; Zhu, J. *Curr. Opin. Chem. Biol.* **2004**, *8*, 492.

(11) Hilgers, M. T.; Ludwig, M. L. *Proc. Nat. Acad. Sci.* **2001**, *98*, 11169.

(12) Ruzheinikov, S. N.; Das, S. K.; Sedelnikova, S. E.; Hartley, A.; Foster, S. J.; Horsburgh, M. J.; Cox, A. G.; McCleod, C. W.; Mekhalfia, A.; Blackburn, G. M.; Rice, D. W.; Baker, P. J. *J. Mol. Biol.* **2001**, *313*, 111.

(13) Lewis, H. A.; Furlong, E. B.; Laubert, B.; Eroshkina, G. A.; Batiyenko, Y.; Adams, J. M.; Bergseid, M. G.; Marsh, C. D.; Peat, T. S.; Sanderson, W. E. *Structure* **2001**, *9*, 527.

(14) Rajan, R.; Zhu, J.; Hu, X.; Pei, D.; Bell, C. E. *Biochemistry* **2005**, *44*, 3745.

(15) Zhu, J.; Knottenbelt, S.; Kirk, M. L.; Pei, D. *Biochemistry* **2006**, *45*, 12195.

(16) Gopishetty, B.; Zhu, J.; Rajan, R.; Sobczak, A. J.; Wnuk, S. F.; Bell, C. E.; Pei, D. *J. Am. Chem. Soc.* **2008**, *131*, 1243.

(17) Zhu, J.; Hu, X.; Dizin, E.; Pei, D. *J. Am. Chem. Soc.* **2003**, *125*, 13379.

(18) Zhu, J.; Patel, R.; Pei, D. *Biochemistry* **2004**, *43*, 10166.

(19) Huang, W.; Gherib, R.; Gauld, J. W. *J. Phys. Chem. B* **2012**, *116*, 8916.

(20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comp. Chem.* **2005**, *26*, 1781.

(21) Molecular Operating Environment; 2012.10 ed.; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2012.

(22) 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., J. A. ; 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. P., 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, Inc., CT: Wallingford 2009.

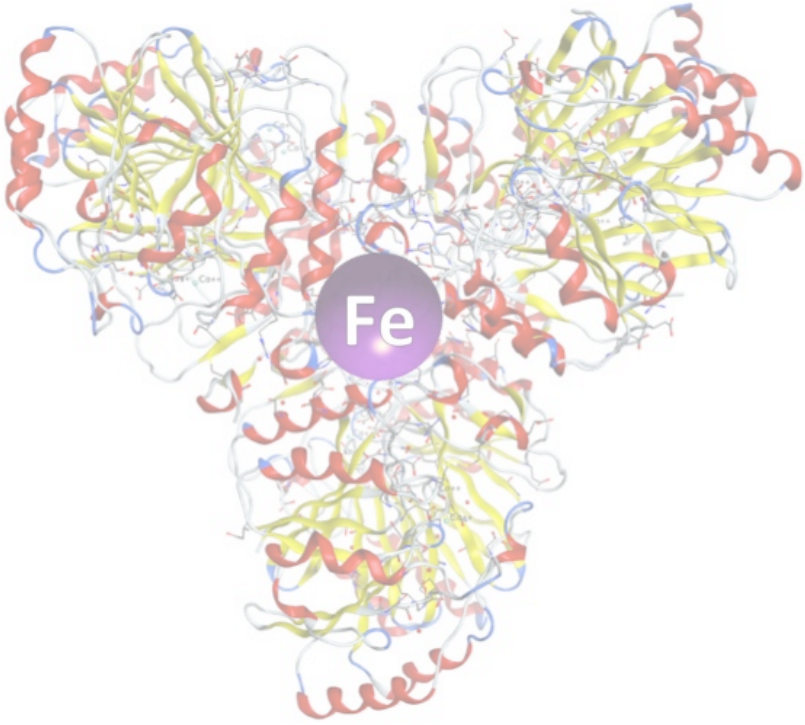
(23) Morokuma, K. *Bull. Korean Chem. Soc.* **2003**, *24*, 797.

(24) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

Chapter 7

Conclusions



7.1 Conclusions

In this work, we have described at the molecular level using a multi-scale approach five processes involved in iron-based biocatalysis. Using a density functional theory (DFT) cluster approach we have shown that (1) small-scale environments of enzymatic active sites enhance the reactivity of nucleophiles by rendering them more polarizable while (2) destabilizing phosphoester substrates as a catalytic strategy. Using a larger scale QM/MM approach we have shown (3) how iron-heme enzymes are often naturally built with active sites that allow them to perform unnatural catalysis. Subsequently, by using MD simulations and QM/MM we have shown (4) that iron can significantly stabilize negatively charged transition states and intermediates thereby increasing the rate at which enzymatic active site perform H^+ transfer and (5) allow carbohydrate-based substrates to adopt states inherently suited for water mediated long-ranged proton transfer.

In **Chapter 3**, the catalytic mechanism of the binuclear metalloenzyme, glycerophosphodiesterase of *Enterobacter aerogenes* (GpdQ), was investigated using a relatively large-scale DFT-cluster methodology. A statistical analysis of DFT models and the analogous crystal structure suggests, that prior to substrate binding, the active site contained a neutral His81, a bridging hydroxide and terminal water ligand. A free energy assessment of the coordination of a dimethylphosphate (DMP) substrate was performed and revealed that DMP spontaneously displaces both hydroxide and water terminal ligands and coordinate to a ferrous metal ion in the α site and that such process was entropically driven.

The mechanism of GpdQ using a variety of Fe(II)-containing heteronuclear metal ion compositions was subsequently studied by DFT and show that the hydrolysis of DMP by GpdQ proceeds through a two-step mechanism. Such mechanism involves a nucleophilic bridging hydroxide ligand and a protonated His81 that acts as an acid while protonating the methoxide leaving group. The theoretical trends reproduce those acquired experimentally by Daumann et al.[1] who have studied the hydrolysis of bisparanitrophenol

(bpNPP) by GpdQ. A more in-depth analysis suggests that the size of Cd(II) allows $Fe(II)_{\alpha}/Cd(II)_{\beta}$-GpdQ to strain and destabilize the substrate by 20.8 kJ mol^{-1} thereby enhancing its catalytic rate. Furthermore, based on a rationale rooted in hard soft acid base (HSAB) theory,[2] we have suggested that $Fe(II)_{\alpha}/Mn(II)_{\beta}$ and $Fe(II)_{\alpha}/Cd(II)_{\beta}$-GpdQ make the hydroxide bridging ligand a softer nucleophile (compared to $Fe(II)_{\alpha}/Co(II)_{\beta}$ and $Fe(II)_{\alpha}/Zn(II)_{\beta}$-GpdQ) allowing it to bind more strongly to the substrate during the formation of the phosphorus-centered bipyramidal activated complex.

In **Chapter 4**, the ability of different DFT exchange-correlation functionals to model compound I (Cpd I) in coral allene oxide synthase (cAOS) has been investigated in conjunction with an ONIOM QM/MM approach. A small basis set assessment indicates that the use of B3LYP/6-31G to optimize Cpd I in catalase-type active sites (as done in some previous computational studies) may not be stringent enough to adequately model its structural and electronic properties. Our assessment indicates BLYP does not predict the doublet and quartet spin states to be degenerate as typically seen in Cpd I/P450. While B3LYP[3] models doublet and quartet Cpd I/cAOS as a square pyramidal complex, other functionals (i.e. BLYP,[4] B3LYP*,[5] M06[6] and M06-L[6,7]) model it as an octahedral complex. M06 favors the sextet spin state by 13.3 kJ mol^{-1}, unlike all other considered functionals which prefer doublet and quartet spin states. The structural and electronic properties predicted by B3LYP and M06-L differ from one another; while B3LYP predicts a 5-membered coordination complex with an electron radical localized on the side-chain of Tyr353, M06-L predicts a 6-membered coordination complex with an electron radical delocalized between the porphyrin moiety and Tyr353 side-chain.

In **Chapter 5**, we have studied via QM/MM the catalytic mechanism of cAOS when oxidized by iodosylbenzene to form Cpd I as proposed experimentally.[8] We have modeled the hydroxylation and epoxidation reactions of the latter with arachidonic acid thereby forming 8R-hydroxyeicosatetraenoic acid (8R-HETE) and 8R,9S-epoxyeicosatetraenoic acid (8R,9S-EET) respectively. The hydroxylation involves an

initial proton-coupled electron transfer (PCET) with His67 shuttling a proton from the substrate to the oxo-ligand coordinated to Fe(IV). Concertedly, an open-shell Tyr353 oxidizes the substrate. The PCET is rate limiting and involves multi-state reactivity with the involvement of both doublet and quartet spin-states with free energy barriers of 93.8 and 94.5 kJ mol^{-1} respectively. Epoxidation involves the direct association of the oxo-ligand to the substrate while undergoing multi-state reactivity with energy barriers of 50.6 and 67.3 kJ mol^{-1} respectively for doublet and quartet spin-surfaces. While the doublet species directly collapses to the product (8R,9S-EET), the substrate in the quartet barrierlessly rearranges to a stable aldehyde byproduct. Compared to a previous QM/MM study on the reactivity of Cpd I/P450 to cyclohexene,[9] our results suggest that Cpd/cAOS slows down the rate of hydroxylation but enhances that of epoxidation.

In **Chapter 6**, we have studied an important part of the mechanism of LuxS via MD simulations and QM/MM. Specifically we have used MD to model the active site's hydrogen-bonding network and the interactions between the catalytic residues and the substrate. A statistical analysis of our MD simulations revealed that the active site may not be inherently flexible to allow Glu57A to shuttle a H^+ from $O3_{Substrate}$ to $O2_{Substrate}$, as suggested experimentally.[10-16] Furthermore, upon the generation of tentative 3-keto intermediate, our MD results suggest that the active site may not be able to adopt the required conformation allowing it to use neither Cys84B nor Glu57A as H^+ shuttles from $C4_{Substrate}$ to $S_{Substrate}$.

Using QM/MM, we modeled a mechanistic pathway involving an active site water performing a long-distance H^+ transfer from $C4_{Substrate}$ to $O2_{Substrate}$. Our results are consistent with experimental studies that suggest that not only is the shift of the carbonyl from C2 to C3 rate-limiting but also that the C4 bound H^+ is lost to the solvent during the mechanism. Both MD and QM/MM suggest that a H^+ transfer from $C3_{Substrate}$ to $S_{Substrate}$ occurs through a Glu57A shuttle. Our results also suggest that the Fe(II) cofactor

enhances the rate at which the active site of LuxS deprotonates the substrate and stabilizes significantly the mechanistic dianion species occurring in the first two biocatalytic steps.

The work herein focused on the catalytic strategies employed by iron-based enzymes. Amongst many things, it has shows how computational chemical tools can generate atomistic insights into their mechanisms. The insights acquired herein can enhance our ability to explain how enzymes dramatically increase the rate of reactions; a substantial topic in modern biochemistry. Not only is it an important theoretical problem whose solution could have significant repercussions in other areas of chemistry, biochemistry and biophysics but can also be used practically. Indeed, insights into this matter can not only enhance our ability to fight diseases by designing more efficient pharmaceutical drugs but also guide the development of biomimetic catalysts that employ the same catalytic strategies as enzymes.

7.2 References

(1) Daumann, L. J.; McCarthy, B. Y.; Hadler, K. S.; Murray, T. P.; Gahan, L. R.; Larrabee, J. A.; Ollis, D. L.; Schenk, G. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 425.

(2) Pearson, R. G. *J. Am. Chem. Soc.* **1963**, *85*, 3533.

(3) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(4) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(5) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48.

(6) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(7) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(8) Boeglin, W. E.; Brash, A. R. *J. Biol. Chem.* **2012**, *287*, 24139.

(9) Cohen, S.; Kozuch, S.; Hazan, C.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 11028.

(10) Gopishetty, B.; Zhu, J.; Rajan, R.; Sobczak, A. J.; Wnuk, S. F.; Bell, C. E.; Pei, D. *J. Am. Chem. Soc.* **2008**, *131*, 1243.

(11) Pei, D.; Zhu, J. *Curr. Opin. Chem. Biol.* **2004**, *8*, 492.

(12) Rajan, R.; Zhu, J.; Hu, X.; Pei, D.; Bell, C. E. *Biochemistry* **2005**, *44*, 3745.

(13) Zhu, J.; Dizin, E.; Hu, X.; Wavreille, A.-S.; Park, J.; Pei, D. *Biochemistry* **2003**, *42*, 4717.

(14) Zhu, J.; Hu, X.; Dizin, E.; Pei, D. *J. Am. Chem. Soc.* **2003**, *125*, 13379.

(15) Zhu, J.; Knottenbelt, S.; Kirk, M. L.; Pei, D. *Biochemistry* **2006**, *45*, 12195.

(16) Zhu, J.; Patel, R.; Pei, D. *Biochemistry* **2004**, *43*, 10166.

Vita Auctoris

Name: Rami Gherib

Place of Birth: Montreal, Quebec, Canada

Year of Birth: 1990

Education

M.Sc. Chemistry 2014

University of Windsor Windsor, Ontario

B.Sc. Biochemistry [H] 2012

University of Windsor Windsor, Ontario