Computational Studies of Dinoflagellate Luciferase and Radical *S*-adenosyl-L-methionine Enzymes

by

Patrick Hugh Donnan

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 10, 2022

Keywords: Mechanistic enzymology, computational chemistry, molecular dynamics, broken-symmetry density functional theory

Approved by

Steven O. Mansoorabadi, Chair, J. Milton Harris Associate Professor, Department of Chemistry and Biochemistry
Eduardus C. Duin, Professor, Department of Chemistry and Biochemistry
Evangelos Miliordos, Associate Professor, Department of Chemistry and Biochemistry
Rafael C. Bernardi, Assistant Professor, Department of Physics
Andrew Adamczyk, Assistant Professor, Department of Chemical Engineering

Abstract

This dissertation seeks to apply computational methods to study two enzymatic systems: dinoflagellate luciferase and radical *S*-adenosyl-L-methionine enzymes. Dinoflagellates are microorganisms, many of which are capable of bioluminescence. Their bioluminescence is 'simple,' requiring only the luciferase enzyme, the luciferin substrate, and molecular oxygen—no other cofactors or cosubstrates are needed. The enzyme is regulated by pH, active at acidic pH and inactive at alkaline pH. To study its pH-dependent dynamics, constant pH accelerated molecular dynamics was employed. This dissertation reports an open, presumed-active conformer of the enzyme which, to date, experimental structural methods have been unable to obtain. Then, using the open conformer of luciferase, molecular dynamics studies of the reaction substrate, product, and proposed intermediates were conducted. Using the results of these studies, this dissertation pieces together the likely stereochemical and regiochemical course of the dinoflagellate luciferase reaction, predicting which amino acid residues of the enzyme are important for catalysis. Turning to radical *S*-adenosyl-L-methionine (SAM) enzymes, this dissertation applies broken-symmetry density functional theory to understand the common catalytic steps utilized by members of this diverse enzyme superfamily, which result in the creation of the strong oxidant 5'-dAdo·. This dissertation suggests that a recently discovered intermediate in the common catalytic mechanism, Ω, is not organometallic but is instead a near-attack conformer of SAM bound to the catalytic iron-sulfur cluster. The dissertation applies the same methodology to the case of a non-canonical radical SAM enzyme, Dph2, which does not form 5'-dAdo·. Predicted paramagnetic reaction intermediates are analyzed and compared to experimental data.

Dedication

To Forannan

Acknowledgements

This dissertation would not have been possible without a great number of people. While I hope to thank them all here, some will be unintentionally omitted; apologies if you are one.

To start, I would like to thank those who contributed to my formation as a researcher during my undergraduate. Prof Francis Robicheaux, Prof Ed Thomas, Kelsie, and Erica all contributed greatly to my development as a scientist. I would like to thank Prof Thomas Pohl, Georg, and Karsten for an incredible research experience in Germany. My time in the Society of Physics Students room is irreplaceable, and I will always treasure the years I spent learning and growing with my fellow SPS members. I would be remiss if I did not thank Paul Harris for his support while applying for scholarships as an undergraduate and upon my return to Auburn as a graduate student. From my brief foray into physics graduate school, I would like to thank Jon, Andy, and Sean for their conversations during coffee breaks, sense of humor, and ability to find free food, respectively.

I would like to thank all members of the Mansoorabadi lab, past and present, for tolerating me, especially at literature review and group meeting. To members past: Phong, Victoria, Ken, Selam, Xingchen, Eta, Trey, and Shuxin, I wish you the best in your current endeavors. To members present: Kenny, Chidinma, Prosenjit, Chelsea, Emily, Luke, and Joshua, I hope you enjoy the rest of your time in the lab and make the most of it! I would especially like to thank Trey for the many interesting conversations (some about science) while we worked in the lab. I would like to thank my committee

members, Prof Evert Duin, Prof Evangelos Miliordos, Prof Rafael Bernardi, and (formerly) Prof Holly Ellis, all of whom contributed to my development as a scientist.

In most dissertations I have read, many people thank their research advisor first in their acknowledgements. However, I felt following that formula would have diminished exactly how important Prof Steve Mansoorabadi was to the completion of this dissertation. To start, without meeting him, I never would have considered pursuing a PhD in biochemistry. Now, I hope to spend the rest of my life working with metalloenzymes. Steve supported me intellectually—his door was always open, and I will never forget the many hours spent working back-and-forth on spin projection math—but also emotionally and mentally—he had unlimited patience when I dealt with personal tragedy and gave me time and support to recover, all during a global pandemic. This dissertation could not have been completed under another advisor.

This work would not have been possible without my mother, brother, and late father. My parents facilitated in me a love of learning and a desire to pursue intellectual challenges for their own sake, without which I would not be writing this dissertation. Thank you both for your tireless support over the years. And Jonathan—thank you for watching the cats whenever we needed, and your constant willingness to help us out.

Thank you to my incredible wife, Alyssa, who has been a constant source of love and encouragement for many years now. Our family, with Hemingway and Fitzgerald, has kept me balanced throughout my doctorate. Finally, to Joan: I hope you read this one day and are inspired to be the best you can be.

Table of Contents

Abstract **2**

Dedication **3**

Acknowledgments **4**

List of Tables **9**

List of Figures **10**

List of Abbreviations **14**

Chapter 1: Introduction **17**

1.1 Modern Enzymology 17

1.2 Dinoflagellate Luciferase 19

1.3 Radical SAM Enzymes 26

1.3.1 Discovery of a Paramagnetic Intermediate 29

1.4 Outline of Following Chapters 31

1.5 References 32

Chapter 2: Computational Methods **37**

2.1 Molecular Dynamics 37

2.1.1 Force fields 38

2.1.2 Solvation and Charge 41

2.1.3 Molecular Dynamics Protocol 42

2.1.4 Advanced Sampling Methods in this Dissertation 45

2.1.5 Molecular Dynamics Analysis Techniques 45

2.2 Quantum Chemical Methods 48

2.2.1 Hartree-Fock 49

2.2.2 Density Functional Theory 49

2.2.3 Broken-Symmetry Density Functional Theory.....50

2.3 Note on Multiscale Modeling 51

2.4 References.....52

Chapter 3: Constant pH Accelerated Molecular Dynamics Investigation of the pH Regulation Mechanism of Dinoflagellate Luciferase 54

3.1 Abstract 54

3.2 Introduction 54

3.3 Computational Methodology 57

3.4 Results and Discussion 59

3.5 References 65

Chapter 4: Stereochemistry and Regiochemistry in Dinoflagellate Bioluminescence 68

4.1 Abstract 68

4.2 Introduction 69

4.3 Computational Methodology 74

4.4 Results and Discussion 80

4.5 References 90

Chapter 5: Broken-Symmetry Density Functional Theory Analysis of the Ω Intermediate in Radical *S*-Adenosyl-L-methionine Enzymes: Evidence for a Near-Attack Conformer over an Organometallic Species 95

5.1 Abstract 95

5.2 Introduction 95

5.3 Computational Methodology 98

5.4 Results and Discussion 100

5.5 References 106

Chapter 6: Computationally and Spectroscopically Consistent Models of Paramagnetic Intermediates in Dph2 Catalysis 112

6.1 Abstract 112

6.2 Introduction 112

6.3 Computational Methodology 117

6.4 Results and Discussion 120

6.5 References 125

Chapter 7: Conclusion and Outlook 129

7.1 Concluding Remarks 129

7.2 Outlook and Future Directions 129

7.3 References 130

Appendix A Supporting Material for Chapter Three 131

Appendix B Supporting Material for Chapter Five 132

List of Tables

Table 4.1 Distance (in Å) between selected atoms and Glu1105 Cδ for LCF•LH_2, LCF•LH^-, and LCF•LO binary complex simulations. 81

Table 4.2 Residues identified as catalytically important in the bioluminescent reaction mechanism, their proposed role in catalysis, and predicted effect of nonfunctional mutagenesis experiments. 89

Table 5.1 Calculated and Experimental Hyperfine Coupling Constants (in MHz) for structural models of Ω. 101

Table 6.1 Calculated hyperfine coupling constants (HFCCs) of atoms with components of the HFCC greater than 10 MHz in the Nε protonated model of intermediate **II**. 122

Table A1. pK_a values for titratable residues calculated over the full 1 μs simulation at pH 8 and 6 reported in Chapter Three. 131

Table B1. Distances between the methyl carbon of SAM and the apical irons (Fe1 and Fe3) of the $[4Fe\text{-}4S]^{1+}$ cluster as the unique iron–SAM sulfur distance is constrained .. 135

List of Figures

Figure 1.1. Hypothesized biosynthetic pathway of dinoflagellate luciferin with enzymes omitted. Chlorophyll *a* is shown with carbon numeration standard for tetrapyrroles, and dinoflagellate luciferin is shown with magenta asterisks at the unresolved stereocenters. For illustration, the C4 chiral carbon is shown in the *S* configuration, while the C15 to C16 double bond is shown in the *E* configuration. 20

Figure 1.2. Example chemical structures of other known luciferins discussed in the text. 22

Figure 1.3. Overview of the reduction and oxidation reactions involving dinoflagellate luciferin. The proposed P630 precursor is reduced to make dinoflagellate luciferin (unknown stereocenters marked with an asterisk; the double bond unknown to be in the *E* or *Z* configuration additionally colored lime green). Dinoflagellate luciferin can be oxidized enzymatically to make oxyluciferin and light, or it can be nonenzymatically air-oxidized at C15. In principle, P630 could be regenerated from the air oxidation product via dehydration.....23

Figure 1.4. AlphaFold predicted structure of full-length dinoflagellate luciferase. 24

Figure 1.5. The only solved crystal structure of LCF, a pH 8 inactive structure of domain 3 (PDB ID: 1VPR). Four intramolecularly conserved histidine residues, previously experimentally implicated in the conformational shift, are shown as sticks. Reprinted with permission from *Biochemistry* **2018**, *57 (3)*, 295–299. Copyright 2018 American Chemical Society. 25

Figure 1.6. Crystal structure of the active site of the radical SAM enzyme HydE (PDB ID: 7O25). SAM and the substrate are shown in cyan and magenta sticks, respectively... 27

Figure 1.7. Sequence similarity network (SSN) for the radical SAM superfamily, showing characterized (colored) and uncharacterized (uncolored) subgroups. Reproduced with permission from *Methods Enzymol.* **2018**, *606*, 1–71. Copyright 2018 Elsevier. 29

Figure 1.8. Reductive SAM cleavage reaction catalyzed by members of the radical SAM superfamily. The proposed organometallic structure of the Ω intermediate is shown. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society. 30

Figure 1.9. Visual overview of dissertation chapters and their organization, as well as methods employed in each.. 32

Figure 3.1 Fraction of LCF domain III residues in the protonated form (fHA) during the CpHaMD simulations at pH 8 (red) and pH 6 (blue). Reproduced with permission from Biochemistry 2018, 57, 295−299. Copyright 2018 American Chemical Society 60

Figure 3.2 Calculated structures of LCF domain III after 1 μs of CpHaMD simulation. (A) Inactive structure at pH 8, showing the close association of H1064 and R1136. Helices α3−α6 and the N- and C-termini are indicated. (B) Two different orientations of the activated structure at pH 6. R1136 forms a salt bridge with E1158, while the N-terminal domain containing the intramolecularly conserved histidine residues reorganizes. (C) Comparison of the N-terminal domain structure at pH 8 and 6. K1094 participates in a hydrogen bonding network that stabilizes the N-terminal helical bundle at pH 6. (D) Comparison of the active site structure at pH 8 and 6, showing an expansion of the β-barrel and repositioning of the putative catalytic base E1105. Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society. 62

Figure 3.3 Root-mean-square fluctuation (RMSF) and inter-residue distance data of LCF domain III from the CpHaMD simulations at pH 8 and 6. (A) Per residue RMSF of LCF domain III calculated over the full 1 μs simulation. (B) Distances between the Cα atoms of H1064 and R1136 during the 1 μs simulation. Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society. 64

Figure 4.1 The bioluminescent reaction of dinoflagellate luciferin, producing oxyluciferin and light (λ_{max} = 474 nm). The luciferase enzyme is shown in a computationally-determined open conformer. Unresolved stereocenters are labeled with magenta asterisks. 70

Figure 4.2 Overview and nomenclature used of substrate, intermediate, and product structures relevant to this chapter. Undetermined stereocenters are left unspecified. 73

Figure 4.3 Comparison of (a) (*E,S*)-LH_2 binding to LCF and (b) (*Z,S*)-LH_2 binding to LCF via closest to average structures. In each case, the average distance between $C13^2$ of the substrate and Cδ of the putative catalytic base, Glu1105, is shown. Protons on the enzyme are omitted for clarity. 81

Figure 4.4 Comparison of (a) (*E,S*)-LO binding to LCF, (b) (*Z,R*)-LO binding to LCF, (c) (*Z,S*)-LO binding to LCF, (d) (*E,S*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105, (e) (Z,*R*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105, and (f) (*Z,S*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105 via closest to average structures. In each case, the average distance between $C13^2$ of the product and Cδ of the putative catalytic base, Glu1105, is shown. We note that only the (*E,S*)-LO•LCF simulation with deprotonated Ty1168 and protonated Glu1105 produced an ordered water molecule in the active site (red circle). Protons on the enzyme are omitted for clarity. 82

Figure 4.5 Comparison of (a) (*E,S*)-LH^- binding to LCF and (b) (*Z,S*)-LH^- binding to LCF via closest to average structures. In each case, the average distance between $C13^2$ of the intermediate and Cδ of the putative catalytic base, Glu1105, is shown. Protons on the enzyme are omitted for clarity. 83

Figure 4.6 Convergence of the closest to average structures of (*E,S*) species with molecular oxygen bound. O_2 is shown in cyan for the LCF•LH_2•O_2 ternary complex, green for the LCF•LH^-•O2 ternary complex starting O2 on the *re* face of the intermediate, and in magenta for the LCF•LH^-•O_2 ternary complex starting O_2 on the *si* face of the intermediate. We note that the A ring of the substrate is twisted in the LCF•LH_2•O_2 simulation relative to the enolate simulations, but the molecular oxygen binding mode is conserved. Protons on the enzyme are omitted for clarity. 84

Figure 4.7 Comparison of (a) (*E,S,R*)-$LHOO^-$, (b) (*E,S,R*)-LHOOH, (c) (*E,S,S*)-$LHOO^-$, and (d) (*E,S,S*)-LHOOH binding to LCF via closest to average structures. The (*E,S,R*) species show remarkably similar binding, facilitated by interactions between the intermediate and the residues Glu1105 and Arg1142. Glu1105 is poised for direct proton transfer to $LHOO^-$ (a), forming LHOOH (b). Protons on the enzyme are omitted for clarity. 86

Figure 4.8 Closest to average structure of the (*E,S,R*)-LOHO− intermediate, highlighting stabilizing interactions with the side chains of Lys1089, Arg1142, and Tyr1168 as well as the backbone of Ser1068. Protons on the enzyme are omitted for clarity. 87

Figure 4.9 Mechanistic proposal for the dinoflagellate bioluminescence reaction. 88

Figure 5.1 Numbering convention used for Fe sites in the broken-symmetry density functional theory (BS-DFT) calculations (a) and isosurface plot of the calculated spin density for the βαβα valence isomer of the crystallographic SAM-bound cluster geometry (b). The β spin density is shown in red and α spin density in blue (surface depicted at ± 0.0035 au after interpolative smoothing). Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.....99

Figure 5.2 Calculated ^{13}C and ^{1}H isotropic hyperfine coupling constants (HFCCs) for SAM near-attack conformer (SAM-NAC) geometries with fixed unique iron−SAM sulfur distances. Calculated values are averaged over all BS valence isomers with $m_S = +1/2$ for each fixed distance, with error bars showing the standard deviation. The average HFCCs for 5′-^{13}C (blue), 5′-C−^{1}H (green), and 4′-^{13}C (black) are shown as circles, and best-fit curves to a decaying exponential equation are shown as lines. The red stars indicate the experimentally determined HFCCs for each curve. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society. 102

Figure 5.3 Relative energy curves for the broken-symmetry (BS) valence isomers of the SAM near-attack conformers (SAM-NACs) as a function of the unique iron−SAM sulfur

distance. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society. 104

Figure 5.4 Proposed structure for the SAM near-attack conformer (SAM-NAC) structure of the Ω intermediate, with calculated HFCCs labeled. Hydrogen atoms (except those on the 5′-C) and coordinated methanethiolate ligands used in the calculations are omitted for clarity. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society. 105

Figure 6.1 Canonical RS and Dph2 enzyme catalysis, showing proposed intermediates for Ω, intermediate **I**, and intermediate **II**. 113

Figure 6.2 Numbering convention used for Fe sites in the broken-symmetry density functional theory (BS-DFT) calculations.....119

Figure 6.3 Comparison of computationally predicted ^{13}C isotropic hyperfine coupling constants for organometallic and SAM-bound [4Fe–4S] cluster models of intermediate I (experimental value 7.8 MHz[4b]). Hydrogen atoms and methanethiol ligands are omitted for clarity.....121

Figure 6.4 Simulated electron paramagnetic resonance (EPR) spectra from the calculated *g*-values and ^{1}H and ^{14}N hyperfine coupling constants (HFCCs) for the Nε protonated model of intermediate II (linewidth = 30 MHz; see Table 1 for HFCCs); plot generated with the EasySpin software package.[29] Inset shows calculated spin density (smoothed at ± 0.0015 au). 123

Figure 6.5 Revised Dph2 catalytic mechanism including a paramagnetic SAM-bound [4Fe–4S] cluster species and the previously proposed Nε protonated His intermediate. 124

Figure B.1 Calculated methyl-^{13}C isotropic hyperfine coupling constants (HFCCs) for SAM near-attack conformer (SAM-NAC) geometries with fixed unique iron–SAM sulfur distances. Calculated values are averaged over all BS valence isomers with $m_s = +1/2$ for each fixed distance with error bars showing the standard deviation. The best fit curve to a decaying exponential equation is shown in black. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society. 136

List of Abbreviations

5'-dAdo·	5'-Deoxyadenosyl Radical
ACP·	2-Amino-2-Carboxypropyl Radical
aMD	Accelerated Molecular Dynamics
ATP	Adenosine Triphosphate
BS-DFT	Broken-Symmetry Density Functional Theory
CG	Conjugate Gradient
CpHaMD	Constant pH Accelerated Molecular Dynamics
CpHMD	Constant pH Molecular Dynamics
Cryo-EM	Cryogenic Electron Microscopy
DCCM	Dynamic Cross-Correlation Mapping
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
EF2	Elongation Factor 2
ENDOR	Electron-Nuclear Double Resonance
EPR	Electron Paramagnetic Resonance
FMN	Flavin Mononucleotide
GB	Generalized Born
GPU	Graphics Processing Unit
HAA	Hydrogen Atom Abstraction
HAB	Harmful Algal Bloom
HFCC	Hyperfine Coupling Constant
KAM	Lysine 2,3-Aminomutase (current)

LAM	Lysine 2,3-Aminomutase (dated)
LBP	Luciferin-binding Protein
LCF	Dinoflagellate Luciferase
LH_2	Dinoflagellate Luciferin
LH^-	Enolate Intermediate in Luciferin Reaction
LO	Oxyluciferin Product in Luciferin Reaction
$LOHO^-$	*gem*-Diolate Intermediate in Luciferin Reaction
LOO^-	Peroxy Intermediate in Luciferin Reaction
LOOH	Hydroperoxy Intermediate in Luciferin Reaction
MCR	Methyl Coenzyme M Reductase
MD	Molecular Dynamics
NAC	Near-Attack Conformer
NMR	Nuclear Magnetic Resonance
OEC	Oxygen Evolving Complex
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PCET	Proton-Coupled Electron Transfer
PES	Potential Energy Surface
PFL-AE	Pyruvate Formate-Lyase Activating Enzyme
PLP	Pyridoxal 5'-Phosphate
QM/MM	Quantum Mechanics/Molecular Mechanics
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid

RS Radical *S*-adenosyl-L-methionine

SAM *S*-adenosyl-L-methionine

SAM-NAC SAM Near-Attack Conformer

SCF Self-consistent Field

SD Steepest Descent

SPL Spore Photoproduct Lyase

TIM Triose Phosphate Isomerase

Chapter 1: Introduction

1.1 Modern Enzymology

The field of enzymology is concerned with explaining the catalytic behavior of enzymes: biological polypeptides generally built from the twenty common, naturally occurring amino acids. Enzymes are essential for life, enhancing many necessary chemical reactions to time scales compatible with biological processes. Enzymes are a type of protein, therefore the product of ribonucleic acid (RNA) translation following transcription from deoxyribonucleic acid (DNA). Because enzymes are comprised only of amino acids, the native chemistry they can catalyze is limited to acid-base and nucleophilic chemistry. As a result, many enzymes employ compounds called cofactors in order to aid catalysis and expand their catalytic scope. In particular, many enzymes use metallocofactors for either catalysis or electron transfer. The two enzymatic systems studied in this thesis are notable for the cofactor usage: one is able to catalyze bioluminescence without any cofactor, while the other utilizes an iron-sulfur cluster to initiate radical chemistry.

Because enzymes are one of the end products of the flow of genetic information, the advancement of genetic technology during the 20th century and into the 21st century has provided enzymologists with an ever-increasing abundance of genomic data and the ability to efficiently conduct heterologous expression and purification of enzymes. Thus, the number of enzymes accessible for laboratory experiments has grown explosively since the mid- to late-1900s. Modern enzymology then employs a wide array of techniques to tease apart and understand how each enzyme functions and probes the

specific mechanisms by which they catalyze chemical reactions. Technological advancements have greatly contributed to these techniques as well, from developing advanced spectroscopic methods to high-throughput screening methods.

Included in the suite of techniques used by enzymologists are computational methods, ranging from dynamic simulations of enzymatic systems to fitting and interpreting experimentally-obtained spectra. Protein structures obtained via X-ray crystallography, nuclear magnetic resonance (NMR), or cryogenic electron microscopy (cryo-EM) can be simulated with classical physical equations to turn static structures into dynamic pictures of enzyme motion. Predicted spectra from quantum mechanical models of enzyme active sites can be compared to experiment to determine accurate reaction intermediates and transition states in catalysis. Simulations combining classical methods with quantum chemistry—so-called multiscale modeling or quantum mechanics/molecular mechanics (QM/MM)—let the enzymologist computationally study the entire enzyme as it facilitates the making and breaking of chemical bonds. While still subject to inherent limits such as computational time and system complexity, computational enzymology has developed into a critical component of modern biochemistry.

This dissertation seeks to apply computational methods to study two diverse enzymatic systems in order to understand the mechanisms of each. The first system is dinoflagellate luciferase (LCF), a bioluminescent enzyme that catalyzes the conversion of a luciferin substrate and molecular oxygen into an oxyluciferin product, water, and blue light. The second system is the common catalytic machinery found in the radical *S*-adenosyl-L-methionine (SAM; RS) enzyme superfamily. RS enzymes are one of nature's

largest, most diverse enzyme superfamilies, yet all members initiate catalysis in the same way: reductive cleavage of a SAM molecule bound to an iron-sulfur cluster. The remainder of this chapter will outline the two systems of study.

1.2 Dinoflagellate Luciferase

Dinoflagellates are eukaryotic microorganisms, found in both marine and freshwater habitats,[1,2] and are the primary source of bioluminescence in the sea.[1] Many species are also capable of photosynthesis, with the ability to either absorb or emit light governed by a circadian rhythm.[3] Dinoflagellate bioluminescence was originally described as the 'phosphorescence of the sea'[4] and gives rise to brilliant displays of blue light, as can be found in the eponymous bioluminescent bays of Puerto Rico. While bioluminescence has convergently evolved in several species,[5] dinoflagellates are thought to use bioluminescence as a clever defense mechanism. The 'burger alarm hypothesis'[6] relies on the fact that dinoflagellate bioluminescence is triggered by physical agitation; when the water around the organism is disturbed, a chain of cellular events (*vide infra*) ultimately results in the production of light. Therefore, when a predator approaches the organism and causes it to glow, the water is illuminated such that the predator's predator can now see, resulting in the chance for the predator's predator to eat the dinoflagellate's predator and keep the dinoflagellate itself safe.

Utilizing a more dangerous defense mechanism, many species of dinoflagellate also produce paralytic toxins, including saxitoxin, brevetoxin, and yessetoxin.[7] When dinoflagellates grow unchecked as a harmful algal bloom (HAB), a phenomenon known as a red tide, toxins reach levels deadly to shellfish and harmful to humans.[8] In addition

to the health risk, red tides also cause significant economic damage; for example, in 1971 an outbreak in Florida was estimated to cause ~$20 million in damages to the tourist industry alone.[8] Understanding dinoflagellate biochemistry provides a possible avenue towards remediation of red tides: if inhibitors specific to dinoflagellate metabolic pathways can be developed, then they could be used to prevent or slow red tides while remaining safe for the surrounding ecosystem. The pathway biosynthesizing the substrate of the bioluminescent reaction and the bioluminescent reaction itself are therefore prime candidates for the development of dinoflagellate-targeted compounds. The pathway is related to the organisms' ability to photosynthesize, which could improve the efficacy of any compounds used to this end, as it would potentially affect both parts of the dinoflagellate's circadian rhythm.

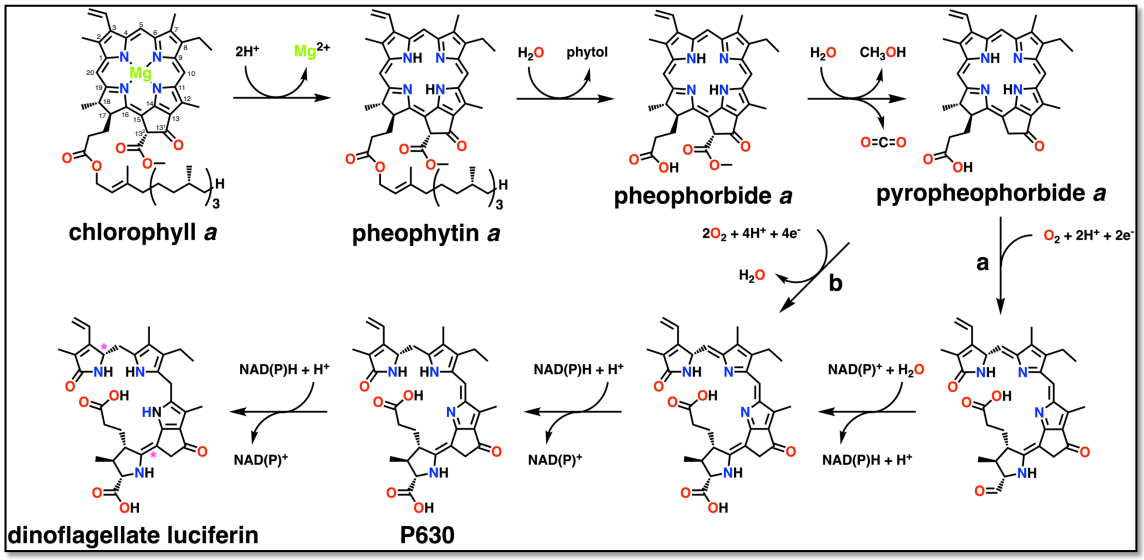

Figure 1.1 Hypothesized biosynthetic pathway of dinoflagellate luciferin with enzymes omitted. Chlorophyll *a* is shown with carbon numeration standard for tetrapyrroles, and dinoflagellate luciferin is shown with magenta asterisks at the unresolved stereocenters. For illustration, the C4 chiral carbon is shown in the *S* configuration, while the C15 to C16 double bond is shown in the *E* configuration.

The biochemical connection between photosynthesis and bioluminescence is found in the aforementioned luciferin substrate—an open chain tetrapyrrole hypothesized to be a metabolite of chlorophyll *a* (Figure 1.1).[9] The metabolism of chlorophyll *a* to dinoflagellate luciferin (LH_2) is proposed to involve several steps, including de-chelation of Mg^{2+}, cleavage of the phytol tail, removal of the methyl ester at $C13^2$, oxidative ring opening, and finally reduction of two double bonds (C5 to C6 and C10 to C11); it is therefore hypothesized that at least five enzymes should be present in the pathway. Importantly, metabolism of chlorophyll *a* to dinoflagellate luciferin requires the introduction of two stereocenters: a chiral carbon at C4 and a double bond between C15 and C16. Neither the *R*/*S* stereochemistry at C4 nor the *E*/*Z* stereochemistry at C15=C16 in dinoflagellate luciferin has been definitively assigned via nuclear magnetic resonance (NMR) spectroscopy.[9] It is unknown which stereoisomer(s) occurs naturally, and therefore the stereo- and regiochemical course of the bioluminescent reaction is unknown.

However, stereocenters notwithstanding, the overall reaction is known and is shown (with other redox reactions involving dinoflagellate luciferin) in Figure 1.3. Dinoflagellate bioluminescence is unique among known bioluminescent reactions as it is 'simple' bioluminescence. Only the enzyme, substrate, and electron acceptor (molecular oxygen) are required, though most species do have a regulatory luciferin-binding protein (LBP).[10] Many other bioluminescent reactions require cofactors such as adenosine triphosphate (ATP) in firefly bioluminescence,[11] flavin mononucleotide (FMN) in bacterial bioluminescence,[12] or another protein (i.e., a green fluorescence protein) to enhance overall light production in colenterazine bioluminescence[13] (structures shown in

Figure 1.2). Like dinoflagellate bioluminescence, fungal bioluminescence requires only the enzyme, substrate, and molecular oxygen, but its reaction results in the decarboxylation of the substrate,[14] whereas the dinoflagellate bioluminescent reaction does not. Fungal bioluminescence is a known dioxygenation reaction; whether

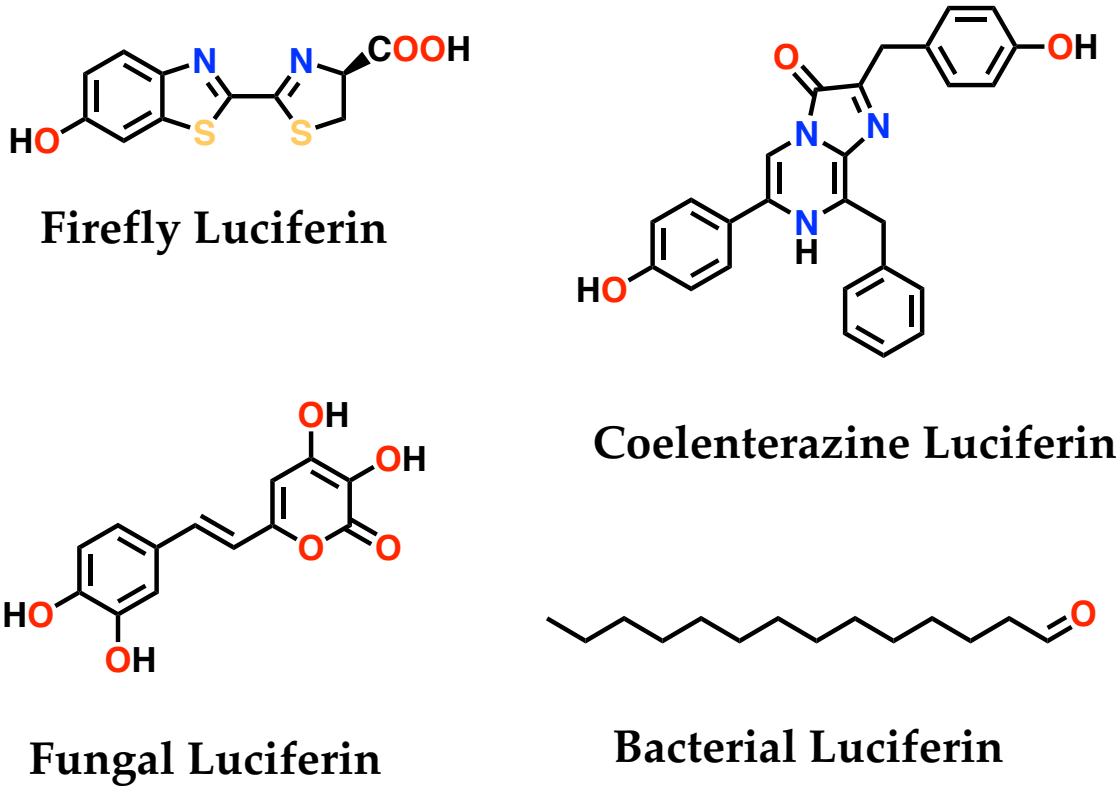

Figure 1.2 Example chemical structures of other known luciferins discussed in the text.

dinoflagellate LCF has mono- or dioxygenase activity is unknown. Thus, how the dinoflagellate bioluminescence system is able to generate a high-energy intermediate capable of relaxing by high-yield emission of light, without decarboxylation, is of fundamental chemical interest. Furthermore, understanding its chemistry is critical for any prospect of developing the dinoflagellate luciferase system for use as an *in vivo* imaging agent, similar to how the firefly bioluminescence system has been extensively

developed.[15] Since nonenzymatic, non-bioluminescent oxidation of dinoflagellate luciferin is known to occur in the presence of air,[9] understanding the enzymatic reaction is especially important for any application of the dinoflagellate luciferase system.

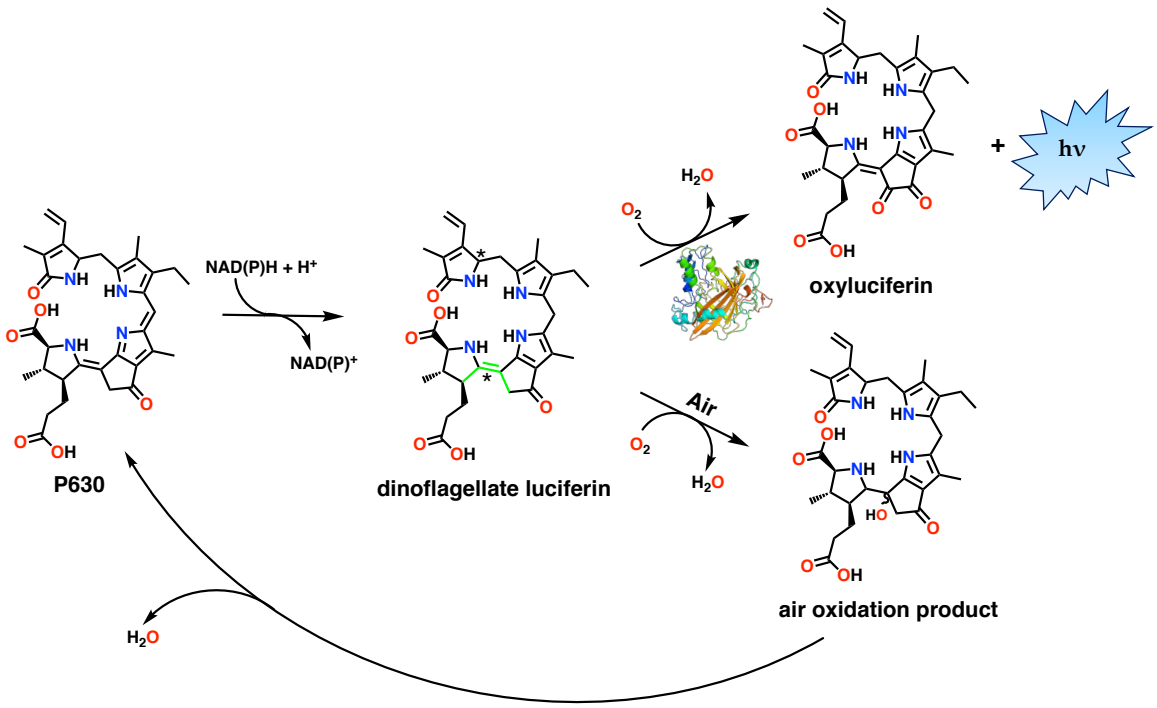

Figure 1.3 Overview of the reduction and oxidation reactions involving dinoflagellate luciferin. The proposed P630 precursor is reduced to make dinoflagellate luciferin (unknown stereocenters marked with an asterisk; the double bond unknown to be in the *E* or *Z* configuration additionally colored lime green). Dinoflagellate luciferin can be oxidized enzymatically to make oxyluciferin and light, or it can be nonenzymatically air-oxidized at C15. In principle, P630 could be regenerated from the air oxidation product via dehydration.

In addition to catalyzing 'simple' bioluminescence, dinoflagellate luciferase is novel due to its pH activation mechanism. LCF is regulated by pH, inactive at pH ~8 and optimally active at pH ~6.[16] LCF, along with LH_2 and LBP, is housed in a specialized organelle termed a scintillon, which is contiguous with the acidic vacuole.[10] The cellular mechanism by which the organism regulates the pH upon physical agitation is thought to operate as follows.[17-20] When shear force is applied to the cellular membrane of the

organism, G-protein coupled receptors in the membrane initiate a signal transduction cascade. While the specifics of the cascade are unknown, it ultimately leads to an increase in calcium levels in the cell. The increase in ion concentration leads to the opening of voltage-gated channels in the vacuole. Upon opening of the channels, the scintillon is flooded with protons, rapidly dropping the pH. The pH drop is hypothesized to both release LH_2 from LBP and activate LCF for catalysis.

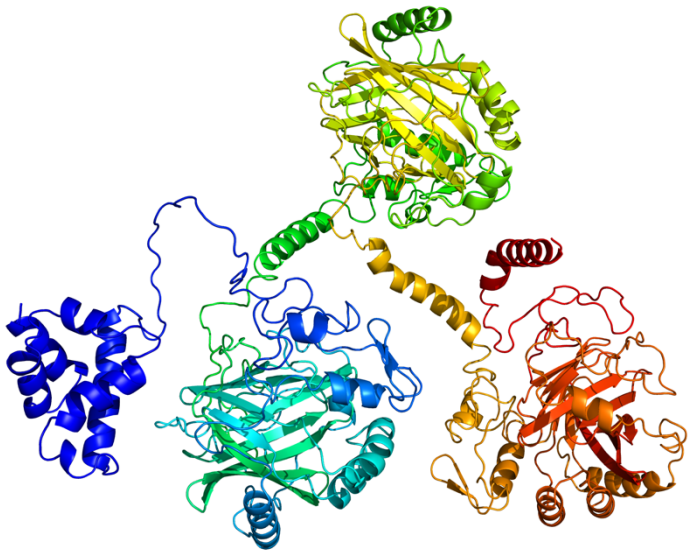

Figure 1.4 AlphaFold[21,22] predicted structure of full-length dinoflagellate luciferase.

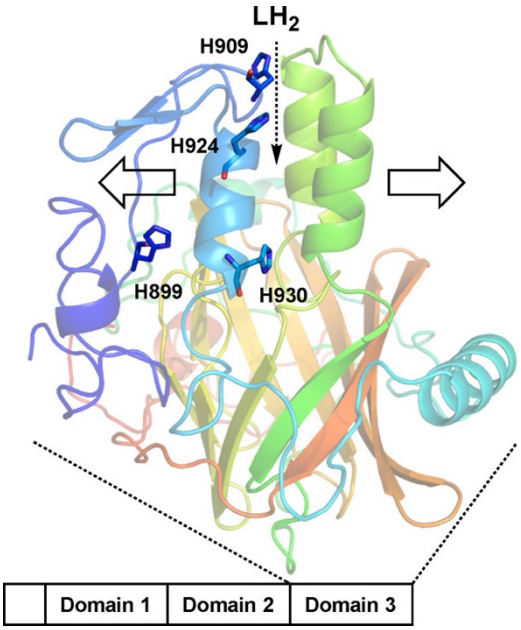

Figure 1.5 The only solved crystal structure of LCF, a pH 8 inactive structure of domain 3 (PDB ID: 1VPR). Four intramolecularly conserved histidine residues, previously experimentally implicated in the conformational shift, are shown as sticks. Reprinted with permission from *Biochemistry* **2018**, *57 (3)*, 295–299. Copyright 2018 American Chemical Society.

How LCF is activated when subject to the pH drop is an active research question. LCF is comprised of an N-terminal domain of unknown function and three homologous, independently catalytic domains (Figure 1.4).[23] Previous research conducted by Hastings and coworkers established a correlation between enzyme activity and the protonation of four N-terminal, intramolecularly conserved histidine residues in each catalytic domain LCF (Figure 1.5).[16] By creating constructs of LCF lacking the N-terminal region and, therefore, three of the His residues, Li *et al.* were able to show a relative increase in the activity of the protein at alkaline pH.[16] Since the His residues would be expected to change protonation state from acidic to alkaline pH, they were postulated to drive the conformational shift of the enzyme. Hastings and coworkers were able to obtain the only crystal structure (to date) of any part of LCF: domain 3 at the inactive pH of 8.[24] While the authors performed a 10 ns molecular dynamics (MD) simulation to explore any

potential conformational changes of the enzyme, no conclusive, large-scale motions were observed. Therefore, an active, open conformer of the enzyme—necessary for studying how LCF catalyzes the bioluminescent reaction—was not determined experimentally or computationally.

To that end, this dissertation uses computational methods to further understand the dinoflagellate bioluminescence system. First, by using advanced sampling molecular dynamics methods, it seeks to determine a potential open, active conformer of the enzyme. The open conformer will then be used to study substrate, intermediate, and product interactions between said ligands and the enzyme using docking and molecular dynamics methods. Because each possible stereoisomer of LH_2 can be 'prepared' and placed into the enzyme active site, a likely stereo- and regiochemical course of the bioluminescent reaction will be predicted based on observed feasibility.

1.3 Radical SAM Enzymes

Radical SAM enzymes were first identified as an enzyme superfamily in 2001, representing one of the earliest successes of genomic enzymology.[25] Genomic enzymology seeks to leverage the ever-increasing datasets of organismal genomes to identify and characterize enzyme families and superfamilies. While little experimental work on radical SAM enzymes had been conducted to that point—primarily that of Perry Frey's group on lysine 2,3-aminomutase (KAM) dating back to the 1980s[26]—the seminal 2001 paper by Sofia *et al.*[25] showed that RS enzymes are not only present but abundant in all kingdoms of life, which has since led to an explosion of research in RS enzymology over the past twenty years and counting.

Radical SAM enzyme are defined as a superfamily by two structural components necessary for the catalytic mechanism.[27-29] The first component is a partial triose phosphate isomerase (TIM) barrel which allows the enzyme to bind *S*-adenosyl-L-methionine. The second is a CX_3CX_2C motif binding a [4Fe4S] cluster, with three irons coordinated by the conserved cysteines and one iron uncoordinated (called the unique iron). SAM can then bidentately coordinate to the unique iron through its carboxy and amide moieties (Figure 1.4). Upon reduction, the cluster can transfer a single electron to SAM resulting in its reductively cleavage, resulting in the 5'-deoxyadenosyl radical (5'-dAdo·), one of nature's most powerful oxidants. In the last common step of RS catalysis, the 5'-dAdo· then abstracts a hydrogen atom from the substrate.

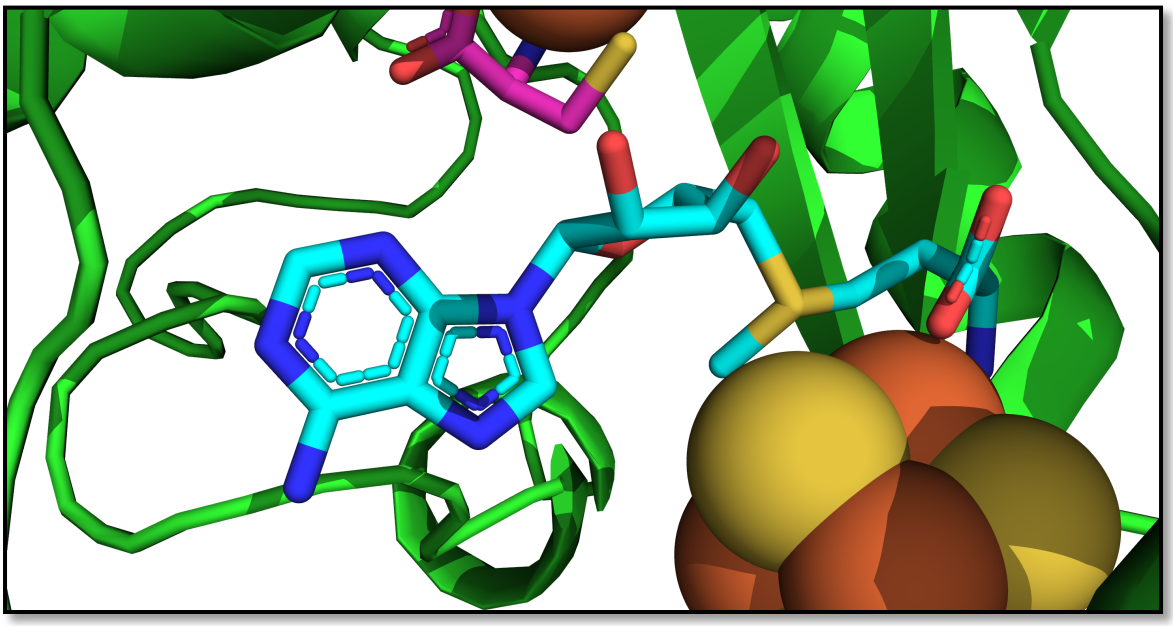

Figure 1.6 Crystal structure of the active site of the radical SAM enzyme HydE (PDB ID: 7O25).[30] SAM and the substrate are shown in cyan and magenta sticks, respectively, and the iron sulfur cluster is shown as spheres.

Following hydrogen atom abstraction (HAA), a wide variety of chemistries can be found in the superfamily, including C-H bond functionalization and complex skeletal

rearrangements.[27-29] While not the topic of the research presented in this dissertation, which focuses on the common catalytic mechanism, Figure 1.5 presents a sequence similarity network (SSN)[31] of the twenty characterized (colored) and twenty-two uncharacterized (uncolored) subgroups of radical SAM enzymes,[32,33] demonstrating the functional diversity of this superfamily. Because such a wide array of enzymes is contained in the superfamily, understanding the common catalytic mechanism is key to understanding the superfamily as a whole.

Several example radical SAM enzymes are discussed below; since the superfamily contains several hundred thousand members only a handful of chemistries will be discussed. Methanogenesis marker protein 10 is a cobalamin-dependent radical SAM methyltransferase that methylates Cδ on Arg285 in the α-subunit of methyl-coenzyme M reductase (MCR), the enzyme which catalyzes the final step of methanogenesis.[34] Other radical SAM methylases, such as ChuW, are instead able to break carbon-carbon bonds by utilizing two molecules of SAM and forming a SAM-adducted intermediate.[35] KAM is a radical SAM enzyme utilizing an additional cofactor, pyridoxal 5'-phosphate (PLP), to achieve isomerization between lysine and β-lysine.[26] ThiC, involved in the thiamine pyrophosphate biosynthesis pathway in bacteria and plants, achieves an incredibly complex skeletal rearrangement to convert aminoimidazole ribonucleotide to hydroxymethylpyrimidine phosphate.[36] Finally, in one of the most straightforward radical SAM reactions, pyruvate formate-lyase activating enzyme (PFL-AE) uses radical SAM chemistry to active another enzyme: pyruvate formate-lyase, a glycyl radical enzyme.[27] For further discussion, the interested reader is further referred to refs. 27-29, comprising several recent reviews and perspectives on radical SAM enzymes.

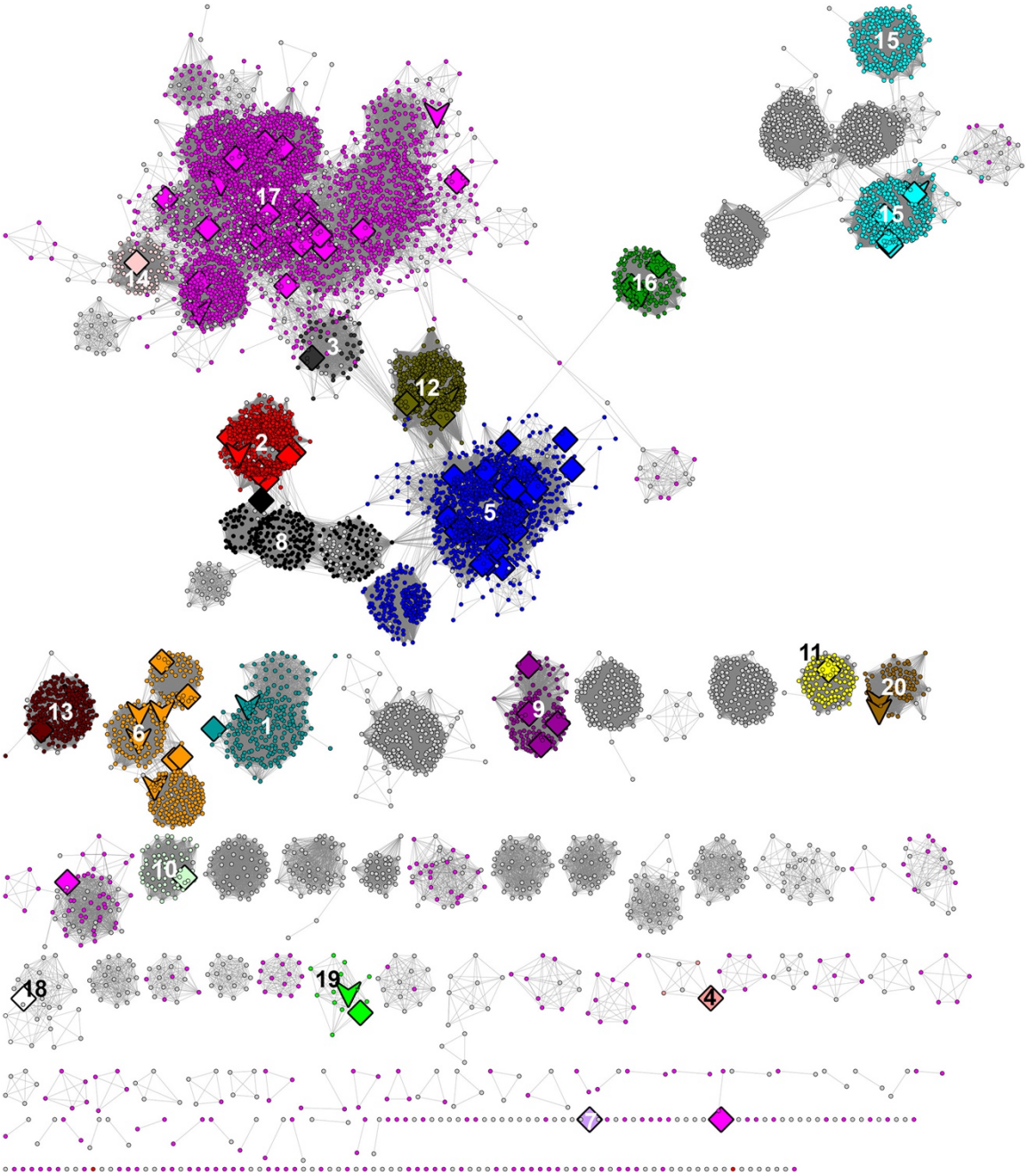

Figure 1.7 Sequence similarity network (SSN) for the radical SAM superfamily, showing characterized (colored) and uncharacterized (uncolored) subgroups. Reproduced with permission from *Methods Enzymol.* **2018**, *606*, 1–71. Copyright 2018 Elsevier.

1.3.1 Discovery of a Paramagnetic Intermediate

Recently, a paramagnetic intermediate was observed common in all tested RS enzymes.[37] First observed in PFL-AE, the intermediate did not appear to be a predicted

pathway intermediate, yet it was found to be catalytically competent.[38] Furthermore, additional experiments found the presence of the intermediate in several more members of the RS superfamily, on which basis Broderick, Hoffman, and coworkers have proposed it to be a universal intermediate in RS catalysis.[37, 39]

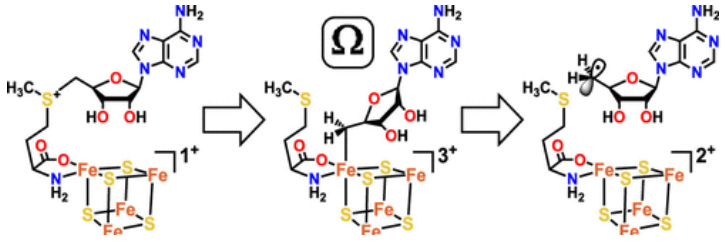

Figure 1.8 Reductive SAM cleavage reaction catalyzed by members of the radical SAM superfamily. The proposed organometallic structure of the Ω intermediate is shown. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.

On the basis of electron paramagnetic resonance (EPR) and electron-nuclear double resonance (ENDOR) spectroscopic experiments, the intermediate was proposed to contain a carbon-iron bond from the 5' carbon of 5'-deoxyadenosine to the unique iron of the cluster and was named Ω (Figure 1.6). The Ω intermediate was therefore proposed to form between initial SAM cleavage and substrate hydrogen abstraction, creating an organometallic species analogous to that of adenosylcobalamin.[39] However, the structure of the intermediate was determined on the basis of EPR and ENDOR spectroscopic experiments alone and was not subject to validation by any other experimental or computational methods. Computational models provide a powerful tool for verifying that proposed species are, in fact, consistent with the observed hyperfine coupling constant (HFCC) values observed in EPR and ENDOR experiments. To that end, this dissertation presents a computational analysis of several possible models of the Ω intermediate in order to determine which model best matches the experimentally observed HFCC values.

In the noncanonical RS enzyme Dph2, SAM is cleaved to form the 2-amino-2-carboxypropyl radical (ACP·), which is then added to a histidine residue on elongation factor 2 (EF2).[40] Again on the basis of EPR and ENDOR spectroscopic experiments, an organometallic intermediate was proposed to occur along the reaction pathway.[41] Similarly to the canonical RS case, this dissertation applies computational methods to study proposed intermediates and discerns which models of intermediates match or disagree with the experimentally determined EPR values for intermediates on the Dph2 reaction pathway.

1.4 Outline of Following Chapters

The remainder of the dissertation is organized as follows, as displayed in Figure 1.7. Chapter 2 provides a brief overview of the computational methods used in the research reported in this dissertation. Chapter 3 reports a constant pH accelerated molecular dynamics (CpHaMD) study of dinoflagellate luciferase. The CpHaMD obtained structure is analyzed and shown to differ from the known crystal structure, providing an open conformer of the enzyme capable of binding substrate. Chapter 4 reports molecular dynamics simulations of dinoflagellate luciferase with its substrate(s), product(s), and reaction intermediates bound. The interactions between the enzymes and its ligands are used to predict the stereo- and regiochemical course of the reaction. Chapter 5 pivots to the research on radical SAM enzymes and reports a broken-symmetry density functional theoretic (BS-DFT) study on proposed models of a catalytic intermediate found in all tested radical SAM enzymes. By comparing computationally derived spectra of

models of the intermediate to known experimental spectra, a revised mechanism of RS catalysis—which does not include an organometallic intermediate—is proposed. Chapter 6 extends the methodology of Chapter 5 to a non-canonical radical SAM enzyme, Dph2, and similarly examines the structure of paramagnetic intermediates in its catalysis. Again, the results prompt a revision of the proposed mechanism not including an organometallic species in catalysis. Chapter 7 provides a brief summary and concluding remarks.

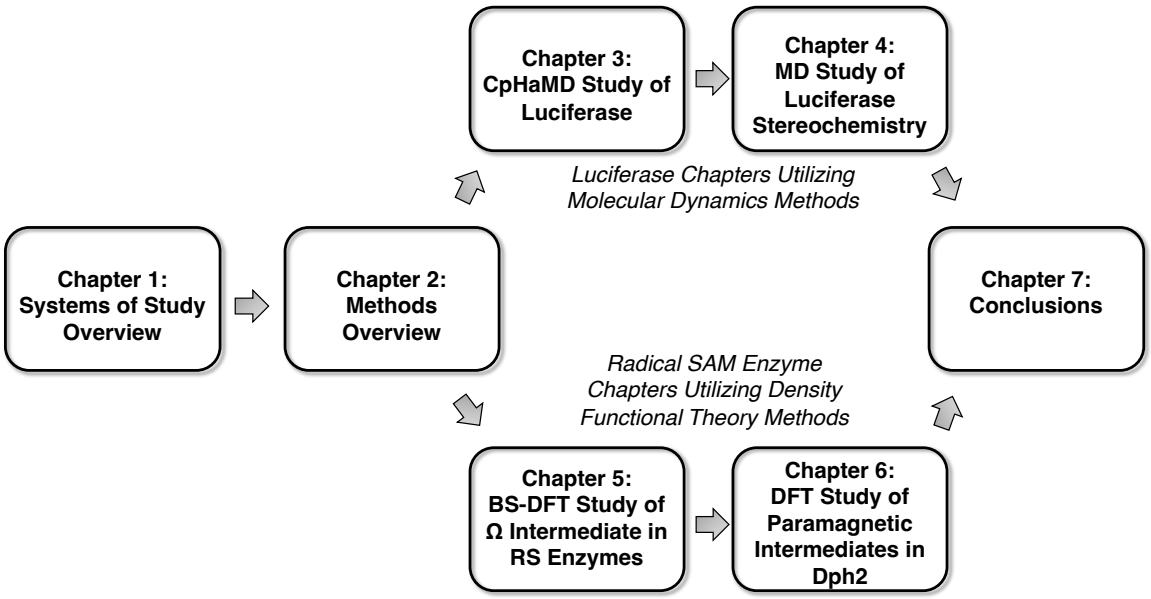

Figure 1.9 Visual overview of dissertation chapters and their organization, as well as methods employed in each.

1.5 References

1. Haddock, S. H.; Moline, M. A.; Case, J. F. Bioluminescence in the Sea. *Annu. Rev. Mar. Sci.* **2010**, *2*, 443–493.

2. Carty, S.; Parrow, M. W. Chapter 17 – Dinoflagellates. In *Aquatic Ecology: Freshwater Algae of North America 2nd ed.*; Wehr, J. D., Sheath, R. G., Kociolek, J. P., Eds.; Academic Press: New York, 2014; pp 773–807.

3. Morse, D.; Milos, P. M.; Roux, E.; Hastings, J. W. Circadian regulation of bioluminescence in Gonyaulax involves translational control. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86 (1)*, 172–176.

4. Clark, R. E. D. Phosphorescence of the Sea. *Nature* **1937**, *139*, 592.

5. Wilson, T.; Hastings, J. W. *Bioluminescence: Living Lights, Lights for Living*; Harvard University Press: Cambridge, Massachusetts, 2013.

6. Abrahams, M. V.; Townsend, L. D. Bioluminescence in Dinoflagellates: A Test of the Burglar Alarm Hypothesis. *Ecology* **1993**, *74 (1)*, 258–260.

7. Wang, D.-Z. Neurotoxins from Marine Dinoflagellates: A Brief Review. *Mar. Drugs* **2008**, *6 (2)*, 349–371.

8. Martin, D. F.; Martin, B. B. Red Tide, Red Terror. *J. Chem. Educ.* **1976**, *53 (10)*, 614–617.

9. Nakamura, H.; Kishi, Y.; Shimomura, O.; Morse, D.; Hastings, J. W. Structure of dinoflagellate luciferin and its enzymic and nonenzymic air-oxidation products. *J. Am. Chem. Soc.* **1989**, *111*, 7607–7611.

10. Morse, D.; Pappenheimer Jr., A. M.; Hastings, J. W. Role of a luciferin-binding protein in the circadian bioluminescent reaction of *Gonyaulax polyedra*. *J. Biol. Chem.* **1989**, *264*, 11822–11826.

11. Vieira, J.; Pinto da Silva, L.; Esteves da Silva, J. C. Advances in the knowledge of light emission by firefly luciferin and oxyluciferin. *J. Photochem. Photobiol. B, Biol.* **2012**, *117*, 33–39.

12. Giuliani, G.; Melaccio, F.; Gozem, S.; Cappelli, A.; Olivucci, M. QM/MM Investigation of the Spectroscopic Properties of the Fluorophore of Bacterial Luciferase. *J. Chem. Theory Comput.* **2021**, *17*, 605–613.

13. Ward, W. W.; Cormier, M. J. In vitro energy transfer in *Renilla* bioluminescence. *J. Phys. Chem.* **1976**, *80 (20)*, 2289–2291.

14. Kaskova, Z. M.; Dörr, F. A.; Petushkov, V. N.; Purtov, K. V.; Tsarkova, A. S.; Rodionova, N. S.; Mineev, K. S.; Guglya, E. B.; Kotlobay, A.; Baleeva, N. S.; Baranov, M. S.; Arseniev, A. S.; Gitelson, J. I.; Lukyanov, S.; Suzuki, Y.; Kanie, S.; Pinto, E.; Di Mascio, P.; Waldenmaier, H. E.; Pereira, T. A.; Carvalho, R. P.; Oliveira, A. G.; Oba, Y.; Bastos, E. L.; Stevani, C. V.; Yampolsky, I. V. Mechanism and color modulation of fungal bioluminescence. *Sci. Adv.* **2017**, *3* (4), e1602847.

15. Liu, S.; Su, Y.; Lin, M. Z.; Ronald, J. R. Brightening up Biology: Advances in Luciferase Systems for *In Vivo* Imaging. *ACS Chem. Biol.* **2021**, *16 (12)*, 2707–2718.

16. Li, L.; Liu, L.; Hong, R.; Robertson, D.; Hastings, J. W. N-Terminal Intramolecularly Conserved Histidines of Three Domains in *Gonylaulax* Luciferase Are

Responsible for Loss of Activity in the Alkaline Region. *Biochemistry* **2001**, *40*, 1844–1849.

17. Valiadi, M.; Iglesias-Rodriguez, D. Understanding Bioluminescence in Dinoflagellates—How Far Have We Come? *Microorganisms* **2013**, *1*, 3–25.

18. von Dassow, P.; I., L. M. The role of Ca^{2+} in stimulated bioluminescence of the dinoflagellate *Lingulodinium polyedrum*. *J. Exp. Biol.* **2002**, *205*, 2971– 2986.

19. Maldonado, E. M.; Latz, M. I. Shear-Stress Dependence of Dinoflagellate Bioluminescence. *Biol. Bull.* **2007**, *212*, 242–249.

20. Chen, A. K.; Latz, M. I.; Sobolewski, P.; Frangos, J. A. Evidence for the role of G-proteins in flow stimulation of dinoflagellate bioluminescence. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2007**, *292*, R2020–R2027.

21. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

22. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679−682.

23. Liu, L., Wilson, T., and Hastings, J. W. Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16555−16560.

24. Schultz, L. W., Liu, L., Cegielski, M., and Hastings, J. W. Crystal structure of a pH-regulated luciferase catalyzing the bioluminescent oxidation of an open tetrapyrrole *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1378−1383.

25. Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **2001**, *29 (5)*, 1097−1106.

26. Baraniak, J.; Moss, M. L.; Frey, P. A. Lysine 2,3-Aminomutase. *J. Biol. Chem.* **1989**, *264 (3)*, 1357−1360.

27. Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical *S*-adenosylmethonine Enzymes. *Chem. Rev.* **2014**, *114 (8)*, 4229−4317.

28. Bridwell-Rabb, J.; Grell, T. A. J.; Drennan, C. L. A Rich Man, Poor Man Story of *S*-Adenosylmethionine and Cobalamin Revisited. *Annu. Rev. Biochem.* **2018**, *87 (1)*, 555−584.

29. Nicolet, Y. Structure−Function relationships of radical SAM enzymes. *Nature Catalysis* **2020**, *3*, 337−350.

30. Rohac, R.; Martin, L.; Liu, L.; Basu, D.; Tao, L.; Britt, R. D.; Rauchfuss, T. B.; Nicolet, Y. Crystal Structure of the [FeFe]-Hydrogenase Maturase HydE Bound to Complex-B. *J. Am. Chem. Soc.* **2021**, *143 (22)*, 8499−8508.

31. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58 (41)*, 4169–4182.

32. Holliday, G. L.; Akiva, E.; Meng, E. C.; Brown, S. D.; Calhoun, S.; Pieper, U.; Sali, A.; Booker, S. J.; Babbitt, P. C. Atlas of the Radical SAM Superfamily: Divergent Evolution of Function Using a "Plug and Play" Domain. *Methods Enzymol.* **2018**, *606*, 1–71.

33. Oberg, N.; Precord, T. W.; Mitchell, D. A.; Gerlt, J. A. RadicalSAM.org: A Resource to Interpret Sequence-Function Space and Discover New Radical SAM Enzyme Chemistry. *ACS Bio. Med. Chem. Au*, **2022**, *2 (1)*, 22–35.

34. Radle, M. I.; Miller, D. V.; Laremore, T. N.; Booker, S. J. Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical *S*-adenosylmethionine methylase that unexpectedly requires cobalamin. *J. Biol. Chem.* **2019**, *294* (31), P11712-P11725.

35. Brimberry, M. A.; Mathew, L.; Lanzilotta, W. Making and breaking carbon-carbon bonds in class C radical SAM methyltransferases. *J. Inorg. Biochem.* **2022**, *226*, 111636.

36. Chatterjee, A.; Hazra, A. B.; Abdelwahed, S.; Hilmey, D. G.; Begley, T. P. A "Radical Dance" in Thiamin Biosynthesis: Mechanistic Analysis of the Bacterial Hydroxymethylpyrimidine Phosphate Synthase. *Angew. Chem. Int. Ed.* **2010**, *49*, 8653−8656.

37. Byer, A. S.; Yang, H.; McDaniel, E. C.; Kathiresan, V.; Impano, S.; Pagnier, A.; Watts, H.; Denler, C.; Vagstad, A. L.; Piel, J.; Duschene, K. S.; Shepard, E. M.; Shields, T. P.; Scott, L. G.; Lilla, E. A.; Yokoyama, K.; Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Paradigm Shift for Radical-*S*-Adenosyl-L-Methionine Reactions: The Organometallic Intermediate Ω is Central to Catalysis. *J. Am. Chem. Soc.* **2018**, *140 (28)*, 8634−8638.

38. Horitani, M.; Shisler, K.; Broderick, W. E.; Hutcheson, R. U.; Duschene, K. S.; Marts, A. R.; Hoffman, B. M.; Broderick, J. B. Radical SAM Catalysis via an organometallic intermediate with an Fe- [5′-C]-deoxyadenosyl bond. *Science* **2016**, *352 (6287)*, 822−825.

39. Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Mechanism of Radical Initiation in the Radical-*S*-Adenosyl-L-Methionine Superfamily. *Acc. Chem. Res.* **2018**, *51 (11)*, 2611−2619.

40. Dong, M.; Zhang, Y.; Lin, H. Noncanonical Radical SAM Enzyme Chemistry Learned from Diphthamide Biosynthesis. *Biochemistry* **2018**, *57 (25)*, 3454–3459.

41. Dong, M.; Kathiresan, V.; Fenwick, M. K.; Torelli, A. T.; Zhang, Y.; Caranto, J. D.; Dzikovski, B.; Sharma, A.; Lancaster, K. M.; Freed, J. H.; Ealick, S. E.; Hoffman, B. M.; Lin, H. Organometallic and radical intermediates reveal mechanism of diphthamide biosynthesis. *Science* **2018**, *359 (6381)*, 1247–1250.

Chapter 2: Computational Methods

This chapter provides an overview of the computational methods employed in this dissertation. While many superb resources exist for understanding both classical and quantum chemical modeling,[1-5] this chapter seeks to provide a useful summary at a level such that any biochemistry graduate student may hopefully benefit from the presentation of the methods. Further references to the techniques discussed are presented in the methods section of Chapters Three, Four, Five, and Six.

2.1 Molecular Dynamics

Molecular dynamics (MD) refers to computational methods which simulate systems using Newton's laws of motion, i.e., 'classical' physics equations, fundamentally based around Newton's Second Law,

$$\vec{F} = m\vec{a} \tag{2.1}$$

where *F* is the force on an object, *m* is the mass of the object, and *a* is the acceleration of the object. MD can be used to study a wide variety of systems, from exotic plasmas[6] to biomolecules.[7] Of particular interest to biochemists, MD can be used to simulate the behavior of proteins, under the limitation that no chemical bonds are created or broken. To a chemist, this may seem a severe defect in the method, but in practice MD provides the enzymologist with the ability to study large- and small-scale conformational changes, solvent effects, allostery, protein-ligand interactions (including substrates, intermediates, products, and inhibitors), and many other enzyme effects. If one wants to study the formation or breaking of chemical bonds in an enzyme active site, MD provides the

foundation for modeling the majority the enzyme in so-called quantum mechanics/molecular mechanics (QM/MM) studies (*vide infra*).

2.1.1 Force fields

MD methods are fundamentally based around solving Newton's second law of motion (Equation 2.1) in clever and efficient ways that allow for simulations of biomolecules on nanosecond to millisecond timescales, i.e., timescales over which chemically and biologically meaningful dynamics can be observed. The most common way to achieve this is through the use of a force field, which determines how atoms in the simulation move. One can relate the force on an object with the spatial change in its potential energy by

$$\vec{F} = -\vec{\nabla}U \tag{2.2}$$

where the ∇ is the gradient operator. For example, for an object released in the air at a variable distance *z* above the ground, the potential energy is determined by gravity and given by

$$U_g = mgz \tag{2.3}$$

where *m* is the mass of the object and *g* is the Earth's gravitational constant. Thus, the force on the object is

$$\vec{F_g} = -\vec{\nabla}mgz = -\frac{d}{dz}\hat{z}\ mgz = -mg\hat{z} \tag{2.4}$$

in the downward direction. Importantly, this simple example shows that the force on a system will act to minimize the potential energy, which is mathematically enforced by the negative sign in Equation 2.2. For any potential energy surface (PES), the force on the

object obeys the same law, meaning that even for a very complicated PES, one can solve the force on an object as long as one can solve the derivatives of the PES.

For enzymes, one can intelligently choose what the form of the PES should be, and therefore one can calculate the force on each atom in the enzyme. Primarily, this involves constraining motion of atoms using chemical knowledge, i.e., the form of the bonded and nonbonded interactions.* Bonded interactions can be modeled with several terms. The first describes the interaction between two bonded atoms as a spring

$$U_{bond}(r) = k(r - r_0)^2 \tag{2.5}$$

where k, the spring constant, is chosen such that it accurately represents the bond oscillation,† and r_0 is chosen to be the equilibrium bond length. Then one can model the three-body bonded term, the bond angles, using the same spring model,

$$U_{angle}(\theta) = k'(\theta - \theta_0)^2 \tag{2.6}$$

where k' is the angle spring constant and θ_0 is the equilibrium angle. The form of the four-body bonded term, called the torsion or dihedral, differs from that of the previous two terms, since the term must be periodic (i.e., a dihedral angle of 0° gives the same structure as a dihedral of 360°). For this reason, the torsional potential energy uses a periodic function, which can be described as a Fourier expansion. However, another dihedral/torsion term is included, called the improper torsion. The improper torsion relies on a different order of selecting the four atoms to generate the two planes used in

* Note that for each of the terms in the force field, a sum over all of the corresponding terms in the system is carried out.

† For example, the k for a C-H bond will be smaller than the k of a C-D bond, since less energy is needed to make a C-H bond oscillate.

calculating the dihedral angle. This potential takes a harmonic form, like the bond distance and angle terms. Together, this results in a term for the torsions of the form

$$U_{torsion}(\omega,\omega') = \sum_n \frac{1}{2}V_n[1 + cos(n\omega - \gamma)] + k''(\omega' - \omega_0)^2 \tag{2.7}$$

where n is the periodicity, V_n is the torsion barrier, ω is the angle between the plane made by bodies one, two, and three with the plane made by bodies two, three, and four, γ is the phase, k'' is the force constant for the harmonic improper term, ω' is the angle between the plane made by bodies two, four, and one with the plane made by bodies two, one, and three, and ω_0 is the equilibrium angle of the harmonic improper term.

A class of interactions give the remaining terms needed to construct a typical MD force field: the non-bonded interactions. Included are the Coulombic ($1/r$) and Lennard-Jones terms (the "12-6" potential), resulting in

$$U_{nonbonded}(i,j) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + A_{ij}\left[\frac{r_{ij}^{12}}{R_{ij}^{12}} - 2\frac{r_{ij}^{6}}{R_{ij}^{6}}\right] \tag{2.8}$$

where i and j are two particles with charges q_i and q_j, respectively, with a distance r_{ij} between them, ε_0 is the permittivity of free space, A_{ij} is the Lennard-Jones well depth, and R_{ij} is the Lennard-Jones equilibrium distance.

As shown, several terms are needed to construct accurate force fields. Yet, by employing chemical knowledge, one can simulation the motion of an enzyme without ever needing to know the exact electronic and nuclear wavefunctions, and therefore simulate the motion of the system on a feasible timescale. Furthermore, force fields can be adapted to the specific system one wants to simulate—proteins,[8] RNA,[9] generalized

biomolecular systems,[10] etc.—allowing for fast, accurate simulations across a wide range of biological systems.

2.1.2 Solvation and Charge

Because enzymatic reactions naturally occur in cellular environments, the effect of solvation must be included when modeling an enzyme with molecular dynamics. There are two primary methods of modeling solvents: implicitly, where a solvent term is added to the force field, or explicitly, where solvent molecules are included in the simulation. Implicit solvent models have the advantage of being much faster than explicit solvent models, but sacrifice accuracy, especially when solvent-enzyme interactions may be important for the system (e.g., ordered water molecules in the active site of the enzyme). While explicit solvent models require more computing time and much more storage memory, they provide a more accurate method of simulating an enzyme system.

Typically, an explicit solvent model will include water molecules and counter-ions used to neutralize the charge of the system. While an implicit solvent model can be used to set a specific salt concentration in the system, this can also be accomplished in an explicit solvent model with the addition of representative ions.

Finally, two other techniques are used in handling the charge of a MD system. First, periodic boundary conditions (PBC) are employed to ensure that the system represents cellular or experimental conditions, creating image shells of the system at its edges. This allows the system, and especially solvent molecules at the edge of the system, to "see" a copy of the system instead of interfacing with vacuum. If a molecule moves through the boundary, it is moved to the opposite side of the system and reintroduced. In this way, solvent molecules remain in the system at the correct density being simulated.

The second technique is the introduction of an interaction cutoff for nonbonded terms. Since the largest nonbonded term decays with $1/r$, by introducing a cutoff where no terms are calculated past the cutoff (for example, ~10 Å cutoff is often employed), computing time is saved as terms with negligible contribution are thus excluded.

2.1.3 Molecular Dynamics Protocol

Most MD simulations begin from either an experimentally determined structure, (i.e., a crystal or cryo-EM structure) or a computationally generated model of the enzyme, which is usually obtained via homology or machine learning methods. While these structures are invaluable sources of information, they are typically determined in noncellular conditions (i.e., high salt concentrations, polyethylene glycol added, etc.).[‡] As a result, to run a MD simulation on a structure, several steps are taken to help ensure the structure is brought to a reasonable set of starting coordinates, which often differs from the original coordinates of the starting structure.

The first step is to create a way for the simulation program to understand how to move the atoms in the enzyme using the chosen force field. This step is called **parameterization**. Typically, any non-biological solvent molecules used to help crystallize the enzyme are removed from the structure, though biological solvent molecules (such as water molecules) are generally kept. At this stage, any potential ligands not included in the structure, such as substrates, products, or cofactors, can be built in modeling software and added to the system. This allows the simulation to include molecules that may be important for function that were not included in the experimental structure, or replace non-native atoms or molecules needed to obtain the structure, such

[‡] In the case of the computationally-generated models, the method (whether homology or machine learning) is still dependent on the body of experimentally-obtained results.

as substrate analogues, non-hydrolyzable ATP, non-native amino acids, or metal centers reconstituted with a non-native metal. Additionally, since very few crystal structures are solved at high enough resolution to include hydrogen atoms, hydrogen atoms are added to the system at this stage. The choice of solvation model is made as this stage, and, if explicit solvent is chosen, water molecules and counterions are added to the system per the user's direction (in terms of the size and shape of the solvation regions surrounding the enzyme). Once the system is prepared as desired, the user can direct the simulation software of choice to parameterize it such that the system can now move according to the chosen force field(s).

The next step is **minimization**. The minimization step is to attempt to ensure that any potential steric clashes are removed and that added atoms and molecules (e.g., the added hydrogen atoms) are in reasonable positions. Minimization uses the PES of the chosen force field, but rather than evolving the system according to Newton's 2nd law, a minimization algorithm is applied to the PES. In the gravity analogy, a stationary ball on top of a hill will remain there until a force acts upon it. However, if its motion was determined by minimization, it would begin to roll down the hill immediately, decreasing its potential energy. The two most common methods for minimization are steepest descent (SD) and conjugate gradient (CG). Briefly, a SD algorithm will always move where the negative gradient of the PES is greatest, but a CG algorithm requires that each step be conjugate[§] to the direction of the previous step. For reasons not discussed here, the CG algorithm generally requires fewer steps to reach a function minimum than SD.

[§] While not mathematically rigorous, for present purposes conjugate simply means a binomial term with sign in the middle changed (e.g., $a - b$ is the conjugate of $a + b$).

Once minimized, the next MD step is **heating**. The heating step takes the minimized system, assumed to be representative of a low temperature, and heats the system to a biologically relevant temperature (for most systems, 300 K is used). The way this is done is putting a 'thermostat' on the system: modifying the equations of motions such that the system is evolving at a certain temperature. For example, a Berendsen thermostat couples the system to a heat bath and then rescales the velocities of the system to ensure they represented the correct Boltzmann distribution for the chosen temperature. More commonly used is the Langevin thermostat,[11] which changes the equations of motion of the system to include a friction (or drag) term as a well as a random kick term that is temperature dependent, serving to simulate the frequency of collisions for a system at the given temperature. Other modifications, such as a barostat to maintain constant pressure, can be applied to the equations of motion depending on the type of statistical ensemble desired. Often, at this stage in the process, the SHAKE algorithm is applied,[12] which keeps bonds between heavy atoms and hydrogen atoms fixed, since these will oscillate on a much faster time scale than heavy atom-heavy atom bonds and require more computing time for the simulation.

Next, the system is subject to **equilibration**. In the equilibration step, the system is briefly simulated just as it will be for the production or main MD calculation. What the equilibration step serves to do is to make sure that the system is not starting a production run in a region of phase space (loosely, coordinates of positions and velocities) too far from equilibrium, which could bias any following analyses. Following equilibrating, production runs[**] of a desired length and with a desired number of replicate runs can be

[**] In this chapter, simulation, calculation, and trajectory are used interchangeably to refer to a MD production run.

conducted. During production runs, various advanced sampling methods can be used to provide trajectories capable of computing a variety of properties not easily accessible by conventional MD (cMD; simply meaning no advanced sampling has been applied to the simulation).

2.1.4 Advanced Sampling Methods in this Dissertation

Brief mention of two advanced sampling methods used in this dissertation, constant pH MD and accelerated MD, is presented here. Constant pH MD (CpHMD)[13] treats the protonation state of residues with titratable side chains as a thermodynamic parameter. This allows for the calculation of a change in free energy associated with protonation/deprotonation, and therefore a probability associated finding each of the two states based on their Boltzmann factors. A Monte Carlo algorithm is applied at a specified number of time steps to each residue that can change protonation states to determine if each keeps or changes its protonation state. Accelerated MD (aMD)[14] applies a biasing potential that decreases the depth of minima in the potential energy surface: when the system is at or below some threshold energy, the biasing potential is applied, helping the system explore more phase space throughout the course of the simulation. Utilizing both at the same time has been developed as constant pH accelerated MD (CpHaMD).[15]

2.1.5 Molecular Dynamics Analysis Techniques

After simulations or sets of trajectories have been obtained, many techniques exist for analysis of the data set. Some commonly applied analysis techniques are discussed below.

Root-Mean-Squared Deviation. Root-mean-squared deviation (RMSD) allows for the comparison of two sets of coordinates and gives a value of the 'distance' between the two structures, using the formula

$$\mathrm{RMSD} = \sqrt{\frac{1}{N_{atoms}} \sum_{i}^{N_{atoms}} (r_i - R_{ref})^2} \tag{2.9}$$

where N_{atoms} is the number of atoms (or could be residues weighted by center of mass, backbone atoms only, etc.) in the simulation, i represents the i^{th} atom in the simulation, r_i is the position of i, and R_{ref} is a reference structure of all the atoms in the simulation for comparison. Notably, since RMSD compares one set of coordinates to another, it is used to compare two static structures to one another. For example, the RMSD between two crystal structures of an enzyme under different conditions, or crystal structures of two enzyme homologs, can be used to help understand structural differences without using dynamic simulations. However, in the case of MD, RMSD is a useful tool, especially when determining if a simulation has converged to an appropriate structure or not. In this case, the reference structure is typically chosen as either the starting coordinates of the simulation or the average coordinates of the simulation. Then, the RMSD for each frame, or time point, of the simulation is computed. If the RMSD converges around a value, one has a good estimate that the simulation is in an expected conformer of the system.

Root-Mean-Squared Fluctuation. Related to RMSD is root-mean-squared fluctuation (RMSF). Instead of comparing two full sets of coordinates, it compares the coordinates of a small part of the system (in the case of enzymes, namely a residue or cofactor) to its positions over the course of the entire trajectory. RMSF is computed by

$$\mathrm{RMSF} = \sqrt{\frac{1}{N_{frames}} \sum_{i}^{N_{frames}} (r_i - R_{average})^2} \qquad (2.10)$$

where N_{frames} is the number of frames in the simulation, i and r_i are as defined above, and $R_{average}$ is the average position for atom i in the simulation. Typically, instead of atoms, mass-weighted average positions of residues will be used. RMSF gives a measure of how dynamic the particle(s) being analyzed is through the simulation. For example, a particle which started at $r = 0$, stayed at $r = 0$ during the simulation, and ended at $r = 0$ would have an RMSF of 0. However, a particle which started at $r = 0$, oscillated between $r = 1$ and $r = -1$ during the simulation, and ended at $r = 0$ would have a nonzero RMSF, even though by comparing the start and end of the trajectory it was identical to the particle that never moved. In this way, RMSF is a useful tool for distinguishing conformationally flexible regions in enzyme MD simulations.

Distance Analysis. A straightforward, but very useful, analytic tool is computing the distance between two sets of coordinates (atoms, residues, side chains, etc.) of interest in a simulation. Distance between two residues can be used to monitor conformation changes of an enzyme, distance between substrate atoms and residue side chains can be used to identify residues potentially important in the catalytic function of the enzyme.

Other Techniques. A variety of other techniques are commonly employed when analyzing the results of molecular dynamics simulations, typically making use of the large amount of data present in a MD simulation in order to identify patterns. Several include principal component analysis (PCA), dynamic cross-correlation mapping (DCCM), and clustering analysis, which can be performed on trajectories or projected trajectories (e.g., on the results of a PCA calculation). In MD, PCA allows for the

identification of which collective motions in a simulation can explain the observed behavior. For example, for an enzyme at equilibrium, one would expect PCA to return the breathing modes of the enzyme as the principal components. DCCM analyses the frames from a trajectory and computes which residues have correlated, anticorrelated, or uncorrelated motion. For example, residues located near one another in secondary structure will often have highly correlated motion (they move in the same direction at the same time), but residues on the opposite side of an enzyme during a conformational change of the enzyme could be expected to have anticorrelated motion (they move in opposite directions at the same time).

2.2 Quantum Chemical Methods

Instead of solving Newton's equations of motions, non-relativistic quantum chemical methods solve either the time-independent Schrödinger equation or the time-dependent Schrödinger equation, with the former being more common. The time-independent Schrödinger equation is given by

$$\hat{H}\Psi = E\Psi \tag{2.11}$$

where H is the Hamiltonian operator, Ψ is the wavefunction, and E is the energy. The Hamiltonian contains the following terms

$$\hat{H} = \hat{T}_{Nuc}(R) + \hat{T}_{Elec}(r) + \hat{V}_{Nuc-Elec}(R,r) + \hat{V}_{Nuc-Nuc}(R) + \hat{V}_{Elec-Elec}(r) \tag{2.12}$$

where T_{Nuc} is the nuclear kinetic energy, T_{Elec} is the electronic kinetic energy, $V_{Nuc\text{-}Elec}$ is the potential energy resulting from interactions between nuclei and electrons, $V_{Nuc\text{-}Nuc}$ is the potential energy between nuclei, $V_{Elec\text{-}Elec}$ is the potential energy between electrons, R contains the nuclear positions, and r contains the electronic positions. For many quantum

chemical systems, the Born-Oppenheimer approximation, where it is assume that the electrons 'move' much faster than the nuclei in the system, is employed. This simplifies the situation such that the nuclear coordinates are assumed to be constant, removing the nuclear kinetic energy term and fixing the nuclear-nuclear potential term to be constant.

2.2.1 Hartree-Fock

One of the earliest methods used to solve the time-independent Schrödinger equation is the Hartree-Fock method, or the self-consistent field (SCF) method. This method allows the user to solve an *N*-body problem, meaning the electronic structure of an atom or molecule with *N* electrons can be solved—which is necessary but not sufficient for almost all biochemical problems. The Hartree-Fock method makes several assumptions, including that the wavefunction can be represented as a single Slater determinant and that a mean-field representation is accurate (electron correlation is not captured by Hartree-Fock). From there, the SCF method works iteratively to determine the wavefunction of a system within a given convergence criterion. Because the method works iteratively, the development of modern computers greatly increased the number of systems that could be studied with the SCF method. Today, many *ab initio* methods that do not make the same assumptions as Hartree-Fock are used in modern quantum chemistry to obtain more accurate results (often referred to collectively as 'post-Hartree' methods); discussion of these methods is beyond the scope of this dissertation.

2.2.2 Density Functional Theory

One of the most widespread methods in modern computational chemistry is density functional theory (DFT). DFT makes use of the two Hohenberg-Kohn theorems, namely, that 1) the external potential of a system of electrons is a unique functional of the

electron density (and therefore the total energy is a unique functional of the electron density), and 2) the lowest energy of the functional corresponds to the ground state density (the ground state energy can be found variationally). While it can be shown that there is one 'true' functional that in principle would always give the correct energy for any electron density, the form of this functional is unknown. Often the functional is split into kinetic, external potential, and exchange-correlation components when it is written:

$$E[\rho] = T[\rho] + V[\rho] + E_{XC}[\rho] \tag{2.13}$$

where ρ is the electron density of the system, and each other term is the respective functional mentioned above.

Many functionals have been developed to try to solve the DFT problem, utilizing approaches that depend only on the density function (local density approximation), approaches that depended on the gradient of the density function (generalized gradient approximation), and even approaches that include a component of Hartree-Fock exchange (hybrid functionals).

2.2.3 Broken-Symmetry Density Functional Theory

Density functional theory behaves very poorly when faced with solving systems with a large number of low-lying states. An important example of this is antiferromagnetically coupled metal clusters displaying low-spin ground states. Two important biological systems that fall into this category are the oxygen evolving complex (OEC) in photosystem II and iron-sulfur clusters found in many metalloenzymes. In order to overcome this limitation, broken-symmetry density functional theory (BS-DFT) was developed beginning in the late 1970s and early 1980s, primarily by L. Noodleman. The broken-symmetry ansatz is that a reasonably accurate electron density can be obtained by

first solving the high-spin wavefunction of the ferromagnetically coupled system. Then, by flipping the spin on certain atoms in the system, intermediate and low-spin wavefunctions whose electron density is a close representation of the physical electron density can be obtained. However, the wavefunctions obtained in BS calculations are not spin eigenfunctions and therefore not physical states, and must be corrected by spin projection in order for comparison to experimental observables. In order to model electron paramagnetic resonance (EPR) and electron-nuclear double resonance (ENDOR) spectra, the spin Hamiltonian is used: it is the spin Hamiltonian that unites broken-symmetry results to physical systems via spin projection. For continued discussion, see Appendix B for the process of determining spin-projection factors linking the phenomenological spin Hamiltonian to the physical Hamiltonian.

2.3 Note on Multiscale Modeling

Finally, this dissertation would be incomplete without mentioning multiscale modeling such as quantum mechanics/molecular mechanics (QM/MM),[1] though no such methods are employed herein. In multiscale modeling, different regions of a system can be treated with different levels of theory; importantly, for enzymes, this means that an enzyme active site can be treated with quantum chemical methods while the rest of the enzyme (and solvent or other molecules present) can be treated with Newton's laws of motion.[16] While the method one uses to include the effects of the classical regions can vary (such as mechanical versus electrostatic embedding), and the atoms or bonds where the treatment of the system changes must be handled correctly, such methods provide powerful tools for simulating the catalytic effect of an enzyme, including on electronic

excited states.[17] Multiscale modeling can also be used to treat one region at a more accurate, but computationally expense, level of quantum theory while treating the region around it still quantum mechanically but at a less computationally intensive level of theory.

2.4 References

1. Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Reading, Massachusetts, 2001.

2. Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons: Hoboken, New Jersey, 2007.

3. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Rev. Ed.; Dover: Mineola, New York, 1996.

4. Neese, F. Prediction of molecular properties and molecular spectroscopy with density functional theory: From fundamental theory to exchange-coupling. *Coord. Chem. Rev.* **2009**, *253*, 526–563.

5. Neese, F.; Atanasov, M.; Bistoni, G.; Maganas, D.; Ye, S. Chemistry and Quantum Mechanics in 2019: Give Us Insight and Numbers. *J. Am. Chem. Soc.* **2019**, *141*, 2814–2824.

6. Pohl, T.; Pattard, T.; Rost, J. M. Kinetic modeling and molecular dynamics simulation of ultracold neutral plasmas including ionic correlations. *Phys. Rev. A* **2004**, *70*, 033416.

7. Hospital, A.; Goñi, J. R.; Orozco, M; Gelpi, J. L. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.

8. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015** *11*, 3696−3713.

9. Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018** *115*, E1346−E1355.

10. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004** *25*, 1157−1174.

11. Xiang, T.-X.; Liu, F.; Grant, D. M. Generalized Langevin equations for molecular dynamics in solution. *J. Chem. Phys.* **1991** *94*, 4463−4471.

12. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977** *23*, 327–341.

13. Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

14. Pierce, L. C.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput*. **2012**, *8*, 2997-3002.

15. Williams, S. L.; de Oliveira, C. A.; McCammon, J. A. Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 560−568.

16. Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *Int. J. Quantum Chem.* **2018**, *118*, e25558.

17. Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115* (12), 6217−6263.

Chapter 3: Constant pH Accelerated Molecular Dynamics Investigation of the pH Regulation Mechanism of Dinoflagellate Luciferase

A version of this chapter was published: Donnan, P. H.; Ngo, P. D.; Mansoorabadi, S. O. *Biochemistry* **2018**, *57 (3)*, 295−299, and associated Supporting Information.[††]

3.1 Abstract

The bioluminescence reaction in dinoflagellates involves the oxidation of an open-chain tetrapyrrole by the enzyme dinoflagellate luciferase (LCF). The activity of LCF is tightly regulated by pH, where the enzyme is essentially inactive at pH ∼8 and optimally active at pH ∼6. Little is known about the mechanism of LCF or the structure of the active form of the enzyme, although it has been proposed that several intramolecularly conserved histidine residues in the N-terminal region are important for the pH regulation mechanism. Here, constant pH accelerated molecular dynamics was employed to gain insight into the conformational activation of LCF induced by acidification.

3.2 Introduction

Bioluminescence, the production of light by living organisms, has arisen independently many times throughout the course of evolution.[1] Consequently, the biological uses of bioluminescence are varied, ranging from communication and courtship to camouflage and predation.[2] The chemistries of bioluminescent reactions are

[††] Note on authorship and contributions: while the dissertation author (P. H. Donnan) is credited as a co-first author on the publication, all calculations and analyses reported in the publication and this chapter were performed by the dissertation author. P. D. Ngo conducted an independent, preliminary CpHaMD study from which no data is reported and no results are used either in the publication or in this dissertation chapter, but which warranted co-first authorship.

also quite diverse, although each involves the enzymatic, luciferase-catalyzed, oxidation of a luciferin substrate with molecular oxygen.[3]

Dinoflagellates are the predominant bioluminescent microorganisms in the sea.[4] Light production by dinoflagellates is proposed to serve as a defense mechanism that attracts predators of dinoflagellate grazers, the so-called burglar alarm hypothesis.[5] The bioluminescence system in dinoflagellates is comprised of an open-chain tetrapyrrolic luciferin (LH_2) derived from the catabolism of chlorophyll, a dinoflagellate luciferase (LCF) that is a divergent member of the lipochalin protein family, and in most species, a luciferin-binding protein (LBP).[6–8] Each of these components is housed together in small, subcellular compartments called scintillons that are contiguous with an acidic vacuole.[9,10]

The dinoflagellate bioluminescence reaction is unique in that it is induced by physical agitation.[11] Shear forces are thought to activate a G-protein-coupled receptor on the surface of the cell, which initiates a signal transduction cascade and leads to an increase in cytoplasmic calcium levels.[12,13] The increase in calcium concentration depolarizes the vacuolar membrane, generating an action potential that opens voltage-gated ion channels and allows the influx of protons into the scintillons.[14] Acidification of the scintillons is then proposed to trigger the release of LH_2 from LBP and the activation of LCF, which leads to the bright flashes of blue light characteristic of dinoflagellate bioluminescence.[15]

The LCFs from most bioluminescent dinoflagellates contain three homologous catalytic domains in a single polypeptide (Figure 1.3).[16] Each individual domain exhibits a sharp pH-rate profile, wherein activity is essentially zero at pH ~8 and optimal at pH ~6. Four intramolecularly conserved histidine residues in the N-terminal region of each

domain have been implicated in the loss of activity in the alkaline region.[17] Mutation of any of these histidine residues to alanine results in an increase in the relative activity of the domains at pH 8.[17]

The crystal structure of domain III from *Lingulodinium polyedrum* was determined in the inactive form at pH 8 (Figure 1.3).[7] The structure consists of a β-barrel that houses the presumed active site and a three-helix bundle cap.[7] The four conserved histidine residues (H899, H909, H924, and H930) are positioned at the interface of the helices in the bundle, which suggests that this structural motif plays an important role in the pH regulation mechanism of LCF.[7]

Disruption of the hydrogen bonding interactions formed by the conserved histidine residues, either by protonation or by mutation to alanine, is proposed to induce a conformational change in the three-helix bundle that activates the enzyme.[7] Consistent with this hypothesis, preliminary molecular dynamics (MD) simulations of the quadruple alanine variant showed an expansion of the three-helix bundle, creating a solvent channel that would allow LH_2 access to the active site (Figure 1).[7] However, these simulations were performed for only 10 ns, which is likely insufficient time to allow for a large-scale conformational change, and it is unclear whether the observed changes will be reflective of those induced by acidification.

To appropriately investigate the pH regulation mechanism of LCF computationally, the protonation state of titratable residues must be allowed to vary in response to proton concentration and the calculations must be performed on sufficiently long timescales (or with enhanced sampling) to allow observation of the conformational transition between the active and inactive forms. One promising approach that addresses

both criteria is to couple constant pH molecular dynamics (CpHMD) with accelerated molecular dynamics (aMD).[18–20] The CpHMD method, as implemented in Amber 16, employs generalized Born solvated molecular dynamics with periodic Monte Carlo sampling of protonation states.[18,21] In aMD, a continuous bias potential is applied when the true potential falls below a chosen threshold energy, which enhances the rate of escape from potential basins, improves sampling of the conformational landscape, and converges to the correct canonical distribution.[19]

Constant pH accelerated molecular dynamics (CpHaMD) was therefore used to investigate the conformational dynamics of LCF domain III at pH 8 and 6. Given the proximity of their free amino acid pK_a values to the pH range of the simulation, the protonation states of all cysteine and histidine residues were made titratable. In addition, aspartate, glutamate, lysine, and tyrosine residues having <15% solvent accessibility (and thus most likely to have microenvironments that deviate significantly from that of the solution) were also allowed to vary during the simulation. After structural minimization, heating, equilibration, and a 10 ns conventional MD simulation to obtain boost parameters, CpHaMD was performed for 1 μs of aMD time.

3.3 Computational Methodology

The crystal structure of dinoflagellate luciferase (LCF) domain III from *Lingulodinium polyedrum* (Protein Data Bank entry 1VPR) was used as the initial structure for the molecular dynamics (MD) investigation.[7] The final 23 residues of the C-terminus, which were not resolved in the crystal structure, were generated using the I-TASSER Online Server.[22–24] The coordinates of the C-terminus were extracted from the

predicted structure and aligned to the crystal structure using PyMOL.[25] Titratable residues were chosen by identifying aspartate, glutamate, lysine, and tyrosine residues with <15% solvent accessibility using Swiss-PdbViewer.[26] Additionally, all histidine and cysteine residues were made titratable since their free amino acid pK_a values are closest to the range of simulated pH conditions. Input structures were parameterized with the Amber ff10 force field and tleap was used to prepare initial structures for Amber simulations. Input files for the constant pH molecular dynamics (CpHMD) simulations were created using the cpinutil program in Amber Tools 17.[21]

Simulations were performed using Amber 16, Amber Tools 16, and Amber Tools 17.[21] Structure minimization, heating, and equilibration were performed with the multisander module of Amber Tools 16 and 17. Minimization was performed over 1000 steps using the conjugate gradient method. Structures were heated over 2 ns from 100 K to 300 K. All simulations used a generalized Born (GB) implicit solvent with a salt concentration of 0.1 M.[18] A cutoff of 1000 Å was used for all non-bonded interactions. The SHAKE algorithm was employed to constrain bonds including bonds to hydrogen.[27] Simulations were run at 300 K using Langevin dynamics with a collision frequency of 5 ps^{-1}.[28] Integration of Newton's equations used a time step of 2 fs.

Starting with equilibration simulations, the CpHMD methodology was used to allow titratable residues to change protonation states.[18] Details of the method, which employs Monte Carlo sampling, can be found elsewhere.[18,21] A Monte Carlo move to change protonation states was attempted every 10 fs in all simulations utilizing CpHMD.

In order to better sample the conformational space of LCF, accelerated molecular dynamics (aMD) methodology was combined with CpHMD.[20,29] Conventional MD

simulations of 10 ns were performed to obtain boost parameters for aMD simulations. Both MD and aMD simulations used the CUDA version of Amber in order to utilize graphics processing unit (GPU) computing. Constant pH accelerated molecular dynamics (CpHaMD) simulations were run for 1000 ns of aMD time.[30,31] This simulation length should give access to the entire conformational space of LCF without error becoming large enough to bias the results.[20,32] The PyMOL and VMD software packages were used to perform visual analysis of results.[25,33] Further data analysis was performed using cpptraj and the cphstats program in Amber Tools 17.[21,34]

3.4 Results and Discussion

Throughout the simulation at pH 8, the four conserved histidine residues implicated in the pH regulation mechanism remain predominantly in the neutral, deprotonated form (Figure 3.1). Interestingly, the C-terminus adopts an α-helical conformation and remains relatively rigidly bound to the top of the β-barrel adjacent to the three-helix bundle, while the N-terminus is highly mobile (Figures 3.2A and 3.3A). This is in contrast to what is observed in the crystal structure, where the C-terminus is disordered and the N-terminus forms a structured loop containing two short antiparallel β-hairpins.[7]

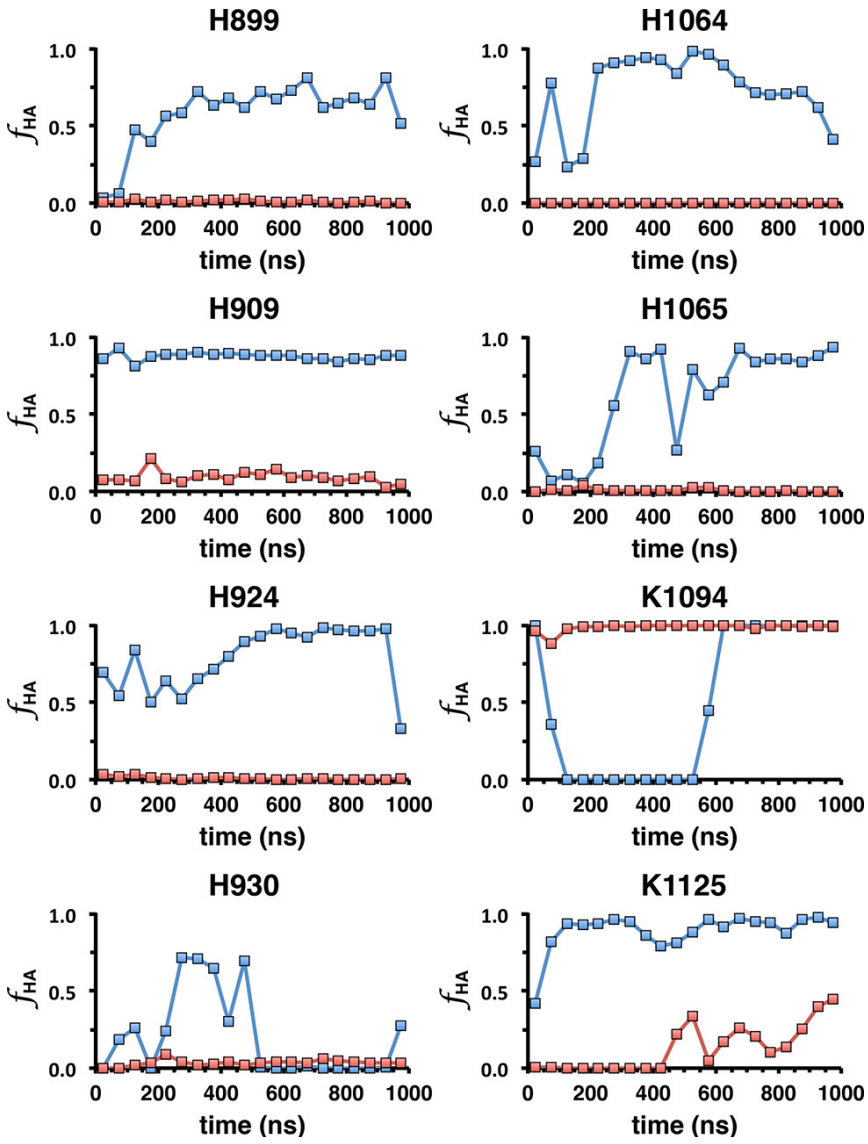

Figure 3.1 Fraction of LCF domain III residues in the protonated form (f_{HA}) during the CpHaMD simulations at pH 8 (red) and pH 6 (blue). Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society.

In addition to the N-terminus, α5 and α6 of the three-helix bundle are also highly mobile, and the region between the glycine-rich motif and the highly conserved catalytic core (which encompasses helices α3 and α4 and wraps around the β-barrel) deviates significantly from its position in the crystal structure (Figures 3.2A and 3.3A).[7] It is unclear whether the flexibility of the N-terminal region of domain III would be retained

in the full-length enzyme (i.e., if it were tethered to domains I and II). However, despite the mobility of the N-terminus and the three-helix bundle, the latter remains tightly associated with the β-barrel throughout the simulation, which is consistent with the hypothesis that this motif restricts access of the substrate to the active site at pH 8. In fact, after the first ~100 ns of the simulation, the gap between the three-helix bundle and the edge of the β-barrel, as defined by the distance between the Cα atoms of H1064 and R1136, decreases from the value of ~13 Å observed in the crystal structure to ~6 Å and remains there for the duration of the simulation (Figures 3.2A and 3.3B).

In contrast, during the simulation at pH 6, the four conserved histidine residues are found in the positively charged, protonated form, although the behavior of each of these residues is distinct (Figure 3.1). H909 is rapidly protonated and is found almost exclusively in this form throughout the simulation. H924 is also predominantly found in the protonated form, although it is more intermittently protonated early in the simulation and the protonated fraction increases after ~300 ns. In contrast, H930 begins the simulation in the deprotonated form and is transiently protonated between ~250 and 500 ns. H899 also starts in the deprotonated form, and the protonated fraction increases after a lag of ~100 ns. Two additional histidine residues, the H1064/H1065 dyad, also display similar behavior, in which they are mostly deprotonated early in the simulation and become protonated after ~250 ns (Figure 3.1).

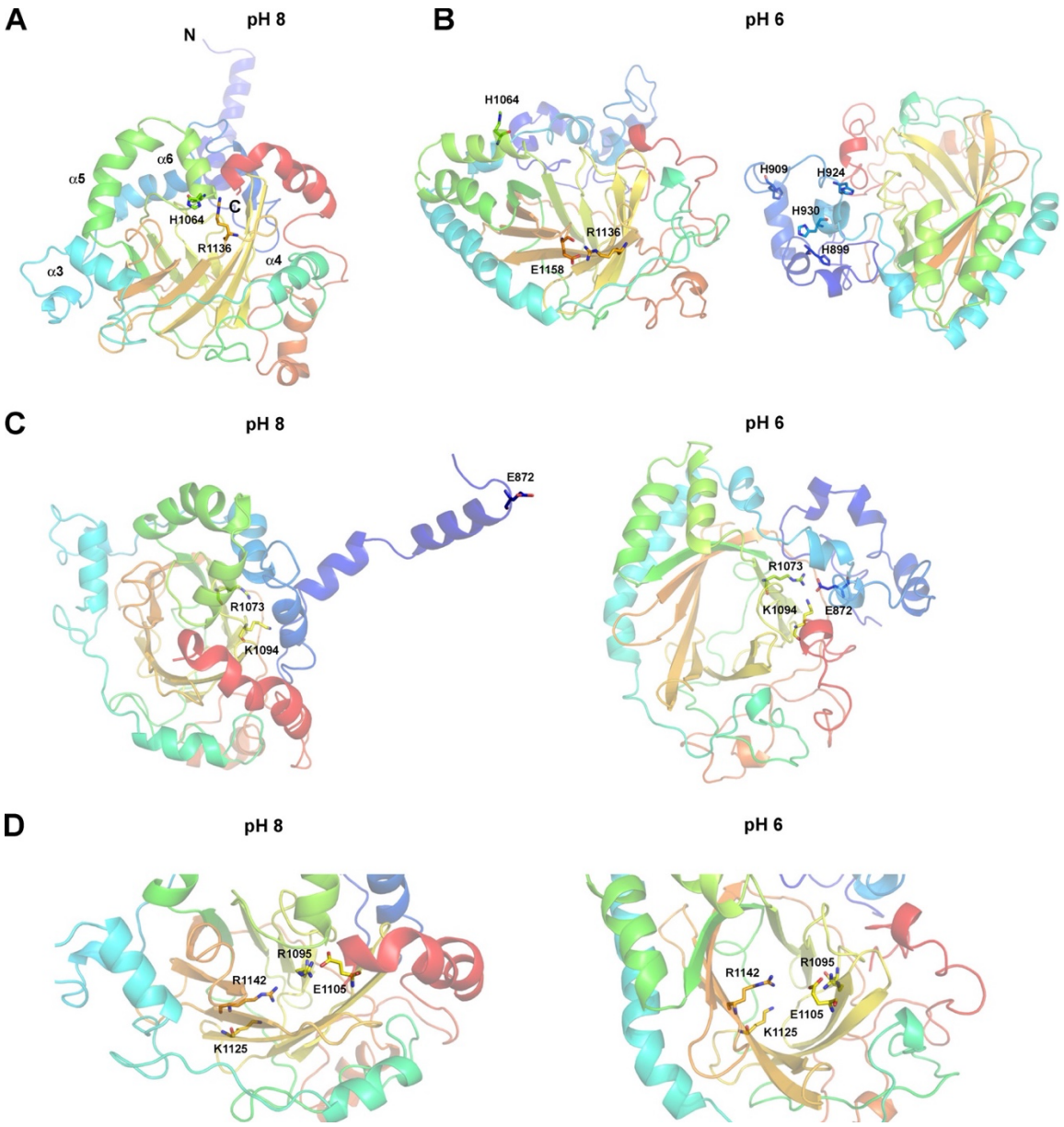

Figure 3.2. Calculated structures of LCF domain III after 1 μs of CpHaMD simulation. (A) Inactive structure at pH 8, showing the close association of H1064 and R1136. Helices α3−α6 and the N- and C-termini are indicated. (B) Two different orientations of the activated structure at pH 6. R1136 forms a salt bridge with E1158, while the N-terminal domain containing the intramolecularly conserved histidine residues reorganizes. (C) Comparison of the N-terminal domain structure at pH 8 and 6. K1094 participates in a hydrogen bonding network that stabilizes the N-terminal helical bundle at pH 6. (D) Comparison of the active site structure at pH 8 and 6, showing an expansion of the β-barrel and repositioning of the putative catalytic base E1105. Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society.

Intriguingly, the delayed protonation of the histidine residues mentioned above correlates with a large-scale conformational change in which the distance between H1064 and R1136 increases dramatically (Figures 3.2B and 3.3B). During the simulation, α5 and α6 from the three-helix bundle tip over and the loop containing the H1064/H1065 dyad is positioned such that the distance between H1064 and R1136 increases to ~30 Å (Figure 3.2B). R1136 also moves away from its position at the top of the β-barrel and forms a salt bridge with E1158 on the adjacent β-strand (Figure 3.2B). The conformational change also involves a decrease in the mobility of the region surrounding the β-barrel containing α3 and α4, an increase in the flexibility of the C-terminus, and the reorganization of the N-terminal domain to form a helical bundle (Figures 3.2B and 3.3A).

A lysine residue (K1094) on the outside of the β-barrel, which is solvent-exposed and in the protonated form at pH 8, rapidly becomes reverse protonated at pH 6 during the first ~600 ns of the simulation (Figures 3.1 and 3.2C). Closer inspection shows that K1094 engages in a hydrogen bond network with R1073 and E872 (Figure 3.2C). R1073 is also on the outside of the β-barrel and moves toward K1094 as the three-helix bundle tips over, while E872 is found near the N-terminus. The interaction of E872 with K1094 and R1073 thus helps to stabilize the reorganized N-terminal domain (Figures 3.2B, 3.2C, and 3.3A).

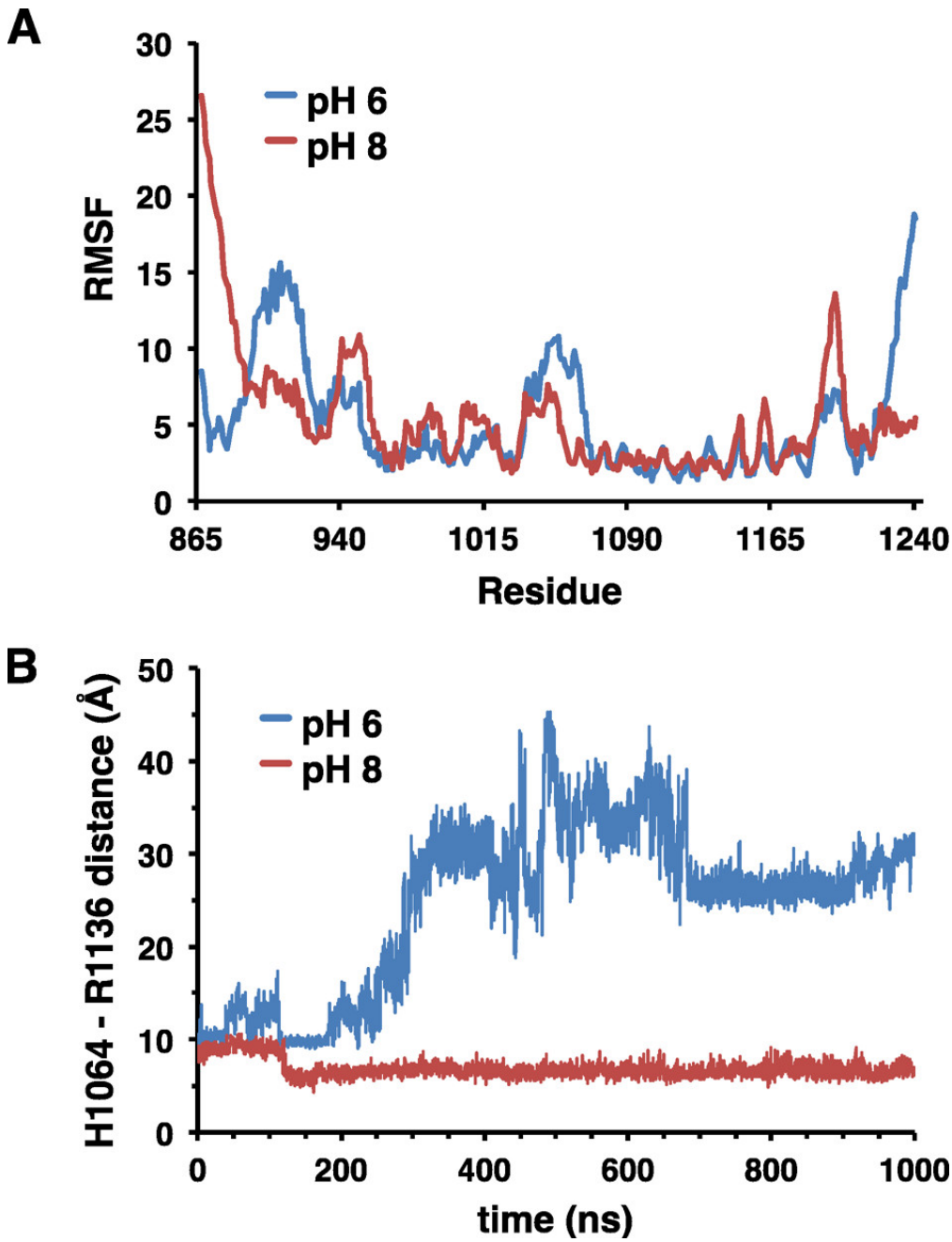

Figure 3.3. Root mean square fluctuation (RMSF) and inter-residue distance data of LCF domain III from the CpHaMD simulations at pH 8 and 6. (A) Per residue RMSF of LCF domain III calculated over the full 1 μs simulation. (B) Distances between the Cα atoms of H1064 and R1136 during the 1 μs simulation. Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society.

Another lysine residue (K1125) found at the bottom of the β-barrel's interior remains predominantly in the neutral, deprotonated form throughout the simulation at pH 8 but is rapidly protonated at pH 6 (Figures 3.1 and 3.2D). This is the only residue whose

protonation state changes within the presumed active site, and this may be important for organizing the active site for binding of the dianionic LH_2 substrate. In particular, there is a notable expansion of the β-barrel, and E1105 moves ~6 Å from the periphery to the center of the active site, where it is positioned near R1095 and R1142 (Figure 3.2D). As it is likely that these two arginine residues in the active site will interact with the carboxylate moieties of the LH_2 chromophore, this conformational change optimally positions E1105 to serve as the catalytic base. The involvement of E1105 in the catalytic mechanism of LCF was proposed in a recent theoretical investigation of dinoflagellate bioluminescence that identified the potential bioluminophore of the reaction.[35]

In summary, CpHaMD was applied to investigate the conformational changes associated with the activation of LCF upon acidification. The protonation of several residues, including the previously identified intramolecularly conserved histidine residues and the H1064/H1065 dyad, correlates with a large-scale conformational change in which the N-terminal domain reorganizes to allow the substrate access to the active site. Concomitantly, the β-barrel expands and a putative active site base, E1105, moves into position where it can initiate catalysis. To the best of our knowledge, this is the first example of CpHaMD being applied to investigate the conformational activation of a pH-regulated enzyme and demonstrates the power of this methodology in gaining insight into enzyme dynamics as a function of pH.

3.5 References

1. Hastings, J. W. *J. Mol. Evol.* **1983**, *19*, 309−321.

2. Wilson, T.; Hastings, J. W. *Bioluminescence: Living Lights, Lights for Living*; Harvard University Press: Cambridge, Massachusetts, 2013.

3. Hastings, J. W. *Gene* **1996**, *173*, 5−11.

4. Haddock, S. H. D.; Moline, M. A.; Case, J. F. *Annu. Rev. Mar. Sci.* **2010**, *2*, 443−493.

5. Abrahams, M.; Townsend, L. *Ecology* **1993**, *74*, 258−260.

6. Topalov, G.; Kishi, Y. *Angew. Chem., Int. Ed.* **2001**, *40*, 3892−3894.

7. Schultz, L. W.; Liu, L.; Cegielski, M.; Hastings, J. W. *Proc. Natl. Acad. Sci. U. S. A*. **2005**, *102*, 1378−1383.

8. Morse, D.; Pappenheimer Jr., A. M.; Hastings, J. W. *J. Biol. Chem.* **1989,** *264*, 11822−11826.

9. Desa, R.; Hastings, J. W.; Vatter, A. E. *Science* **1963**, *141*, 1269−1270.

10. Nicolas, M. T.; Sweeney, B. M.; Hastings, J. W. *J. Cell Sci.* **1987**, *87*, 189−196.

11. Maldonado, E. M.; Latz, M. I. *Biol. Bull.* **2007**, *212*, 242− 249.

12. Chen, A. K.; Latz, M. I.; Sobolewski, P.; Frangos, J. A. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2007**, *292*, R2020−R2027.

13. Von Dassow, P.; Latz, M. I. *J. Exp. Biol.* **2002**, *205*, 2971−2986.

14. Rodriguez, J. D.; Haq, S.; Bachvaroff, T.; Nowak, K. F.; Nowak, S. J.; Morgan, D.; Cherny, V. V.; Sapp, M. M.; Bernstein, S.; Bolt, A.; DeCoursey, T. E.; Place, A. R.; Smith, S. M. *PLoS One* **2017**, *12*, e0171594.

15. Fogel, M.; Hastings, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 690−693.

16. Liu, L.; Wilson, T.; Hastings, J. W. *Proc. Natl. Acad. Sci. U. S. A*. **2004**, *101*, 16555−16560.

17. Li, L.; Liu, L.; Hong, R.; Robertson, D.; Hastings, J. W. *Biochemistry* **2001**, *40*, 1844−1849.

18. Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

19. Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

20. Williams, S. L.; de Oliveira, C. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 560−568.

21. Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; and Kollman, P. A. AMBER 2017, University of California, San Francisco, 2017.

22. Zhang, Y. *BMC Bioinformatics* **2008**, *9*:40.

23. Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5*, 725-738.

24. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2015**, *12*, 7-8.

25. Schrödinger, LLC, PyMOL The PyMOL Molecular Graphics System, Version 1.8.

26. Guex, N.; Peitsch, M. C. Electrophoresis **1997**, *18*, 2714-2723.

27. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys*. **1977**, *23*, 327-341.

28. Xiang, T.; Liu, F.; Grant, D. M. *J. Chem. Phys.* **1991**, *94*, 4463-4471.

29. Pierce, L. C.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput*. **2012**, *8*, 2997-3002.

30. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput*. **2012**, *8*, 1542-1555.

31. Le Grand, S.; Götz, A. W.; Walker, R. C. *Comput. Phys. Commun*. **2013**, *184*, 374-380.

32. Swails, J. M.; York, D. M.; Roitberg, A. E. *J. Chem. Theory Comput*. **2014**, *10*, 1341-1352.

33. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33-38.

34. Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

35. Ngo, P. D.; Mansoorabadi, S. O. *ChemPhotoChem* **2017**, *1*, 383−387.

Chapter 4: Stereochemistry and Regiochemistry in Dinoflagellate Bioluminescence

4.1 Abstract

Dinoflagellate luciferase catalyzes the bioluminescent oxidation of an open-chain tetrapyrrole substrate using molecular oxygen and producing light, an oxidized product, and water. Dinoflagellate bioluminescence is 'simple' bioluminescence, requiring no other cofactors or cosubstrates. The tetrapyrrole substrate, called dinoflagellate luciferin, has two unresolved stereocenters, one at the C4 chiral carbon and one at the C15=C16 double bond. Different bioluminescent reaction mechanisms have been proposed for each substrate stereochemistry. Importantly, luciferase is pH-dependent, optimally active at pH ~6 and inactive above pH 8. Mechanistic studies of the bioluminescent reaction have been hindered by the lack of a crystal structure of the active form of the enzyme; however, recent computational studies have produced a proposed active conformer. Here, using the calculated open conformer of the enzyme, molecular dynamics simulations were conducted in order to understand the binding of each potential stereoisomer of the luciferin substrate and oxyluciferin product, as well as potential stereoisomers of proposed reaction intermediates implicated in the catalytic mechanism. The enzymatic interactions of each were investigated and implications for the stereochemical course of the bioluminescent mechanism are discussed. Our results predict that the bioluminescent reaction utilizes the (*E,S*) stereoisomer of luciferin, and molecular oxygen can bind before or after deprotonation. The putative catalytic base, Glu1105, then abstracts the pro-*S* hydrogen from C13^2 of the substrate to create an enolate intermediate, which radically couples to oxygen resulting in superoxide addition to the *re* face of the enolate

intermediate. Proton abstraction by the substrate then produces a hydroperoxy intermediate facilitated by the deprotonation of Glu1105. Each intermediate is stabilized by hydrogen bonding to a key residue, Arg1142. The hydroperoxy species undergoes radical rearrangement to an open-shell singlet *gem*-diolate state which emits a photon. Finally, we predict that the post-emission *gem*-diolate species is stereospecifically protonated by Tyr1168 resulting in the elimination of water, which predicts that dinoflagellate luciferase is a monooxygenase. We identify further residues facilitating each step in the proposed mechanism as candidates for mutagenesis experiments.

4.2 Introduction

Dinoflagellates are marine and freshwater eukaryotic microorganisms. Many species are bioluminescent upon physical agitation;[1,2] including all species, dinoflagellates comprise the primary source of marine bioluminescence.[1] As in other bioluminescent systems, the bioluminescent reaction is an enzyme-catalyzed oxidation of a luciferin substrate with molecular oxygen.[3] In dinoflagellates, the luciferin substrate (LH_2) is an open-chain tetrapyrrole postulated to be a chlorophyll *a* catabolite[4,5] and, in many species, is regulated by a circadian rhythm[6] leading to photosynthesis during the day and bioluminescence at night.[7] The enzyme catalyzing bioluminescence, dinoflagellate luciferase (LCF), is housed with LH_2 in specialized organelles called scintillons that are contiguous with acid-housing vacuoles.[6,8,9] Upon physical agitation, G protein-coupled receptors in the cell membrane are activated, triggering a signal transduction cascade that ultimately increases cytoplasmic calcium ion levels.[10,11] The calcium ion concentration serves to depolarize the membrane of the acid-containing

vacuole, which in turns opens voltage-gated ion channels via an action potential.[12] Finally, protons flood the scintillon through the ion channels, leading to a rapid decrease in scintillon pH.[12] The acidification of the scintillon activates the enzyme,[13] which has optimal enzymatic activity at pH ~6.[14]

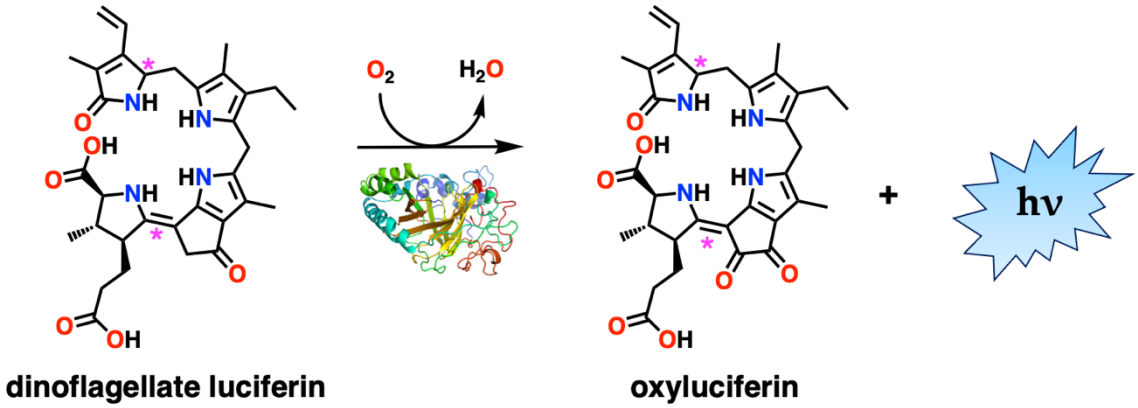

Figure 4.1 The bioluminescent reaction of dinoflagellate luciferin, producing oxyluciferin and light (λ_{max} = 474 nm). The luciferase enzyme is shown in a computationally-determined open conformer. Unresolved stereocenters are labeled with magenta asterisks.

In the scintillon, many species bind LH_2 to a luciferin-binding protein (LBP),[15] and though little is known about the binding mechanism, it is thought that release of LH_2 by LBP is also triggered by the pH drop upon physical agitation. While the binding and release of LH_2 by LBP are active areas of study, the process of LCF activation by acidification is relatively well understood. In most species of bioluminescent dinoflagellates, LCF is comprised of three homologous catalytic domains and a N-terminal domain.[16] Crystal structures of LCF have proven difficult to obtain; only the inactive (pH 8) form of domain III from *Lingulodinium polyedrum* has been solved (Protein Data Bank entry 1VPR).[17] Using site-directed and N-terminal deletion mutagenesis, four intramolecularly conserved histidine residues in the N-terminal region

of each domain (His899, His909, His924, and His930 in domain III) were implicated as contributing to the inactivity of the protein at alkaline pH.[14] Mutation of any was shown to increase the relative luciferase activity at pH 8, and deletion of the N-terminus through residue 927 (removing three implicated His residues) restored essentially full enzymatic activity at pH 8.[14] It was therefore proposed that the protonation and deprotonation of the His residues drives a conformational shift in the enzyme, leading to active or inactive conformers, respectively.

In order to explore the possibility of a large-scale conformational shift driven by acidification, constant pH accelerated molecular dynamics (CpHaMD) simulations of LCF have been conducted.[18,19] Simulations revealed an open conformer at pH 6 where the α5 and α6 helices have tipped away from the central β barrel of the enzyme. The N-terminus, containing the four His residues, was shown to rearrange and three of the four His residues were shown to maintain opposite protonation states in the pH 8 and 6 simulations.[18] Additional His residues, the His1064-His1065 dyad, were implicated in the conformational shift, and Glu1105, previously postulated as the putative catalytic base,[17,20] was shown to adopt a conformation amenable to catalysis by pointing into the center of the β barrel at pH 6.[18] Thus, both computational and experimental progress has been made towards understanding LCF activation by acidification.

The biosynthesis and bioluminescent chemistry of the substrate, however, remain relatively poorly understand. LH_2 has two unresolved stereocenters, at the C4 chiral carbon and the C15=C16 double bond, and it is unknown if the double bond stereochemistry of LH_2 is retained or converted during the bioluminescent reaction (Figure 4.1). In the proposed precursor, chlorophyll *a*, the chiral carbon is yet to be

reduced and hydrogen addition to either the *re* or *si* face could occur; the double bond is present and adopts the *Z* configuration. However, in order to derive LH_2 from chlorophyll *a*, the δ bridge of the porphyrin ring must be oxidatively cleaved, and cleavage could be accompanied by isomerization. Crystal structures of LH_2 and the oxyluciferin product (LO) would help determine the stereochemical course of the reaction, but, to date, neither has been obtained. An *E*-*Z* isomerization during either biosynthesis or bioluminescent chemistry, which has been proposed as a possible mechanism,[20] is physically feasible. Furthermore, dinoflagellate luciferin and its biosynthetic precursor, P630, are known to contain a peptide linkage,[21] but the location and chemical nature of the linkage are unknown. Thus, many open questions remain in the structural understanding of dinoflagellate bioluminescence. Even with limited experimental data, computational tools have provided avenues for studying the bioluminescent reaction. Computational methods have thus far predicted both the bioluminophore and the open, catalytic conformer of the enzyme.[18–20,22] However, predictions of the mechanism were different for the *E* and *Z* isomers, and only a mimic of an active site base was included, with no further residue interactions considered.[20] With the predicted open conformer of the enzyme accessible, studying interactions between ligands (including substrate, intermediates, and product) and the enzyme is now feasible. Here, we report molecular dynamics (MD) simulations of all possible stereoisomers of LH_2 bound to LCF. We extend simulations to include all possible stereoisomers of the LO product with LCF, including simulations with either a protonated or deprotonated Tyr1168; on the basis of our calculations, we show that the *S* isomer of each binds most favorably to LCF. We then show calculations of the two *S* isomers of a proposed enolate reaction intermediate (LH^-) interacting with LCF. To

explore whether deprotonation and oxygen binding are ordered, we simulated ternary complexes of both LH_2 and LH^- with molecular oxygen added to the active site pocket, predicting that the events are not necessarily ordered and O_2 will bind to the same pocket regardless of whether LH_2 or LH^- is in the active site. Further reaction intermediates

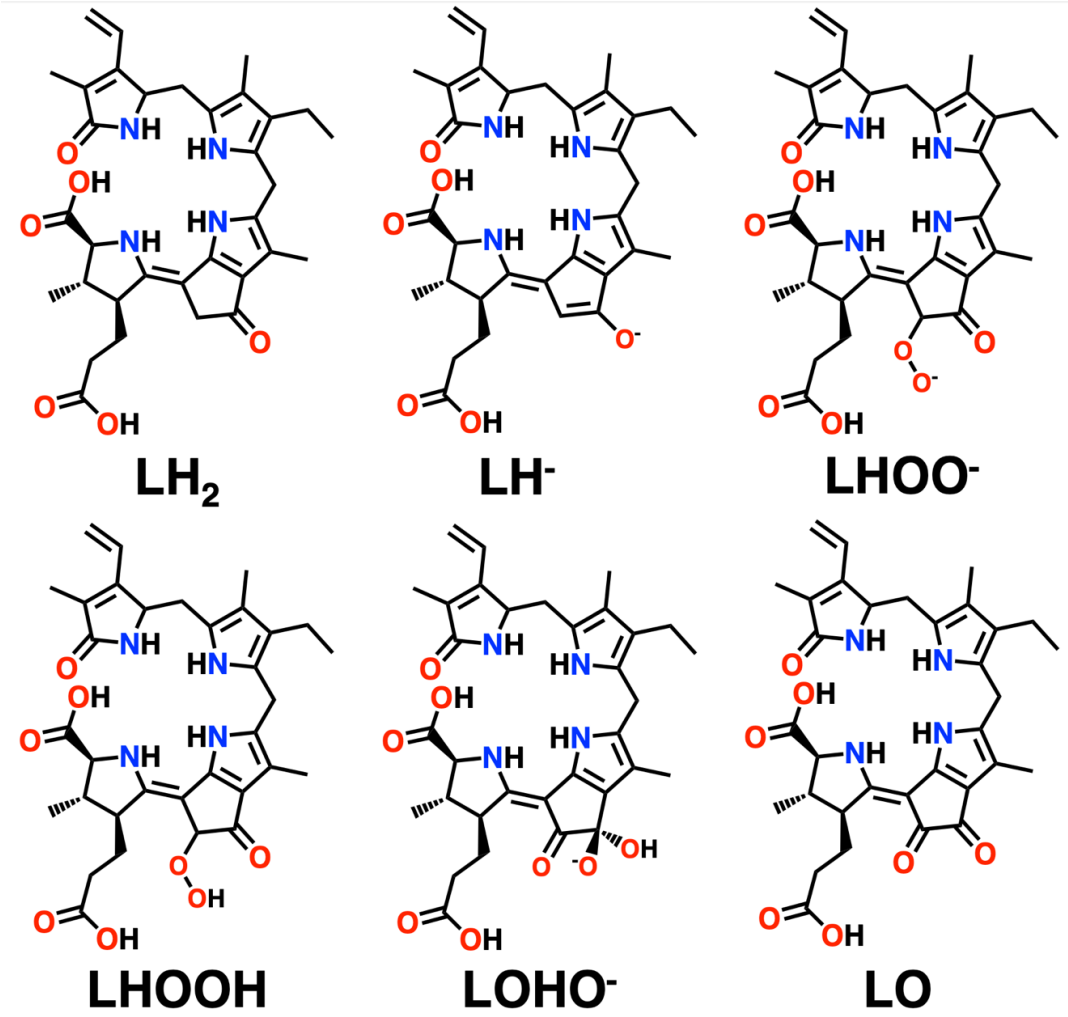

Figure 4.2 Overview and nomenclature used of substrate, intermediate, and product structures relevant to this chapter. Undetermined stereocenters are left unspecified.

Binary complexes of three subsequent proposed reaction intermediates (Figure 4.2), including a peroxy anion formed after O_2 addition (LOO^-), a subsequent hydroperoxy species (LOOH), and finally a *gem*-diolate intermediate ($LOHO^-$) that has been suggested as the post-emission product, were studied in different stereoisomers in order

to predict the stereo- and regiochemical course of the reaction, including predicting that LCF should catalyze monooxygenation, not dioxygenation. We identify interactions with key active site residues, including Glu1105, Arg1142, Tyr1168, Ser1068, and Lys1089, and propose functions for each in the catalytic mechanism. In sum, our simulations provide a picture of the stereochemical and regiochemical course of the dinoflagellate bioluminescence reaction.

4.3 Computational Methodology

All molecular dynamics (MD) simulations were performed using Amber18, AmberTools18, and AmberTools19.[23] The starting structures of were built from our previously obtained constant pH accelerated molecular dynamics (CpHaMD) structure of *Lingulodinium polyedrum* LCF domain III at pH 6 after 1000 ns of accelerated MD simulation time,[18] originally built from PDB entry 1VPR.[17] All residue protonation states were determined by experimental results[14] combined with computationally implicated residues[18] at pH 6. Unless otherwise specified (*vide infra*), residues were set to the free solution amino acid protonation state. His899, His909, His924 were selected as doubly protonated since they were experimentally implicated in the conformational shift. However, on the basis of calculated pK_a values, His904 was also doubly protonated while His930 was singly protonated.[18] The His1064-His1065 dyad was set to the doubly protonated state for both His residues.[18] Two Cys residues, Cys972 and Cys980, were deprotonated on the basis of calculated pK_a values.[24] For the current investigation, the remaining His residues were selected as either δ protonated (His886 and His1101) or ε protonated (His930 and His1027) such that hydrogen bonding interactions were

maximized in the starting structure, as determined by visual inspection. During previous CpHaMD simulations, these His residues were either predominantly found in the deprotonated state or found to have highly variable protonation states,[18] and we therefore defaulted to the free amino acid solution protonation state, even though some had depressed pK_a values. Similar reasoning was used in keeping both Lys1094 and Lys1125 protonated, though these residues had variable protonation states in previous calculations.[18]

Initial LH_2 substrate structures were prepared using the Avogadro software package.[25,26] The AutoDockVina software package[27] was used to determine docked poses of LH_2 with LCF for use as starting binary complexes in MD simulations. LH_2 bonds were allowed to be flexible while enzymatic bonds were not. Docked poses of every stereoisomer (*E,R*; *E,S*; *Z,R*; and *Z,S*) were calculated. The AutoDockVina search box was $30 \times 30 \times 30$ Å^3 in volume and centered at the origin in the β barrel of the protein so as to cover the region containing the putative active site, which is exposed in the open conformer. An increased exhaustiveness parameter, 20, was used. For each stereoisomer, simulation poses were chosen according to several criteria: the AutoDockVina best predicted structures, the structure minimizing the distance between $C13^2$ on LH_2 and the Cδ of Glu1105 (the putative catalytic base[18]), structures with the two terminal carboxylate moieties of LH_2 pointing into the β barrel of the enzyme, and finally structures maximizing hydrogen bonding interactions between the substrate and the enzyme. For every stereoisomer except (*E,R*)-LH_2, the chosen binary complex was AutoDockVina's top predicted pose; for the (*E,R*)-LH_2 stereoisomer another pose (the

fifth best by AutoDockVina's scoring algorithm; binding affinity < 1 kcal mol^{-1} difference from top predicted pose) was chosen for simulation.

The LH_2 substrates were then parameterized for MD simulation using the RESP charge fitting method.[28] Initial binary complex structures were prepared with the tleap module of Amber. The protein was parameterized with the ff14sb force field[29] while the substrate was parameterized with the GAFF2 force field.[30] The binary complex was solvated in a box of TIP3P water molecules[31] with a minimum distance of 20 Å between any protein residue and the edge of the box. Chloride counter ions were added to neutralize the system.

Parameterized structures were minimized with 500 steps of conjugate gradient algorithm followed by 500 steeps of a steepest descent algorithm. A 10 Å interaction cutoff was used, and protein backbone atoms were restrained by a harmonic restraint (force constant = 10 kcal mol^{-1} $Å^{-2}$). Periodic boundary conditions were employed. Following minimization, a fast simulated annealing/heat shock algorithm with velocity rescaling was employed in order to help ensure that the enzyme could accommodate the substrate during later MD steps. The algorithm repeatedly started the system at high temperature and then cooled to low temperature over the course of a cycle. Ten 20 ps cycles were conducted with two varying parameters,

$$T = \begin{cases} 800 & t \leq 5 \text{ ps} \\ 800 - \frac{700}{13}(t-5) & 5 \text{ ps} < t \leq 18 \text{ ps} \\ 0 & t > 18 \text{ ps} \end{cases} \quad (1)$$

and

$$\tau = \begin{cases} 0.4 & t \leq 5 \text{ ps} \\ 4 & 5 \text{ ps} < t \leq 18 \text{ ps} \\ 1 & 18 \text{ ps} < t \leq 19 \text{ ps} \\ 0.1 - 0.05(t - 19) & t > 19 \text{ ps} \end{cases} \quad (2)$$

where T is the temperature of the system (in Kelvin), t is the simulation time (in ps), and τ is the time constant of heat bath coupling (in ps). A Berendsen thermostat[32] was employed for this step. No backbone restraints were employed. An extended 20 Å interaction cutoff with periodic boundary conditions was used. Several variations of the algorithm were explored, but subjecting the system to this protocol was found to lead to substrate remaining bound in later MD steps.

Heating was conducted from 100 K to 300 K over 2 ns using a Langevin thermostat[33] with time constant $\tau = 2$ ps and collision frequency $\Gamma = 5$ ps^{-1}. Bonds to hydrogen atoms were fixed using the SHAKE algorithm.[34] An interaction cutoff of 10 Å with periodic boundary conditions was used. Atypically, no backbone restraints were used, as it was found to facilitate substrate binding after the simulated annealing/heat shock procedure. We note that the removal of backbone restraints allowed the enzyme to adjust to accommodate the substrate, and that because the enzyme is starting in a conformer determined at 300 K,[18] restraints are less important than for a MD simulation beginning from a crystal structure. A short equilibration run of 2 ns was then conducted before production MD. Equilibration and production runs used the same parameters as the heating simulations, except that the temperature was fixed at 300 K. For production MD, five independent trajectories of 100 ns each for every species studied (500 total ns per species) were conducted and analyzed.

For product simulations, substrate docked poses were modified by removal of the hydrogen atoms on $C13^2$ and replacement by an oxygen atom. The carbon-oxygen bond was then minimized, while keeping all other atoms fixed, using the Gaussian16 software package at the B3LYP/6-31G* level of theory.[35–39] The procedure then followed the same as substrate simulations, starting with RESP charge fitting, including for product simulations with a deprotonated Tyr1168 and protonated Glu1105.

For enolate (LH^-) simulations, representative poses from substrate simulations were modified to make starting structures. Because deprotonation of LH_2 to LH^- is expected change several bond lengths, the intermediate molecule was re-minimized using the ORCA software package at the B3LYP/def2-SVP def2/J RIJCOSX level of theory.[40–43] For the (*E,S*) isomer, a CPCM model with $\varepsilon = 20$ was used to aid with convergence. Glu1105, as the putative base, was protonated for the simulations. The resulting LCF•LH^- binary complexes then followed the same procedure as the substrate simulations, starting with RESP charge fitting.

For ternary complex simulations adding molecular oxygen (i.e., LCF•LH_2•O_2 and LCF•LH^-•O_2), representative frames were chosen from LCF•LH_2 and LCF•LH^- binary simulations. The *R* stereoisomers were excluded from these (and all further) calculations on the basis of the substrate and product simulations. For the LCF•LH_2•O_2 calculations, the initial oxygen position was between the two $C13^2$ hydrogen atoms, while for the LCF•LH^-•O_2) case, calculations were conducted starting oxygen on both the *re* and *si* faces of the sp^2 hybridized $C13^2$ carbon. Molecular O_2 was parameterized with the RESP method.[28] The MD protocol followed the same conditions as previous, with two exception. First, the heat shock/simulation annealing step was now omitted, and the

heating was carried out without backbone restraints. Second, the production runs were 10 ns (for 50 total ns of data) due to the fast motion of the oxygen molecule. Oxygen was observed to diffuse from the active site during the heating of (*Z,S*)-LH_2 when started from the *re* face, and this regio- and stereochemistry was not considered further.

The starting structures for the LCF•LOO^- binary complexes used the starting coordinates of the LCF•LH^- binary complexes, with the LOO^- initial geometries minimized from the LH^- geometries at the same level of theory used for the LH^- geometries. Similarly, the starting geometries for the LOOH and $LOHO^-$ binary complexes with LCF were determined in the same fashion but using the LOO^- initial coordinates. The $LOHO^-$ geometry optimizations used the CPCM solvent model with ε = 20 to facilitate convergence. Glu1105 was protonated in the LOO^- and $LOHO^-$ simulations and deprotonated in the LOOH simulations. The MD protocol followed that of the substrate, product, and enolate simulations, except again the heat shock/simulation annealing step was omitted, and the heating was carried out without backbone restraints. Parameterization of another predicted intermediate, a dioxo-bridged species,[20] was attempted but optimizations failed to converge to a reasonable geometry regardless of the method used.

The cpptraj software package in AMBERTools was used for trajectory analysis.[44] Visual Molecular Dynamics[45] and PyMol[46] were used for visual inspection of trajectories. In order to generate representative coordinates for analysis, the average structure across all replicate simulations was calculated. Because the average structure is not a true coordinate set found in the simulations, and therefore can display problems such as distorted bond lengths, the root-mean-squared deviation (RMSD) between each

simulation frame and the average structure was calculated. The frame with lowest RMSD, i.e., the closest to average structure, was then used as a representative set of coordinates for analysis.

4.4. Results and Discussion

We first investigated substrate and product binding to LCF. When discussing stereoisomers, we specify each species with the C15=C16 configuration (*E* or *Z* followed by the C4 configuration (*R* or *S*). For example, the substrate stereoisomer with *E* configuration at C15=C16 and *R* configuration at C4 is (*E,R*)-LH_2.

LCF•LH2 and LCF•LO Binary Complexes. Simulations of the LCF•LH_2 and LCF•LO binary complexes revealed differences in binding for each stereoisomer. While each starting structure had the carboxylate moieties of the D pyrrole ring pointed in direction of the barrel, each stereoisomer adopted different binding positions. In Table 4.1, we consider the distance of the $C13^2$ atom, which is oxidized in the bioluminescent reaction (Figure 4.1), to the δ carbon of the putative catalytic base, Glu1105 over the course of five independent trajectories for each stereoisomer (see Computational Methods for details). For the substrate calculations (representative poses shown in Figure 4.3), a

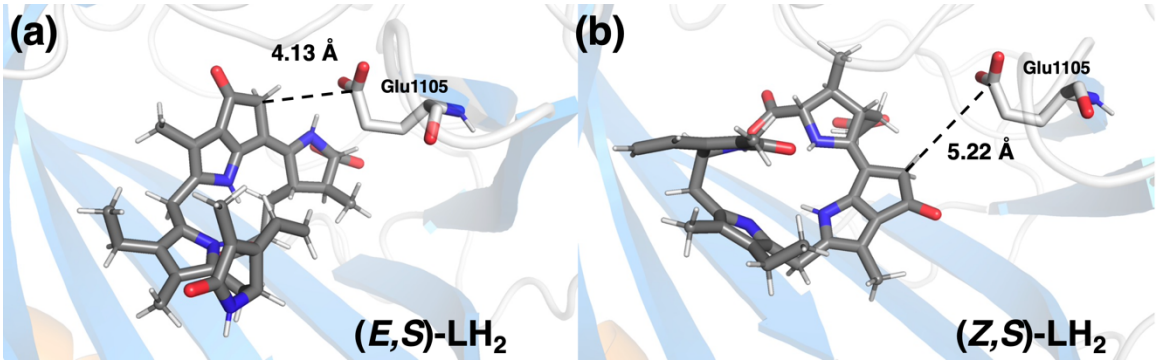

Figure 4.3 Comparison of (a) (*E,S*)-LH_2 binding to LCF and (b) (*Z,S*)-LH_2 binding to LCF via closest to average structures. In each case, the average distance between $C13^2$ of the substrate and Cδ of the putative catalytic base, Glu1105, is shown. Protons on the enzyme are omitted for clarity.

clear preference for the (*E,S*) stereoisomer was shown. While both *Z* stereoisomers also had reasonable average distances, (*E,R*) did not bind in a manner conducive to proton abstraction.

Table 4.1 Distance (in Å) between selected atoms and Glu1105 Cδ for LCF•LH_2, LCF•LH^-, and LCF•LO binary complex simulations.

Stereoisomer	(*E,R*)	(*E,S*)	(*Z,R*)	(*Z,S*)
LH_2 $C13^2$ Carbon	13.33 ± 0.82	4.13 ± 0.34	6.32 ± 0.75	5.22 ± 0.68
LH_2 $C13^2$ pro-*R* Hydrogen	N. D.	4.59 ± 0.43	N. D.	4.26 ± 0.72
LH_2 $C13^2$ pro-*S* Hydrogen	N. D.	3.19 ± 0.35	N. D.	5.49 ± 0.66
LH^- $C13^2$ Carbon	N. D.	4.48 ± 0.29	N. D.	9.27 ± 1.70
LO $C13^2$ Carbon	11.90 ± 2.46	6.30 ± 1.98	6.55 ± 1.44	6.23 ± 0.89
LO $C13^2$ Carbon[a]	11.67 ± 4.29	9.61 ± 2.95	5.83 ± 0.48	4.82 ± 0.55

[a]For simulations with Glu1105 protonated and Tyr1168 deprotonated.

Initial product simulations did not discriminate between (*E,S*), (*Z,R*), and (*Z,S*), with each having an average distance slightly over 6 Å (Table 4.1). Once, again the (*E,R*) isomer was not found to bind in the putative active site. Upon analysis, a tyrosine residue, Tyr1186, was found near the carbonyl on $C13^2$. As no other candidate residues were close, we postulated that Tyr1186 could be responsible for protonation and elimination of water from the post-emission product.[20] Therefore, we conducted simulations with

Tyr1186 deprotonated and Glu1105 protonated to simulate the product after elimination of water but before the active site could reset to the initial protonation state.

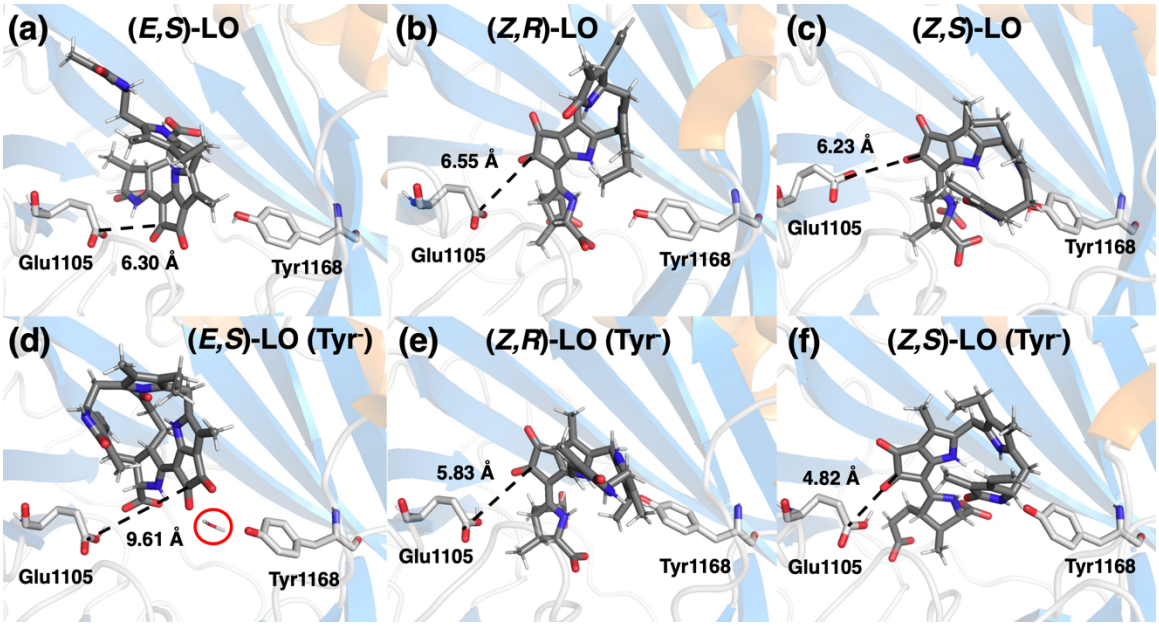

Figure 4.4 Comparison of (a) (*E,S*)-LO binding to LCF, (b) (*Z,R*)-LO binding to LCF, (c) (*Z,S*)-LO binding to LCF, (d) (*E,S*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105, (e) (Z,*R*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105, and (f) (*Z,S*)-LO binding to LCF with deprotonated Ty1168 and protonated Glu1105 via closest to average structures. In each case, the average distance between C13^2 of the product and Cδ of the putative catalytic base, Glu1105, is shown. We note that only the (*E,S*)-LO•LCF simulation with deprotonated Ty1168 and protonated Glu1105 produced an ordered water molecule in the active site (red circle). Protons on the enzyme are omitted for clarity.

The results of the deprotonated Tyr1186/protonated Glu1105 product simulations revealed difference in the binding of the (*E,S*), (*Z,R*), and (*Z,S*) stereoisomers (Figure 4.4). While the average distance between C13^2 and Glu1105 was increased for the (*E,S*) stereoisomer, it was the only simulations which displayed a closest to average structure included an ordered water molecule in the active site. The ordered water molecule, between the product and Tyr1186, is both responsible for the increased difference in average distance between C13^2 and Glu1105 and provides evidence that Tyr1186 is responsible for stereospecific protonation to eliminate water from the post-emission

product, $LOHO^-$. The deprotonated Tyr1186/protonated Glu1105 product simulations provide the evidence for the favorability of (*E,S*) as the biological stereoisomer of luciferin. On the basis of substrate and product simulations, only the (*E,S*) and (*Z,S*) stereoisomers were considered further.

LCF•LH^- Binary Complex. In order to determine the stereochemistry of hydrogen atom abstraction to form LH^-, we analyzed the average distances between the pro-*S* and pro-*R* hydrogen atoms of $C13^2$ to the Glu1105 Cδ atom for (*E,S*) and (*Z,S*), the most favorable stereoisomers. The results are reported in Table 4.1, and each stereoisomer shows a preference for abstraction of a different prochiral hydrogen by Glu1105. Due to the sizable resolution in average distance between the two prochiral hydrogen atoms (> 1 Å in each case), this abstraction is expected to be stereospecific.

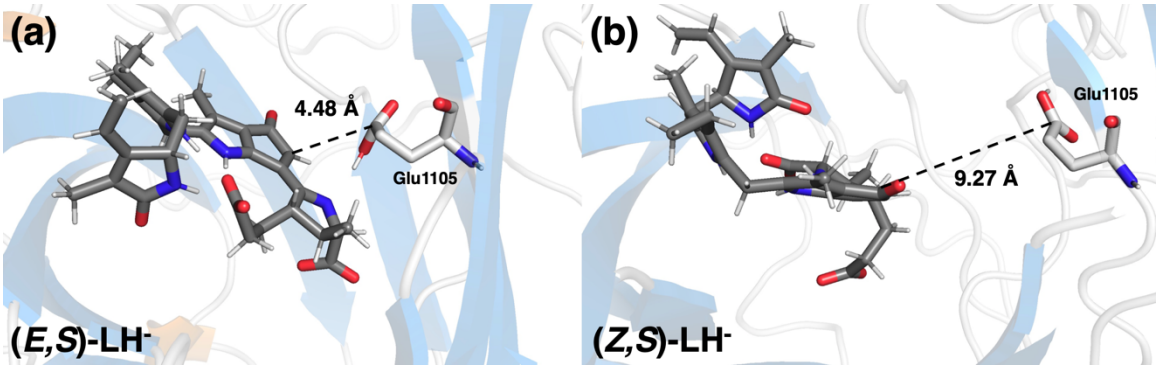

Figure 4.5 Comparison of (a) (*E,S*)-LH^- binding to LCF and (b) (*Z,S*)-LH^- binding to LCF via closest to average structures. In each case, the average distance between $C13^2$ of the intermediate and Cδ of the putative catalytic base, Glu1105, is shown. Protons on the enzyme are omitted for clarity.

However, when LH^- was simulated with LCF, the (*E,S*) stereoisomer remained bound in the active site pocket while the (*Z,S*) moved away (Figure 4.5) and the average distance between C132 and the δ carbon of Glu1105 (now protonated) increased significantly (Table 4.1). While not necessarily ruling out (*Z,S*) as a likely stereoisomer on its own, the simulation provided more evidence favoring (*E,S*) over (*Z,S*).

Ternary Complexes with O_2. Ternary complex simulations with molecular oxygen were conducted for *S* stereoisomers of both LH_2 and LH^- with molecular oxygen. If deprotonation and oxygen binding are ordered, a difference would be expected in O_2 with LH_2 and LH^- for a given stereoisomer. In the case of LH^-, O_2 could add to either the *re* or *si* face of the sp^2 hybridized $C13^2$; simulations were conducted starting O_2 from both sides in order to explore both possibilities. For (*Z,S*), O_2 was not found to bind well to the active site pocket, regardless of whether LH_2 or LH^- was bound.

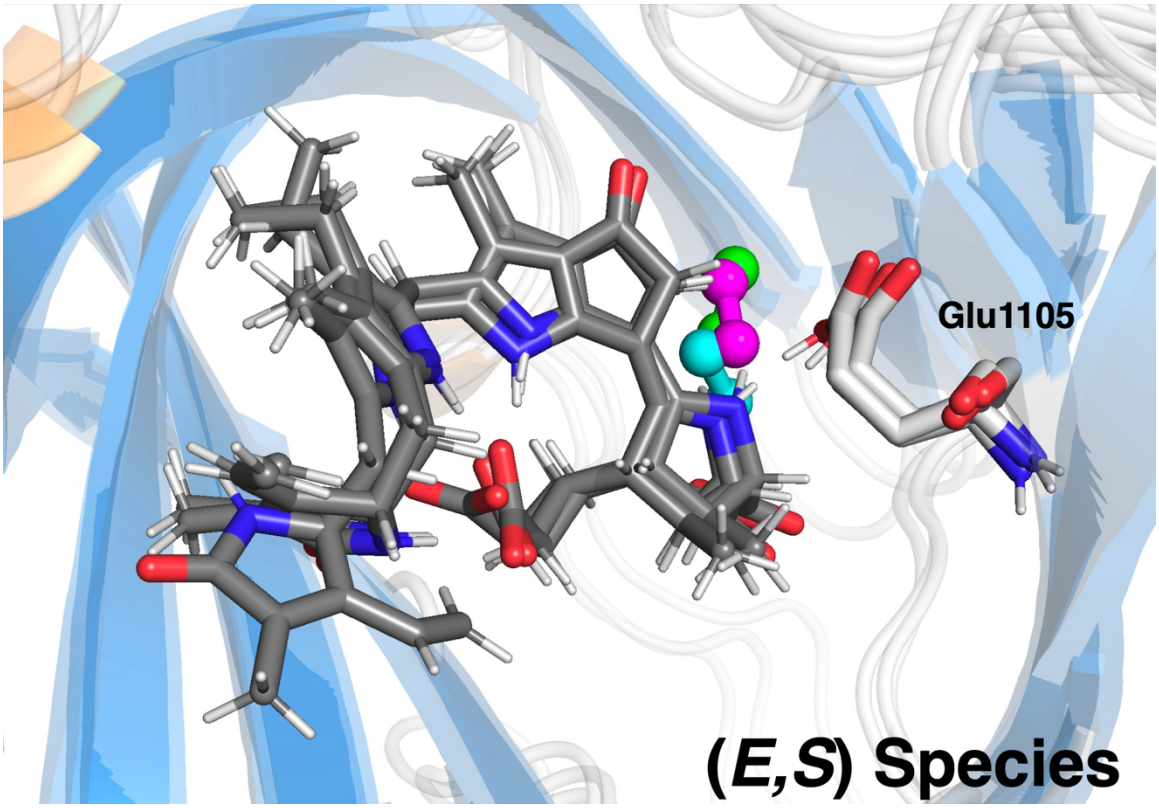

Figure 4.6 Convergence of the closest to average structures of (*E,S*) species with molecular oxygen bound. O_2 is shown in cyan for the LCF•LH_2•O_2 ternary complex, green for the LCF•LH^-•O2 ternary complex starting O_2 on the *re* face of the intermediate, and in magenta for the LCF•LH^-•O_2 ternary complex starting O_2 on the *si* face of the intermediate. We note that the A ring of the substrate is twisted in the LCF•LH_2•O_2 simulation relative to the enolate simulations, but the molecular oxygen binding mode is conserved. Protons on the enzyme are omitted for clarity.

The results for the (*E,S*) species, however, are striking. In each case, the same oxygen binding mode is reached, on the *re* face of the newly formed double bond, indicating that deprotonation state does not affect oxygen binding. The convergence of the O_2 binding mode across all (*E,S*) simulations provides strong evidence for (*E,S*) as the biological stereoisomer, as it was the only stereoisomer to show consistent results across LH_2, LH^-, and LO simulations and the trend continued upon inclusion of O_2.

LCF•LHOO$^-$ and LCF•LHOOH Binary Complexes. While O_2 always bound to the *re* face of LH^-, all of (*E,S,R*)-$LHOO^-$, (*E,S,R*)-LHOOH, (*E,S,S*)-$LHOO^-$, and (*E,S,S*)-LHOOH were simulated since the addition of O_2 introduces a new stereocenter that can be *R* (resulting from *re* addition) or *S* (resulting from *si* addition). Unsurprisingly, the binding modes again reveal a clear preference for one stereoisomer (Figure 4.7). For both $LHOO^-$ and LHOOH, the (*E,S,R*) stereoisomer adopts the same binding pose, facilitated by interactions with Glu1105 and Arg1142. In the (*E,S,S*) complex, Arg1142 is prohibited from stabilizing the structures, resulting in different binding poses between $LHOO^-$ and LHOOH and preventing Glu1105 from facilitating proton transfer.

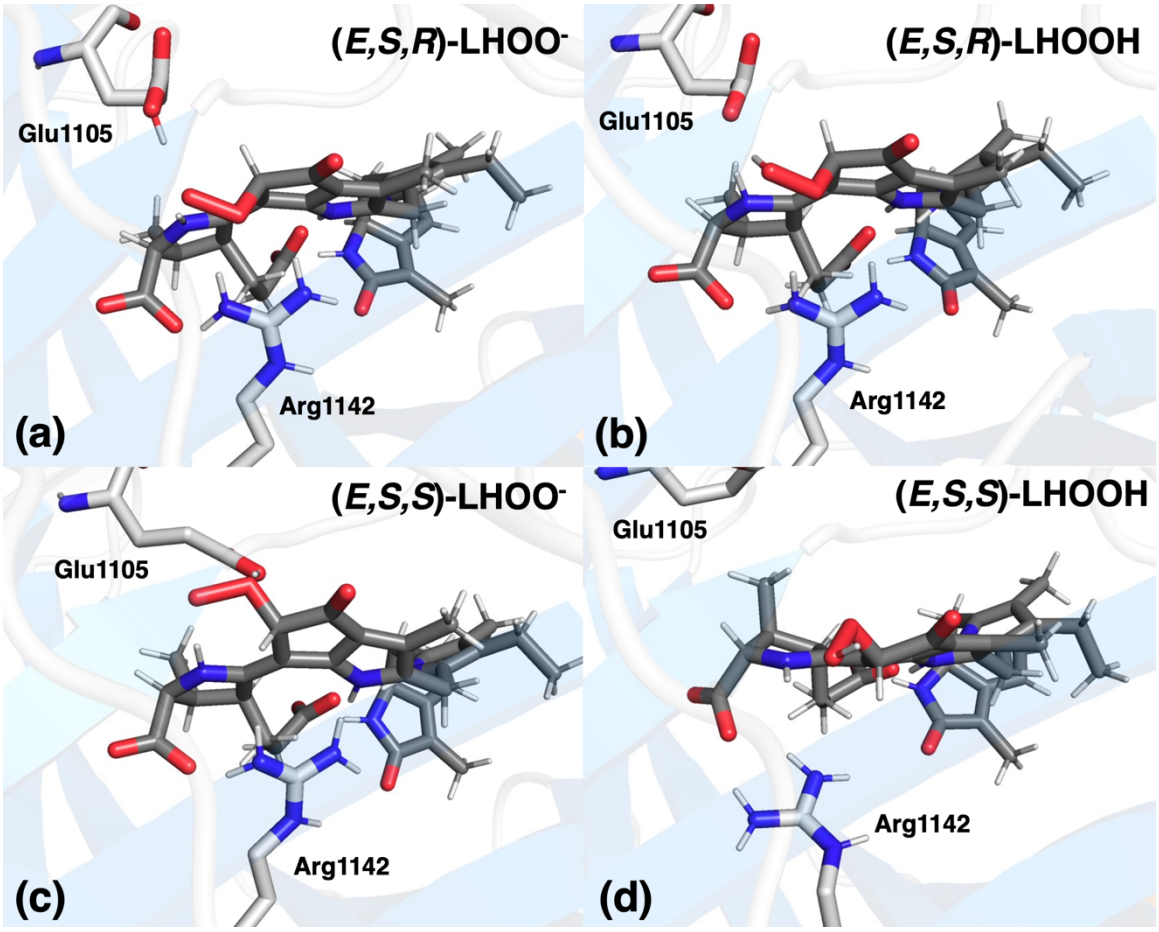

Figure 4.7 Comparison of (a) (*E,S,R*)-LHOO$^-$, (b) (*E,S,R*)-LHOOH, (c) (*E,S,S*)-LHOO$^-$, and (d) (*E,S,S*)-LHOOH binding to LCF via closest to average structures. The (*E,S,R*) species show remarkably similar binding, facilitated by interactions between the intermediate and the residues Glu1105 and Arg1142. Glu1105 is poised for direct proton transfer to LHOO$^-$ (a), forming LHOOH (b). Protons on the enzyme are omitted for clarity.

LCF•LOHO$^-$ Binary Complex. After emission, the next intermediate is proposed to be a *gem*-diolate species (LOHO$^-$).[20] While both (*E,S,R*)-LOHO$^-$ and (*E,S,S*)-LOHO$^-$ were simulated, (*E,S,R*)-LOHO$^-$ results are presented here since (*E,S,R*) was found to maintain the same binding pose in the LCF•LHOO$^-$ and LCF•LHOOH binary complexes (Figure 4.7).

The LCF•(*E,S,R*)-LOHO$^-$ binary complex potentially provides insight into the bioluminophore, as it is the species structurally closest to it. Several interactions are

worth highlighting between LCF and (*E,S,R*)-LOHO$^-$ (Figure 4.8). Tyr1168 is positioned to stereospecifically protonate and eliminate water from the intermediate. Ser1068 and Lys1089 are interacting with the opposite face of the intermediate in stabilizing hydrogen bonding interactions. Finally, Arg1142 interacts with the C13^1 carbonyl and Tyr1168, indicating that it may be important in stabilizing the bioluminophore.

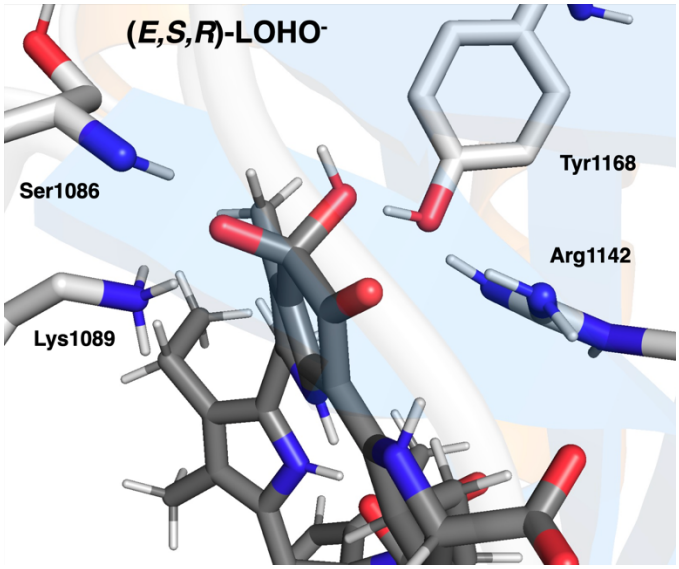

Figure 4.8 Closest to average structure of the (*E,S,R*)-LOHO$^-$ intermediate, highlighting stabilizing interactions with the side chains of Lys1089, Arg1142, and Tyr1168, as well as the backbone of Ser1068. Protons on the enzyme are omitted for clarity.

Mechanism. Taken together, the results support the following proposed mechanism for dinoflagellate bioluminescence (Figure 4.9). (*E,S*)-LH_2 and O_2 bind to the enzyme, with LH_2 possibly being deprotonated by Glu1105 (if LH_2 is not deprotonated when O_2 binds, then deprotonation is the next step). Deprotonation is stereospecific and removes the pro-*S* hydrogen from C13^2. O_2 then attacks the sp^2 hybridized C13^2 from the *re* face to generate (*E,S,R*)-LOO$^-$. After protonation by Glu1105 to form (*E,S,R*)-LOOH, C13^2 is deprotonated and electronic rearrangement leads to the creation of a high-energy,

open-shell singlet. The high-energy intermediate emits a photon and decays to (*E,S,R*)-LOHO$^-$, which then undergoes stereospecific protonation by Tyr1168 to eliminate water.

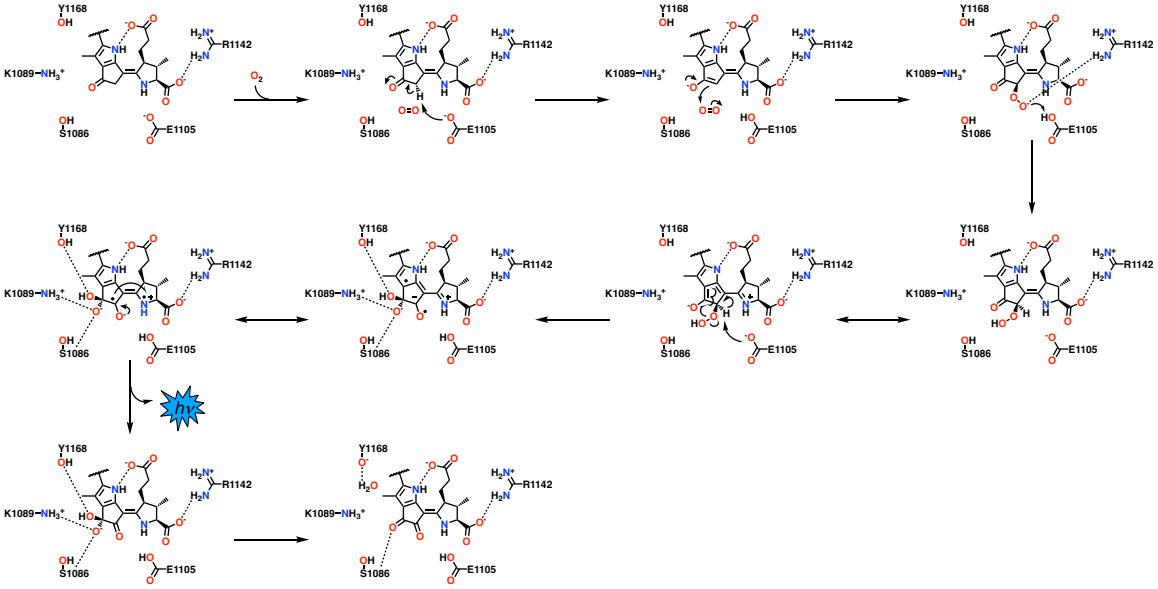

Figure 4.9 Mechanistic proposal for the dinoflagellate bioluminescence reaction.

Mutagenesis Candidate Residues. Using the mechanistic predictions from our calculations, roles for several residues in the β barrel have been postulated (*vida supra*). Table 4.2 compiles the results of said predictions and proposes mutagenesis experiments for each. If completed, such a suite of experiments would strongly corroborate the proposed mechanism (Figure 4.9). More conservative substitutions for Glu1105 and Arg1142, such as to Asp and Lys, respectively, should retain functionality but reduce activity and potentially lead to color-shifted emission.

Table 4.2 Residues identified as catalytically important in the bioluminescent reaction mechanism, their proposed role in catalysis, and predicted effect of mutagenesis experiments.

Residue	Proposed Role	Predicted Effect of Mutation
S1068	Intermediate binding and stabilization	Reduced activity
K1089	Intermediate stabilization	Reduced activity with color change
E1105	Catalytic base	No activity
R1142	Intermediate binding and stabilization	No activity
Y1168	Stereospecific protonation	Mixed mono/di-oxygenation

Identifying residues that will red-shift the emitted photon by changing the electrostatic and hydrogen bonding environment of the active is key to developing the dinoflagellate luciferase system for cellular imaging. While the system brings the unique aspect of pH-dependent activation, it is disadvantaged due to the emission being in the blue end of the visible spectrum. Visible emission is necessary for the postulated biological function of dinoflagellate bioluminescence, the burglar alarm hypothesis, but visible wavelengths due not penetrate tissue well. By red-shifting the emission with mutated luciferase, potentially into the infrared regime, the enzyme system could be adapted for much more robust use in biological imaging.

Critically, this study illustrates the ability of molecular dynamics to ‘freeze’ the making and breaking of chemical bonds and study the interactions of potentially short-lived reaction intermediates with their catalytic enzyme. By doing so, residues important for catalysis, and their conformations, can be studied and used to inform mutagenesis experiments and to help ensure the robustness of quantum mechanics/molecular mechanics (QM/MM) calculations conducted on the system. The dinoflagellate luciferase bioluminescent reaction provides an excellent model system for examining reaction intermediate interactions using molecular dynamics.

In summary, we have employed molecular dynamics techniques to study the stereochemical and regiochemical course of the dinoflagellate bioluminescence reaction. By simulating each possible stereoisomer of the substrate, product, and postulated intermediates, we highlight binding poses and interactions that provide a model of both which stereoisomer is most likely the substrate in vivo and how the stereo- and regiochemical course of the reaction proceeds. We have shown that (*E,S*) is the most likely candidate for substrate stereochemistry on the basis of molecular dynamics simulations of proposed intermediates, using the simulations to propose a mechanism of how dinoflagellate luciferase is able to generate light. Our results help to provide insight into one of nature's least-understood bioluminescent systems.

4.5 References

1. Haddock, S. H.; Moline, M. A.; Case, J. F. Bioluminescence in the Sea. *Ann. Rev. Mar. Sci.* **2010** *2*, 443–493.

2. Maldonado, E. M.; Latz, M. I. Shear-Stress Dependence of Dinoflagellate Bioluminescence. *Biol. Bull.* **2007** *212*, 242–249.

3. Wilson, T.; Hastings, J. W. Bioluminescence: Living Lights, Lights for Living; Harvard University Press: Cambridge, Massachusetts, 2013.

4. Nakamura, H.; Kishi, Y.; Shimomura, O.; Morse, D.; Hastings, J. W. Structure of dinoflagellate luciferin and its enzymic and nonenzymic air-oxidation products. *J. Am. Chem. Soc.* **1989** *111*, 7607–7611.

5. Topalov, G.; Kishi, Y. Chlorophyll Catabolism Leading to the Skeleton of Dinoflagellate and Krill Luciferins: Hypothesis and Model Studies. *Angew. Chem. Int. Ed.* **2001** *40*, 3892–3894.

6. Johnson, C. H.; Inouó, S.; Flint, A.; Hastings, J. W. Compartmentalization of algal bioluminescence: autofluorescence of bioluminescent particles in the dinoflagellate Gonyaulax as studied with image-intensified video microscopy and flow cytometry. *J. Cell Biol.***1985** *100*, 1435–1446.

7. Johnson, C. H.; Hastings, J. W. The Elusive Mechanism of the Circadian Clock: The quest for the chemical basis of the biological clock is beginning to yield tantalizing clues. *Am. Sci.* **1986** *74*, 29–37.

8. De Sa, R.; Hastings, J. W.; Vatter, A. E. Luminescent "Crystalline" Particles: An Organized Subcellular Bioluminescent System. *Science* **1963** *141*, 1269–1270.

9. Nicolas, M. T.; M., S. B.; W., H. J. The ultrastructural localization of luciferase in three bioluminescent dinoflagellates, two species of Pyrocystis, and Noctiluca, using antiluciferase and immunogold labelling. *J. Cell. Sci.* **1987** *87*, 189–196.

10. von Dassow, P.; I., L. M. The role of Ca^{2+} in stimulated bioluminescence of the dinoflagellate *Lingulodinium polyedrum*. *J. Exp. Biol.* **2002** *205*, 2971– 2986.

11. Chen, A. K.; Latz, M. I.; Sobolewski, P.; Frangos, J. A. Evidence for the role of G-proteins in flow stimulation of dinoflagellate bioluminescence. *Am. J. Physiol. Regul. Integr. Comp. Physiol*. 2007, 292, R2020– R2027.

12. Rodriguez, J. D.; Haq, S.; Bachvaroff, T.; Nowak, K. F.; Nowak, S. J.; Morgan, D.; Cherny, V. V.; Sapp, M. M.; Bernstein, S.; Bolt, A.; DeCoursey, T. E.; Place, A. R.; Smith, S. M. E. Identification of a vacuolar proton channel that triggers the bioluminescent flash in dinoflagellates. *PLoS ONE* **2017** *12*, 1–24.

13. Fogel, M.; Hastings, J. W. Bioluminescence: mechanism and mode of control of scintillon activity. *Proc. Natl. Acad. Sci. U.S.A.* **1972** *69*, 690–693.

14. Li, L.; Liu, L.; Hong, R.; Robertson, D.; Hastings, J. W. N-Terminal Intramolecularly Conserved Histidines of Three Domains in Gonyaulax Luciferase Are Responsible for Loss of Activity in the Alkaline Region. *Biochemistry* **2001** *40*, 1844–1849.

15. Morse, D.; Pappenheimer Jr., A. M.; Hastings, J. W. Role of a luciferin-binding protein in the circadian bioluminescent reaction of *Gonyaulax polyedra*. *J. Biol. Chem.* **1989** *264*, 11822–11826.

16. Liu, L.; Wilson, T.; Hastings, J. W. Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide. *Proc. Natl. Acad. Sci. U.S.A.* **2004** *101*, 16555–16560.

17. Schultz, L. W.; Liu, L.; Cegielski, M.; Hastings, J. W. Crystal structure of a pH-regulated luciferase catalyzing the bioluminescent oxidation of an open tetrapyrrole. *Proc. Natl. Acad. Sci. U.S.A.* **2005** *102*, 1378–1383.

18. Donnan, P. H.; Ngo, P. D.; Mansoorabadi, S. O. Constant pH Accelerated Molecular Dynamics Investigation of the pH Regulation Mechanism of Dinoflagellate Luciferase. *Biochemistry* **2018** *57*, 295–299.

19. Kamerlin, N.; Delcey, M. G.; Lindh, R. Unravelling the mechanism of pH-regulation in dinoflagellate luciferase. *Int. J. Biol. Macromol.* **2020** *164*, 2671–2680.

20. Ngo, P. D.; Mansoorabadi, S. O. Investigation of the Dinoflagellate Bioluminescence Mechanism: Chemically Initiated Electron Exchange Luminescence or Twisted Intramolecular Charge Transfer? *ChemPhotoChem* **2017** *1*, 383–387.

21. Fresneau, C.; Blanot, D.; Hill, M.; Dupaix, A.; Arrio, B. Demonstration of the peptidic nature of luciferin from *Pyrocystis lunula*. *Int. J. Pept. Res. Ther.* **1988** *31*, 126–129.

22. Wang, M.-Y.; Liu, Y.-J. Theoretical Study of Dinoflagellate Bioluminescence. *Photochem. Photobiol.* **2017** *93*, 511–518.

23. Case, D. et al. AMBER 2019, University of California, San Francisco. 2019.

24. Ngo, P. D. A Story About Light and Energy: Mechanistic and Biosynthetic Investigations of Bioluminescence and Methanogenesis. Ph.D. thesis, Auburn University, Auburn, AL, 2016.

25. Avogadro: an open-source molecular builder and visualization tool. Version 1.20. http://avogadro.cc/.

26. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2012** *4*, 17.

27. Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comp. Chem.* **2010** *31*, 455–461.

28. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993** *97*, 10269–10280.

29. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem, Theory Comput.* **2015**, *11*, 3696–3713.

30. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comp. Chem.* **2004** *25*, 1157–1174.

31. Jorgensen, W. L.; Madura, J. D. Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. *J. Am. Chem. Soc.* **1983** *105*, 1407–1413.

32. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

33. Xiang, T.-X.; Liu, F.; Grant, D. M. Generalized Langevin equations for molecular dynamics in solution. *J. Chem. Phys.* **1991** *94*, 4463–4471.

34. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977** *23*, 327–341.

35. Frisch, M. J. et al. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

36. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993** *98*, 5648–5652.

37. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988** *37*, 785–789.

38. Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971** *54*, 724–728.

39. Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972** *56*, 2257–2261.

40. Neese, F. The ORCA program system. *WIREs Comp. Mol. Sci.* **2012**, 2, 73–78.

41. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005** *7*, 3297–3305.

42. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006** *8*, 1057–1065.

43. Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009** *356*, 98–109.

44. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013** *9*, 3084–3095.

45. Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996** *14*, 33–38.

46. Schrödinger, LLC, PyMOL The PyMOL Molecular Graphics System, Version 1.8.

Chapter 5: Broken-Symmetry Density Functional Theory Analysis of the Ω Intermediate in Radical *S*-Adenosyl-L-methionine Enzymes: Evidence for a Near-Attack Conformer over an Organometallic Species

A version of this chapter was published: Donnan, P. H.; Mansoorabadi, S. O. *J. Am. Chem. Soc.* **2022**, *144 (8)*, 295−299, and associated Supporting Information.

5.1 Abstract

Radical *S*-adenosyl-L-methionine (SAM) enzymes are found in all domains of life and catalyze a wide range of biochemical reactions. Recently, an organometallic intermediate, Ω, has been experimentally implicated in the 5′-deoxyadenosyl radical generation mechanism of the radical SAM superfamily. In this work, we employ broken-symmetry density functional theory to evaluate several structural models of Ω. The results show that the calculated hyperfine coupling constants (HFCCs) for the proposed organometallic structure of Ω are inconsistent with the experiment. In contrast, a near-attack conformer of SAM bound to the catalytic [4Fe−4S] cluster, in which the distance between the unique iron and SAM sulfur is ∼3 Å, yields HFCCs that are all within 1 MHz of the experimental values. These results clarify the structure of the ubiquitous Ω intermediate and suggest a paradigm shift reversal regarding the mechanism of SAM cleavage by members of the radical SAM superfamily.

5.2 Introduction

Utilizing *S*-adenosyl-L-methionine (SAM) and iron−sulfur clusters, radical SAM (RS) enzymes are one of nature's most preeminent ways of generating radicals for use in biological catalysis. First identified as an enzyme superfamily in 2001,[1] RS enzymes have

since been found in every domain of life and have been shown to catalyze an ever-increasing set of chemical reactions, from unactivated C−H bond functionalization to complex skeletal rearrangements.[2] Central to canonical RS chemistry is the radical generation mechanism: SAM bound to a reduced $[4Fe{-}4S]^{1+}$ cluster is reductively cleaved to form the 5′-deoxyadenosyl radical (5′-dAdo·) and methionine bound to an oxidized $[4Fe{-}4S]^{2+}$ cluster (Figure 1.6). It was previously thought that the 5′-dAdo· was formed directly through inner-sphere electron transfer within the SAM-$[4Fe{-}4S]^{1+}$ complex. However, recent electron paramagnetic resonance (EPR) and electron−nuclear double resonance (ENDOR) spectroscopic investigations have identified a transient intermediate common to many enzymes in the RS superfamily. Initially found in pyruvate formate-lyase activating enzyme (PFL-AE), the so-called Ω intermediate (Figure 1.6) was then found in many other RS enzymes, including lysine-2,3-aminomutase (LAM), spore photoproduct lyase (SPL), and the [FeFe]-hydrogenase maturation protein HydG.[3−5] A similar intermediate was later found in the noncanonical RS enzyme Dph2 involved in diphthamide biosynthesis.[6]

The detection of Ω, in tandem with the discovery that SAM can be photolyzed to generate 5′-dAdo· in the absence of a substrate, has influenced a growing number of recent experiments, including studies of ChuW, SuiB, ArsS, HydG, and SPL.[7,8] Fundamental studies of RS enzymes have continued to provide evidence of or support for Ω.[7,9] Synthetic models of the organometallic Ω structure have been prepared and have demonstrated Fe−C bond homolysis reactivity.[10] Computational investigations of Ω, on the contrary, have been sparse. A quantum mechanics/molecular mechanics (QM/ MM) investigation of the transition states between a SAM- bound $[4Fe{-}4S]^{1+}$ cluster and Ω has

been conducted, and it was suggested that organometallic Ω could be a possible shunt product of PFL-AE due its large rate-limiting transition state (>16.7 kcal/mol).[11]

Yet the rates of Ω formation and decay in PFL-AE were shown to be catalytically competent, and its pervasiveness among characterized RS enzymes indicates that it is not a shunt product.[4] Turning to the initial experiments used to determine the structure of Ω may help provide a bridge between the QM/MM-derived reaction energetics indicating a possible shunt product and the experimental observations of catalytic competency. Well-defined g values and, more importantly, hyperfine coupling constants (HFCCs) were determined for Ω, and these values can be compared with theoretical estimates from structural models to provide evidence for or against Ω as an organometallic species. Here we report broken-symmetry density functional theory (BS-DFT) calculations of different models of the Ω intermediate to obtain 5′-^{13}C, 4′-^{13}C, methyl-^{13}C, and 5′- C−^{1}H HFCCs for comparison with experiment. BS-DFT has been a powerful tool in the study of the electronic structure of antiferromagnetically coupled metal centers in biological systems, including iron−sulfur clusters, the oxygen-evolving complex of photosystem II, and the FeMo-cofactor of nitrogenase (including predicting the identity of ligand X ahead of experiments).[12−14] Utilizing the fundamental BS-DFT methodology established by Noodleman and coworkers, combined with the more recent HFCC calculation approach developed by Pantazis, Neese, and coworkers, we carried out geometry optimization and single-point energy calculations of the proposed organometallic structure of Ω and of SAM bound to a [4Fe−4S]$^{1+}$ cluster in the conformation observed in the crystal structure of PFL-AE.[15−18] Because it is unknown which low-spin BS state/valence isomer

corresponds best to the physical state of Ω in the active site of RS enzymes, we report an average value for the spin-projected HFCCs of the BS valence isomers with $m_S = +1/2$.

5.3 Computational Methodology

Initial *S*-adenosyl-L-methionine (SAM)-bound [4Fe-4S]$^{1+}$ cluster models were built from the crystal structure of the pyruvate formate-lyase activating enzyme (PFL-AE) complex with SAM (PDB ID: 3CB8).[18] Since no structures are available for the proposed organometallic Ω intermediate, the input structure was built from the PFL-AE structure using the Avogadro software package and subjected to preliminary geometry optimization in the Gaussian16 software package at the HF/STO-3G level of theory.[19-21] All other calculations were performed with the ORCA software package (version 4.2.1, unless otherwise noted).[17a,17b] All calculations were performed with the CPCM solvent model with a dielectric constant ε = 20 to model protein interior.[22,23] Coordinating cysteine residues were modeled as methanethiolate ligands with the carbons constrained to their crystallographic positions for all geometry optimizations. The geometry optimizations of the crystallographic SAM-bound cluster and the proposed organometallic structure were otherwise unconstrained.[‡‡] As reported for other broken-symmetry density functional theory (BS-DFT) calculations in the literature, the BP86 functional was used for geometry optimizations.[17c,17d,24] The Ahlrichs def2-SVP basis set, with the def2-J auxiliary basis set, was used on all atoms except Fe and S.[17f,17g] Fe and S atoms used def2-TZVP with the def2-J auxiliary basis set.[17f] For unique iron-SAM sulfur constrained distance calculations, the same methodology was used except the RIJCOSX

[‡‡] These calculations were performed with ORCA version 4.0.1.

approximation was employed and S atoms used the def2-SVP basis set.[17f,17j] A tighter integration grid during self-consistent field (SCF) iterations with an increased grid for final energy evaluation was used for all calculations (Grid5 and FinalGrid6, respectively), along with the TIGHTSCF setting.

The numbering convention used for the Fe sites is shown in Figure 5.1. Using this convention, we can refer to broken-symmetry (BS) states/valence isomers by whether they are assigned majority α or β spin on each site. For example, the βαβα state carries majority β spin on Fe sites 1 and 3 and majority α spin on sites 2 and 4. An isosurface plot of the calculated spin density of this BS valence isomer is shown in Figure 5.1.

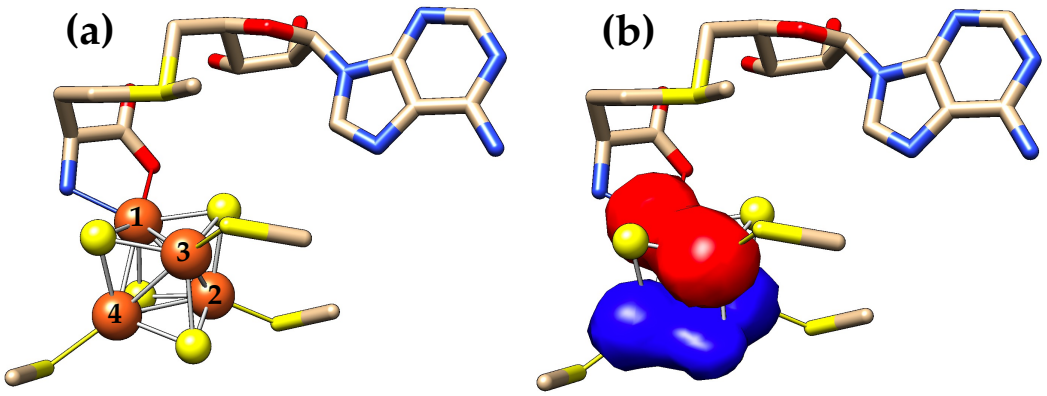

Figure 5.1. Numbering convention used for Fe sites in the broken-symmetry density functional theory (BS-DFT) calculations (a) and isosurface plot of the calculated spin density for the βαβα valence isomer of the crystallographic SAM-bound cluster geometry (b). The β spin density is shown in red and α spin density in blue (surface depicted at ± 0.0035 au after interpolative smoothing). Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.

Geometry optimization was performed on the high-spin (αααα) and all low-spin (S = 1/2) BS valence isomers (ααββ, αβαβ, αββα, ββαα, βαβα, and βααβ) for each structural model, and the lowest energy geometry was chosen for single point calculations. We note this differs slightly from the previous strategy outlined for oligonuclear transition-metal complexes, which used the high-spin geometry for single

point calculations.[16] Following the methodology of Pantazis *et al.*,[16] all single point calculations were carried out with the TPSSh functional.[17e] The EPR-II basis set was used on all atoms except S and Fe.[17h] The Ahlrichs def2-TZVP basis was used on S atoms, and the CP(PPP) basis was used on Fe atoms.[17f,17i] The size of the integration grid was increased to 7 for Fe atoms.

Hyperfine coupling constants (HFCCs) calculated with BS-DFT must be corrected by spin projection. We utilized the spin projection procedure developed by Pantazis *et al.* that builds upon the BS-DFT methodology of Noodleman and coworkers.[15,16] For a diagrammatic and informative discussion of this method, we refer the reader to reference 25.

5.4 Results and Discussion

Calculations of the organometallic species revealed HFCCs for 5′-^{13}C and 4′-^{13}C that were an order of magnitude too large to be consistent with the experimentally observed values (Table 3.1). Similarly, the calculated 5′-C−^{1}H HFCC is larger than the experimental value (Table 3.1). In contrast, the corresponding HFCCs of the geometry-optimized, crystallographic SAM-bound [4Fe−4S]$^{1+}$ cluster are five to eight times too small to reproduce the experimental values. Because neither the proposed organometallic structure nor the crystallographic SAM-bound cluster has HFCCs consistent with experiment, we then calculated structures of SAM near-attack conformers (SAM-NACs) by constraining the distance between the unique iron of the cluster and the sulfur atom of SAM.[26]

Table 5.1. Calculated and Experimental Hyperfine Coupling Constants (in MHz) for structural models of Ω.[a]

Structural Model	5′-^{13}C	4′-^{13}C	methyl-^{13}C	5′-C–^{1}H[b]	Reference
Experiment	9.4	~0.5	~0.5	7-8	3,4
Proposed Organometallic Structure	140.4 ± 12.6	4.4 ± 2.9	0.03 ± 0.01	11.7 ± 14.1	This work
Crystallographic SAM-bound Cluster	1.7 ± 1.6	0.1 ± 0.1	1.1 ± 0.5[c]	1.0 ± 1.2	This work
SAM-NAC Structure (3.0 Å Fe-S Distance)	8.8 ± 1.0	0.6 ± 0.05	1.3 ± 0.8[c]	6.3 ± 0.2	This work

[a]The average and standard deviation of the absolute value of HFCCs from maximal site spin, m_s = +1/2 BS determinants are reported (see text).

[b]The larger HFCC of the two 5′-C–^{1}H's are reported, as would be measured experimentally.

[c]Due to close proximity to both apical Fe atoms, this methyl group is treated as a bridging ligand during spin projection.

In the optimized crystallographic geometry, this distance is 3.53 Å; distance constraints of 3.2, 3.0, 2.8, and 2.6 Å were employed for the NAC calculations. When analyzed after single-point energy calculations on geometries optimized at each fixed distance, the NACs gave 5′-^{13}C, 4′-^{13}C, and 5′- C–^{1}H hyperfine couplings smaller than the organometallic structure but larger than the crystallographic SAM-bound cluster. The experimentally determined HFCCs lie within the range of these calculated values, which vary exponentially with the Fe−S distance (Figure 5.2). For example, the 5′-^{13}C HFCC increases from 4.5 MHz at 3.2 Å to 13.7 MHz at 2.8 Å, bracketing the experimental value of 9.4 MHz.

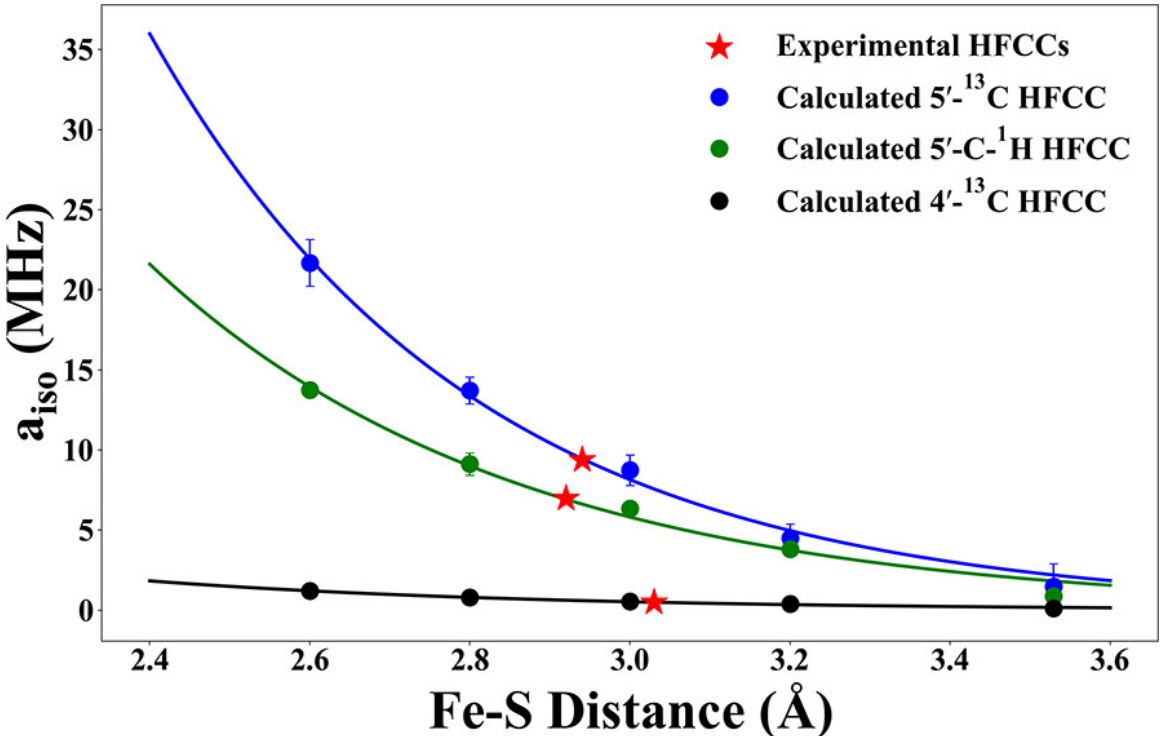

Figure 5.2. Calculated ^{13}C and ^{1}H isotropic hyperfine coupling constants (HFCCs) for SAM near-attack conformer (SAM-NAC) geometries with fixed unique iron−SAM sulfur distances. Calculated values are averaged over all BS valence isomers with $m_S = +1/2$ for each fixed distance, with error bars showing the standard deviation. The average HFCCs for 5′-^{13}C (blue), 5′-C−^{1}H (green), and 4′-^{13}C (black) are shown as circles, and best-fit curves to a decaying exponential equation are shown as lines. The red stars indicate the experimentally determined HFCCs for each curve. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.

Assuming that the experimentally detected species is a SAM-NAC, we can use the best-fit curves for each HFCC to determine a unique iron−SAM sulfur distance that corresponds to the experimental values. For the 5′-^{13}C, 4′-^{13}C, and 5′-C−^{1}H HFCCs, the estimated Fe−S distance for Ω is self- consistent, ranging from 2.92 to 3.03 Å (Figure 5.2). Thus, of all of the structural models considered, a SAM-NAC structure with an Fe−S distance of 3.0 Å most closely matches the experimental isotropic HFCCs of Ω (Table 3.1). Moreover, the calculated axial dipolar coupling for the 5′-^{13}C of this structure, 2T = 6.0 ± 0.6 MHz, also closely matches the experimental value of ~5.3 MHz,

whereas the value for the organometallic model is again an order of magnitude too large ($2T = 44.0 \pm 8.8$ MHz). The only calculated HFCC whose best-fit curve did not predict an Fe−S bond distance of ~3 Å for Ω was that for the methyl-^{13}C of SAM (Figure B1). However, this atom is unique in that it is approximately equidistant to each of the Fe atoms in the apical face of the cluster in the SAM-NAC structures (Table B1) and must therefore be treated similar to a bridging ligand during spin projection. Because the spin projection factors of the apical Fe atoms are opposite in sign, the net result is a cancellation effect that greatly diminishes the calculated methyl-^{13}C HFCC. This explains the unexpectedly small experimental methyl-^{13}C HFCC that was taken as evidence of 5′-C−S bond cleavage and the migration of the methionine side chain away from the cluster upon Ω formation.

The other experimental observation that was critical for the initial assignment of Ω as an organometallic species was the relatively large experimental 5′-^{13}C HFCC.[3] For this model, the 5′-C is alpha to the spin-bearing unique iron. β-Hyperfine interactions are also known to give rise to large HFCCs in organic radicals, depending on the dihedral angle of the relevant bond relative to the p orbital of the spin-bearing carbon. Similarly, the large isotropic 5′-^{13}C HFCC of the SAM-NAC structure can be rationalized as a β-hyperfine interaction/ hyperconjugative effect arising from the alignment of the 5′-C−S bond of SAM with the d orbital(s) of the adjacent unique iron. This gives rise to a more modest 5′-13C HFCC in the SAM-NAC structure that is consistent with the experimental value for Ω, as opposed to the very large α coupling predicted to be observed with the organometallic structure.

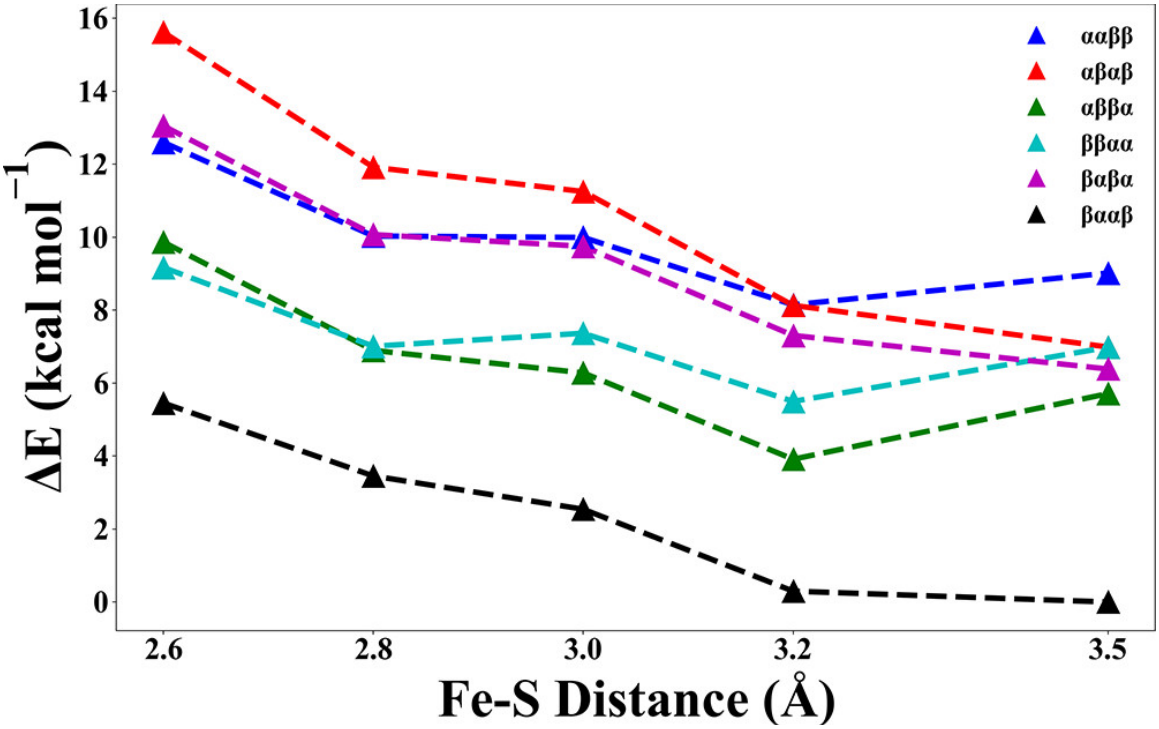

Figure 5.3. Relative energy curves for the broken-symmetry (BS) valence isomers of the SAM near-attack conformers (SAM-NACs) as a function of the unique iron−SAM sulfur distance. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.

To verify the energetic feasibility of the SAM-NAC structure, we examined the relative energy curves of the system as the distance between the unique iron and SAM sulfur was constrained (Figure 5.3). We found that as the distance was shortened from the optimized crystallographic distance of 3.53 Å to the predicted distance of 3 Å, the energy increased by <3 kcal mol^{-1} for each low-spin BS valence isomer. As the Fe−S distance was further decreased from 3.0 to 2.8 Å, the change in energy was even smaller, <1 kcal mol^{-1}. This provides a low-energy plateau in the potential energy surface and suggests that there could be a local minimum within the region corresponding to the SAM-NAC structure of Ω that can be stabilized within the active site of RS enzymes.

Indeed, Rohac *et al*. recently provided compelling evidence for this by determining the structure of the RS enzyme HydE after its reaction was reductively

triggered in crystallo.[27] Electron density within the active site was consistent with a 50−50 mixture of cleaved and uncleaved SAM.[27] Post cleavage, the sulfur of the liberated methionine formed a 2.87 Å coordinate bond with the unique iron, whereas the Fe−S distance for the intact SAM-bound cluster was 3.08 Å, which is fully consistent with our predicted SAM-NAC structure of Ω. Moreover, no density was attributable to an organometallic structure containing a 5′-C−Fe bond or an altered methionine geometry.[27]

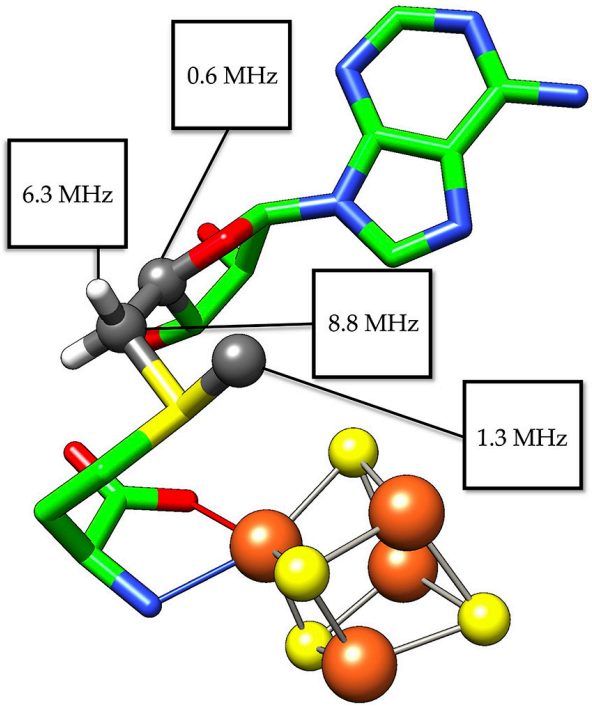

Figure 5.4. Proposed structure for the SAM near-attack conformer (SAM-NAC) structure of the Ω intermediate, with calculated HFCCs labeled. Hydrogen atoms (except those on the 5′-C) and coordinated methanethiolate ligands used in the calculations are omitted for clarity. Reproduced with permission from *J. Am. Chem. Soc.* **2022**, *144 (8)*, 3381–3385. Copyright 2022 American Chemical Society.

Thus, we offer a new proposed structure for the Ω intermediate of RS catalysis: a SAM-NAC with an Fe−S bond distance of ~3 Å (Figure 5.4). Such a species is consistent with all known experimental and computational evidence and reconciles previous findings that an organometallic Ω is likely to be a shunt product, whereas Ω itself is known to be catalytically competent. Our model of Ω as a SAM-NAC is

consistent with the Frey catalytic model of 5′-dAdo· formation by RS enzymes and can be thought of as the structure along the reaction coordinate wherein the nonbonding electron pair of the sulfonium of SAM is brought in close proximity to the unique iron of the cluster.[28] Such a model also obviates the need for methionine gymnastics after SAM cleavage to allow 5′-dAdo· coupling to the $[4Fe{-}4S]^{2+}$ cluster and the requirement of having to pass through the same transition state twice to effect substrate hydrogen atom abstraction. Our reassignment of Ω as a NAC of SAM bound to the catalytic [4Fe−4S] cluster thus has significant implications for the radical generation mechanism of RS enzymes and should help guide future experimental and computational studies of this important enzyme superfamily.

5.5 References

1. Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **2001**, *29 (5)*, 1097−1106.

2. (a) Frey, P. A.; Hegeman, A. D.; Ruzicka, F. J. The Radical SAM Superfamily. *Crit. Rev. Biochem. Mol. Bio.* **2008**, *43 (1)*, 63−88. (b) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical *S*-adenosylmethonine Enzymes. *Chem. Rev.* **2014**, *114 (8)*, 4229−4317. (c) Landgraf, B. J.; McCarthy, E. L.; Booker, S. J. Radical *S*-Adenosylmethionine Enzymes in Human Health and Disease. *Annu. Rev. Biochem.* **2016**, *85 (1)*, 485−514. (d) Bridwell-Rabb, J.; Grell, T. A. J.; Drennan, C. L. A Rich Man, Poor Man Story of *S*-Adenosylmethionine and Cobalamin Revisited. *Annu. Rev. Biochem.* **2018**, *87 (1)*, 555−584. (e) Nicolet, Y. Structure−Function relationships of radical SAM enzymes. *Nature Catalysis* **2020**, *3*, 337−350.

3. Horitani, M.; Shisler, K.; Broderick, W. E.; Hutcheson, R. U.; Duschene, K. S.; Marts, A. R.; Hoffman, B. M.; Broderick, J. B. Radical SAM Catalysis via an organometallic intermediate with an Fe- [5′-C]-deoxyadenosyl bond. *Science* **2016**, *352 (6287)*, 822−825.

4. Byer, A. S.; Yang, H.; McDaniel, E. C.; Kathiresan, V.; Impano, S.; Pagnier, A.; Watts, H.; Denler, C.; Vagstad, A. L.; Piel, J.; Duschene, K. S.; Shepard, E. M.; Shields,

T. P.; Scott, L. G.; Lilla, E. A.; Yokoyama, K.; Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Paradigm Shift for Radical-*S*-Adenosyl-L-Methionine Reactions: The Organometallic Intermediate Ω is Central to Catalysis. *J. Am. Chem. Soc.* **2018**, *140 (28)*, 8634−8638.

5. Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Mechanism of Radical Initiation in the Radical-*S*-Adenosyl-L-Methionine Superfamily. *Acc. Chem. Res*. **2018**, *51 (11)*, 2611−2619.

6. Dong, M.; Kathiresan, V.; Fenwick, M. K.; Torelli, A. T.; Zhang, Y.; Caranto, J. D.; Dzikovski, B.; Sharma, A.; Lancaster, K. M.; Freed, J. H.; Ealick, S. E.; Hoffman, B. M.; Lin, H. Organometallic and radical intermediates reveal mechanism of diphthamide biosynthesis. *Science* **2018**, *359 (6381)*, 1247−1250.

7. Yang, H.; McDaniel, E. C.; Impano, S.; Byer, A. S.; Jodts, R. J.; Yokoyama, K.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. The Elusive 5′-Deoxyadenosine Radical: Captured and Characterized by Electron Paramagnetic Resonance and Electron Nuclear Double Resonance Spectroscopies. *J. Am. Chem. Soc*. **2019**, *141 (30)*, 12139−12146.

8. (a) Mathew, L. G.; Beattie, N. R.; Pritchett, C.; Lanzilotta, W. N. New Insight into the Mechanism of Anaerobic Heme Degradation. *Biochemistry* **2019**, *58 (46)*, 4641−4654. (b) Balo, A. R.; Caruso, A.; Tao, L.; Tantillo, D. J.; Seyedsayamdost, M. R.; Britt, D. R. Trapping a cross-linked lysine−tryptophan radical in the catalytic cycle of the radical SAM enzyme SuiB. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118 (21)*, No. e2101571118. (c) Cheng, J.; Ji, W.; Ma, S.; Ji, X.; Deng, Z.; Ding, W.; Zhang, Q. Characterization and Mechanistic Study of the Radical SAM Enzyme ArsS Involved in Arsenosugar Biosynthesis. *Angew. Chem., Int. Ed.* **2021**, *60 (14)*, 7570−7575. (d) Impano, S.; Yang, H.; Shepard, E. M.; Swimley, R.; Pagnier, A.; Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. S-Adenosyl-L-ethionine is a Catalytically Competent Analog of S-Adenosyl-L-methionine (SAM) in the Radical SAM Enzyme HydG. *Angew. Chem., Int. Ed.* **2021**, *60 (9)*, 4666−4672. (e) Pagnier, A.; Yang, H.; Jodts, R. J.; James, C. D.; Shepard, E. M.; Impano, S.; Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Radical SAM Enzyme Spore Photoproduct Lyase: Properties of the Ω Organometallic Intermediate and Identification of Stable Protein Radicals Formed during Substrate-Free Turnover. *J. Am. Chem. Soc.* **2020**, *142 (43)*, 18652−18660.

9. (a) Miller, S. A.; Bandarian, V. Analysis of Electrochemical Properties of *S*-Adenosyl-L-methionine and Implications for Its Role in Radical SAM Enzymes. *J. Am. Chem. Soc.* **2019**, *141 (28)*, 11019−11026. (b) Yang, H.; Impano, S.; Shepard, E. M.; James, C. D.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. Photoinduced Electron Transfer in a Radical SAM Enzyme Generates an *S*-Adenosylmethionine Derived Methyl Radical. *J. Am. Chem. Soc.* **2019**, *141 (40)*, 16117−16124. (c) Sayler, R. I.; Stich, T. A.; Joshi, S.; Cooper, N.; Shaw, J. T.; Begley, T. P.; Tantillo, D. J.; Britt, R. D. Trapping and Electron Paramagnetic Resonance Characterization of the 5′ dAdo· Radical in Radical *S*-Adenosyl Methionine Enzyme Reaction with a Non-Native Substrate. *ACS Cent. Sci.*

2019, *5 (11)*, 1777−1785. (d) Impano, S.; Yang, H.; Jodts, R. J.; Pagnier, A.; Swimley, R.; McDaniel, E. C.; Shepard, E. M.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. Active-Site Controlled, Jahn−Teller Enabled Regioselectivity in Reductive S−C Bond Cleavage of *S*-Adenosylmethionine in Radical SAM Enzymes. *J. Am. Chem. Soc*. **2021**, *143 (1)*, 335−348.

10. (a) Ye, M.; Thompson, N. B.; Brown, A. C.; Suess, D. L. M. A Synthetic Model of Enzymatic $[Fe_4S_4]$−Alkyl Intermediates. *J. Am. Chem. Soc.* **2019**, *141 (34)*, 13330−13335. (b) Brown, A. C.; Suess, D. L. M. Reversible Formation of Alkyl Radicals at $[Fe_4S_4]$ Clusters and Its Implications for Selectivity in Radical SAM Enzymes. *J. Am. Chem. Soc.* **2020**, *142 (33)*, 14240−14248.

11. Zhao, C.; Li, Y.; Wang, C.; Chen, H. Mechanistic Dichotomy in the Activation of SAM by Radical SAM Enzymes: QM/MM Modeling Deciphers the Determinant. *ACS Catal.* **2020**, *10*, 13245−13250.

12. (a) Noodleman, L. A Model for the Spin States of High-Potential Iron-Sulfur $[Fe_4S_4]^{3+}$ Proteins. *Inorg. Chem*. **1988**, *27 (20)*, 3677−3679. (b) Noodleman, L. Exchange Coupling and Resonance Delocalization in Reduced $[Fe_4S_4]^{+}$ and $[Fe_4Se_4]^{+}$ Clusters. 2. A Generalized Nonlinear Model for Spin-State Energies and EPR and Hyperfine Properties. *Inorg. Chem.* **1991**, *30 (2)*, 256−264. (c) Torres, R. A.; Lovell, T.; Noodleman, L.; Case, D. A. Density Functional and Reduction Potential Calculations of Fe_4S_4 Clusters. *J. Am. Chem. Soc.* **2003**, *125 (7)*, 1923−1936. (d) Blachly, P. G.; Sandala, G. M.; Giammona, D. A.; Bashford, D.; McCammon, J. A.; Noodleman, L. Broken-Symmetry DFT Computations for the Reaction Pathway of IspH, an Iron−Sulfur Enzyme in Pathogenic Bacteria. *Inorg. Chem.* **2015**, *54 (13)*, 6439−6461.

13. (a) Schinzel, S.; Schraut, J.; Arbuznikov, A. V.; Siegbahn, P. E. M.; Kaupp, M. Density Functional Calculations of ^{55}Mn, ^{14}N, and ^{13}C Electron Paramagnetic Resonance Parameters Support an Energetically Feasible Model System for the S2 State of the Oxygen-Evolving Complex of Photosystem II. *Chem. Eur. J.* **2010**, *16 (34)*, 10424−10438. (b) Pantazis, D. A.; Ames, W.; Cox, N.; Lubitz, W.; Neese, F. Two Interconvertible Structures that Explain the Spectroscopic Properties of the Oxygen-Evolving Complex of Photosystem II in the S2 State. *Angew. Chem., Int. Ed.* **2012**, *51 (39)*, 9935−9940. (c) Beal, N. J.; Corry, T. A.; O'Malley, P. J. A Comparison of Experimental and Broken-Symmetry Density Functional Theory (BS- DFT) Calculated Electron Paramagnetic Resonance (EPR) Parameters for Intermediates Involved in the S2 to S3 State Transition of Nature's Oxygen Evolving Complex. *J. Phys. Chem. B* **2018**, *122 (4)*, 1394−1407. (d) Yang, K. R.; Lakshmi, K. V.; Brudvig, G. W.; Batista, V. S. Is Deprotonation of the Oxygen-Evolving Complex of Photosystem II during the S1→S2 Transition Suppressed by Proton Quantum Delocalization? *J. Am. Chem. Soc.* **2021**, *143 (22)*, 8324− 8332.

14. Lovell, T.; Li, J.; Liu, T.; Case, D. A.; Noodleman, L. FeMo Cofactor of Nitrogenase: A Density Functional Study of States MN, MOX, MR, and MI. *J. Am. Chem. Soc.* **2001**, *123 (49)*, 12392−12410. (b) Pelmenschikov, V.; Case, D. A.;

Noodleman, L. Ligand-Bound *S* = 1/2 FeMo-Cofactor of Nitrogenase: Hyperfine Interaction Analysis and Implication for the Central Ligand X Identity. *Inorg. Chem.* **2008**, *47 (14)*, 6162−6172. (c) Benediktsson, B.; Bjornsson, R. QM/MM Study of the Nitrogenase MoFe Protein Resting State: Broken-Symmetry States, Protonation States, and QM Region Convergence in the FeMoco Active Site. *Inorg. Chem.* **2017**, *56 (21)*, 13417− 13429. (d) Raugei, S.; Seefeldt, L. C.; Hoffman, B. M. Critical computational analysis illuminates the reductive-elimination mechanism that activates nitrogenase for N2 reduction. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115 (45)*, E10521−E10530.

15. Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J.-M. Orbital interaction, electron delocalization and spin coupling in iron- sulfur clusters. *Coord. Chem. Rev.* **1995**, *144*, 199−244.

16. Pantazis, D. A.; Orio, M.; Petrenko, T.; Zein, S.; Bill, E.; Lubitz, W.; Messinger, J.; Neese, F. A New Quantum Chemical Approach to the Magnetic Properties of Oligonuclear Transition-Metal Complexes: Application to a Model for the Tetranuclear Manganese Cluster of Photosystem II. *Chem. Eur. J.* **2009**, *15 (20)*, 5108−5123.

17. (a) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78. (b) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1327. (c) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33 (12)*, 8822− 8824. (d) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38 (6)*, 3098−3100. (e) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119 (23)*, 12129. (f) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005***, 7 (18)*, 3297−3305. (g) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8 (9)*, 1057−1065. (h) Rega, N.; Cossi, M.; Barone, V. Development and validation of reliable quantum mechanical approaches for the study of free radicals in solution. *J. Chem. Phys.* **1996**, *105 (24)*, 11060. (i) Neese, F. Prediction and interpretation of the ^{57}Fe isomer shift in Mössbauer spectra by density functional theory. *Inorg. Chim. Acta* **2002**, *337*, 181−192. (j) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree−Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree−Fock exchange. *Chem. Phys.* **2009**, *356 (1−3)*, 98−109.

18. Vey, J. L.; Yang, J.; Li, M.; Broderick, W. E.; Broderick, J. B.; Drennan, C. L. Structural basis for glycyl radical formation by pyruvate-formate lyase activating enzyme. *Proc. Natl. Acad. U.S.A.* **2008**, *105 (42)*, 16137−16141.

19. (a) Avogadro: an open-source molecular builder and visualization tool. Version 1.20. https://avogadro.cc. (b) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.;

Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An advanced semantics chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2012**, *4*, 17.

20. Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.

21. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed., rev.; Dover, 1996; pp. 115–122; 184–185.

22. Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102 (11)*, 1995–2001.

23. Blachy, P. G.; Sandala, G. M.; Giammona, D. A.; Liu, T. Bashford, D.; McCammon, J. A.; Noodleman, L. Use of Broken-Symmetry Density Functional Theory to Characterize the IspH Oxidized State: Implications for IspH Mechanism and Inhibition. *J. Chem. Theory Comput.* **2014**, *10 (9)*, 3871–3884.

24. (a) Han, W.-G.; Liu, T.; Lovell, T.; Noodleman, L. Active Site Structure of Class I Ribonucleotide Reductase Intermediate X: A Density Functional Theory Analysis of Structure, Energetics, and Spectroscopy. *J. Am. Chem. Soc*. **2005**, *127 (45)*, 15778–15790. (b) Orio, M.; Pantazis, D. A.; Petrenko, T.; Neese, F. Magnetic and Spectroscopic Properties of Mixed Valence Manganese(III, IV) Dimers: A Systematic Study Using Broken Symmetry Density Functional Theory. *Inorg. Chem.* **2009**, *48 (15)*, 7251–7260. (c) Cox, N.; Ames, W.; Epel, B.; Kulik, L. V.; Rapatskiy, L.; Neese, F.; Messinger, J.; Wieghardt, W.; Lubitz, W. Electronic Structure of a Weakly Antiferromagnetically Coupled MnIIMnIII Model Relevant to Manganese Proteins: A Combined EPR, ^{55}Mn-ENDOR, and DFT Study. *Inorg. Chem.* **2011**, *50 (17)*, 8238–8251. (d) Beal, N. J.; Corry, T. A.; O'Malley, P. J. Comparison between Experimental and Broken Symmetry Density Functional Theory (BS-DFT) Calculated Electron Paramagnetic Resonance (EPR) Parameters of the S2 State of the Oxygen-Evolving Complex of Photosystem II in Its Native (Calcium) and Strontium-Substituted Form. *J. Phys. Chem. B* **2017**, *121 (50)*, 11273– 11283.

25. Orio, M.; Pantazis, D. Successes, challenges, and opportunities for quantum chemistry in understanding metalloenzymes for solar fuels research. *Chem. Commun.* **2021**, *57*, 3952– 3974.

26. Bruice, T. C.; Benkovic, S. J. Chemical Basis for Enzyme Catalysis. *Biochemistry* **2000**, *39 (21)*, 6267−6274.

27. Rohac, R.; Martin, L.; Liu, L.; Basu, D.; Tao, L.; Britt, R. D.; Rauchfuss, T. B.; Nicolet, Y. Crystal Structure of the [FeFe]-Hydrogenase Maturase HydE Bound to Complex-B. *J. Am. Chem. Soc.* **2021**, *143 (22)*, 8499−8508.

28. Cosper, N. J.; Booker, S. J.; Ruzicka, F.; Frey, P. A.; Scott, R. A. Direct FeS Cluster Involvement in Generation of a Radical in Lysine 2,3-Aminomutase. *Biochemistry* **2000**, *39 (51)*, 15668−15673.

Chapter 6: Computationally and Spectroscopically Consistent Models of Paramagnetic Intermediates in Dph2 Catalysis

6.1 Abstract

Dph2 is a non-canonical radical *S*-adenosyl-L-methionine (SAM) which reductively cleaves SAM to form a 3-amino-3-carboxypropyl radical (ACP·) instead of the canonical cleavage resulting in the production of the 5′-deoxyadenosyl radical (dAdo·). ACP· then adds to a histidine residue on elongation factor 2. Two paramagnetic intermediates, intermediate **I** and intermediate **II**, have been identified and characterized by electron paramagnetic and electron-nuclear double resonance spectroscopies in the Dph2 reaction. Here, we apply density functional methods to models of both intermediates. For intermediate **I**, we consider the previously proposed organometallic model as well as a SAM-bound [4Fe–4S] cluster model, finding that the SAM-bound cluster model accurately predicts the observed ^{13}C hyperfine coupling constant (HFCC), while the organometallic model predicts a ^{13}C HFCC much too large to match experiment. For intermediate **II**, we consider three protonation states of the histidine moiety and find a doubly protonated histidine residue model best matches the observed ^{13}C, ^{1}H, and ^{14}N HFCCs. Our results show that previous computational studies of the formation of an organometallic intermediate should be reconsidered and call for a revision of the organometallic intermediate paradigm in the non-canonical radical SAM enzyme Dph2.

6.2 Introduction

Radical S-adenosyl-L-methionine (SAM; RS) enzymes constitute one of nature's largest enzyme superfamilies and catalyze an incredible diversity of chemical reactions, from inert C-H bond functionalization to difficult skeletal rearrangements to DNA repair.[1] RS enzymes utilize SAM bound to a reduced $[4Fe–4S]^{1+}$ cluster to initiate catalysis. So-called canonical RS enzymes proceed via reductive cleave of SAM to form the 5′-deoxyadenosyl radical (dAdo·) and an oxidized $[4Fe–4S]^{2+}$ cluster with methionine bound (Figure 6.1). Substrate hydrogen abstraction by the powerful oxidant dAdo· then provides the first step in the numerous chemistries facilitated by the enzyme superfamily. In contrast, there is one identified RS enzyme, Dph2, that does not reductively cleave SAM to form dAdo·. Rather, in non-canonical Dph2 catalysis, reductive cleavage of SAM results in the formation of the 3-amino-3-carboxypropyl radical (ACP·).[2] Found in the diphthamide biosynthetic pathway, Dph2 uses ACP· to catalyze the addition of a SAM-derived 3-amino-3-carboxypropane moiety to a histidine residue of elongation factor 2 (EF2; Figure 6.1).[2]

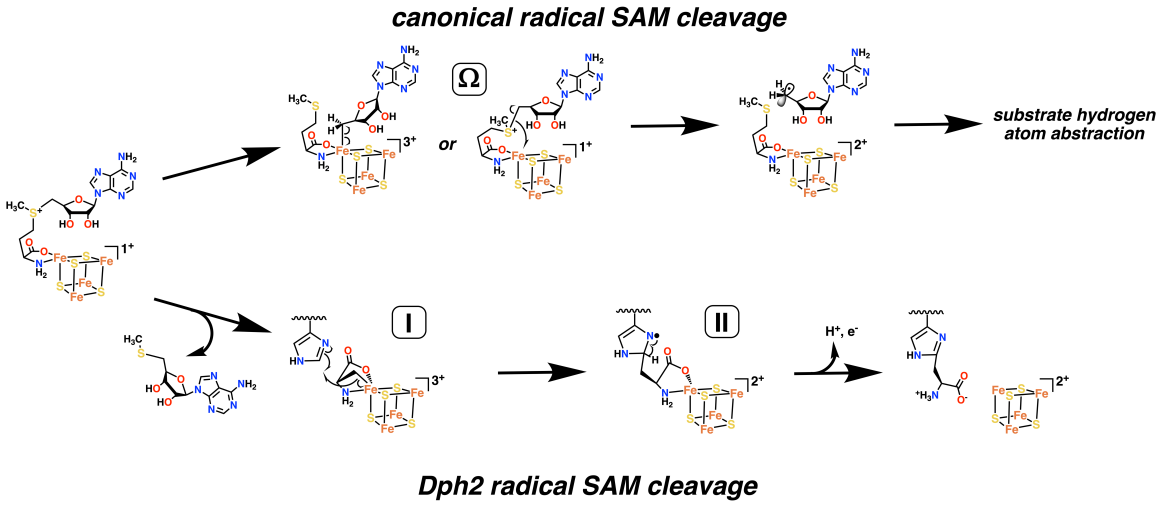

Figure 6.1. Canonical RS and Dph2 enzyme catalysis, showing proposed intermediates for Ω, intermediate **I**, and intermediate **II**.

Recently, both canonical RS enzymes and Dph2 have had organometallic intermediates implicated in their catalytic mechanisms after initial SAM cleavage occurs.[3,4] In canonical RS enzymes, this was proposed to result from the addition of the dAdo· to the $[4Fe–4S]^{2+}$ cluster, creating an organometallic species dubbed the Ω intermediate (Figure 6.1). Ω was then proposed to undergo adenosylcobalamin-like organometallic bond cleavage to re-form the dAdo· radical, which can then abstract a hydrogen atom from the substrate. In Dph2, ACP· has been proposed to bind to the $[4Fe–4S]^{2+}$ cluster in a similar fashion, creating a species called intermediate **I** (Figure 6.1);[4] the organometallic bond is then cleaved and the organic radical then adds to the histidine substrate. In both cases, electron paramagnetic resonance (EPR) and electron-nuclear double resonance (ENDOR) spectroscopies were used to identify and characterize Ω and I.[3,4] The difference in which SAM sulfur-carbon bond is cleaved has been proposed to result from an active-site controlled pseudo-Jahn-Teller effect on the sulfonium moiety of SAM.[5] However, the analysis was not extended to the S-Cγ cleavage occurring in Dph2, instead focusing on enzymes that form dAdo· (spore photoproduct lyase)[6] or methyl radical (HydE and HydG)[7] when photolyzed in the absence of substrate. To the best of our knowledge, it is unknown if Dph2 can create ACP· when photolyzed in the absence of substrate. However, it is known that in Dph2 catalysis another paramagnetic species forms in the mechanism, intermediate **II**, which was identified as occurring after **I** in the mechanism and proposed to be the structure shown in Figure 6.1.[4b]

The discovery of Ω and **I** has driven computational interest in studying the formation and structure of each proposed organometallic species. Notably, computational methods can be used to predict spectroscopic parameters, from which models of

intermediates that are consistent with between experiment and calculation can be derived. Toward this end, recent work in our group used broken-symmetry density functional theory (BS-DFT) to study the structure of the proposed organometallic structure of Ω.[8] By comparing experimentally determined hyperfine coupling constants (HFCCs) to those predicted by model structures, we found that a near-attack conformer (NAC) of SAM proximal to the reduced [4Fe–4S]$^{1+}$ cluster fit experimental HFCCs much more robustly than the HFCCs of the proposed organometallic structure, leading to the proposal of a SAM-NAC model of Ω (Figure 6.1). For example, the isotropic 5′-^{13}C HFCC in Ω has an experimental value of 9.4 MHz;[3a] from calculations, a value >140 MHz is expected for an organometallic model.[8] In contrast, a NAC model predicts a 5′-^{13}C HFCC of 8.8 MHz, within 1 MHz of experiment.[8] Our findings were in line with a computational analysis of the reaction mechanism which had proposed that the organometallic Ω structure could be a reaction shunt product due to a large predicted transition barrier of formation (>16.7 kcal mol^{-1}).[9]

However, several broken-symmetry computational analyses showed that **I** does not have a prohibitive barrier of formation,[9,10] and therefore an organometallic intermediate is potentially more energetically feasible on the Dph2 reaction pathway. Importantly, each of these studies utilized single-reference broken-symmetry valence isomers to attempt to draw conclusions about reaction energetics. In order to accurately determine reaction energetics for iron-sulfur cluster systems, post-DFT methods must be used due to the large number of low-lying states present in iron-sulfur clusters.[11] As discussed elsewhere,[11a] energetic analysis performed along a single broken-symmetry valence isomer cannot accurately capture reaction thermodynamics in the absence of

calibration to a known physical state of the system (via Mössbauer or other spectroscopic techniques). While the results of previous studies on intermediate **I** therefore remain inconclusive, the categorically lower barriers of formation than those of Ω do warrant spectroscopic calculations of **I** for comparison to experiment. In contrast to energetics, spectroscopic properties of broken-symmetry valence isomers can be compared to experimental spectra to predict models of the physical state. Determining which broken-symmetry valence isomer best matches experimental spectra allows that valence isomer to then be used as a physical model of the system, such as has been done for the BS7 state of FeMoco in nitrogenase.[12]

Using this reasoning, the magnitude of the measured ^{13}C hyperfine coupling constant (HFCC) thus casts doubt on the possibility of an organometallic intermediate. Experimentally, the ^{13}Cγ isotropic hyperfine coupling constant of **I** was determined to be 7.8 MHz,[4b] aligning very closely with the predicted 5′-^{13}C HFCC of the SAM-NAC model of Ω (8.8 ± 1.0 MHz) but far removed from the predicted 5′-^{13}C HFCC for the organometallic model (140.4 ± 12.6 MHz).[8] Therefore, further investigation of possible models of **I** is not only warranted but indeed necessary for understanding the catalytic mechanism of Dph2. A model of intermediate **I** consistent with spectroscopic results must be obtained in order to inform both meaningful broken-symmetry mechanistic calculations and further spectroscopic experiments. Here, we present results of computationally and spectroscopically consistent structures of intermediates **I** and **II** in the Dph2 reaction mechanism, considering both an organometallic model and SAM-bound [4Fe–4S] cluster model of **I** for comparison to experiment.

6.3 Computational Methodology

Following our earlier methodology,[8] we employed the BS-DFT methodology established by Noodleman and coworkers[13] and its extension to the prediction of HFCCs of generic oligonuclear complexes by Pantazis, Neese, and coworkers[14] in structures containing iron-sulfur clusters (*vide infra*). Structures with only organic atoms were treated with standard DFT methodology (*vide infra*).

Initial intermediate **I** models of both *S*-adenosyl-L-methionine (SAM) bound to a reduced $[4Fe4S]^{1+}$ and the proposed organometallic intermediate were built from the crystal structure of Dph2 with SAM bound (PDB ID: 6BXN).[4b] The model of the proposed organometallic intermediate was then edited with the Avogadro software package in order to create the iron-carbon bond and remove the 5′-methylthioadensione (MTA) moiety.[15] The models for intermediate **II**, including all protonation states, were built in the Avogadro software package.[15] The protonation states considered were protonated on either the δ or ε nitrogen atom or the doubly protonated species with both δ and ε nitrogen atoms protonated. All geometry optimizations and single point calculations were performed with the ORCA software package, version 4.2.1.[16] All calculations used the CPCM solvent model with a dielectric constant ε = 20 to model protein interior.[17] For geometry optimizations of both models of intermediate **I**, coordinating cysteine residues were capped as methanethiolate ligands with carbons constrained to their crystallographic positions. For intermediate **II**, the His residue of elongation factor 2 (EF2) was capped at the Cβ position, and the iron-sulfur cluster coordinating the 3-amino-3-carboxypropyl (ACP) moiety was excluded.

Intermediate I Calculation Details. Geometry optimizations were performed using the BP86 functional[18] using Ahlrichs def2-SVP basis set[19] with the def2/J auxiliary basis set[20] on all atoms excluding Fe. Fe atoms used Ahlrichs def2-TZVP basis set[19] with the def2-J auxiliary basis set[20] and had the size of their integration grids set to 7 during all steps. An increased integration grid was used during SCF steps with a further increased grid used for final energy evaluation (Grid6 and Finalgrid7, respectively, in ORCA syntax). The RIJCOSX approximation was employed during geometry optimization,[21] with tighter convergence criterion applied (TIGHTSCF in ORCA syntax). The numbering convention chosen for each Fe site is shown in Figure S1 and is used to refer whether the site carries majority α or β spin in a given broken-symmetry (BS) valence isomer (e.g., ααββ has majority α spin on sites 1 and 2 and majority β spin on sites 3 and 4). Geometry optimizations were performed on the high-spin (αααα) and all m_s = +1/2 BS valence isomers (ααββ, αβαβ, and αββα valence isomers for the SAM-bound cluster model and ααββ, αβαβ, βααβ for the organometallic model).[22] The lowest-energy geometry was then used for all single point calculations, which was the αββα determinant for the SAM-bound cluster model and the αβαβ determinant for the organometallic model.

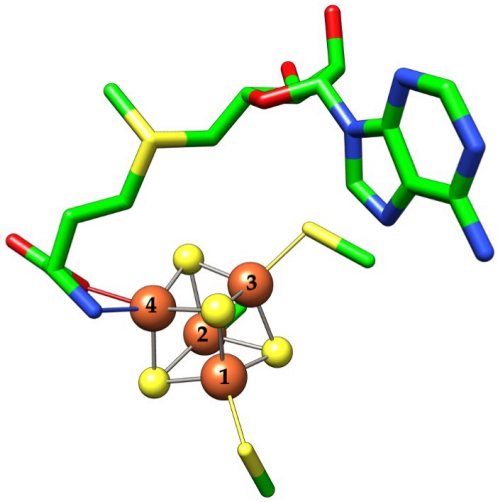

Figure 6.2. Numbering convention used for Fe sites in the broken-symmetry density functional theory (BS-DFT) calculations.

Single point calculations used the TPSSh functional[23] and the ZORA Hamiltonian to account for relativistic effects.[24] The SARC/J auxiliary basis set[25] was used in combination with the AutoAux functionality of ORCA to generate all auxiliary basis sets. Fe atoms used the CP(PPP) basis set,[26] S atoms used the ZORA-def2-TZVP basis set,[19] and all other atoms used the EPR-II basis set.[27] The size of the integration grid was increased to 7 for all atoms, and tighter SCF convergence criterion was applied. Hyperfine coupling constants (HFCCs) were then determined using the spin projection method developed for oligonuclear metal clusters by Pantazis *et al.*,[14] as described in application to radical SAM iron-sulfur clusters by our previous work.[8]

Intermediate II Calculation Details. Geometry optimizations were performed using the BP86 functional[18] using Ahlrichs def2-SVP basis set[19] with the def2/J auxiliary basis set[20] on all atoms with the RIJCOSX approximation employed.[21] Tighter SCF convergence criterion were required, and a slightly increased integration grid was used with an increased integration grid used in final energy evaluation (Grid5 and Finalgrid6,

respectively, in ORCA syntax). For single point calculations, the TPSSh functional[23] with the EPR-II basis set[27] was used, retaining the same grid specifications.

6.4 Results and Discussion

Considering both an organometallic and SAM-bound cluster model for **I**,[28] we find that an organometallic model (Figure 6.3) predicts a ^{13}C HFCC far too large to be consistent with experiment: 136.9 ± 3.1 MHz. Just as in the canonical RS case, a value with this magnitude would be expected if the carbon atom was participating in an organometallic bond with the unique iron of the iron-sulfur cluster due to a significant transfer of spin density to the carbon atom. However, the SAM-bound [4Fe–4S] cluster model (Figure 6.3) predicts a ^{13}C HFCC of 7.3 ± 3.0 MHz, matching the experimental value of 7.8 MHz. Unsurprisingly, the SAM-bound cluster model of intermediate **I** also matches closely with the related NAC model of Ω, giving evidence that if a NAC model correctly explains the observed spectra of Ω, a similar result—of SAM bound to a reduced $[4Fe–4S]^{1+}$ cluster—should be expected for **I**.

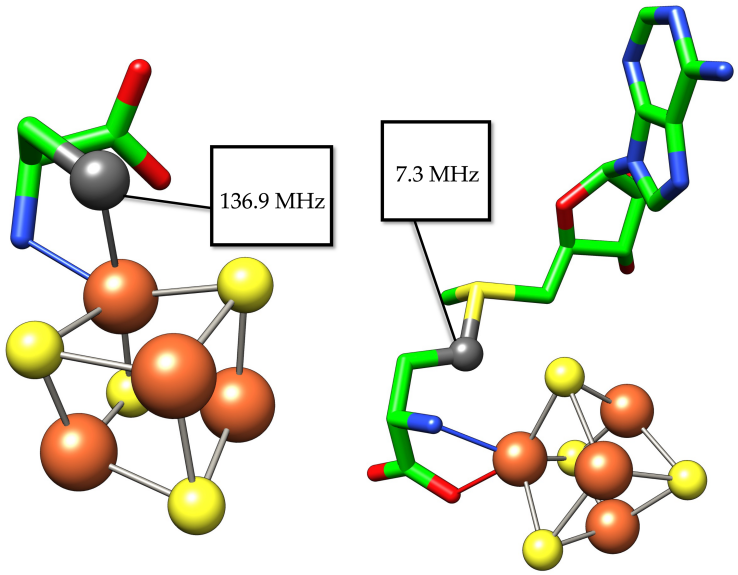

Figure 6.3. Comparison of computationally predicted ^{13}C isotropic hyperfine coupling constants for organometallic and SAM-bound [4Fe–4S] cluster models of intermediate I (experimental value 7.8 MHz[4b]). Hydrogen atoms and methanethiol ligands are omitted for clarity.

For intermediate **II**, several considerations arise when modeling the observed spectra. First, the experimentally determined ^{14}N hyperfine will have contributions from both the Nδ and Nε atoms of the His residue of EF2. The reported simulated EPR spectra used one nitrogen HFCC to match the experimental EPR spectra,[4b] but by calculating the contributions from both we can include both in a simulated EPR spectra built from the DFT results. Further, the possibility of tautomerization, where either Nε is protonated and the radical is localized on Nδ (shown in Figure 6.1) or where Nδ is protonated and the radical is localized on Nε (not shown), exists, and contributions from each tautomer could contribute to the observed HFCC. Additionally, the doubly protonated state can be examined, as a doubly protonated species can be a substitute model for the case where each nitrogen atom is participating in hydrogen bonding interactions.

Table 6.1. Calculated hyperfine coupling constants (HFCCs) of atoms with components of the HFCC greater than 10 MHz in the Nε protonated model of intermediate **II**.

Atom	A_1 (MHz)	A_2 (MHz)	A_3 (MHz)
^{13}C	60.6	60.7	73.1
^{1}H-Cε	134.6	137.4	141.1
^{1}H-Nε	-10.6	14.9	-20.8
^{1}H-Cδ	-11.7	-32.3	-48.0
^{14}Nε	-1.1	-1.3	35.1
^{14}Nδ	-4.4	-4.6	43.3

For the best model of intermediate **II** (*vide infra*), the isotropic ^{13}C HFCC shows excellent agreement between the proposed model and the experimental value: calculated to be 65 MHz and experimentally determined to be 59 MHz.[4b] However, for the ^{1}H and ^{14}N HFCCs, we note the experimental values were determined by spectral fitting using only one ^{1}H HFCC (A = 122, 122, 103 MHz; ref. 4b, Figure S4) and one ^{14}N HFCC (A = 4, 4, 54 MHz; ref. 4b, Figure S4). A second, exchangeable ^{1}H HFCC was determined by a D_2O ENDOR experiment (A = ~15 MHz; ref 4b, Figure S2). However, the linewidth of the simulation used to assign the larger ^{1}H HFCC and the ^{14}N HFCC was 8 MHz[4b]—meaning that the exchangeable ^{1}H HFCC would be large enough to be observable in the EPR spectra at this linewidth. We find that by instead including every calculated ^{1}H and ^{14}N HFCC with a component whose absolute value is greater than 10 MHz (Table 6.1) and increasing the linewidth of the simulated spectra, from 8 MHz to 30 MHz, we can reproduce the experimental spectra without this inconsistency. The atoms included in our simulated spectra were Nδ, Nε, H-Cε, H-Nε, and H-Cδ (Table 6.1; all nitrogen and hydrogen atoms on the imidazole moiety). By simulating EPR spectra for each

protonation state model of intermediate **II** including the HFCCs from these atoms, we find consistent results with the assignment given previously that a protonated Nε model best matches the observed spectra (Figure 6.4).[4b]

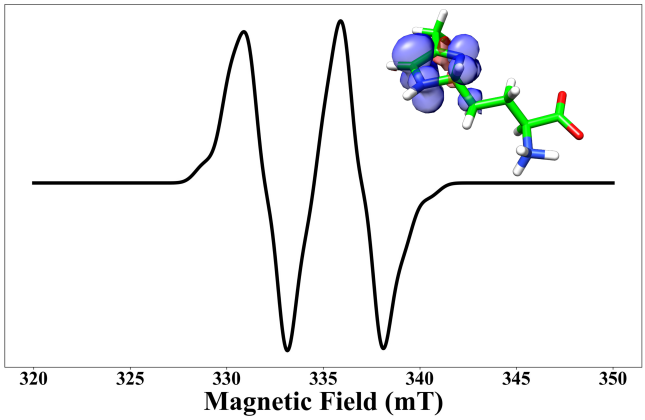

Figure 6.4. Simulated electron paramagnetic resonance (EPR) spectra from the calculated *g*-values and ^{1}H and ^{14}N hyperfine coupling constants (HFCCs) for the Nε protonated model of intermediate II (linewidth = 30 MHz; see Table 1 for HFCCs); plot generated with the EasySpin software package.[29] Inset shows calculated spin density (smoothed at ± 0.0015 au).

Combining the results of models of **I** and **II**, we suggest a revised mechanism to that proposed in the literature, where a SAM-bound [4Fe–4S] cluster structure replaces that of the proposed organometallic intermediate (Figure 6.5).[2b,4b,9,11] In particular, we note that our model accounts for the increased value of the ^{13}C HFCC in intermediate **II** relative to intermediate **I** as the spin density moves from the [4Fe–4S] cluster to the His residue. In the organometallic model, the carbon atom is alpha to the spin center, whereas in **II** it is located beta to the spin center, regardless of tautomeric state. As a result, if an organometallic intermediate did occur, one would expect its ^{13}C HFCC value to be similar to or larger than the 13C of **II**, as both would have meaningful contribution from Fermi contact in the hyperfine coupling term. Instead, the value is observed to increase by an order of magnitude between **I** and **II**. In the case of a SAM-bound [4Fe–4S] cluster

structure of intermediate **I**, one would expect both a smaller HFCC value in **I** and for the value to increase between **I** and **II**, as is observed experimentally. Importantly, a

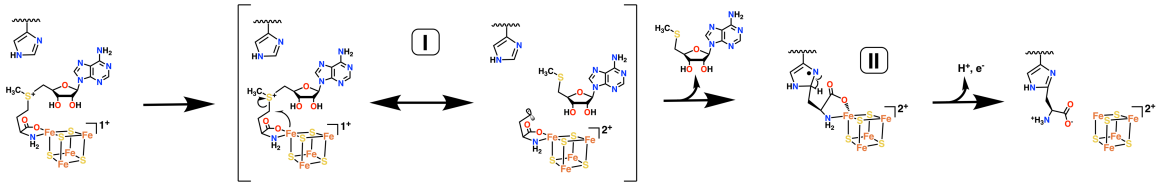

Figure 6.5. Revised Dph2 catalytic mechanism including a paramagnetic SAM-bound [4Fe–4S] cluster species and the previously proposed Nε protonated His intermediate.

model optimized from the crystal structure alone is enough to explain the observed ^{13}C HFCC for intermediate **I**. In the case of canonical radical SAM cleavage, a computational model required a near-attack conformer model to accurately model experimental HFCCs. For intermediate **I**, which adopts a different SAM conformer in order to facilitate unique S-Cγ cleavage, an optimized geometry from the crystal structure alone suffices to explain the observed ^{13}C HFCC.

In summary, our results provide evidence that the noncanonical RS enzyme Dph2 does not use an organometallic intermediate during catalysis, in line with what was recently proposed for canonical RS enzymes.[8] Comparison to experiment suggests that a SAM-bound [4Fe–4S] cluster model better explains the observed ^{13}C HFCC in intermediate **I**, prompting a revision of the proposed catalytic mechanism. As in canonical RS catalysis, an organometallic intermediate would require the enzyme to access the same transition state twice, both during organometallic bond formation and cleavage. A SAM-bound [4Fe–4S] cluster model of **I** calls for revision of computationally predicted mechanisms,[9,10] to date all of which have employed single broken-symmetry valence isomers to determine mechanisms without experimental calibration. Because a single broken-symmetry determinant cannot accurately represent

the multireference character of the oligonuclear system, choosing a broken-symmetry determinant that accurately matches known spectroscopic parameters of the system is key towards accurate computational analysis of any reaction with a multinuclear iron cluster, such as the BS7 state of FeMoco in nitrogenase[12] or the recently proposed models for oxidized and intermediate structures in soluble methane monooxygenase.[30] Our SAM-bound cluster model provides an important step towards reaching unity between spectroscopy and computation in Dph2 catalysis. Finally, we find the proposed model of intermediate **II** matches well between calculation and experiment. Taken together, our results provide key implications for understanding the mechanism a unique member of the radical S-adenosyl-L-methionine superfamily of enzymes.

6.5 References

1. (a) Frey, P. A.; Hegeman, A. D.; Ruzicka, F. J.; The Radical SAM Superfamily. *Crit. Rev. Biochem. Mol. Bio.* **2008**, *43* (1), 63–88. (b) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical *S*-adenosylmethonine Enzymes. *Chem. Rev.* **2014**, *114* (8), 4220–4317. (c) Landgraf, B. J.; McCarthy, E. L.; Booker, S. J. Radical S-Adenosylmethionine Enzymes in Human Health and Disease. *Annu. Rev. Biochem.* **2016**, *85* (1), 485–514. Bridwell-Rabb, J.; Grell, T. A. J.; Drennan, C. L. A Rich Man, Poor Man Story of *S*-Adenosylmethionine and Cobalamin Revisited. *Annu. Rev. Biochem.* **2018**, *87* (1), 555–584. (e) Nicolet, Y. Structure–Function relationships of radical SAM enzymes. *Nature Catalyis* **2020,** *3*, 337–350.

2. (a) Dong, M.; Horitani, M.; Dzikovski, B.; Freed, J. H.; Ealick, S. E.; Hoffman, B. M.; Lin H. Substrate-Dependent Cleavage Site Selection by Unconventional Radical SAM Enzymes in Diphthamide Biosynthesis. *J. Am. Chem. Soc.* **2017**, *139* (16), 5680–5683. (b) Dong, M.; Zhang, Y.; Lin, H. Noncanonical Radical SAM Enzyme Chemistry Learned from Diphthamide Biosynthesis. *Biochemistry* **2018**, *57* (25), 3454–3459.

3. (a) Horitani, M.; Shisler, K.; Broderick, W. E.; Hutcheson, R. U.; Duschene, K. S.; Marts, A. R.; Hoffman, B. M.; Broderick, J. B. Radical SAM Catalysis via an organometallic intermediate with an Fe-[5´-C]-deoxyadenosyl bond. *Science* **2016**, *352* (6287), 822–825. (b) Byer, A. S.; Yang, H.; McDaniel, E. C.; Kathiresan, V.; Impano, S.; Pagnier, A.; Watts, H.; Denler, C.; Vagstad, A. L.; Piel, J.; Duschene, K. S.; Shepard, E. M.; Shields, T. P.; Scott, L. G.; Lilla, E. A.; Yokoyama, K.; Broder-ick, W. E.; Hoffman, B. M.; Broderick, J. B. Paradigm Shift for Radical-S-Adenosyl-L-Methionine Reactions:

The Organometallic Intermediate Ω is Central to Catalysis. *J. Am. Chem. Soc.* **2018**, *140* (28), 8634–8638. (c) Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Mechanism of Radical Initiation in the Radical-*S*-Adenosyl-L-Methionine Superfamily. *Acc. Chem. Res.* **2018**, *51* (11), 2611–2619.

4. (a) Dong, M.; Horitani, M.; Dzikovski, B.; Pandelia, M.-E.; Krebs, C.; Freed, J. H.; Hoffman, B. M.; Lin, H. Organometallic Complex Formed by an Unconventional Radical *S*-Adenosylmethionine Enzyme. *J. Am. Chem. Soc.* **2016**, *138* (31), 9755–9758. (b) Dong, M.; Kathiresan, V.; Fenwick, M. K.; Torelli, A. T.; Zhang, Y.; Caranto, J. D.; Dzikovski, B.; Sharma, A.; Lancaster, K. M.; Freed, J. H.; Ealick, S. E.; Hoffman, B. M.; Lin, H. Organometallic and radical intermediates reveal mechanism of diphthamide biosynthesis. *Science* **2018**, *359* (6381), 1247–1250.

5. Impano, S.; Yang, H.; Jodts, R. J.; Pagnier, A.; Swimley, R.; McDaniel, E. C.; Shepard, E. M.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. Active-Site Controlled, Jahn–Teller Enabled Regioselectivity in Reductive S–C Bond Cleavage of S-Adenosylmethionine in Radical SAM Enzymes. *J. Am. Chem. Soc.* **2021**, *143* (1), 335–348.

6. Pagnier, A.; Yang, H.; Jodts, R. J.; James, C. D.; Shepard, E. M.; Impano, S.; Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Radical SAM Enzyme Spore Photoproduct Lyase: Properties of the Ω Organometallic Intermediate and Identification of Stable Protein Radicals Formed during Substrate-Free Turnover. *J. Am. Chem. Soc.* **2020**, *142* (43), 18652–18660.

7. Yang, H.; Impano, S.; Shepard, E. M.; James, C. D.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. Photoinduced Electron Transfer in a Radical SAM Enzyme Generates an *S*-Adenosylmethionine Derived Methyl Radical. *J. Am. Chem. Soc.* **2019**, *141* (40) 16117–16124.

8. Donnan, P. H.; Mansoorabadi, S. O. Broken-Symmetry Density Functional Theory Analysis of the Ω Intermediate in Radical *S*-Adenosyl-L-Methionine Enzymes: Evidence for a Near-Attack Conformer over an Organometallic Species *J. Am. Chem. Soc.* **2022**, *144* (8), 3381–3385.

9. Zhao, C.; Li, Y.; Wang, C.; Chen, H. Mechanistic Dichotomy in the Activation of SAM by Radical SAM Enzymes: QM/MM Modeling Deciphers the Determinant. *ACS Catal.* **2020**, *10*, 13245–13250.

10. (a) Zhou, S.; Wei, W.-J.; Liao, R.-Z. QM/MM Study of the Mechanism of the Noncanonical S-Cγ Bond Scission in S-Adenosylmethionine Catalyzed by the CmnDph2 Radical Enzyme. *Top. Catal.* **2021**, DOI: 10.1007/s11244-021-01420-5. (b) Feng, J.; Shaik, S.; Wang, B. Spin-Regulated Electron Transfer and Exchange-Enhanced Reactivity in Fe_4S_4-Mediated Redox Reaction of the Dph2 Enzyme During the Biosynthesis of Diphthamide. *Angew. Chem.* **2021**, *133* (37), 20593–20599. (c) Feng, J.; Shaik, S.; Wang, B. Spin-Regulated Electron Transfer and Exchange-Enhanced

Reactivity in Fe_4S_4-Mediated Redox Reaction of the Dph2 Enzyme During the Biosynthesis of Diphthamide. *Angew. Chem. Int. Ed.* **2021**, *60* (37), 20430–20436.

11. (a) Orio, M.; Pantazis, D. Successes, challenges, and opportunities for quantum chemistry in understanding metalloenzymes for solar fuels research. *Chem. Commun.* **2021**, *57* 3952–3974. (b) Sharma, S.; Sivalingam, K.; Neese, F.; G. K.-L. Low-energy spectrum of iron-sulfur clusters directly from many-particle quantum mechanics. *Nat. Chem.* **2014**, *6* (10), 927–933. (c) Cho, D.; Rouxel, J. R.; Mukamel, S.; Chan, G. K.-L.; Li, Z. Stimulated X-ray Raman and Absorption Spectroscopy of Iron–Sulfur Dimers. *J. Phys. Chem. Lett.* **2019**, *10* (21), 6664–6671.

12. Pelmenschikov, V.; Case, D. A.; Noodleman, L. Ligand-Bound *S* = ½ FeMo-Cofactor of Nitrogenase: Hyperfine Interaction Analysis and Implication for the Central Ligand X Identity. *Inorg. Chem.* **2008**, *47* (14), 6162–6172. (b) Benediktsson, B.; Bjornsson, R. QM/MM Study of the Nitrogenase MoFe Protein Resting State: Broken-Symmetry States, Protonation States, and QM Region Convergence in the FeMoco Active Site. *Inorg. Chem.* **2017**, *56* (21), 13417–13429. (c) Raugei, S.; Seefeldt, L. C; Hoffman, B. M. Critical computational analysis illuminates the reductive-elimination mechanism that activates nitrogenase for N_2 reduction. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (45), E10521–E10530.

13. Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J.-M.; Orbital interaction, electron delocalization and spin coupling in iron-sulfur clusters. *Coord. Chem. Rev.* **1995**, *144*, 199–244.

14. Pantazis, D. A.; Orio, M.; Petrenko, T.; Zein, S.; Bill, E.; Lubitz, W.; Messinger, J.; Neese, F. A New Quantum Chemical Approach to the Magnetic Properties of Oligonuclear Transition-Metal Complexes: Application to a Model for the Tetranuclear Manganese Cluster of Photosystem II. *Chem. Eur. J.* **2009**, *15* (20), 5108–5123.

15. a) Avogadro: an open-source molecular builder and visualization tool. Version 1.20. https://avogadro.cc. (b) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An advanced semantics chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2012**, *4*, 17.

16. (a) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78. (b) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *8*, e1327.

17. Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001.

18. (a) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33* (12), 8822–8824. (b) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.

19. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.

20. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1057–1065.

21. Neese, F.; Wennmoh, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *365* (1–3), 98–109.

22. The difference arises due to the oxidation state of the cluster between the two models: $[4Fe4S]^{1+}$ in the SAM-bound cluster model and formally $[4Fe4S]^{3+}$ in the organometallic model.

23. Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119* (23), 12129.

24. van Lethe, E.; Baerends, E. J.; Snijders, J. G. Relativistic regular two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597.

25. Pantazis, D. A.; Neese, F. All-electron scalar relativistic basis sets for the 6p elements. *Theor. Chem. Acc.* **2012**, *131*, 1292.

26. Neese. F. Prediction and interpretation of the ^{57}Fe isomer shift in Mössbauer spectra by density functional theory. *Inorg. Chim. Acta* **2002***, 337*, 181–192.

27. Rega, N.; Cossi, M.; Barone, V. Development and validation of reliable quantum mechanical approaches for the study of free radicals in solution. *J. Chem. Phys.* **1996**, *105* (24), 11060.

28. Because it is unknown which broken-symmetry determinant best models the physical state for either the organometallic and SAM-bound cluster models of **I**, we present spin-projected average absolute HFCC values of the $m_s = +1/2$ broken-symmetry valence isomers.

29. Stoll, S.; Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **2006**, *178* (1), 42–55.

30. Schulz, C. E.; Castillo, R. G.; Pantazis, D. A.; DeBeer, S.; Neese, F. Structure-Spectroscopy Correlations for Intermediate Q of Soluble Methane Monooxygenase: Insights from QM/MM Calculations. *J. Am. Chem. Soc.* **2021**, *143* (17), 6560–6577.

Chapter 7: Conclusions and Outlook

7.1 Concluding Remarks

This dissertation has attempted to shed light on two enzymatic systems: bioluminescence in dinoflagellates and the common catalytic mechanism in radical *S*-adenosyl-L-methionine enzymes, including the non-canonical case of Dph2. For each system, different computational methods were employed: molecular dynamics for understanding LCF and its pH-dependent conformational changes and ligand interactions, quantum chemical methods for studying the radical SAM system and helping to determine consistency between experimentally observed spectroscopic values and those calculated from models. The work presented here highlights the variety of computational techniques that can be used to help solve research questions of biochemical interest. Importantly, this dissertation has intended to provide either testable predictions or otherwise presented results consistent with experiment, as any computational work in biochemistry must be thoroughly grounded in experiment. However, computational biochemistry has become an invaluable tool to the modern enzymologist, and this dissertation aims to provide several case studies testifying to that fact.

7.2 Outlook and Future Directions

The author leaves several avenues of research open at the conclusion of this dissertation. First, with the open conformer of LCF and the predicted stereochemical course of the reaction in hand, a QM/MM study of the bioluminescent reaction is in order. While residues important for catalysis have been proposed, identifying the nature

of the interactions stabilizing the open-shell singlet pre-emission can only be done with QM/MM. The proposed mechanism must be tested by both QM/MM calculations and by the experimental predictions made on the basis of this work. Much experimental validation of the results presented here is yet to be conducted.

The exact structure of Ω is still under debate,[1-4] and the results presented here have hopefully provided a convincing alternative to the initially proposed organometallic structure. More structural and spectroscopic studies will help resolve the structure of the intermediate, along with recent progress in synthetic iron-sulfur cluster chemistry. It is the author's hope that the work presented here will help uncover the true structure of a universal reaction intermediate in one of nature's largest enzyme superfamilies.

7.3 References

1. Donnan, P. H.; Mansoorabadi, S. O. Broken-Symmetry Density Functional Theory Analysis of the Ω Intermediate in Radical *S*-Adenosyl-L-Methionine Enzymes: Evidence for a Near-Attack Conformer over an Organometallic Species *J. Am. Chem. Soc.* **2022**, *144* (8), 3381–3385.

2. Lundahl, M. N.; Sarksian, R.; Yang, H.; Jodts, R. J.; Pagnier, A.; Smith, D. F.; Mosquera, M. A.; van der Donk, W. A.; Hoffman, B. M.; Broderick, W. E.; Broderick, J. B. Mechanism of Radical *S*-Adenosyl-L-methionine Adenosylation: Radical Intermediates and the Catalytic Competence of the 5′-Deoxyadenosyl Radical. *J. Am. Chem. Soc.* **2022**, *144* (11), 5087–5098.

3. Ye, M; Brown, A. C.; Suess, D. L. M. Reversible Alkyl-Group Migration between Iron and Sulfur in [Fe_4S_4] Clusters. *J. Am. Chem. Soc.* **2022**, *144* (29), 13184–13195.

4. Ho, M. B.; Jodts, R. J.; Kim, Y.; McSkimming, A.; Suess, D. L. M.; Hoffman, B. M. Characterization by ENDOR Spectroscopy of the Iron–Alkyl Bond in a Synthetic Counterpart of Organometallic Intermediates in Radical SAM Enzymes. *J. Am. Chem. Soc.* **2022**, Article ASAP. DOI: 10.1021/jacs.2c07155.

Appendix A: Supporting Material for Chapter Three

Table A1. pK_a values for titratable residues calculated over the full 1 μs simulation at pH 8 and 6 reported in Chapter Three. Reproduced with permission from *Biochemistry* **2018**, *57*, 295−299. Copyright 2018 American Chemical Society.

Residue	pH 8	pH 6
E872	4.2	-
H886	6.1	5.4
Y890	10.5	10.8
H899	6.0	6.1
H904	5.6	6.3
H909	7.0	6.9
H924	5.9	6.6
Y925	11.9	10.8
C929	-	12.7
H930	6.6	5.4
K935	10.2	9.5
C946	10.9	11.0
D968	-	-1.5
C972	5.6	4.7
D978	3.2	1.1
C980	7.0	3.4
C991	10.8	10.3
Y996	12.0	8.7
D998	3.6	1.2
K999	9.9	-
C1005	10.5	-
H1027	5.2	4.9
D1063	-	5.0
H1064	4.3	6.4
H1065	5.9	6.2

Residue	pH 8	pH 6
Y1071	12.9	10.1
D1079	3.4	2.8
C1081	8.0	-
D1084	2.5	0.8
K1089	8.9	9.6
K1094	9.9	6.0
H1101	4.1	2.8
Y1102	11.1	12.2
Y1104	8.1	6.7
E1105	-	4.0
K1125	7.2	6.9
E1131	-	4.3
C1137	12.1	-
E1141	-	-0.4
E1158	-	-0.1
Y1168	12.7	-
K1170	10.1	9.2
D1186	2.4	4.5
Y1204	11.1	11.8
D1210	-	-
D1212	5.8	-
K1216	7.2	9.5
D1218	2.4	-
E1227	4.4	3.9

Appendix B: Supporting Material for Chapter Five

A version of this appendix was published: Donnan, P. H.; Mansoorabadi, S. O. *J. Am. Chem. Soc.* **2022**, *144 (8)*, 295−299, and associated Supporting Information.

B.1. Discussion of Broken-Symmetry DFT and Spin Projection

In this section, we briefly outline the procedure used to extract SPFs from BS-DFT calculations. The process involves calculation of the single point energies of several BS states, which are then related to each other through the Heisenberg-Dirac-van Vleck (HDV) Hamiltonian. While BS states are not spin eigenstates, they can be used to obtain the relevant J couplings between spin centers, which allows determination of HDV Hamiltonian matrix elements. Diagonalization of the HDV energy matrix returns the spin eigenstates, which then allows calculation of the site expectation values for $\hat{S}_z$, from which SPFs can be determined.

The HDV Hamiltonian operator in the strong-coupling limit is given by

$$\hat{H} = -2\sum_{i<j} J_{ij}\vec{S}_i \cdot \vec{S}_j = -\sum_{i<j} J_{ij}\left[(\hat{S}_{ij}^2 - \hat{S}_i^2 - \hat{S}_j^2)\right] \tag{S1}$$

which includes all pairwise exchange coupling interactions (J_{ij}) between spin centers. Calculation of the expectation value of the HDV Hamiltonian for each BS valence isomer requires the evaluation of matrix elements of the form

$$\langle i\ j|\,\hat{H}\,|i\ j\rangle = -J_{ij}\left[S_{ij}(S_{ij}+1) - S_i(S_i+1) - S_j(S_j+1)\right] \tag{S2}$$

for spin centers i and j. For like spin states (αα or ββ), these matrix elements are relatively straightforward to calculate, since the spins couple to form a pure high-spin S_{ij} state. To evaluate matrix elements for mixed-spin states (αβ or βα), one must first express the states in terms of a coupled basis with well-defined S_{ij} values using the appropriate Clebsch-Gordon coefficients. For example, for two high-spin ferrous centers, the state $|\alpha\beta\rangle$ can be written as $\left|S_i\ m_i\,;\,S_j\ m_j\right\rangle = |2\ 2;\ 2\ {-2}\rangle$ in the site basis, which can be expressed in terms of the coupled basis $\left|S_{ij}\ m_{ij}\right\rangle$ as

$$|\alpha\beta\rangle = \frac{1}{\sqrt{5}}|0\ 0\rangle + \sqrt{\frac{2}{5}}|1\ 0\rangle + \sqrt{\frac{2}{7}}|2\ 0\rangle + \frac{1}{\sqrt{10}}|3\ 0\rangle + \frac{1}{\sqrt{70}}|4\ 0\rangle . \tag{S3}$$

The corresponding matrix element is then given by

$$\begin{aligned}\langle\alpha\beta|\, J_{ij}\vec{S}_1\cdot\vec{S}_2\,|\alpha\beta\rangle = -J_{12}\times\Bigg[&\left(\frac{1}{\sqrt{5}}\right)^{*}\frac{1}{\sqrt{5}}\Big[0(0+1)-2(2+1)-2(2+1)\Big]\\ &+\left(\sqrt{\frac{2}{5}}\right)^{*}\sqrt{\frac{2}{5}}\Big[1(1+1)-2(2+1)-2(2+1)\Big]\\ &+\left(\sqrt{\frac{2}{7}}\right)^{*}\sqrt{\frac{2}{7}}\Big[2(2+1)-2(2+1)-2(2+1)\Big]\\ &+\left(\frac{1}{\sqrt{10}}\right)^{*}\frac{1}{\sqrt{10}}\Big[3(3+1)-2(2+1)-2(2+1)\Big]\\ &+\left(\frac{1}{\sqrt{70}}\right)^{*}\frac{1}{\sqrt{70}}\Big[4(4+1)-2(2+1)-2(2+1)\Big]\Bigg] = 8J_{12}.\end{aligned} \tag{S4}$$

Using this procedure, the single point energies of the BS states for structural models with [4Fe-4S]$^{1+}$ clusters (i.e., the crystallographic SAM-bound cluster and the SAM-NAC structures), which contain one high-spin ferric S = 5/2 iron (assigned to the unique iron, Fe1) and three high-spin ferrous S = 2 irons, are related to the J couplings by:

$$E_{\alpha\alpha\alpha\alpha} = -10J_{12} - 10J_{13} - 10J_{14} - 8J_{23} - 8J_{24} - 8J_{34} \tag{S5}$$
$$E_{\alpha\alpha\alpha\beta} = -10J_{12} - 10J_{13} + 10J_{14} - 8J_{23} + 8J_{24} + 8J_{34} \tag{S6}$$
$$E_{\alpha\alpha\beta\alpha} = -10J_{12} + 10J_{13} - 10J_{14} + 8J_{23} - 8J_{24} + 8J_{34} \tag{S7}$$
$$E_{\alpha\beta\alpha\alpha} = 10J_{12} - 10J_{13} - 10J_{14} + 8J_{23} + 8J_{24} - 8J_{34} \tag{S8}$$
$$E_{\beta\alpha\alpha\alpha} = 10J_{12} + 10J_{13} + 10J_{14} - 8J_{23} - 8J_{24} - 8J_{34} \tag{S9}$$
$$E_{\alpha\alpha\beta\beta} = -10J_{12} + 10J_{13} + 10J_{14} + 8J_{23} + 8J_{24} - 8J_{34} \tag{S10}$$
$$E_{\alpha\beta\alpha\beta} = 10J_{12} - 10J_{13} + 10J_{14} + 8J_{23} - 8J_{24} + 8J_{34} \tag{S11}$$
$$E_{\alpha\beta\beta\alpha} = 10J_{12} + 10J_{13} - 10J_{14} - 8J_{23} + 8J_{24} + 8J_{34} \tag{S12}$$

A similar set of equations can be obtained for structural models with [4Fe-4S]$^{3+}$ clusters (i.e., the proposed organometallic structure), which contain three high-spin ferric S = 5/2 irons and one high-spin ferrous S = 2 iron (assigned to Fe4 after charge and spin analysis):

$$E_{\alpha\alpha\alpha\alpha} = -\frac{25}{2}J_{12} - \frac{25}{2}J_{13} - 10J_{14} - \frac{25}{2}J_{23} - 10J_{24} - 10J_{34} \tag{S13}$$

$$E_{\alpha\alpha\alpha\beta} = -\frac{25}{2}J_{12} - \frac{25}{2}J_{13} + 10J_{14} - \frac{25}{2}J_{23} + 10J_{24} + 10J_{34} \tag{S14}$$

$$E_{\alpha\alpha\beta\alpha} = -\frac{25}{2}J_{12} + \frac{25}{2}J_{13} - 10J_{14} + \frac{25}{2}J_{23} - 10J_{24} + 10J_{34} \tag{S15}$$

$$E_{\alpha\beta\alpha\alpha} = \frac{25}{2}J_{12} - \frac{25}{2}J_{13} - 10J_{14} + \frac{25}{2}J_{23} + 10J_{24} - 10J_{34} \tag{S16}$$

$$E_{\beta\alpha\alpha\alpha} = \frac{25}{2}J_{12} + \frac{25}{2}J_{13} + 10J_{14} - \frac{25}{2}J_{23} - 10J_{24} - 10J_{34} \tag{S17}$$

$$E_{\alpha\alpha\beta\beta} = -\frac{25}{2}J_{12} + \frac{25}{2}J_{13} + 10J_{14} + \frac{25}{2}J_{23} + 10J_{24} - 10J_{34} \tag{S18}$$

$$E_{\alpha\beta\alpha\beta} = \frac{25}{2}J_{12} - \frac{25}{2}J_{13} + 10J_{14} + \frac{25}{2}J_{23} - 10J_{24} + 10J_{34} \tag{S19}$$

$$E_{\beta\alpha\alpha\beta} = \frac{25}{2}J_{12} + \frac{25}{2}J_{13} - 10J_{14} - \frac{25}{2}J_{23} + 10J_{24} + 10J_{34} \tag{S20}$$

We then apply singular value decomposition (SVD) to solve the over-constrained system of equations to obtain best fit estimates of the *J* couplings. Once the *J* couplings are known, the full HDV Hamiltonian matrix elements can be calculated using the following equation and a basis set containing every possible combination of site m_s values:

$$\begin{aligned}
&\langle S_1 m_1' S_2 m_2' S_3 m_3' S_4 m_4' | \, \hat{H} \, | S_1 m_1 S_2 m_2 S_3 m_3 S_4 m_4 \rangle = \\
&-2[J_{12} m_s^1 m_s^2 + J_{13} m_s^1 m_s^3 + J_{14} m_s^1 m_s^4 + J_{23} m_s^2 m_s^3 + J_{24} m_s^2 m_s^4 + J_{34} m_s^3 m_s^4] \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime} m_s^4} - \\
&\Big[J_{12} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1+1)][S^2(S^2+1) - m_s^2(m_s^2-1)]} \times \delta_{m_s^{1\prime}(m_s^1+1)} \delta_{m_s^{2\prime}(m_s^2-1)} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{13} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1+1)][S^3(S^3+1) - m_s^3(m_s^3-1)]} \times \delta_{m_s^{1\prime}(m_s^1+1)} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime}(m_s^3-1)} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{14} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1+1)][S^4(S^4+1) - m_s^4(m_s^4-1)]} \times \delta_{m_s^{1\prime}(m_s^1+1)} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime}(m_s^4-1)} + \\
&J_{23} \sqrt{[S^2(S^2+1) - m_s^2(m_s^2+1)][S^3(S^3+1) - m_s^3(m_s^3-1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime}(m_s^2+1)} \delta_{m_s^{3\prime}(m_s^3-1)} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{24} \sqrt{[S^2(S^2+1) - m_s^2(m_s^2+1)][S^4(S^4+1) - m_s^4(m_s^4-1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime}(m_s^2+1)} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime}(m_s^4-1)} + \\
&J_{34} \sqrt{[S^3(S^3+1) - m_s^3(m_s^3+1)][S^4(S^4+1) - m_s^4(m_s^4-1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime}(m_s^3+1)} \delta_{m_s^{4\prime}(m_s^4-1)} + \\
&J_{12} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1-1)][S^2(S^2+1) - m_s^2(m_s^2+1)]} \times \delta_{m_s^{1\prime}(m_s^1-1)} \delta_{m_s^{2\prime}(m_s^2+1)} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{13} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1-1)][S^3(S^3+1) - m_s^3(m_s^3+1)]} \times \delta_{m_s^{1\prime}(m_s^1-1)} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime}(m_s^3+1)} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{14} \sqrt{[S^1(S^1+1) - m_s^1(m_s^1-1)][S^4(S^4+1) - m_s^4(m_s^4+1)]} \times \delta_{m_s^{1\prime}(m_s^1-1)} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime}(m_s^4+1)} + \\
&J_{23} \sqrt{[S^2(S^2+1) - m_s^2(m_s^2-1)][S^3(S^3+1) - m_s^3(m_s^3+1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime}(m_s^2-1)} \delta_{m_s^{3\prime}(m_s^3+1)} \delta_{m_s^{4\prime} m_s^4} + \\
&J_{24} \sqrt{[S^2(S^2+1) - m_s^2(m_s^2-1)][S^4(S^4+1) - m_s^4(m_s^4+1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime}(m_s^2-1)} \delta_{m_s^{3\prime} m_s^3} \delta_{m_s^{4\prime}(m_s^4+1)} + \\
&J_{34} \sqrt{[S^3(S^3+1) - m_s^3(m_s^3-1)][S^4(S^4+1) - m_s^4(m_s^4+1)]} \times \delta_{m_s^{1\prime} m_s^1} \delta_{m_s^{2\prime} m_s^2} \delta_{m_s^{3\prime}(m_s^3-1)} \delta_{m_s^{4\prime}(m_s^4+1)} \Big],
\end{aligned} \tag{S21}$$

where δ_{ij} is the Kronecker δ between states i and j. Calculation of the HDV Hamiltonian matrix elements followed by diagonalization returns the physical spin ladder of the system. The two lowest energy eigenvectors of the diagonalized Hamiltonian are for the $S_T = 1/2$ Kramer's doublet, from which the SPFs, c_i, for site i are obtained by calculation of the site expectation values of $\hat{S}_z$,

$$c_i = \frac{\langle \hat{S}_z \rangle_i}{S_T}. \tag{S22}$$

The SPFs are then applied to correct calculated HFCCs for nuclei associated with each site i. In the present case, the 5′-^{13}C, 4′-^{13}C, and 5′-C–^{1}H HFCCs for each structural model of Ω (as well as the methyl-^{13}C HFCC for the organometallic model) were corrected using the SPF for the unique iron (Fe1). For the methyl-^{13}C HFCCs of the SAM-bound [4Fe-4S]$^{1+}$ cluster calculations, it was found that the distance between the methyl carbon of SAM and the two apical Fe atoms (Fe1 and Fe3) was comparable, ranging from 4.49 Å and 4.14 Å, respectively, in the crystallographic SAM-bound cluster to 3.74 Å and 3.76 Å, respectively, in the 2.6 Å Fe-S distance SAM-NAC structure (Table B1). We therefore treated the methyl carbon in the same manner as a bridging ligand when determining the appropriate SPF for this HFCC. We calculated a distance-weighted average of the SPFs of these two Fe sites and then applied the resulting average SPF to correct the calculated methyl-^{13}C HFCC. The average and standard deviation of the absolute value of the isotropic component of these hyperfine tensors are then reported in Table B1 for comparison with experiment.

Table B1. Distances between the methyl carbon of SAM and the apical irons (Fe1 and Fe3) of the [4Fe-4S]$^{1+}$ cluster as the unique iron–SAM sulfur distance is constrained.

Fe1–S Distance (Å)	Fe1–methyl-C Distance (Å)	Fe3–methyl-C Distance (Å)
3.53	4.49	4.14
3.20	4.20	3.99
3.00	4.05	3.94
2.80	3.89	3.87
2.60	3.74	3.76

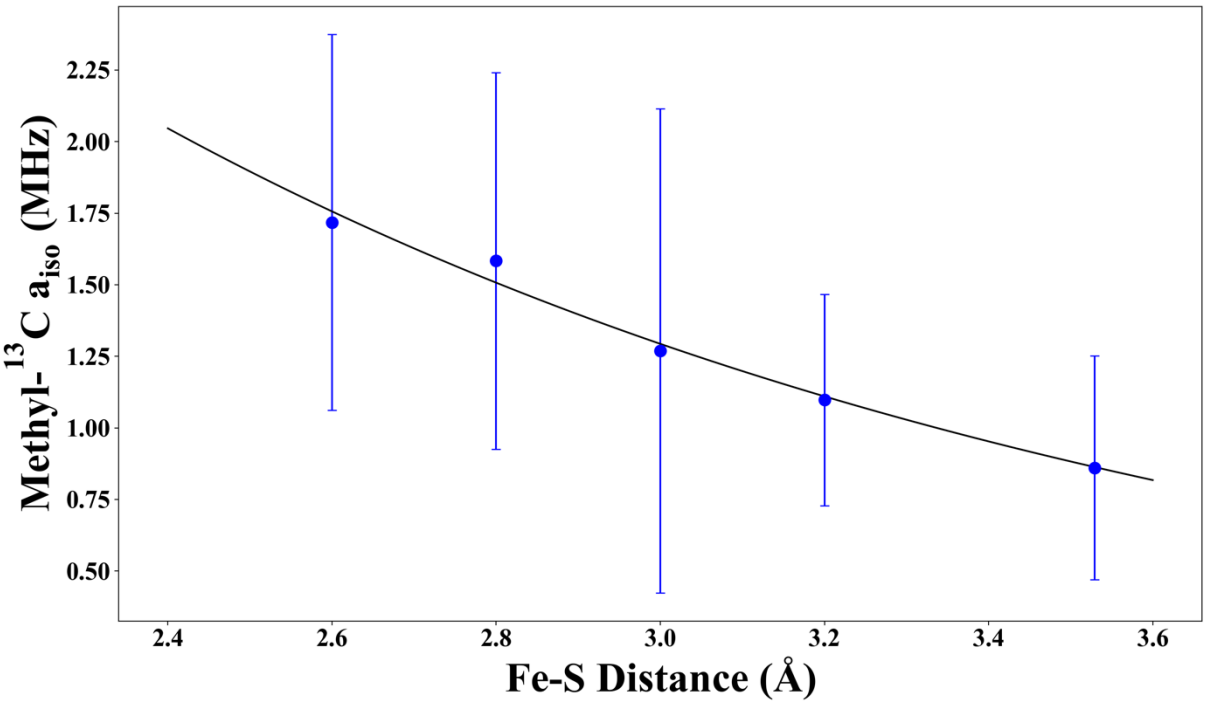

Figure B1. Calculated methyl-^{13}C isotropic hyperfine coupling constants (HFCCs) for SAM near-attack conformer (SAM-NAC) geometries with fixed unique iron–SAM sulfur distances. Calculated values are averaged over all BS valence isomers with m_s = +1/2 for each fixed distance with error bars showing the standard deviation. The best fit curve to a decaying exponential equation is shown in black.