2018

Computational Insights into Sulphur-Crosslinking Enzymatic Mechanisms

Taqred Hussain Alnakhli
University of Windsor

Computational Insights into Sulfur-Crosslinking Enzymatic Mechanisms

By

Taqred H. Alnakhli

A Thesis

Submitted to the Faculty of Graduate Studies

Through the Department of Chemistry and Biochemistry

In Partial Fulfillment of the Requirements for

the Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

Computational Insights into Sulfur-Crosslinking Enzymatic Mechanisms

by

Taqred Alnakhli

APPROVED BY:

_____

L. Porter
Department of Biological Sciences

_____

B. Mutus
Department of Chemistry & Biochemistry

_____

J. Gauld, Advisor
Department of Chemistry & Biochemistry

March 29, 2018

Declaration of Co-Authorship

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

Chapter3 was done in collaboration with Mr. Wanlei Wei under the supervision of Prof. James W. Gauld.

Chapter4 was done in collaboration with Miss Sahar Nikoo under the supervision of Prof. James W. Gauld.

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

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

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this

thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Sulfur plays a diverse array of important roles in biochemical systems. This is due to its ability to possess a range of oxidation states, engage in a variety of bonding environments (e.g., thiols, disulfide bonds, hypervalent species), and reversibly participate in redox chemistry. In this thesis we have examined two biochemically important system that involve sulfur and exploit its rich chemistry.

In Chapter 3 we examine the hypobromous acid mediated formation of the ancient but critical inter-protein sulfilimine (S=N) crosslink of Collagen IV. Previously it was proposed to form either via a halosulfonium or haloamine intermediate, with the pathway via the former being preferred. Using a density functional theory-chemical cluster (QM-cluster) approach we show that formation of a haloamine intermediate is thermodynamically favoured, with it lying much lower in energy than the proposed halsulfonium intermediate. However, its barrier to formation is higher than for the halosulfonium intermediate. Hence, formation of the sulfilimine bond via the latter is kinetically favoured.

In chapter 4, we examine possible mechanisms by which the iron-dependent enzyme thiazole synthase forms a life-essential thiazole ring-containing metabolite. In particular, comparison of available experimental crystal structures with QM-cluster and QM/MM optimized enzyme-substrate structures shows that the substrate is unlikely to be a glycyl-imine. Rather, it is more likely a -C(CH_3)=N-containing species. Furthermore, water density analysis of the substrate-bound active site suggests that it may be the conduit for mechanistically required protons to enter the active site. In addition, our results suggest that the loss of the iron ion from the active site may occur earlier in the mechanism than proposed, before formation of the proposed thione intermediate.

These results provide new insights into these important bonds and species and lay the ground work for further computational and experimental studies.

Dedication

I dedicate this work to my family.

Acknowledgements

Very special thanks to my supervisor Dr. James Gauld for welcoming me in his incredible group and for all the efforts he made and advices he is giving me during my study. I could not have imagined having better advisor for my master study. He was more than happy to help anytime I need him.

Beside my supervisor I also like to thank my committee members, Dr. Bulent Mutus and Dr. Lisa Porter for taking the time to read my thesis. Also would like to thank Marlene Bezaire on the organizing and guiding me during all this time.

I would like to thank all members of the Gauld group Paul Meister, Mohamed Aboelnga, Travis Dewolfe, Wanlei Wei, Anupom Roy, everyone was amazing helping and supporting all this time. Special thanks to Sahar Nikoo for all the time we spent together chatting and working together on the research.

I also want to thank the Saudi Cultural Bureau for giving me the opportunity and all the support to pursue my study abroad.

Last, I would like to finish with thanking all my friends and my family for their support. I am grateful to my wonderful mother Najiah Alnakhli, my father Hussain Alnakhli and all my sisters and brothers. Especial thanks to my brother Ali Alnakhli who spent all these years with me far away from his family.

Table of Contents

List of Figures

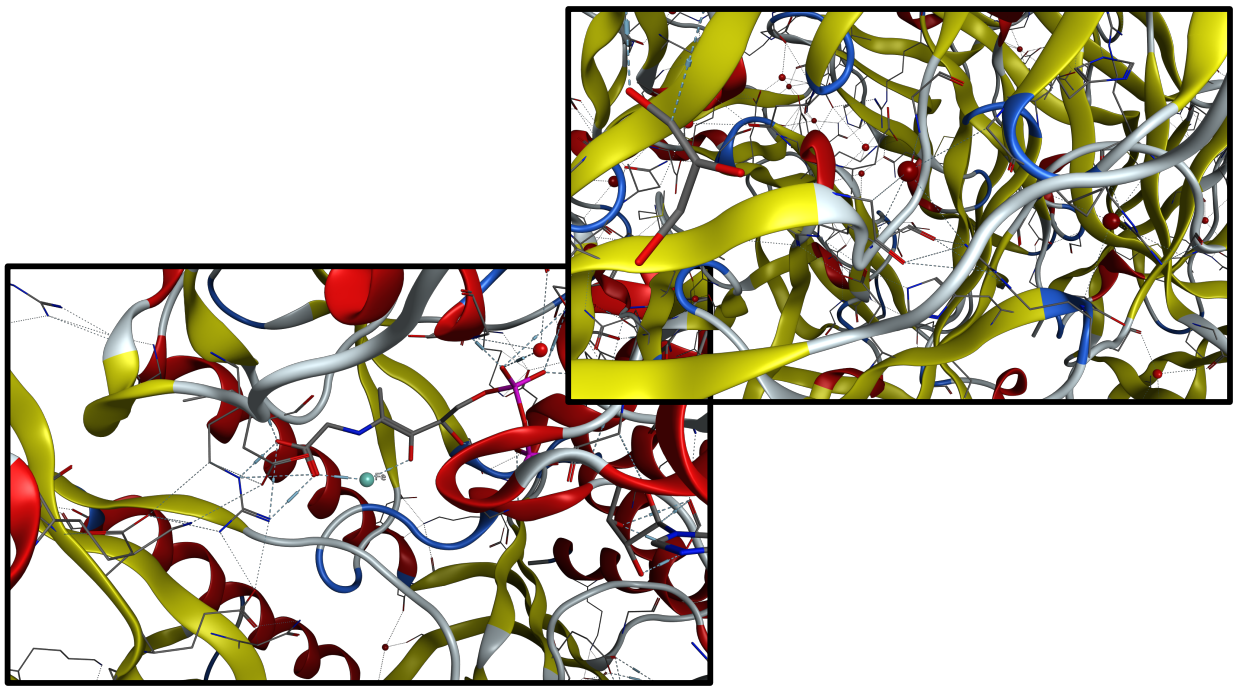
List of Abbreviations and Symbols

FAD	Flavin adenine dinucleotide
NAD	Nicotinamide adenine dinucleotide
Sec	Selenocysteine
ADT	Adenylated thiazole
DFT	Density functional theory
PES	Potential energy surfaces
MD	Molecular dynamics
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
MM	Molecular mechanics
NAMD	Nanoscale molecular dynamics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
AMBER	Assisted Model Building with Energy Refinement
ONIOM	Our own n-layered integrated molecular orbital and molecular mechanics
IMOMM	Integrated molecular orbital and molecular mechanics
PC	Product complex
RC	Reactant complex
TS	Transition state/structure
vdW	Van der Waals
IEFPCM	Integrated equation formalism polarizable continuum model
ZPVE	Zero point vibrational energy corrections

NBO	Natural bond order
QT-AIM	Quantum theory of atoms in molecules
PDB	Protein data bank
MOE	Molecular Operator Environment
RMSD	Root mean square deviation

CHAPTER 1:

Introduction



1.1 Biomolecules

Biomolecules is often used as an all-encompassing term for the molecules found in or synthesized in cells. However, it is common to divide biomolecules into four main categories based on their predominant component(s): (i) amino acids and proteins, (ii) carbohydrates, (iii) nucleic acids, and (iv) lipids (Figure 1).[1] Each biomolecule category differs in their structure and physical properties; however, all are important for cell and organism functions.[1] Further functionalization is often achieved through incorporation of other atoms or ions such as metal ions (e.g. heme groups), or conjugating biomolecules together, even from different categories (e.g. glycoproteins).[2-4]

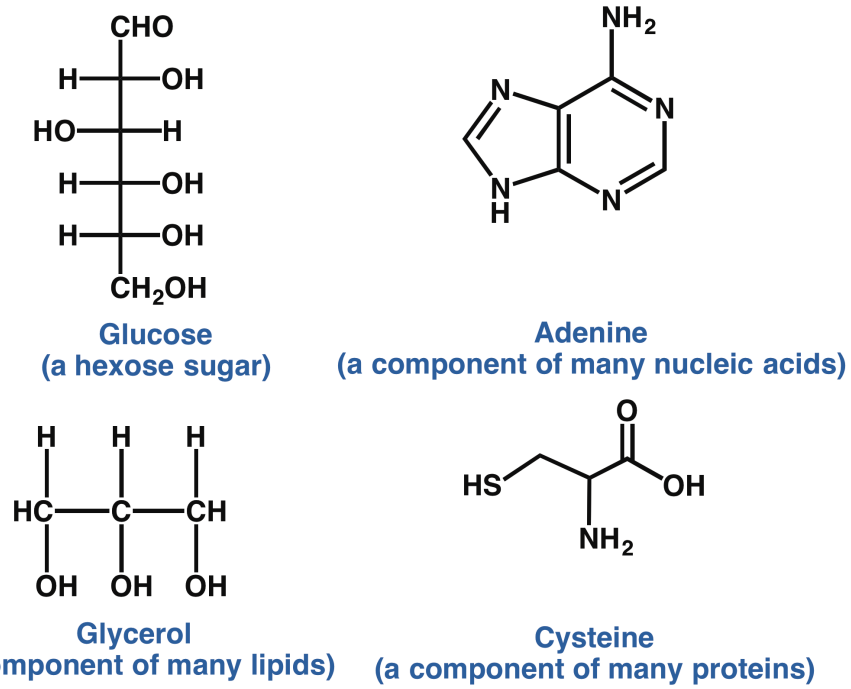


Figure 1.1. Illustration of example monomers or components of each of the four main categories of biomolecules: carbohydrates, nucleic acids, lipids, and proteins.

1.2 Amino Acids and Proteins:

Of the four major categories above, amino acids and proteins are arguably one of the most studied. Amino acids are, in their own right, important metabolites in many physiological processes. This is highlighted by, for example, that high levels of proline in the blood can result in hyperprolinemia; a physiological disorder that can result in seizures or neurological disorders.[5-6]

Meanwhile, proteins are polymers built from their constituent monomeric amino acids.[7-8] Importantly, they are central to many key physiological processes and in fact are often referred to as the workhorses of cells.[9] For example, they are critical for maintaining both cellular and organism structure.[10-12] Furthermore, many proteins function as transporters that carry, for instance, essential molecules to their site of need. For example, hemoglobin is a protein that is responsible in mammals for transporting O_2 from the lungs to the tissues, and also transports CO_2 from the tissues to the lungs.[13] Other proteins are involved in storage such as myoglobin which stores O_2 in tissues, while others act as receptors that bind their targeted signaling molecules and/or modify their shapes based on messages received from inside or outside the cell.[14]

As noted above proteins are polymers formed from the condensing of their constituent monomer amino acids together via amide bonds.[7] There are 22 encoded proteinogenic amino acids, with most organisms having 20 amino acids encoded within their genetic codes (Figure 2).[15]

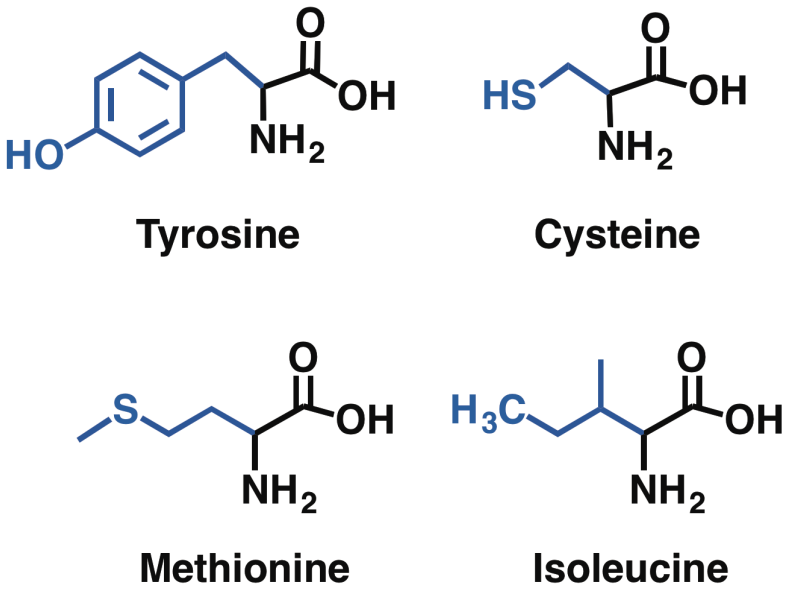
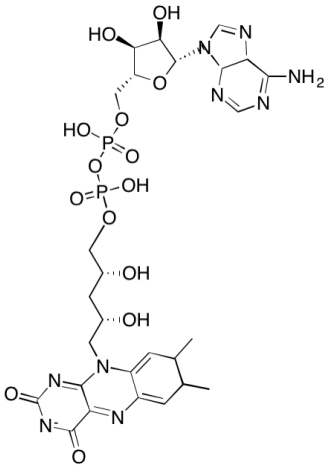
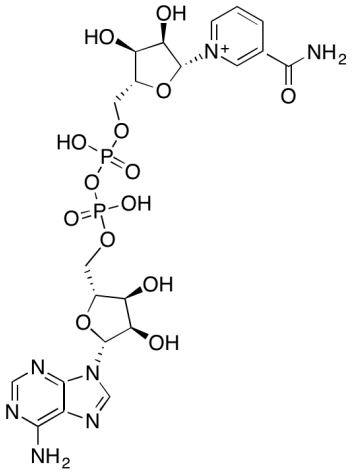


Figure 1.2. Examples of some amino acids in proteins.

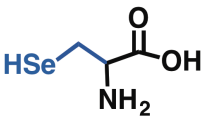
All of these amino acids only contain Carbon, Nitrogen, Oxygen and Hydrogen except Cysteine and Methionine which also contain a sulfur. The sequence of amino acids is also known as the proteins primary structure.[15] Importantly, it dictates the proteins secondary (e.g. α-helices and β-sheets) and tertiary structures.[16] Thus, every protein has its own unique sequence and misfolding of a protein can have significant health consequences (e.g. prion proteins).[17] It is due in part to this huge variety of possible structures that enables proteins to take on their numerous and diverse roles. Furthermore, many proteins expand their possible functionality through post-translational modification (e.g. phosphorylation) or by incorporating cofactors and coenzymes such as heme groups and metal ions as detailed below (Figure 3).[18-20]



Flavin adenine dinucleotide



Nicotinamide adenine dinucleotide



Selenocysteine

Figure 1.3. Schematic illustration of two common enzyme cofactors, Flavin adenine dinucleotide (FAD) and Nicotinamide adenine dinucleotide (NAD), and the non-standard amino acid selenocysteine (Sec).

This enables proteins to also expand their functional roles. Two of the main physiological roles of proteins are as catalysts or structural components of cells and organisms.

1.3 Structural Proteins:

As noted above, many proteins have a structural role inside a cell or, alternatively external to a cell.[10, 12] For example, all cells contain a cytoskeleton that functions to at least in part to maintain cell shape and control placement of organelles.[21] In eukaryotic cells the cytoskeleton is composed of three different proteins. In particular, actin is a common structural globular protein in all eukaryotic cells that is also able to self-associate into polymeric filaments.[22] In one or both of these forms it is important for muscle contraction, where it interacts with another protein (myosin) that "walks" along the actin filaments.[1, 23-24] Within, for example,

organisms, many proteins are found in the extracellular matrix where they help to bestow structure and strength to organs and the organism.[25] The commonest structural protein in the connective tissues of animals is the collagen, a fibrous triple-helix protein.[1] Meanwhile, many bacteria produce biofilms which is generally a conglomeration composed of nucleic acids, carbohydrates, and proteins.[26] Notably, such biofilms also help to protect the bacterial cells embedded within them and thus can render pathological bacteria more resistant to antibiotic treatment.[27] Meanwhile, in higher organisms such as humans, tissues are often use an extracellular matrix that is critical to correct maintenance of their structure and function. Essential to the proper role of such an extracellular matrix are proteins such as Collagen IV (see Chapter 3).

1.4 Catalytic Proteins: Enzymes

Enzymes play crucial roles in the biochemistry of all organisms. In particular, they ensure that many of the essential reactions in cells and organisms occur at life-sustaining rates. In broad terms, they achieve their remarkable rate-enhancing abilities through lowering the barrier(s) of a reaction or by catalyzing lower-energy alternative pathways (Figure 4).[28]

More specifically, however, the fundamental chemical principles that lie at the origin of the catalytic power of enzymes has been at times attributed to numerous sources including, for example, transition structure stabilization, reactant destabilization, desolation, electrostatic and active site pre-organization.[29] As a result, enzyme catalyzed reactions often proceed thousands of times faster than without an enzyme.

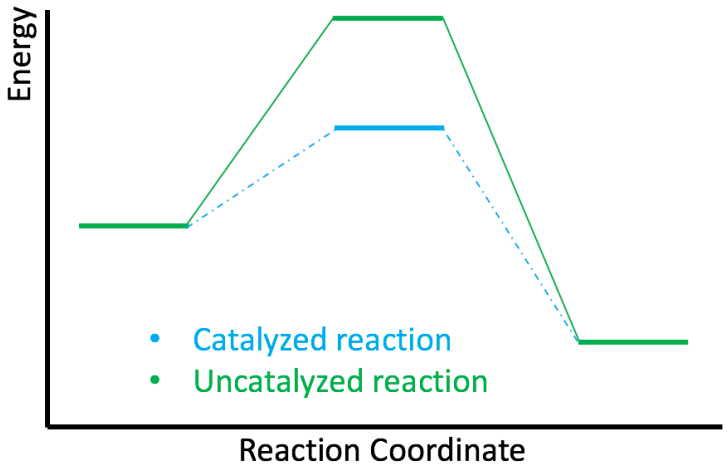


Figure 1.4. Illustration of the effect of enzymes on the reaction energy where the blue line represents the catalyzed reaction and the green line represent the uncatalyzed reaction.

Enzymes are often classified and named based on the type of reaction they catalyze. For example, proteases or peptidases hydrolytically cleave peptide bonds.[30] There are many classes of proteases that exist in different species, which in fact accomplish the exact reaction.[31] Meanwhile, other enzymes known as transferases catalyze the transfer of a group from one substrate component to another. Enzymes may also be named based on the type of cofactor or co-enzyme that they may require in order to achieve their catalytic function. A common group of enzymes are those that require a metal ion for catalysis; they are referred to as metalloenzymes.[32]

1.5 Metalloenzymes.

Metal-containing proteins, metalloproteins, have diverse functions in biological systems from transport, storage to transduction proteins. Indeed, it has been

estimated that half of all proteins fall within this category.[33] This diversity of roles is often due to the chemical abilities of the metal ion(s) they contain. For instance, metals can participate in electron transfer processes, ligand binding, or perhaps more passively as structural components.[3] Ordinarily, the metal ion(s) coordinate to the protein via an oxygen, nitrogen or sulfur within an amino acid residue of the protein's polypeptide chain.[34]

Metalloenzymes are those enzymes that contain a metal ion that is essential to their catalytic function. While many metal ions are found in enzymes, including Copper and Zinc, one of the commonest types of metalloenzymes are those in which an iron is involved.[35] It has the ability to bind to more than one ligand due to its unoccupied d orbitals. Its oxidation states range from +2 to +6 and, depending on the ligand that iron binds to, its electronic spin states and the biological redox potentials can vary markedly. For example, the heme-containing P450 enzymes, which are central to numerous physiological and metabolic processes including nitric oxide (NO) production, involve an iron that's oxidation state varies from +2 to +4 over the course of their catalytic mechanism.[36] While in some of iron-containing enzymes the metal ion may be redox active, in others it is redox passive. For example, formation of Thiamin diphosphate, a metabolite found in all living cells, requires a non-heme iron containing enzyme protein.[37]

1.6 Computational enzymology:

There are numerous experimental techniques available for the study of enzymes including X-ray crystallography, mutagenesis, and kinetic measurements. Computational chemistry is an alternative approach in which computers are used to model the properties and reactions of chemical species.[38] With the development

of new computational methods and ever more-powerful computers, it can now be applied to massive chemical molecules or complexes. The application of such methods to the study of enzymes is known as computational enzymology.[39] The latter is now well-established as an important tool for the study of enzymes as it can be applied as readily to elucidating highly-reactive or transient species (e.g., reaction transition structures) as stable, long-lived species.[28] Furthermore, using a multi-scale computational enzymology approach one can, for instance, study the dynamic nature of an entire enzyme-ligand complex, or the mechanistic impact of mutation of a single enzyme active site atom. Importantly, this can now be done with such accuracy and reliability that it can be used to either complement or guide experimental studies.[28]

1.7 References:

1. Lamond, A. I., Molecular biology of the cell, 4th edition. *Nature.* **2002,** *417*, 383-383.

2. Gogoi, P.; Chandravanshi, M.; Mandal, S. K.; Srivastava, A.; Kanaujia, S. P., Heterogeneous Behavior of Metalloproteins Toward Metal Ion Binding and Selectivity: Insights From Molecular Dynamics Studies. *J. Biomol. Struct. Dyn.* **2016,** *34*, 1470-1485.

3. Bowman, S. E. J.; Bridwell-Rabb, J.; Drennan, C. L., Metalloprotein Crystallography: More Than a Structure. *Acc. Chem. Res.* **2016,** *49*, 695-702.

4. Shu, X. G.; Su, J. H.; Du, K. J.; You, Y.; Gao, S. Q.; Wen, G. B.; Tan, X. S.; Lin, Y. W., Rational Design of Dual Active Sites in a Single Protein Scaffold: a Case Study of Heme Protein in Myoglobin. *ChemistryOpen.* **2016,** *5*, 192-196.

5. Duarte, M.; Afonso, J.; Moreira, A.; Antunes, D.; Ferreira, C.; Correia, H.; Marques, M.; Sequeira, S., Hyperprolinemia as a Clue in The Diagnosis of a Patient with Psychiatric Manifestations. *Brain Dev.* **2017,** *39*, 539-541.

6. Ferreira, A. G. K.; Scherer, E. B.; da Cunha, A. A.; Manfredini, V.; Biancini, G. B.; Vanzin, C. S.; Vargas, C. R.; Wyse, A. T. S., Hyperprolinemia Induces DNA, Protein and Lipid Damage in Blood of Rats: Antioxidant Protection. *Int. J. Biochem. Cell Biol.* **2014,** *54*, 20-25.

7. Chou, P. Y., Prediction of Protein Structural Classes from Amino Acid Compositions. In *Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman, G. D., Ed. Springer US: Boston, MA, 1989; pp 549-586.

8. Kidera, A.; Konishi, Y.; Oka, M.; Ooi, T.; Scheraga, H. A., Statistical Analysis of the Physical Properties of the 20 Naturally Occurring Amino Acids. *J. Protein. Chem.* **1985,** *4*, 23-55.

9. Frauenfelder, H.; Fenimore, P. W.; Chen, G.; McMahon, B. H., Protein Folding is Slaved to Solvent Motions. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 15469-15472.

10. Branden, C.; Tooze, J., Introduction to Protein Structure. **1991**.

11. Nagarajan, R.; Chothani, S. P.; Ramakrishnan, C.; Sekijima, M.; Gromiha, M. M., Structure Based Approach for Understanding Organism Specific Recognition of Protein-RNA Complexes. *Biology Direct.* **2015,** *10*.

12. Nagarajan, R.; Archana, A.; Thangakani, A. M.; Jemimah, S.; Velmurugan, D.; Gromiha, M. M., PDBparam: Online Resource for Computing Structural Parameters of Proteins. *Bioinform. Biol. Insights.* **2016,** *10*, 73-80.

13. Giardina, B.; Messana, I.; Scatena, R.; Castagnola, M., The Multiple Functions of Hemoglobin. *Crit. Rev. Biochem. Mol. Biol.* **1995,** *30*, 165-196.

14. Silverstein, T. P.; Kirk, S. R.; Meyer, S. C.; Holman, K. L. M., Myoglobin Structure and Function: a Multiweek Biochemistry Laboratory Project. *Biochem. Mol. Biol. Educ.* **2015,** *43*, 181-188.

15. Wu, G., Amino acids: metabolism, functions, and nutrition. *Amino Acids.* **2009,** *37*, 1-17.

16. Zhang, C. T.; Chou, K. C., An Optimization Approach to Predicting Protein Structural Class from Amino-Acid-Composition. *Protein. Sci.* **1992,** *1*, 401-408.

17. Kupfer, L.; Hinrichs, W.; Groschup, M. H., Prion Protein Misfolding. *Curr. Mol. Med.* **2009,** *9*, 826-835.

18. Wang, Y. C.; Peterson, S. E.; Loring, J. F., Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Res.* **2014,** *24*, 143-160.

19. Huang, G. H.; Li, X. M., A Review of Computational Identification of Protein Post-Translational Modifications. *Mini-Rev. Org. Chem.* **2015,** *12*, 468-480.

20. Loew, G. H.; Harris, D. L., Role of the Heme Active Site and Protein Environment in Structure, Spectra, and Function of the Cytochrome p450s. *Chem. Rev.* **2000,** *100*, 407-419.

21. Jaeken, L., A New List of Functions of the Cytoskeleton. *IUBMB Life.* **2007,** *59*, 127-133.

22. Fletcher, D. A.; Mullins, R. D., Cell Mechanics and the Cytoskeleton. *Nature.* **2010,** *463*, 485-492.

23. Tang, D. D.; Gerlach, B. D., The Roles and Regulation of the Actin Cytoskeleton, Intermediate Filaments and Microtubules in Smooth Muscle Cell Migration. *Respir. Res.* **2017,** *18*, 12.

24. Salter, B.; Pray, C.; Radford, K.; Martin, J. G.; Nair, P., Regulation of Human Airway Smooth Muscle Cell Migration and Relevance to Asthma. *Respir. Res.* **2017,** *18*, 15.

25. Turoverova, L. V.; Khotin, M. G.; Yudintseva, N. M.; Magnusson, K.-E.; Blinova, M. I.; Pinaev, G. P.; Tentler, D. G., Analysis of Extracellular Matrix Proteins Pproduced by Cultured Cells. *Cell and Tissue Biol.* **2009,** *3*, 497.

26. Friedman, L.; Kolter, R., Genes Involved in Matrix Formation in Pseudomonas Aeruginosa PA14 Biofilms. *Mol. Microbiol.* **2004,** *51*, 675-690.

27. Høiby, N.; Bjarnsholt, T.; Givskov, M.; Molin, S.; Ciofu, O., Antibiotic Resistance of Bacterial Biofilms. *Int. J. of Antimicrob. Agents.* **2010,** *35*, 322-332.

28. Carvalho, A. T.; Barrozo, A.; Doron, D.; Kilshtain, A. V.; Major, D. T.; Kamerlin, S. C., Challenges in Computational Studies of Enzyme Structure, Function and Dynamics. *J. Mol. Graph. Model.* **2014,** *54*, 62-79.

29. Giraldo, J.; Roche, D.; Rovira, X.; Serra, J., The Catalytic Power of Enzymes: Conformational Selection or Transition State Stabilization? *FEBS Lett.* **2006,** *580*, 2170-2177.

30. López-Otín, C.; Bond, J. S., Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **2008,** *283*, 30433-30437.

31. Shen, H.-B.; Chou, K.-C., Identification of Proteases and Their Types. *Anal. Biochem.* **2009,** *385*, 153-160.

32. Ulmer, D. D.; Vallee, B. L., Structure and Function of Metalloenzymes. In *Bioinorganic Chemistry*, Am. Chem. Soc. 1971, 100, 187-218.

33. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., Design of Functional Metalloproteins. *Nature.* **2009,** *460*, 855-862.

34. Smith, A. M., CHAPTER 1 Interaction of Metal Ions with Proteins as a Source of Inspiration for Biomimetic Materials. In *Functional Metallosupramolecular Materials*. The Royal Society of Chemistry. 2015, 1-31.

35. Eser, B. E.; Zhang, X.; Chanani, P. K.; Begley, T. P.; Ealick, S. E., From Suicide Enzyme to Catalyst: The Iron-Dependent Sulfide Transfer in Methanococcus Jannaschii Thiamin Thiazole Biosynthesis. *J. Am. Chem. Soc.* **2016,** *138*, 3639-3642.

36. Poulos, T. L., Heme Enzyme Structure and Function. *Chem. Rev.* **2014,** *114*, 3919-3962.

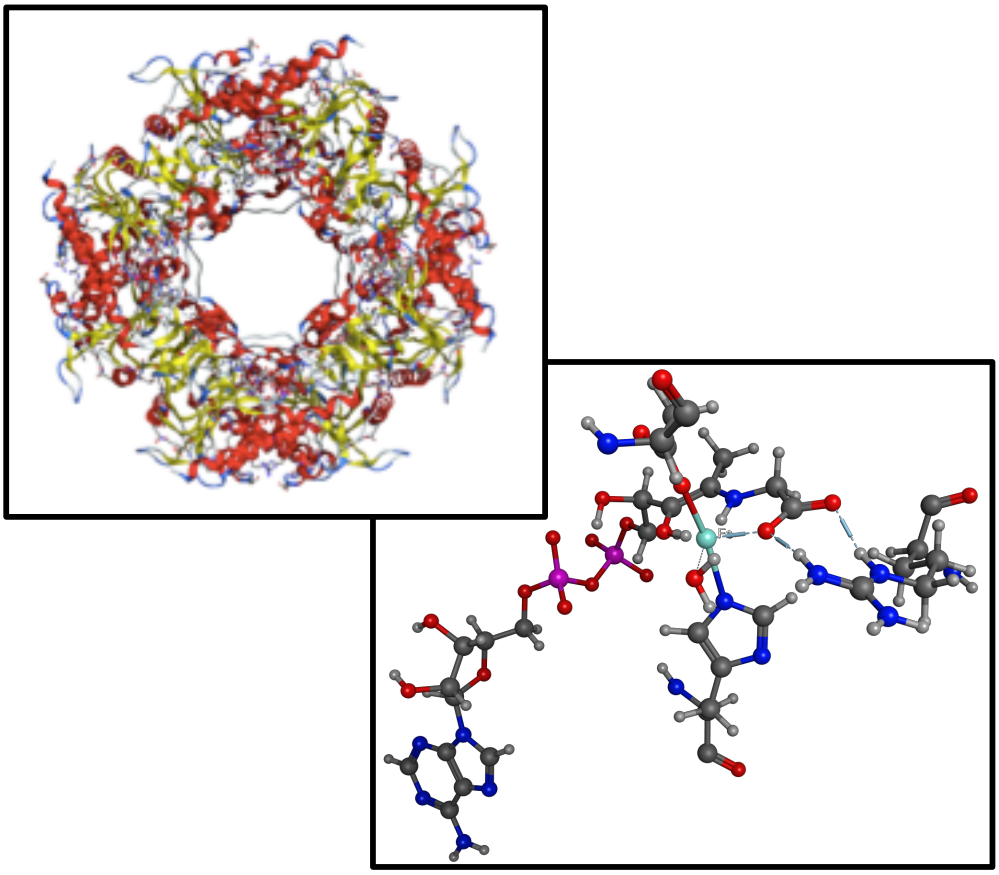
37. Zhang, X.; Eser, B. E.; Chanani, P. K.; Begley, T. P.; Ealick, S. E., Structural Basis for Iron-Mediated Sulfur Transfer in Archael and Yeast Thiazole Synthases. *Biochemistry.* **2016,** *55*, 1826-1838.

38. Bruice, T. C., Computational Approaches: Reaction Trajectories, Structures, and Atomic Motions. Enzyme Reactions and Proficiency. *Chem. Rev.* **2006,** *106*, 3119-3139.

39. Gherib, R.; Dokainish, H.; Gauld, J., Multi-Scale Computational Enzymology: Enhancing Our Understanding of Enzymatic Catalysis. *Int. J. Mol. Sci.* **2013,** *15*, 401-422.

CHAPTER 2:

Computational Approaches



2.1 Introduction:

Modern computational chemistry is the use of computers to model the properties and reactions of molecular species and systems. More specifically, it is the use of computers to apply the mathematical equations and theorems of quantum chemistry; which is itself the application of quantum mechanics to chemical problems.[1]

A major impetus for the development of computational methods has been the need for investigate the properties and mechanisms of systems that be challenging to traditional experimental means or beyond their limits.[2] For example, many of the species that occur within organic tissues, such as reactive sulfur species, cannot be tested or are difficult to be measured experimentally.[3] To be able to address such challenges, over time researchers have explored and developed a range of new computational methods that can be used on every larger chemical system. In addition, many of these methods can now be used alongside experiments and in many cases, can guide experiments and set benchmark values.[2]

Indeed, a variety of methods is now available that can be used for systems of differing complexities and sizes, and as such can be applied to the widest breadth of chemistry. For example, it has enabled researchers to design new therapeutic drugs and advanced materials, as well as investigate and elucidate the mysteries of processes inside organisms, in some cases without the need for experiment.[4] One particular area in which computational chemistry has had tremendous impact is enzymology; this application is commonly referred to as computational enzymology.[4-5-6] In such systems, its strengths come to the fore in that it can almost as readily and accurately applied to reactive and transient species such as transitions structures as it can to long lived stable species such as products.[5-6]

Computational enzymology methods range from quite approximate to highly accurate.[1-7] Ab initio methods, for example, are based on quantum mechanics and only use the fundamental physical constants in their solution.[7] In contrast, other methods use empirical parameters to, for example, simplify the equations or integrals that need to be solved.[1] Such methods are commonly referred to as semi-empirical.[1]

In this thesis, computational chemistry has been used to study biomolecular and enzymatic problems. Many of these methods and approaches used have been reviewed in numerous excellent reviews and books.[1-7-8] Hence, only a brief overview of some of the key features pertinent of thesis are discussed herein.

2.2 The Foundational Equation of Quantum Chemistry:

At the heart of quantum chemistry lies the time-independent Schrödinger Equation can be deceptively simply written as in **Equation 2.1.**[1-7]

$$\hat{H}\psi = E\psi \qquad \textbf{(2.1)}$$

where $\hat{\boldsymbol{H}}$ is the Hamiltonian Operator that acts on the wave function ψ that completely describes the chemical system of interest, to give the energy ***E*** of that system multiplied by ψ.

The wave function is itself a function of both the nuclear coordinates of each of the nuclei in the system and the Cartesian coordinates and the spin of each electron in the system.[1-7] In fact, determining the exact solution of the wave function is only possible for a single electron system.[1,7] Similarly, for many systems the

Hamiltonian is unknown or too complicated to be solved for exactly.[7] Hence, there is a need to use approximate methods to obtain useful chemical insights. Each of these methods uses different approximations or concepts in order to obtain approximate solutions to the above **Equation 2.1**. As a result, each method has it strengths but also its limitations.

2.3 Molecular Mechanics(MM):

One common approach is to treat atoms or groups of atoms using classical mechanics. The resulting type of methods are referred to as molecular mechanics (MM), or MM force fields.

This approach, is amongst the simplest computational methods available.[9] For instance, they treat an atom, composed of its nucleus and electrons, as a sphere with a certain set of properties (e.g. charge). Structural parameters are also treated classically. For example, bonds are modelled as springs between two hard spheres. As a result, however, MM methods can not calculate electronic properties and processes. Therefore, electronic spectra and bond breaking and making formation can not be modelled. Rather, MM force fields can and have been used extensively to model the conformations and dynamic (see Molecular Dynamics) nature of, for example, proteins and nucleic acids.[10] In addition, due to their low computational costs they can be applied to considerably larger systems. That is, MM methods are capable of giving insights into the non-covalent interactions between molecules in small systems as well as in proteins with thousands of atoms.

There are many different force fields that have been developed and used over the last several years. Of the variety of force fields available, CHARM,

AMBER and MMFF94 have perhaps been amongst the most extensively used for proteins.[11-12] For each, the total energy of the system is generally considered as being composed of contributions as outlined in **Equation 2.2**.

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{van\ der\ waals} + U_{coulomb} \quad \textbf{(2.2)}$$

where the terms describe the contributions from, in turn, bond stretching, bending, dihedral bonds, and Van der Waal's and electrostatic interactions.[11]

In this thesis the most common forcefield used is Amber96 for which the total energy can be written as in **Equation 2.3**:[13]

$$E_{total} = \sum_{bond} k_r(r - r_{eq}) + \sum_{angles} k_\theta(r - r_\theta) + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad \textbf{(2.3)}$$

where the first two-terms use harmonic potentials for bond stretching and bending respectively, the third term is the dihedral potential, and the last term is the Van der Waals (described by what is commonly simply referred to as a 12-6 potential) and a Coulombic electrostatic component respectively.[11-13]

2.4 Molecular Dynamic (MD) Simulations:

For some systems one may want to examine or take into account the dynamic behaviour that can occur in biochemical systems. Molecular dynamics (MD) simulations are an effective tool for studying or modelling such behaviours

including conformational interconversions, interactions between molecules, and ligand or therapeutic drug binding.[14]

These methods are based upon solving Newton's equations of motion.[15] Since the first MD simulations in ~1977, which were on small proteins and less than 10 ps in duration, there have been numerous advances that have greatly expanded their utility, and importantly, accuracy.[16-17] Indeed, it is now possible to perform MD simulations on massively large multi-protein complexes, or membrane fragments with their associated proteins.[10] Furthermore, simulations on such complexes can be run for 100's of ns in duration and can be analyzed to monitor a single distance or those of a large ensemble (e.g. active site groups) or an entire complex over time.[18] Nowadays, in computational enzymology, it is an important technique that enables researchers to confidently determine the most likely stable confirmation(s) of a system which can then be used, if required, in further studies (e.g. QM-cluster or QM/MM studies; see below).

In general, prior to beginning the MD simulation, an appropriate X-Ray crystal structure (e.g. high resolution; as complete as possible) is chosen of the protein. In some cases, the structure is then modified by mutating residues, or replacing missing atoms (e.g. hydrogens), or by binding a desired ligand. The entire complex is then solvated (e.g. by water if we wish to mimic the environment inside a cell). The entire solvated complex is then energy minimized. The resulting structure is then used as the starting structure for the subsequent MD simulation.

In an MD simulation, as noted above, Newton's equations of motion and other required information is used to calculate the structure of the complex as it varies over a period of time.[16-17] Hence, one must select a time-step (e.g. 2 fs) and a series of structures are generated at each step for the duration of the simulation.

As this is potentially a computationally expensive process an MM forcefield is typically used to describe the complex. An added advantage of MD simulations is that one can vary the temperature and thus one can allow for thermal affects such as conformational variation at room or body temperature (e.g. 298 K).

In this thesis the MD programs NAMD and AMBER have been used, with the chosen forcefield being Amber12 in order to obtain thermally equilibrated solvated structures.[13]

2.5 Quantum Mechanics (QM) Methods:

In Quantum Mechanical methods one takes a different approach to solving the Schrödinger Equation (**Equation 2.1**); namely, we attempt to solve it using only the fundamental physical constants and invoking only a few rigorous mathematical approximations.[1-7] It should be noted that for this thesis both wave function-based methods and density functional theory (DFT) methods are considered as QM methods.

The field-free molecular Hamiltonian operator in **Equation 2.1** is more fully represented in **Equation 2.4**:

$$\widehat{H} = \frac{-\hbar^2}{2m_e}\sum_{electrons\ (i)} \nabla_i^2 - \frac{\hbar^2}{2m_N}\sum_{nuclei\ (N)} \nabla_N^2 + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{r_{AB}} - \sum_{i,A} \frac{Z_A}{r_{iA}} \qquad \textbf{(2.4)}$$

where the terms in order of occurrence are the:

kinetic energy of the electrons,

kinetic energy of the nuclei,

electron-electron interactions,

nuclei-nuclei interactions,

nuclei-electron interactions.

That is, the Hamiltonian, and similarly wave functions, depend on both the electron and nuclear components. This Hamiltonian and wave function is too complicated to be solved. However, by recognizing the fact that in chemistry we are generally interested in electronic properties we can simplify the problem by invoking the Born-Oppenheimer (BO) Approximation.[1-7]

The BO Approximation can be simply worded that as the nuclei are so much heavier than the electrons and thus move so much more slowly, that to a first-order approximation, from the perspective of the electrons the nuclei are stationary. This leads to the following simplified equation and the electronic Hamiltonian ($\hat{H}_{el}$) **in Equation 2.5**:

$$\hat{H} = \hat{H}_{el} + V_{NN} = \frac{-\hbar^2}{2m_e}\sum_{electrons\ (i)} \nabla_i^2 + \sum_{i>j} \frac{1}{r_{ij}} - \sum_{i,A} \frac{Z_A}{r_{iA}} \qquad \textbf{(2.5)}$$

where V_{NN} is now a constant. This approximation also impacts the wave function by essentially making it now only dependent on the electronic coordinates.

In wave function based methods (e.g. MP2) we also invoke the *Molecular Orbital Approximation* which can be stated as 'the motion of the electrons are assumed to be independent of each other".[7] This means that we can also write the wave function as a product of 1-electron orbitals.

In DFT methods we instead invoke the Hohenberg-Kohn and Kohn-Sham Theorems which collectively state that from the density of a system we can obtain all physical properties and the wave function, and that the density of an n-electron

system can be divided into 1-electron density packets.[1]

For biochemical systems DFT methods such as B3LYP and M062X, have become the methods of choice due to their cheaper computational cost yet often similar reliability and accuracy compared to wave function based methods.[19] In practical terms while wave function based methods may be applied to systems of up to 20-30 atoms, DFT can be applied to systems of up to 200 atoms.[20]

For both wave function and DFT methods one must also choose a basis set, a set of basis functions. These are used to construct and/or describe the wave function or density. In theory, the larger the basis set, the more complete it is, and hence the more accurate the result. Common examples of basis sets used in this thesis include 6-31G(d) and 6-311+G(2df,p).

For enzymes and biochemical systems, however, we often need to use a chemical model that is composed of thousands of atoms. Consequently, we must use methods that combine two or more methods in some manner so as to be able to study such systems.

2.6 Quantum Mechanics/Molecular Mechanics (QM/MM):

The 2013 Nobel Prize in Chemistry was given to Arieh Warshel, Michael Levitt and Martin Karplus for their work in the development of hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods which Warshel and Levitt had first proposed in 1979.[21-22-23] Later, in 1995 Maseras et al. introduced what is known as Integrated Molecular Orbital + Molecular Mechanics (IMOMM) which considered to be an alternative to QM/MM hybrid method.[23]

The common feature of these methods, however, is that a chemical system is divided into several layers (e.g., like an onion), each of which can be treated using

a different computational method.[24] As a result, one can now computationally study systems of thousands of atoms. Notably, this approach takes advantage of the fact that a given method such as MM can reliably and accurately model long-range effects or steric of, for instance, the environment surrounding an enzyme active site, while QM methods can describe bond making and breaking processes that occur within the active site.[23] In general one could think of a QM/MM energy as being written as in **Equation 2.6**:

$$E_{QM/MM} = E_{QM(Model)} + E_{MM(Environment)} + E_{QM-MM} \quad \textbf{(2.6)}$$

where the energy of the whole system ($E_{QM/MM}$) is the sum of the energy of the model (also known as reactive or high) region (E_{QM}) that is described by a QM method, the energy of the environment (E_{MM}) described using an MM method, and the interaction between the QM and MM-described regions (E_{QM-MM}).

In this thesis we have exclusively used the ONIOM QM/MM method. It is a two-layer approach that combines the accuracy of a QM methods and the speed of MM approaches.[21] QM methods are applied to atoms in the enzyme's active site where reactions take place, whereas the rest is treated using an MM approach (Figure 2.1). As a result, that region in which electronic changes are occurring and are important for obtaining geometries, energies and electronic properties are described using an appropriate QM method. Whereas on the other hand the rest is treated with the MM approaches that are better for describing non-bonded interactions such as van der Waals interaction.

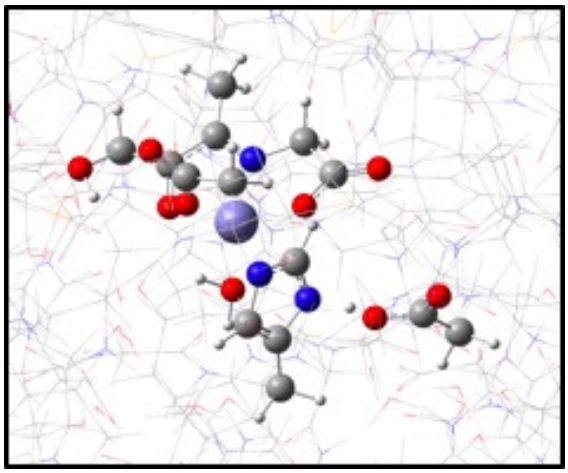


Figure 2.1. Illustration of an ONIOM(QM/MM) approach where the ball and sticks indicate atoms placed in the QM-region and wire indicates atoms in the MM-layer.

ONIOM(QM/MM) is also known as a subtractive scheme as the energy is calculated using the general formula in **Equation 2.7**:[23]

$$E_{QM/MM} = E_{QM(model)} + E_{MM(real)} - E_{MM(model)} \quad \textbf{(2.7)}$$

where the energy of the whole system ($E_{QM/MM}$) equals the energy of the model (the reactive) region ($E_{QM(model)}$) plus the energy of the entire system ($E_{MM(real)}$) minus the MM energy of the model region ($E_{MM(model)}$).

2.7 Relative Energies and Potential Energy Surfaces (PESs):

It is now well established in computational practice that optimized structures can generally be obtained using relative modest or small basis sets.[25] However, to obtain accurate energies one must use a large or larger basis set. To reduce the potential computational cost of optimizing structures using large basis sets or

higher levels of theory, a common approach is to perform single-point energy calculations as illustrated in **Equation 2.8**:[26-27]

Method 2/Basis Set 2//Method 1/Basis set 1 **(2.8)**

This indicates that the optimized geometry was obtained using Method 1/Basis Set 1. Then, using this geometry, a single point energy calculation was performed at the Method 2/Basis Set 2 level of theory. Additional energy corrections can also be made to the above formula such as Zero-Point Vibrational Energy Corrections (ZPVE) or Gibb's Free Energy Corrections.

In addition, as a result of the Born-Oppenheimer Approximation we do not explicitly solve the entire reaction surface. Rather, we obtain the optimized structures and energies of chemically key points along the reaction such as reactant complexes, intermediates, transition structures, and product complexes. Then, plotting the energies of these species we can represent their relative energies in the form of a potential energy surface (Figure 2.2). Such surfaces, as our group has noted before, can provide very accurate insights into the chemical mechanism being investigated as they clearly illustrate both the interconnections and reaction relationships between species, and a reactions or mechanisms thermochemistry.[1-27]

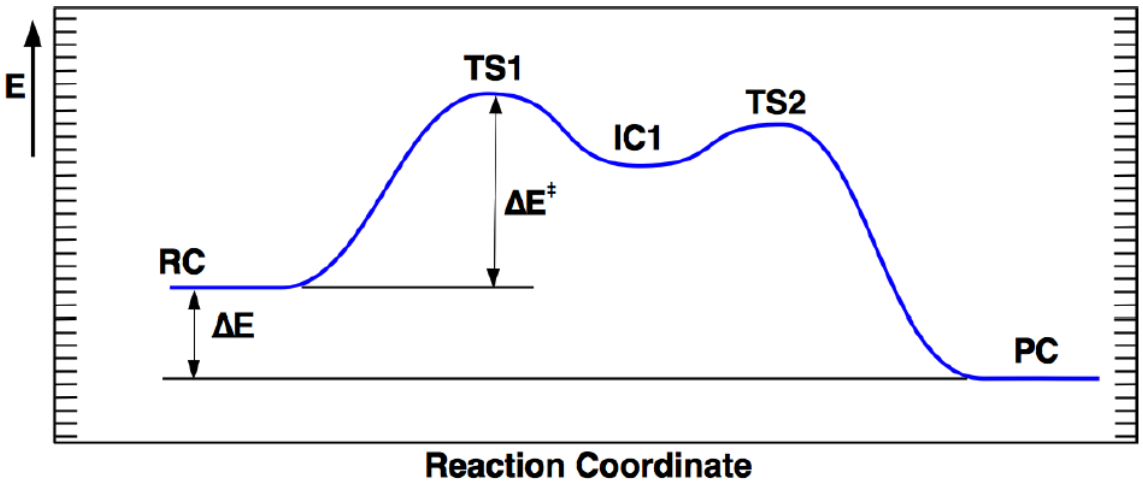


Figure 2.2. A Potential Energy Surface (PES) plot of energy versus reaction coordinate illustrating the relationships between chemically key structures and points along a reaction such as the reactive complex (RC), transition structures (TS1 and TS2), an intermediate complex (IC), and the final product complex (PC). Note: key thermochemical features such as barrier heights (ΔE) and overall reaction energies (ΔE) are also indicated.

2.8 References:

1. Cramer, C. J., *Essential of Computational Chemistry: Theories and Models*. John Wiley & Sons Ltd. 2nd edition ed.; USA, 2004.

2. Gherib, R.; Dokainish, H. M.; Gauld, J. W., *Int. J. Mol. Sci.* **2013,** *15*, 401-422.

3. Paulsen, C. E.; Carroll, K. S., Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery. *Chem. Rev.* **2013,** *113*, 4633-4679.

4. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J., Computational Enzymatic Catalysis - Clarifying Enzymatic Mechanisms with The Help of Computers. *Phys. Chem. Chem. Phys.* **2012,** *14*, 12431-12441.

5. Himo, F., Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **2017,** *139*, 6780-6786.

6. Bushnell, E. A. C.; Llano, J.; Eriksson, L. A.; Gauld, J. W., Mechanisms of Mutagenic DNA Nucleobase Damages and Their Chemical an Enzymatic Repairs Investigated by Quantum Chemical Methods. In *Selected Topics in DNA Repair*, Chen, C. C., Ed. InTech: 2011; Vol. 1, pp 389-414.

7. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A., *Ab initio molecular orbital theory*. Wiley: New York, 1986.

8. Tunon, I.; Moliner, V., *Simulating Enzyme Reactivity*. The Royal Society of Chemistry: UK, 2017.

9. Maseras, F.; Morokuma, K., IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comp. Chem.* **1995,** *16*, 1170-1179.

10. Karplus, M., Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.* **2002,** *35*, 321-323.

11. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for The Simulation of Proteins, Nucleic Acids, and Organic Molecules (vol 117, pg 5179-5197, 1995). *J. Am. Chem. Soc.* **1996,** *118*, 2309-2309.

12. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comp. Chem.* **2009,** *30*, 1545-1614.

13. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984,** *106*, 765-784.

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

15. Vishveshwara, S.; Ghosh, A.; Hansia, P., Intra and Inter-Molecular Communications Through Protein Structure Network. *Curr. Protein Pept. Sci.* **2009,** *10*, 146-160.

16. Alder, B. J.; Wainwright, T. E., Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **1957,** *27*, 1208-1209.

17. Karplus, M.; Petsko, G. A., Molecular dynamics simulations in biology. *Nature* **1990,** *347*, 631-639.

18. Siegbahn, P. E. M., Modeling Aspects of Mechanisms for Reactions Catalyzed by Metalloenzymes. *J. Comp. Chem.* **2001,** *22*, 1634-1645.

19. Chattaraj, P. K.; Pérez, P.; Zevallos, J.; Toro-Labbé, A., Ab initio SCF and DFT studies on solvent effects on intramolecular rearrangement reactions. *J. Phys. Chem. A* **2001,** *105*, 4272-4283.

20. Li, X.; Cai, Z.; Sevilla, M. D., DFT calculations of the electron affinities of nucleic acid bases: dealing with negative electron affinities. *J. Phys. chem. A* **2002,** *106*, 1596-1603.

21. Groenhof, G., Introduction to QM/MM Simulations. In Bimolecular Simulations Methods and Ptotocols, Monticelli, L., Ed. Humana Press, Totowa, NJ: 2013; Vol. 924.

22. Mulholland, A. J.; Lyne, P. D.; Karplus, M., Ab Initio QM/MM Study of The Citrate Synthase Mechanism. A low-Barrier Hydrogen Bond is Not Involved. *J. Am. Chem. Soc.* **2000***, 122,* 534-535.

23. Chung, L. W.; Sameera, W. M.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K., The ONIOM Method and Its Applications. *Chem. Rev.* **2015,** *115*, 5678-5796.

24. Senn, H. M.; Thiel, W., QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* **2009,** *48*, 1198-1229.

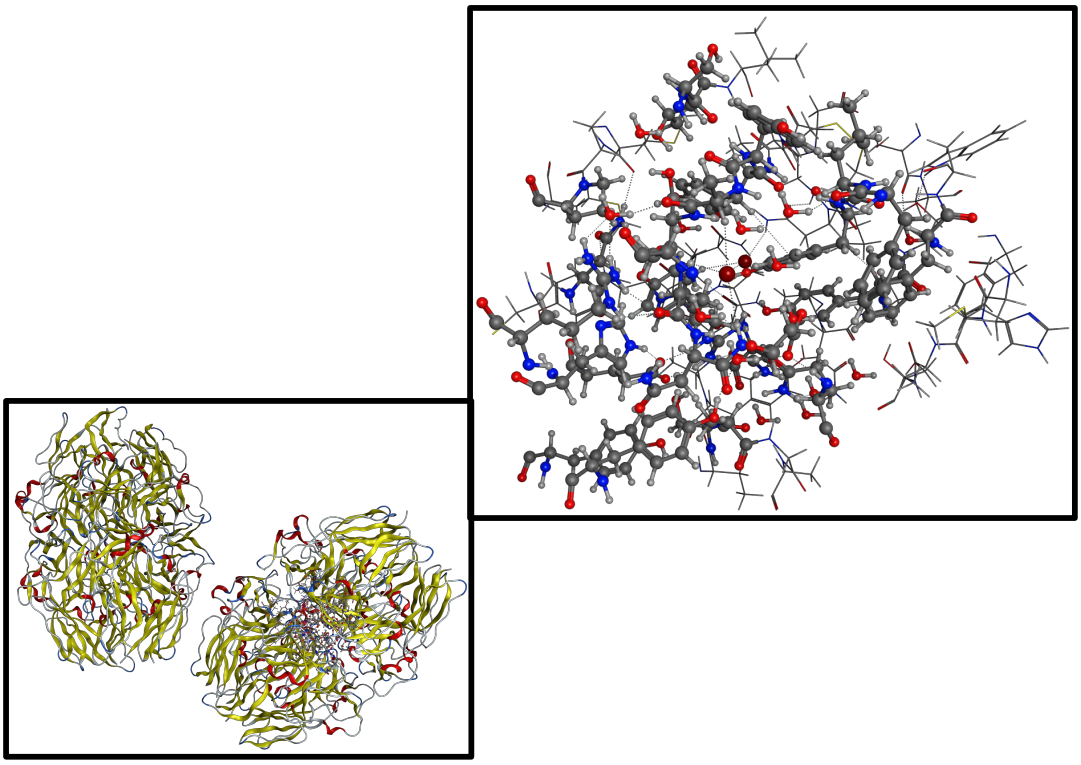
25. Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O. N.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J.; Zeng, Q.; Case, D. A.; Eastman, P.; Wang, L.-P.; Essex, J. W.; Head-Gordon, M.; Pande, V. S.; Ponder, J. W.; Shao, Y.; Skylaris, C.-K.; Todorov, I. T.; Tuckerman, M. E.; Head-Gordon, T., Advanced Potential Energy Surfaces for Molecular Simulation. *J. Phys. Chem. B* **2016,** *120*, 9811-9832.

26. Alcamí, M.; Luna, A.; Mó, O.; Yáñez, M.; Tortajada, J., Theoretical survey of the potential energy surface of ethylenediamine + Cu^+ reactions. *J. Phys. Chem. A* **2004,** *108*, 8367-8372.

27. Bushnell, E. A. C.; Huang, W.; Gauld, J. W., Applications of Potential Energy Surfaces in the Study of Enzymatic Reactions. *Adv. Phys. Chem.* **2012,** *2012*, 1-15.

CHAPTER 3:

Mechanism of Sulfilimine Bond Formation in Collagen IV Using QM/MM and QM-Cluster Approaches



3.1 Introduction:

Collagen IV, discovered in 1966,[1] is a critical constitutive molecular component of basement membranes.[2] The latter is an important extracellular matrix that, for example, acts as a supportive scaffold for the epithelial and endothelial cells and also facilitates interactions between their cell receptors and macromolecules.[3-4] In fact, it has been shown to be essential for embryogenesis process and wound healing.[5-8]

It is also now known collagen IV, in particular its improper structure or function, plays a central role in several diseases including immune-related rheumatological (i.e., affect joints) and dermatological (i.e., affect muscles and skin) diseases.[9-10] In addition, it is also thought to be the antigen that induces Goodpasture's syndrome; an autoimmune disorder caused by mistaken attacks on the basement membrane in tissues of the lungs and kidneys.[11-12] This can ultimately lead to, for instance, kidney failure. Moreover, Alport syndrome is a genetic disorder in collagen IV that impacts the eyes, hearing, and kidneys.[13-14] Sufferers experience vision impairment, hearing, and declining kidney function.[5-8]

Collagen IV comprises six homologous but distinct polypeptides that are referred to as α1(IV) to 6(IV). Three chains combine into a select few types of trimers, a protomer, which then aggregating end to end to form a hexamer.[15] It is known that Collagen IV undergoes several post-translational modifications (e.g., hydroxylation of lysine's) of which one of the most important, structurally, is the formation of sulfilimine (S=N) crosslinks between Methionine (Met93) and Lysine (Lys211) residues in different collagen IV protomers.[16-17] This covalent crosslink is

rare in biochemistry; so far only being known to occur naturally in collagen IV. It was previously unknown how such bonds are formed. Recently, however, in vitro studies identified that the bromine and chlorine hypohalous acids, hypobromous (HOBr) and hypoclorous (HOCl), produced by the enzyme peroxidasin can induce sulfilimine formation, although their exact roles remaining unclear or unknown.[8, 18] Interestingly, the efficiency of HOBr facilitated sulfilimine formation was experimentally measured to be 50,000-fold greater than when HOCl was used.[18] Based on their experimental studies they proposed a mechanism of sulfilimine formation, mediated by hypohalous acids (HOX; X = Br, Cl) which is shown below in Figure 3.1.[19]

More specifically, the bromine or chlorine of the hypohalous acid first attacks at either the Met93 Sulphur or side-chain nitrogen of Lys211 to form a halosulfonium or haloamine, respectively. Subsequently, the now activated Met93 sulfur or Lys211 side-chain nitrogen is now susceptible to attack from the sidechain heteroatom center of the other residue. Of these two possible reaction pathways, they concluded that reaction of the hypohalous acid with Met93 was the dominate pathway. It was also noted that the formation of the halosulfonium may also explain the observed formation of a sulfoxide minor product.

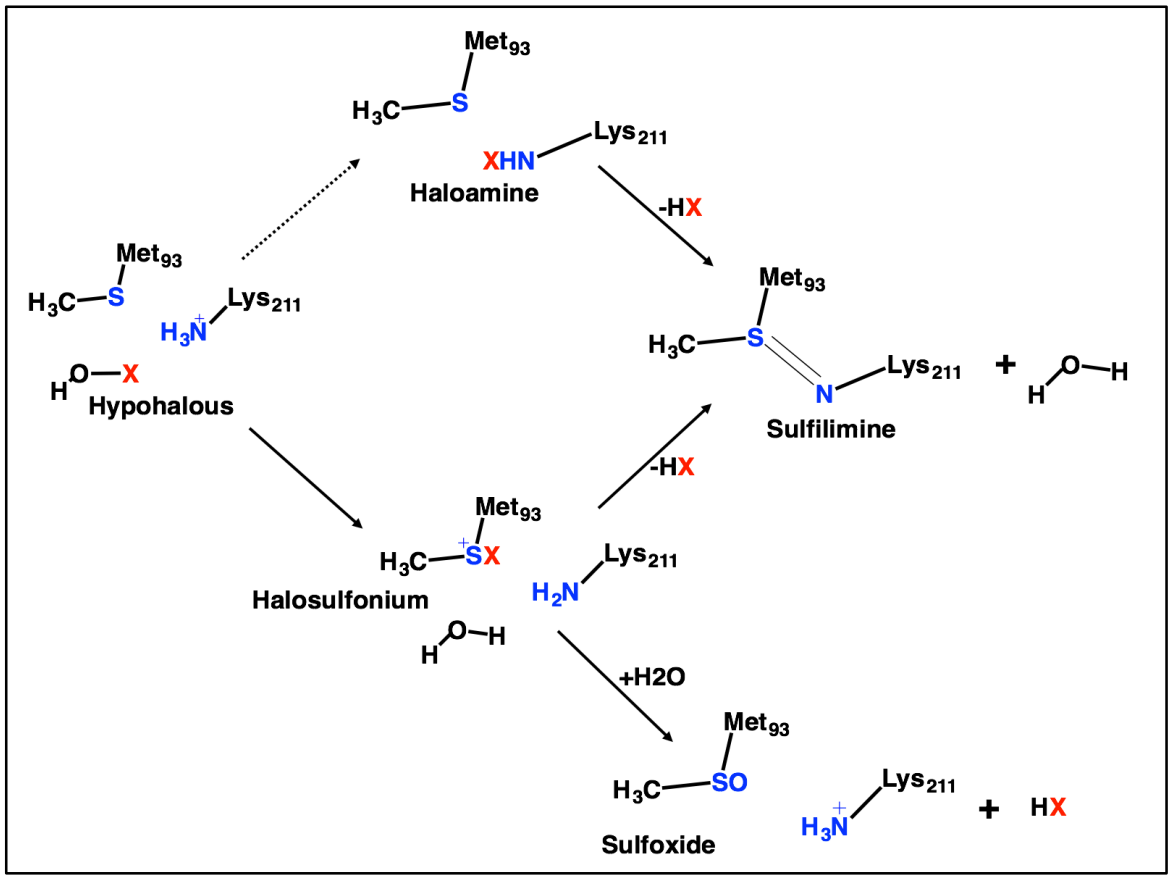


Figure 3.1. The proposed[19] mechanism for formation of the sulfilimine bond in Collagen IV.

To date, and to the best of our knowledge, there has been no previous computational investigation into this important and poorly understood biochemical mechanism. However, in 2009, Pichierri[20] performed a detailed computational study on the nature of the S-N bond in sulfilimine's such as collagen IV.[20] Based on their results obtained in part via both Natural Bond Order (NBO) analysis and Quantum Theory of Atoms in Molecules (QT-AIM), they concluded that the S–N distance in such bonds should be around 1.6 Å and that it is in fact a polarized (strongly towards the nitrogen) covalent single bond.[20]

In this study, we have performed a detailed multi-scale computational approach to investigate the proposed[19] mechanism for sulfilimine formation in Collagen IV as facilitated by hypobromous acid. More specifically, we have examined the two proposed pathways for sulfilimine bond formation; attack of bromine on either the side-chain Sulphur of Met93 to form a halosulfonium cation intermediate or the side-chain nitrogen of Lys211 to form a haloamine intermediate. In addition, we have also examined the reaction of the halosulfonium intermediate with water to produce sulfoxide.

3.2 Computational methods:

Molecular Dynamic Simulations (MD): MD simulations were performed, by Wanlei Wei (a former group member), using the Molecular Operator Environment (MOE) performing two different crystal structures of two different proteins that were taken from the Protein Data Bank with the PDB codes (PDB ID: 1M3D and 1T60). NAMD program were utilized during the simulations on the two crystal structures.[21] The structures were hydrogenated to place the missing atoms and solvated with water. The structures were then minimized using the force field AMBER12 and submitted for an initial 1 ns run followed by a 15 ns production run.[22] After the simulations, the resultant structures were analyzed and a representative structure of the most populated cluster were obtained to provide a template structure for further calculations using QM and QM/MM approaches.

Quantum Mechanics (QM)-Cluster and Quantum Mechanics/Molecular Mechanics (QM/MM) Studies: A representative structure was derived from the MD simulations and for all QM-cluster and QM/MM calculations the Gaussian 09 software was

used.[23] For the QM-cluster calculations the M06-2X/6-31G(d,p) level of theory was used with the surrounding protein and solvent environment modelled using the IEF-PCM method Using the same level of theory M062X/6-31G(d,p). The M06-2X method was used because of its ability to more accurately describe non-covalent long-range interactions.[24-25] The chemical cluster model used consisted of 27 atoms in total and included just a truncated model of the side-chains of Met93, Lys211, Glu214 and one water molecule. For the two layer ONIOM(QM/MM) calculations the M06-2X/6-31G(d,p) level of theory was used to describe the QM-region while the remainder of the protein and aqueous solvent, the MM region, was modeled using the AMBER96 force field. The QM layer consisted of 79 atoms in total including the side-chains of Met93 and Lys211, Glu214 and 8 water molecules.

3.3 Results and Discussion:

MD results. First, as noted in the Computational Methods, a 15 ns MD simulation was performed to, in part, explore key distances and interactions between residues, in particular those involving Met93 and Lys211, in the reactive site. The average positions of Lys211 and Met93, and distance between their mechanistically key side-chains, in the absence of HOBr, is shown in Figure 3.2.

As it can be seen, the side-chain protonated amine of Lys211 forms a weak electrostatic interaction with the side-chain sulfur of Met93 with the shortest $_{Lys211}NH^{\cdots}S_{Met93}$ having a distance of 3.09 Å. Meanwhile, the distance between the Lys211 amine nitrogen and Met93 sulfur is approximately 3.56 Å. While these distances indicate a weak interaction between the two mechanistically key groups,

they do suggest that even initially reasonably positioned with respect to each other for potential future interaction and/or reaction.

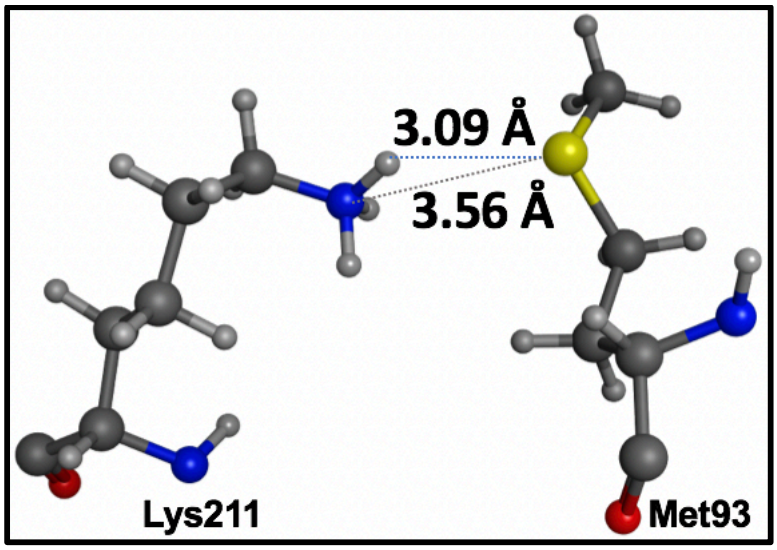


Figure 3.2. Representative structures from most populated cluster in the MD analysis showing the position of Met93 and Lys211.For clarity, the other residues and waters are omitted.

To explore possible and proposed sulfilimine bond formation reactions involving HOBr, as noted in the Computational Methods section a suitable representative structure was then selected and hypobromous acid (HOBr) manually added and positioned in proximity to the side-chains of Met93 and Lys211. It is important to note that except for the addition of hypobromous acid, no additional changes were made on the structures after the simulation.

QM-cluster results. As noted above, it has been proposed that the HOBr can react with Met93, transferring its Br^+ component to the Met93 sulfur center. Obtained

structures of the reactant complex, intermediate and transition structure are shown in Figure 3.3 indicating the distances observed.

As can be seen, in the initial reactant complex (**RC**), the nitrogen of the Lys211 amine (N_{Lys211}), sulfur of Met93 (S_{Met93}) and bromine moiety of the HOBr all lie in proximity to each other though with long interaction distances of approximately 3.5 Å between each centre.

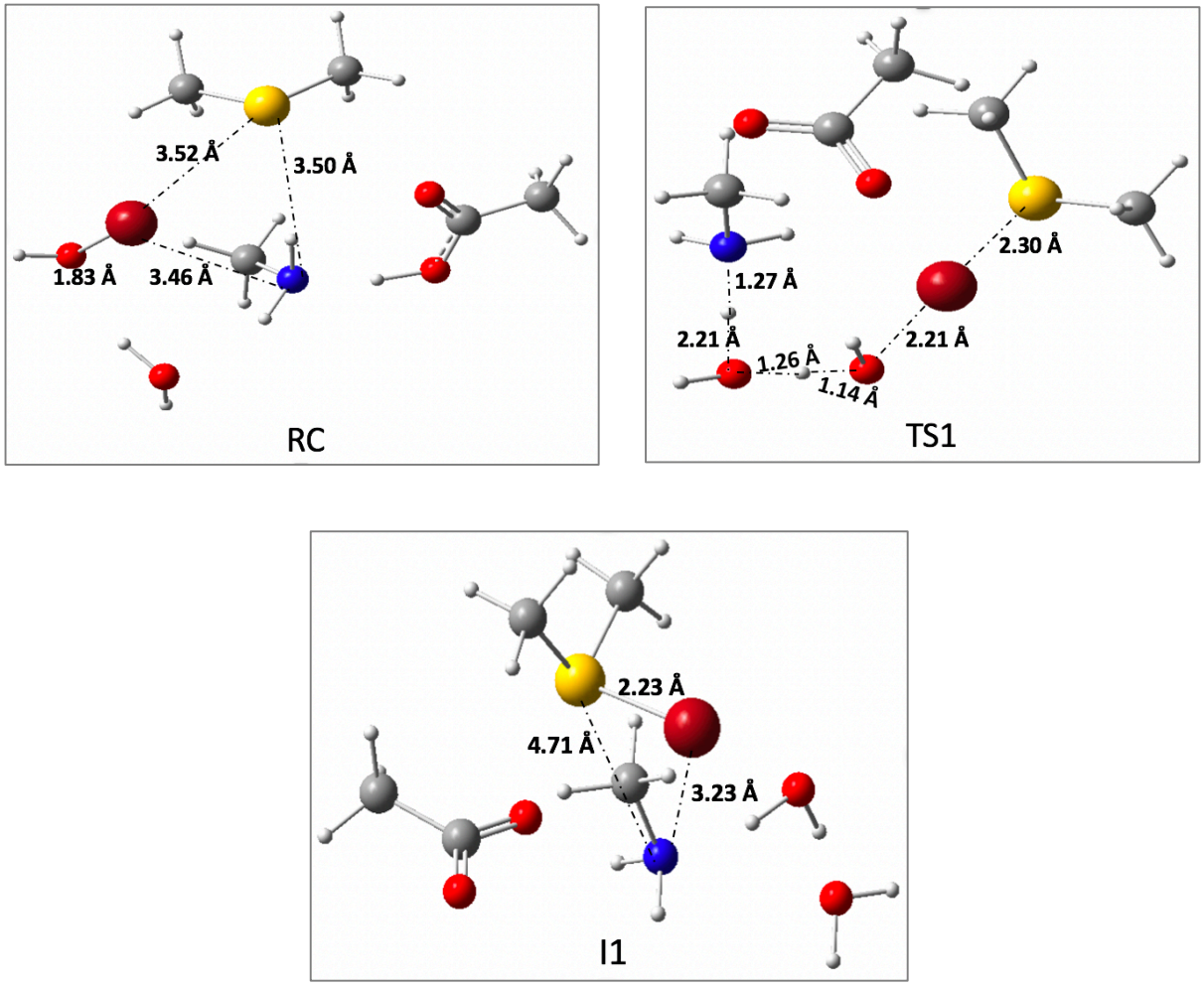


Figure 3.3. Illustration of the optimized initial reactant complex (**RC**), halosulfonium intermediate (**I1**) and corresponding transition structure (**TS1**) obtained using the

present QM-cluster (see Computational Methods) approach. Selected key distances are shown (Angstrom).

The reaction and energies obtained halosulfonium formation obtained using the present QM-cluster method are shown in Figure 3.4. In particular, it can be seen that as the bromine center of HOBr attacks at the Met93 Sulfur centre a proton is transferred via a bridging water molecule onto the oxygen of the HOBr (Figure 3.3: **TS1**). This reaction is predicted to be quite facile, proceeding via **TS1** with a barrier of just 14.8 kJ/mol to give the resulting halosulfonium intermediate **I1**. The latter lies lower in energy than the initial RC by 13.0 kJ/mol, indicating that the reaction is exothermic. In **I1** the $_{Met93}$S–Br bond distance is 2.23 Å indicating formation of a strong bond. Meanwhile, the $_{Met93}$SBr$\cdots$N$_{Lys211}$ distance is still reasonably long at 3.23 Å (Figure 3.4).

We also examined the proposed alternative pathway for formation of the sulfilimine crosslink via a haloamine intermediate (**I2**). The mechanism obtained is shown in Figure 3.5 while the optimized structures are given in Figure 3.6. In this pathway the Lys211 amine nitrogen instead nucleophilically attacks at the bromine center of the HOBr moiety. This proceeds via **TS2** at a cost of 39.0 kJ/mol to give the haloamine intermediate **I2** lying 27.5 kJ/mol lower in energy than **RC***. As observed during halosulfonium formation, formation of the Br-N$_{Lys211}$ bond occurs concomitantly with transfer of a proton onto the oxygen of the HOBr moiety. Similar to the case in the halosulfonium **I1**, the length of the $_{Lys211}$N–Br bond in **I2** (1.90 Å; Figure 3.6) suggests that it is a reasonably strong bond. It should be noted that in this reaction process two bridging waters were required to transfer the proton onto the oxygen of the BrOH moiety during reaction. Furthermore, while the haloamine

intermediate **I2** lies lower in energy than **RC*** than does the halosulfonium **I1** (see, Figure 3.4), the barrier for formation of **I2** is predicted to be almost three times greater than for **I1**. That is, the experimentally observed pathway may be kinetically controlled.

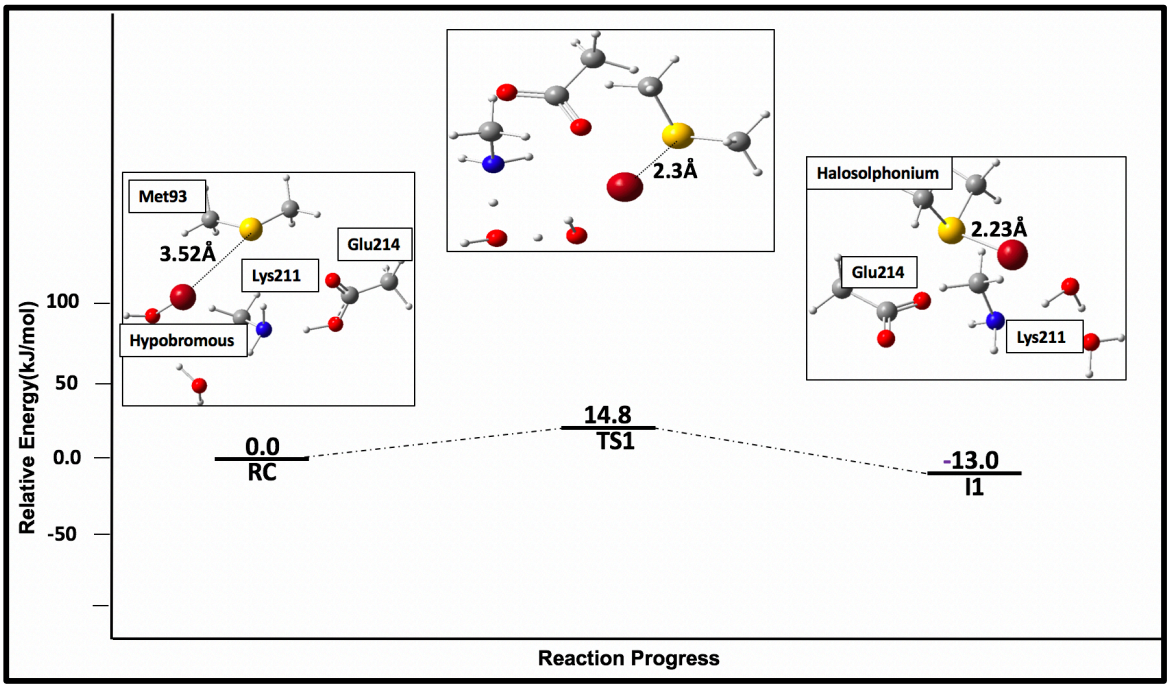


Figure 3.4. Potential energy surface (kJ/mol) obtained using the QM-cluster approach for formation of the proposed halosulfonium intermediate.

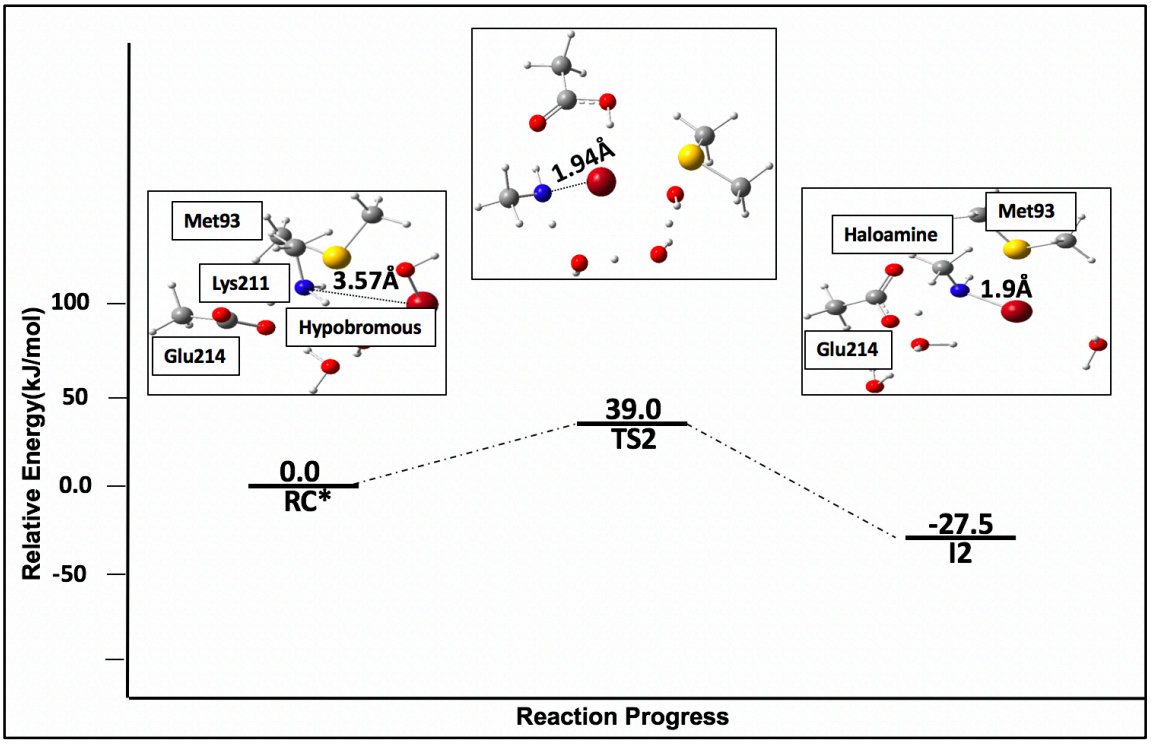


Figure 3.5. Potential energy surface (kJ/mol) obtained using the QM-cluster approach for formation of the proposed haloamine intermediate.

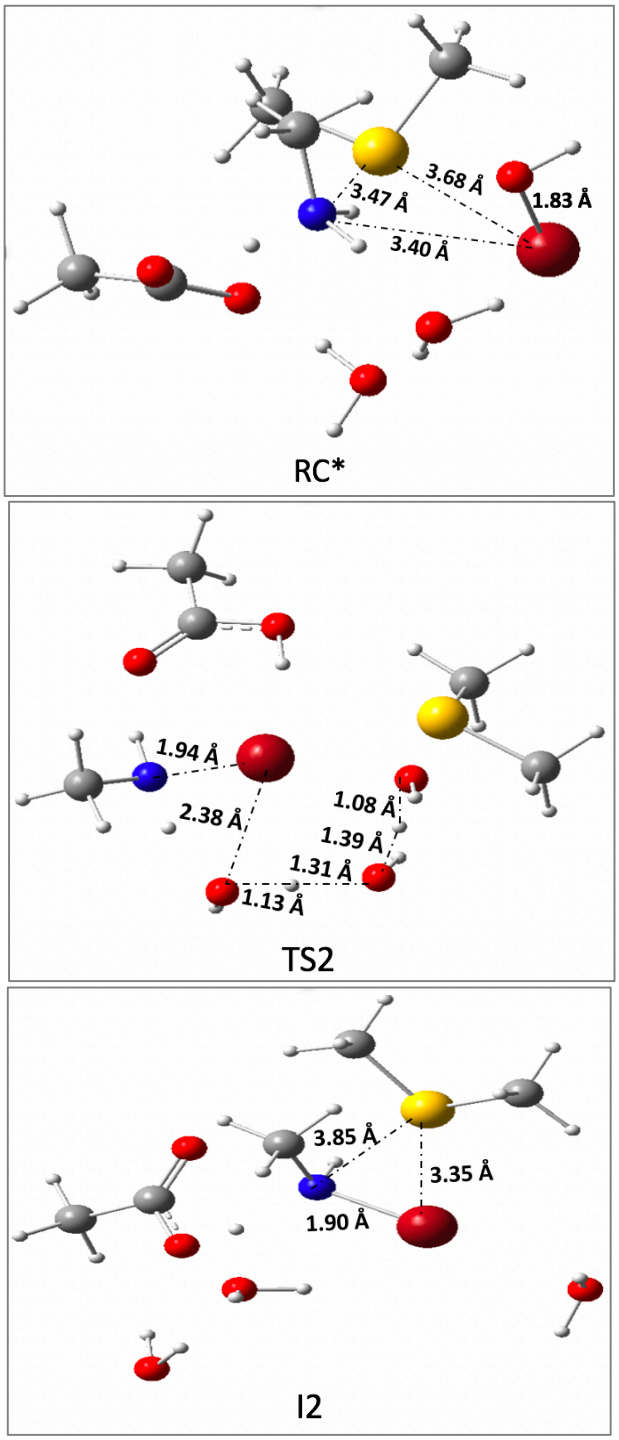


Figure 3.6. Illustration of the optimized initial reactant complex (**RC***), haloamine intermediate (**I2**) and corresponding transition structure (**TS2**) obtained using the

present QM-cluster (see Computational Methods) approach. Selected key distances are shown (Angstrom).

QM/MM results. For this component of the study we first re-examined the relative energies of the halosulfonium (**I1'**) and haloamine (**I2'**) relative to the corresponding QM/MM optimized reactant complex (**RC'**). The energies obtained are shown in Figure 3.7 while the optimized structures of the initial reactant complex (RC') and halosulfonium (I1') and haloamine (I2') intermediates are shown in Figure 3.8.

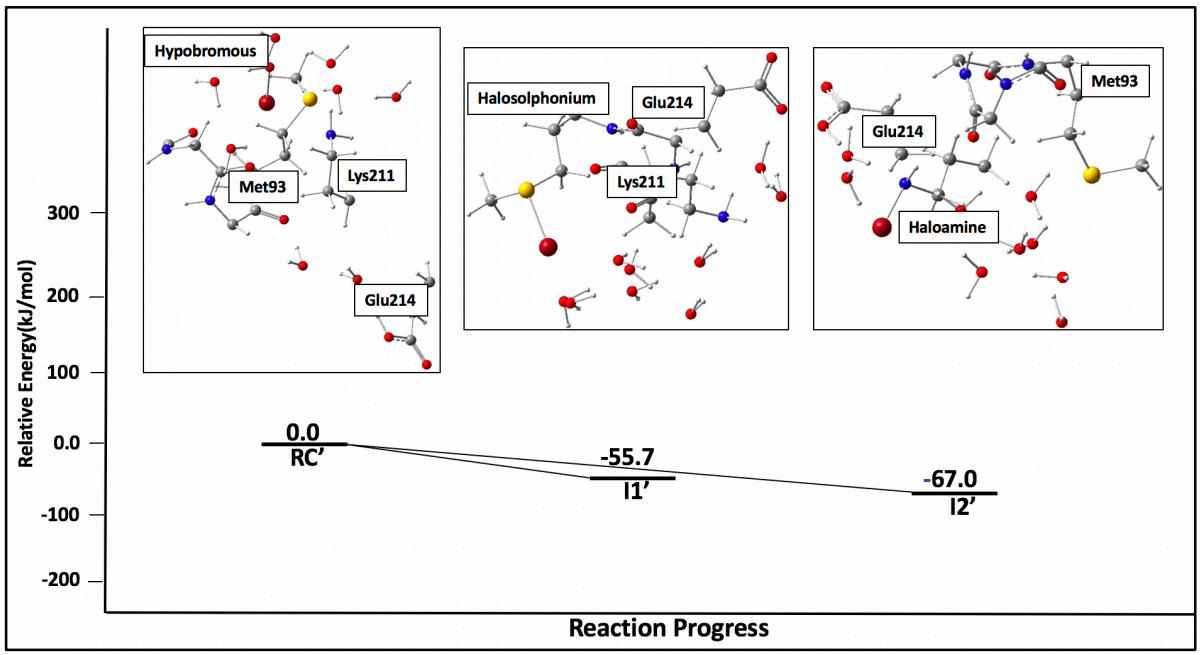


Figure 3.7. The relative energies (kJ/mol) of the initial reactant complex (**RC'**), halosulfonium (**I1'**), and haloamine (**I2'**) obtained at the present QM/MM level of theory (see Computational Methods). For clarity, the MM layer has been omitted.

As can be seen, similar trends in relative energies are observed. Both the halosulfonium and haloamine intermediates lie lower in energy than **RC'**. In contrast, however, they now lie markedly lower in energy by 55.7 and 67.0 kJ/mol with the haloamine again being the most exothermic intermediate by 11.3 kJ/mol.

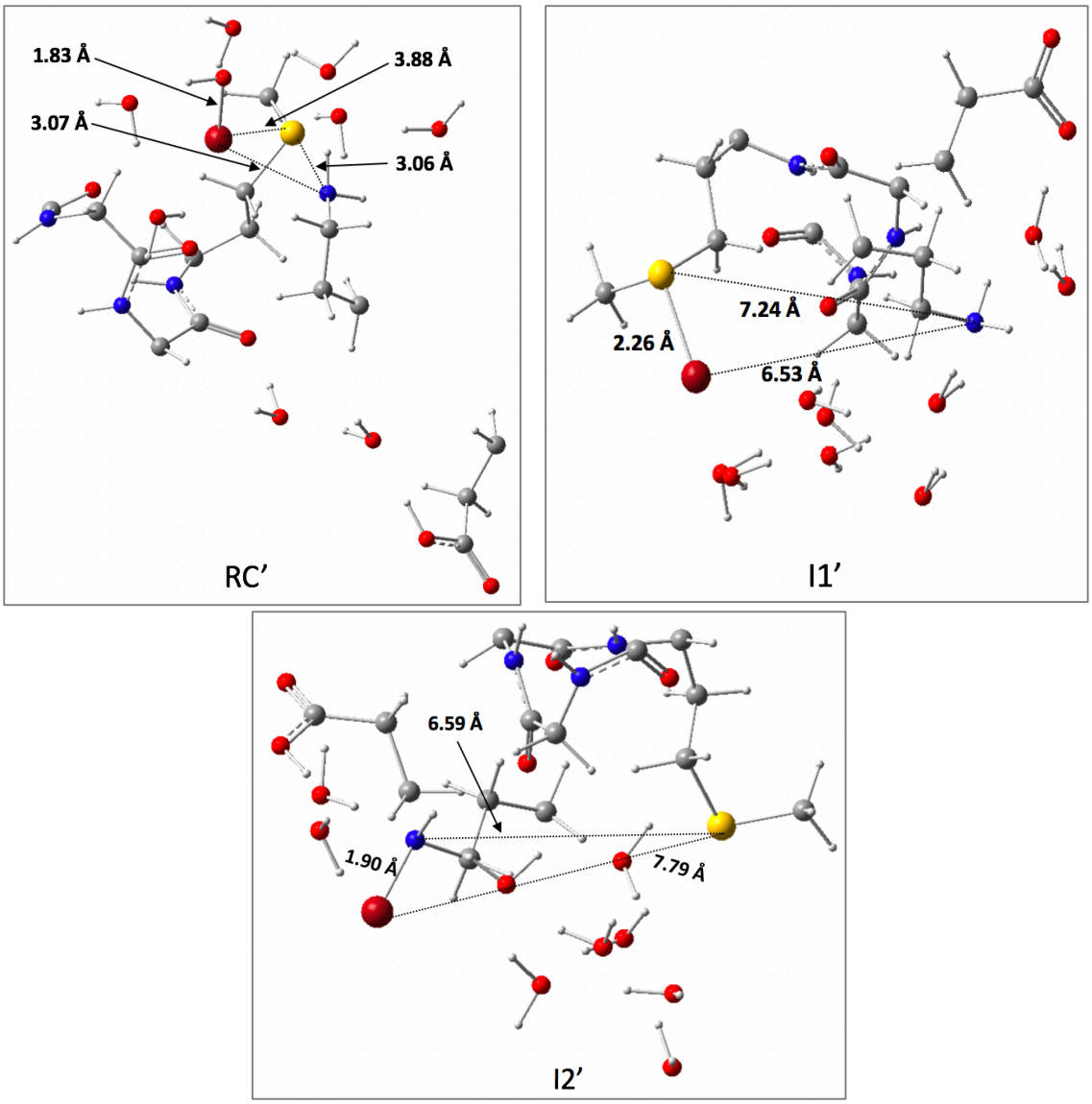


Figure 3.8. Illustration of the QM/MM obtained optimized initial reactant complex (**RC'**), and halosulfonium (**I1'**) and haloamine intermediate (**I2'**) (see Computational Methods) approach. Selected key distances are shown (Angstrom).

In the QM/MM initial reactant complex **RC'**, the bromine of the HOBr moiety is markedly closer to the Lys211 nitrogen than Met93 sulfur with distances of 3.07 and 3.88 Å, respectively (Figure 3.8). Meanwhile, the $_{Met93}S\cdots N_{Lys211}$ distance is shorter than observed in the QM-cluster model at 3.06 Å. In the halosulfonium intermediate (**I1'**), the $_{Met93}S–Br$ distance is only marginally longer by 0.03 Å than observed in the QM-cluster model at 2.26 Å (see Figure 3.3). However, the $_{Lys211}N\cdots S$ and $_{Lys211}N\cdots Br$ distances are now considerably longer at 7.24 and 6.53 Å, respectively. While these distances are large, the marked flexibility in the side chains of Met93 and Lys211 would allow them to come close enough to react as observed experimentally. Similar trends are also observed for the haloamine intermediate (**I2'**). The $_{Lys211}N\cdots Br$ distance is very close to that obtained using the QM-cluster model (1.90 Å). However, the $_{Met93}S\cdots N_{Lys211}$ and $_{Met93}S\cdots Br$ distances are both very long at 6.59 and 7.79 Å, respectively.

Atoms in Molecules (AIM) Analysis of the Densities: The molecular graphs obtained using AIM are shown in Figure 3.9 with bond critical point (BCP) density values given in the Appendices. The lines between atoms indicate bonds and/or paths of interaction while the value of the density at the bond critical gives insights into the strength and type of interaction.

In **RC** one can see that the HOBr moiety is positioned such that its Br component is direct in towards the amine and sulfur of the Lys211 and Met93, respectively. In fact, one can see a bond path between the Br and S with a density at its BCP of just 0.009. Clearly, this indicates that it is a weak interaction as one would expect between two closed-shell centres with in which Br has a slight positive dipole and electron rich sulfur. Indeed one can also observe an interaction

pathway between the sulfur and an oxygen of the carboxylic group of Glu214 with a similar density at its BCP of 0.007. Hence, while the S…Br interaction is weak, it does indicate that the two centres are able to "see each other" which is a key requirement for a direct reaction. Notably, one can see that the single water molecule forms a bridge of interaction pathways between the Lys211 amine and oxygen of the HOBr moiety; this is the proton transfer that is required to occur during addition of Br to either the Met93 sulfur or Lys211 nitrogen. The value of the density at the BCP of the $_{Lys211}NH^{\cdots}OH_2$ interaction, clearly a hydrogen bond interaction, is 0.020. This value also underscores the comparative weakness of the $Br^{\cdots}S_{Met93}$ and $Br^{\cdots}N_{Lys211}$ interactions with their values of 0.009 (for both).

In **RC*** the HOBr is positioned slightly differently to that observed in **RC**. However, we again can see interaction pathways between the Br and sulfur centres with a density at the BCP of 0.011. This is a negligible increase from that observed in **RC**. However, it is interesting to note that no interaction pathway is observed between the Br and $_{Lys211}N$ centres. This suggests that the reaction of these two centres likely require some rearrangement(s) to enable such a possibility. This may also partly explain the higher barrier to formation of **I2** compared to **I1**. An interaction network of hydrogen bonds can be seen between the HOBr oxygen and Lys211 amine as well as the carboxylic acid group Glu214; both potential proton donors in haloamine formation.

In the halosulfonium **I1** species the newly formed S-Br bond has a density at its BCP of 0.119. It is noted that while this is markedly larger than observed in **RC**, as one would expect, it is still less than that observed for other covalent bonds within the complex. For example, the two S–C bonds have densities at their bond critical points of 0.183 (for both). This suggests that the S-Br bond is weaker than

that of such covalent bonds, possibly possessing some ionic character. Meanwhile in the haloamine intermediate **I2**, the newly formed N–Br bond has a density at its BCP of 0.159. This is more density than in the S–Br bond suggesting that the N–Br bond is likely stronger. This, collectively this may also help explain the observation that within the present computational models the haloamine intermediate is thermodynamically preferred, but kinetically disfavored, compared to the halosulfonium.

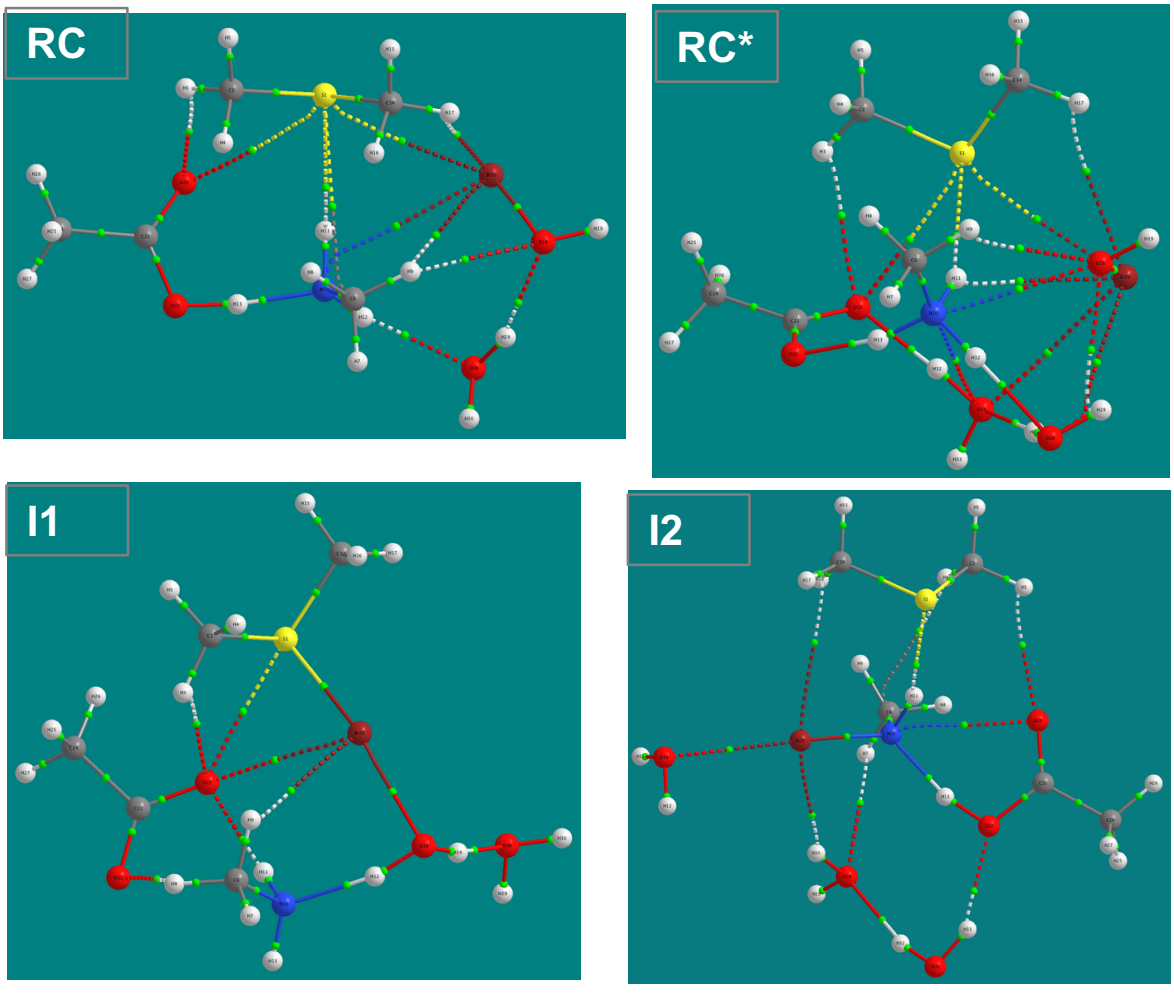


Figure 3.9. Molecular graphs obtained by an AIM analyses of the two reactant complexes **RC** and **RC***, the halosulfonium (**I1**) and haloamine (**I2**) intermediates.

3.4 Conclusions:

A multi-scale computational approach, complementarily applying molecular dynamics (MD) simulations, QM-cluster, and QM/MM methods, has been used to investigate the first step in formation of the Collagen IV sulfinimine cross-link and to gain insights into the nature and energetics of the key putative halosulfonium and haloamine intermediates.

Significantly, using both QM-cluster and QM/MM-based approaches, the haloamine intermediate is predicted to lie markedly lower in energy relative to the initial reactant complex than the halosulfonium intermediate. However, at least within the QM-cluster approach, the barrier to formation of the halosulfonium intermediate is significantly lower at just 14.8 kJ/mol. In contrast, the barrier to formation of the haloamine is 39.0 kJ/mol. Hence, it is predicted to be magnitudes of order slower than the former pathway.

Thus, these results suggest that sulfinimine bond formation may be kinetically controlled to proceed via a halosulfonium rather than the thermodynamically preferred haloamine intermediate.

Further studies, possibly using electronic embedding within the ONIOM QM/MM formalism, should shed further light on these two key intermediates as well as the final stage of the overall reaction, formation of the sulfinimine bond, and formation of the side-product sulfoxide.

3.5 References:

1. Khoshnoodi, J.; Pedchenko, V.; Hudson, B. G., Mammalian Collagen IV. *Microsc. Res. Tech.* **2008,** *71*, 357-370.

2. Kanazawa, Y.; Ikegami, K.; Sujino, M.; Koinuma, S.; Nagano, M.; Oi, Y.; Onishi, T.; Sugiyo, S.; Takeda, I.; Kaji, H.; Shigeyoshi, Y., Effects of Aging on Basement Membrane of the Soleus Muscle During Recovery Following Disuse Atrophy in Rats. *Exp. Gerontol.* **2017,** *98*, 153-161.

3. Lebleu, V.; Macdonald, B.; Kalluri, R., *Structure and Function of Basement Membranes*. 2007; Vol. 232, p 1121-9.

4. Brown, K. L.; Cummings, C. F.; Vanacore, R. M.; Hudson, B. G., Building Collagen IV Smart Scaffolds on the Outside of Cells. *Protein Sci.* **2017,** *26*, 2151-2161.

5. Urbano, J. M.; Torgler, C. N.; Molnar, C.; Tepass, U.; Lopez-Varea, A.; Brown, N. H.; de Celis, J. F.; Martin-Bermudo, M. D., Drosophila Laminins Act as Key Regulators of Basement Membrane Assembly and Morphogenesis. *Dev.* **2009,** *136*, 4165-4176.

6. Naduk-Kik, J.; Hrabec, E., The Role of Matrix Metalloproteinases in the Pathogenesis of Diabetes Mellitus and Progression of Diabetes Retinopathy. *Postep. Hig. Med. Dosw.* **2008,** *62*, 442-450.

7. Matsubayashi, Y.; Louani, A.; Dragu, A.; Sanchez-Sanchez, B. J.; Serna-Morales, E.; Yolland, L.; Gyoergy, A.; Vizcay, G.; Fleck, R. A.; Heddleston, J. M.; Chew, T. L.; Siekhaus, D. E.; Stramer, B. M., A Moving Source of Matrix Components Is Essential for De Novo Basement Membrane Formation. *Curr. Biol.* **2017,** *27*, 3526-+.

8. McCall, A. S. The Role of Peroxidasin in Basement Membrane Physiology and Human Disease. Vanderbilt Univ. 2015.

9. Ero-Tolliver, I. A.; Hudson, B. G.; Bhave, G., The Ancient Immunoglobulin Domains of Peroxidasin Are Required to Form Sulfilimine Cross-links in Collagen IV. *J. Biol. Chem.* **2015,** *290*, 21741-21748.

10. Tanjore, H.; Kalluri, R., The Role of Type IV Collagen and Basement Membranes in Cancer Progression and Metastasis. *Am. J. Pathol.* **2006,** *168*, 715-717.

11. Vanacore, R.; Pedchenko, V.; Bhave, G.; Hudson, B. G., Sulphilimine Cross-Links in Goodpasture's Disease. *Clin. Exp. Immunol.* **2011,** *164*, 4-6.

12. Pedchenko, V.; Bondar, O.; Fogo, A. B.; Vanacore, R.; Voziyan, P.; Kitching, A. R.; Wieslander, J.; Kashtan, C.; Borza, D. B.; Neilson, E. G.; Wilson, C. B.; Hudson, B. G., Molecular Architecture of the Goodpasture Autoantigen in Anti-GBM Nephritis. *N. Engl. J. Med.* **2010,** *363*, 343-354.

13. Omachi, K.; Miyakita, R.; Fukuda, R.; Kai, Y.; Suico, M. A.; Yokota, T.; Kamura, M.; Shuto, T.; Kai, H., Long-Term Treatment with EGFR Inhibitor Erlotinib Attenuates Renal Inflammatory Cytokines but not Nephropathy in Alport Syndrome Mouse Model. *Clin. Exp. Nephrol.* **2017,** *21*, 952-960.

14. Yokota, T.; Omachi, K.; Suico, M. A.; Kojima, H.; Kamura, M.; Teramoto, K.; Kaseda, S.; Kuwazuru, J.; Shuto, T.; Kai, H., Bromide Supplementation Exacerbated the Renal Dysfunction, Injury and Fibrosis in a Mouse Model of Alport Syndrome. *PLoS One.* **2017,** *12*, 16.

15. Cummings, C. F.; Hudson, B. G., Lens Capsule as a Model to Study Type IV Collagen. *Connective tissue. res.* **2014,** *55*, 8-12.

16. Ortega, N.; Werb, Z., New Functional Roles for Non-Collagenous Domains of Basement Membrane Collagens. *J. Cell. Sci.* **2002,** *115*, 4201-4214.

17. Bhave, G.; Cummings, C. F.; Vanacore, R. M.; Kumagai-Cresse, C.; Ero-Tolliver, I. A.; Rafi, M.; Kang, J. S.; Pedchenko, V.; Fessler, L. I.; Fessler, J. H.; Hudson, B. G., Peroxidasin Forms Sulfilimine Chemical Bonds Using Hypohalous Acids in Tissue Genesis. *Nat. Chem. Biol.* **2012,** *8*, 784-790.

18. McCall, A. S.; Cummings, C. F.; Bhave, G.; Vanacore, R.; Page-McCaw, A.; Hudson, B. G., Bromine Is an Essential Trace Element for Assembly of Collagen IV Scaffolds in Tissue Development and Architecture. *Cell.* **2014,** *157*, 1380-1392.

19. Ronsein, G.; Winterbourn, C.; Di Mascio, P.; Kettle, A., Cross-Linking Methionine and Amine Residues with Reactive Halogen Species. *Free Radical Biol. Med.* **2014,** *76*, S158-S158.

20. Pichierri, F., Theoretical Characterization of the Sulfilimine Bond: Double or Single?. *Chem. Phys. Lett.* **2010,** *487*, 315-319.

21. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984,** *106*, 765-784.

22. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996,** *118*, 2309-2309.

23. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.;

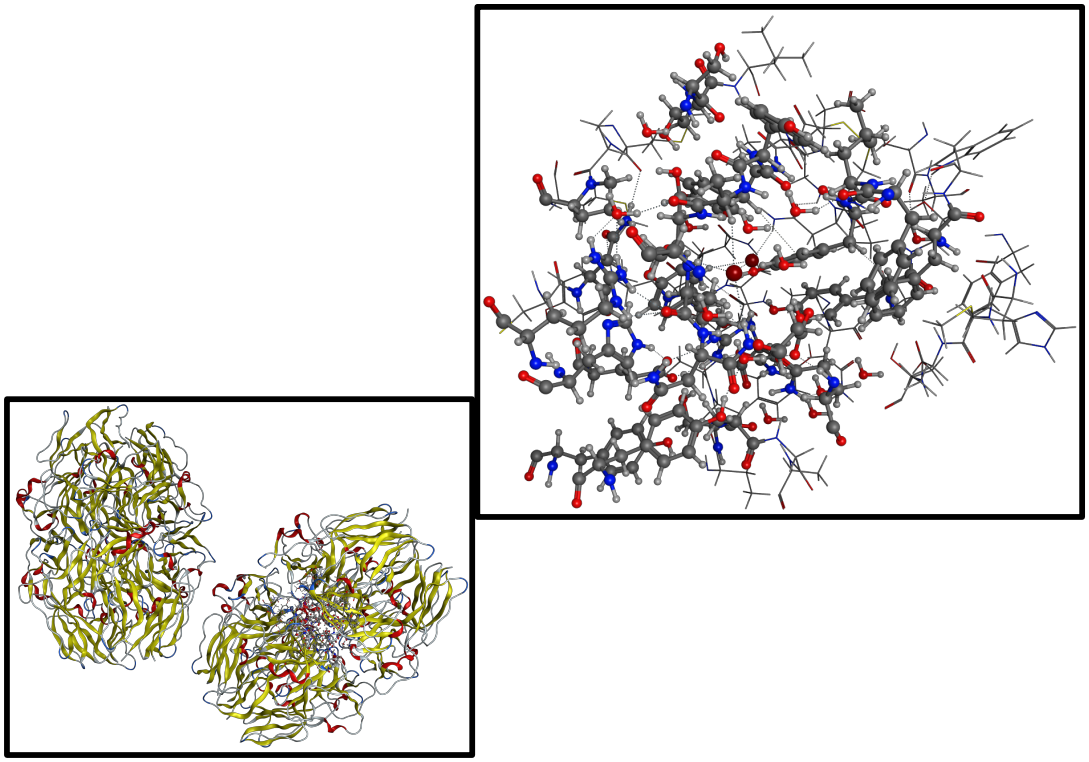
Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz de Montellano, P. R.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision E.01; Gaussian Inc.: Pittsburgh, PA, 2013.

24. Walker, M.; Harvey, A. J. A.; Sen, A.; Dessent, C. E. H., Performance of M06, M06-2X, and M06-HF Density Functionals for Conformationally Flexible Anionic Clusters: M06 Functionals Perform Better than B3LYP for a Model System with Dispersion and Ionic Hydrogen-Bonding Interactions. *J. Phys. Chem. A.* **2013,** *117*, 12590-12600.

25. Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D., Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theor. Comp.* **2008,** *4*, 1996-2000.

CHAPTER 4:

Computational Insights into The Active Site and Catalytic Mechanism of Archaeal and Yeast Thiazole Synthetases



4.1 Introduction:

Thiamine, more commonly referred to as vitamin B1, is an essential cofactor found in many cells and organisms.[1-2] As shown in Figure 4.1 it consists of a thiazole-derivative and pyrimidine-derivative cross-linked via a methylene bridge, and as a result is a water-soluble biomolecule. Physiologically, it plays critical roles in a variety of important biochemical reactions including amino acid branching and carbohydrate metabolism.[1, 3] Indeed, its central roles in many such processes is underscored by the fact that thiamine deficiency is fatal if left untreated.[4]

Although first discovered in 1911 by Casimir Funk, it was first isolated in 1926 from yeast.[5-8] Since then, a number of phosphorylated thiamine derivatives have been identified in organisms. More specifically, mono-, di- and tri-phosphorylated forms have been observed, as well as some further adenosine derivatives. However, the diphosphorylated derivative is generally held to be the most active form of thiamine, the physiological roles of the others appearing to be less common and not well understood.[9-10]

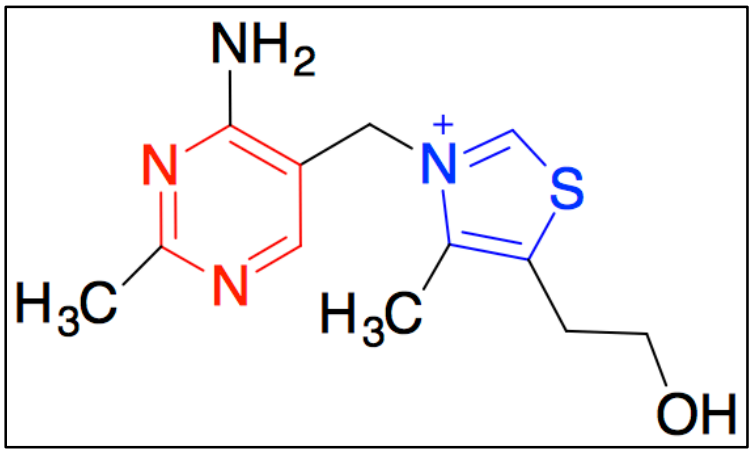


Figure 4.1. Illustration of the core structure of Thiamine with the thiazole ring highlighted in blue and pyrimidine ring highlighted in red.

Despite its critical roles, biosynthesis of thiamine is only found to occur in prokaryotes and some eukaryotes such as fungi and plants.[2] Humans, for example, are unable to synthesize thiamine and therefore must obtain it from external sources through their diet.[10-11] A common strategy is employed by organisms in which thiamine biosynthesis occurs. Specifically, they independently synthesize a pyrimidine precursor and a thiazole precursor that are then coupled together to form the thiamine cofactor (Figure 4.1).[12-14]

However, on the other hand, the biosynthetic pathway for the thiazole precursor varies between bacteria (prokaryotes) and yeast (eukaryotes).[15] Bacteria require five different enzymes and several substrates; 1-deoxy-D-xylulose-5-phosphate, cysteine, and glycine or tyrosine.[16-17] In contrast, yeast, require just one enzyme: thiazole synthase (THI4). The latter requires its glycine substrate, a nicotinamide adenine dinucleotide (NAD) cofactor, and a sulfur donor. In the specific case of yeast the latter is a conserved Cys205 residue.[16]

In some species, Thi4 is believed to be a suicide or single-turnover enzyme. More specifically, in these species, such as yeast, the enzyme uses a cysteinyl residue as the required sulfur source in thiazole formation.[18] For example, in Thi4 from the yeast Saccharomyces cerevisiae (ScThi4) the sulfur obtained from Cys205 is used to form adenylated thiazole (ADT) from glycine (Figure 4.2).[17] As a result, the enzyme can be thought of as a cosubstrate and after a single-turnover the enzyme is inactivated. This also underscores the paramount importance of thiamine in biochemical systems in that some organisms are willing to expend considerable energy and resources to create a molecular machine that works just once.

ScThi4 shares some active site similarities with the iron-dependent Methanococcus jannaschii Thi4 (MjThi4). Significantly, however, based on structural analysis on both enzymes, MjThi4 lacks the presence of a cysteinyl (Cys205) residue. In its place it instead has a histidyl residue.[16] As a result MjThi4 lacks an internal sulfur source and instead requires an exogenous sulphide source.[13, 16] This also enables MjThi4 to perform multiple turnovers; it is not a one-time catalyst.

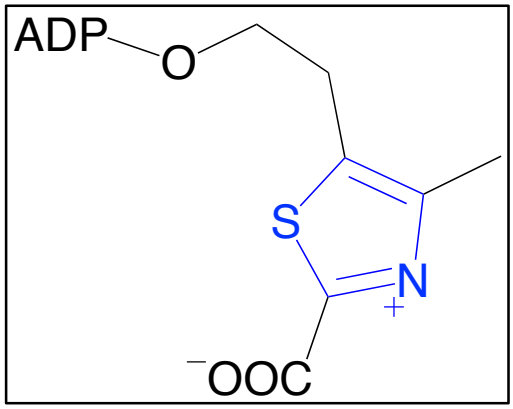


Figure 4.2. Illustration of the adenylate thiazole (ADT) formed by thiazole synthases from a glycine, sugar, and cysteine or other sulfur source.

Based on numerous experimental studies including X-ray crystallographic structures and comparison with the prokaryotic biosynthetic pathway, a mechanism for ADT biosynthesis in eukaryotes has recently been proposed and is shown in Figure 4.3.[16]

In particular, it is proposed that the mechanism proceed via two half-reactions. The first half-reaction begins with thiolate ($^{-}$SH), derived form an exogenous source (a cofactor) entering the Fe(II)-containing active site. In particular, it ligates to the Fe(II) centre and breaks a substrate carbonyl···Fe(II)

($C_{carb}O_{carb}\cdots Fe(II)$) ligation. The substrates $C_{carb}=O_{carb}$ group twists and repositions itself, but remains near the reactive region. Indeed, in the next step the sulfur nucleophilically attacks at the substrates C_{carb} centre while its proton transfers onto the substrates carbonyl oxygen (O_{carb}) to form a hydroxyl. In the final step of the first-half-reaction the newly formed $-O_{carb}H$ hydroxyl group is protonated and lost as H_2O, while the newly formed C_{carb}–S bond becomes a thioketone (i.e., forms a $C_{carb}=S$ group). Notably, no suitable acid group appears to be positioned near the leaving hydroxyl.

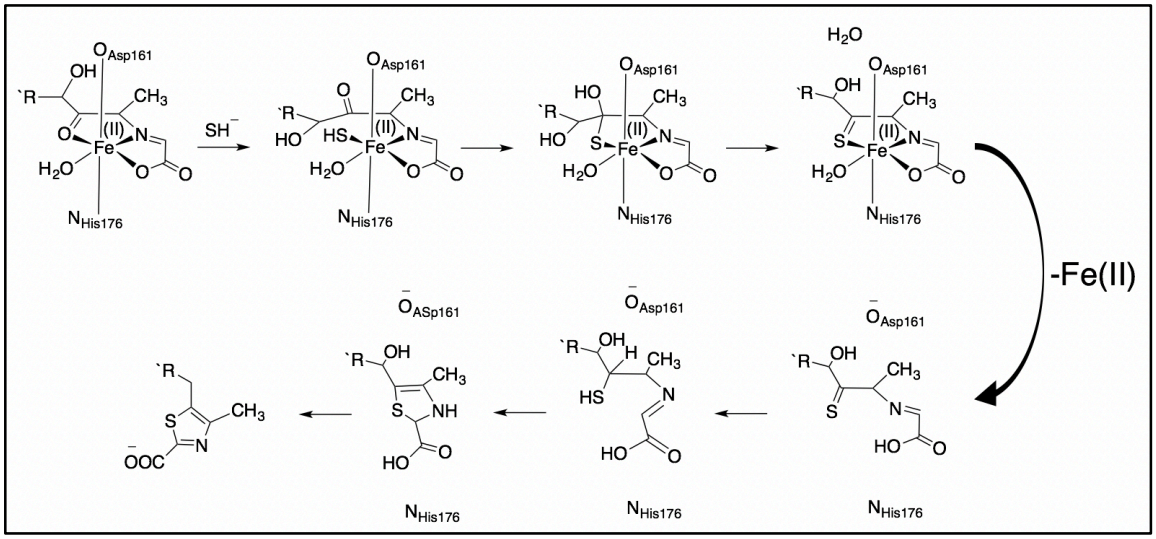


Figure 4.3. Schematic illustration of the proposed mechanism of thiazole formation.

The second half-reaction begins upon loss of the Fe(II) centre and apparently neutralization of both the active site histidyl (His176) and substrate carboxylate. In the first half-reaction both groups were implied to be deprotonated (anionic). The loss of the Fe(II) centre allows the thioketone intermediate to

undergo a conformational change by rotation around its C–N bond. The thioeketone group is then converted to a thiol, presumably by protonation, while the substrates nearby methylated carbon centre apparently obtains a hydride. The thiol sulfur, now presumably being a better nucleophile, is able to attack at the now spatially nearby N=CH– carbon to form a cyclic intermediate. The latter then loses two hydrogens to form adenylate thiazole. Notably, with the same article the authors proposed a different mechanism for related thiazole synthases in which the sulfide instead binds to the substrates N=CH– carbon centre to form a thiol derivative. The sulfur centre then nucleophilically attacks at the substrates C_{carb}=O_{carb} carbon centre to eventually form the ADT product. Obviously, many key questions and features of the active site and mechanism remain unclear or unknown.

In this chapter, we have used a multi-scale computational approach to gain insights into several critical features of the active site. More specifically, molecular dynamics (MD) simulations, QM-cluster and QM/MM methodologies have been used to examine the water distribution in the active site, the potential identity of mechanistically required acid and base groups, as well as the overall mechanism (both the Fe(II)-facilitated first half-reaction and the non-metallo second half-reaction).

4.2 Computational Methods:

Molecular Dynamics (MD) Simulations: The Molecular Operating Environment (MOE) program was used to prepare all MD simulation calculations.[19] For the initial starting structures a suitable X-ray crystal structure of an octamer of the thiazole synthase from Methanococcus igeus (PDB ID: 4Y4N) with a resolution of 2.1 Å and

both Fe(II) and organic non-sulfur substrate, 2-[(E)-[(4R)-5-[[[(2R,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-bis(oxidanyl)oxolan-2-yl]methoxy-oxidanyl-phosphoryl]oxy-oxidanyl-phosphoryl]oxy-4-oxidanyl-3-oxidanylidene-pentan-2-ylidene]amino]-ethanoic acid, bound within the active site.[16] It is noted that the structure of the substrate as determined by the PDB differed from its illustration in the original literature; specifically, the placement of the imine double bond. Missing hydrogens were added using the default protocol in MOE as were the protonation states of ionizable residues. The enzyme was then solvated with a 6 Å layer of water and minimized using the AMBER12 force field.[20] The default settings of MOE were used in the MD simulation which was itself performed using the NAMD program. The solvated complex was then submitted for a 1 ns equilibration MD simulation. It should be noted as part of these MD studies the protonation state of the active site His176 bound to Fe(II) was also investigated. Furthermore, for the second half-reaction (see Figure 4.3) the Fe(II) was omitted from the active site and a 1 ns MD simulation was performed as above. For the Fe(II)-free intermediate-bound active site analyses were performed on the distribution of the aqueous solvent molecules to gain insights into the impact on solvent distribution and networking after loss of the Fe(II). This is also a suitable method to investigate the surrounding environment for proton transfer during the reaction.

QM-Cluster Investigations: For this component of the study all calculations were performed using the Gaussian 09 program.[21] The initial chemical models for the QM-cluster calculations were derived from a representative structure of the most populated cluster obtained after cluster analysis of the MD simulations. For the initial reactant complex (**RC**) of the first half-reaction two chemical models were

examined. In one the Fe(II)-ligated histidyl was modelled as anionic, as shown in the literature,[12] while in the other it was modelled as neutral. The remaining intermediates were then modeled using a neutral histidine. Optimized geometries and harmonic vibrational frequencies were obtained using both the B3LYP and B3LYP* (HF exchange reduced to 15% due to the presence of the Fe(II) centre) methods in combination with the 6-31G(d) basis set on all atoms except Fe(II) for which the LANL2DZ effective core potential basis set was used.[17,18] Relative energies were obtained by performing single point energy calculations, on the above optimized structures, at the B3LYP/6-311+G(2df,p) level of theory.

For the second half-reaction two models were again considered which differed in the protonation state of the active site histidyl (His176). In this case it was either neutral or protonated. The possibility of a proton transfer from His176 onto sulfur going from **I5** to **I6** was examined as well as the potential for proton transfer from sulfur to the intermediates' nitrogen, bridging through Asp161 in **I7**. For the second half-reaction, all optimized structures and harmonic vibrational frequencies were obtained at the B3LYP/6-31G(d) level of theory. In addition, the surrounding protein environment was included via use of the IEFPCM solvation method formalism with a dielectric constant of 4.0 to mimic the protein's internal polarity.[22]

QM/MM investigations: All QM/MM calculations were performed within the ONIMO formalism as implemented in the Gaussian 09 program.[21] Similarly, the initial chemical models for the QM/MM calculations were again derived from an appropriate representative structure selected after cluster analysis of the MD simulations.

For each half-reaction, a different chemical model was used. For the first half-reaction, the chemical model contained all residues up to 31 Å away from the Fe(II) center while for the second half-reaction all residues up to 40 Å from the center of the active site (the approximate position of the now lost Fe(II) ion) were included. Both models allowed enough flexibility for the QM-region residues and groups for any position changes they may require, expected or unexpected, while ensuring the overall structural integrity of the chemical models.

For the first half-reaction, the QM layer consisted of Fe(II), the initial proposed imine substrate, the residues believed to be integral to substrate binding and possibly the mechanism (Arg236, His176, Asp161, and Glu198) and one water molecule. This region was described using B3LYP method with the same basis set as for the QM-cluster calculations due again in part to their documented accuracy and low computational cost. The remainder of the protein was described using the AMBER96 force field as implemented in Gaussian 09. For the second half-reaction, several models were used that were the same as above but without the Fe(II) centre. They differed from each, however, in the protonation state of the active site His176; it was either protonated or neutral while in a third model the Glu198 was placed into the MM layer in order to help better understand its role.

4.3 Results and Discussion:

4.3. The First Half-Reaction.

MD results. An analysis was performed on the MD simulation. For example, a plot of the calculated root mean square deviation (RMSD) versus time (ps) for the selected active site residues is shown in Figure 4.4. From this data, a

representative conformation was selected based on clustering analysis to find groups of similar structure and then pick the one nearest to the center.

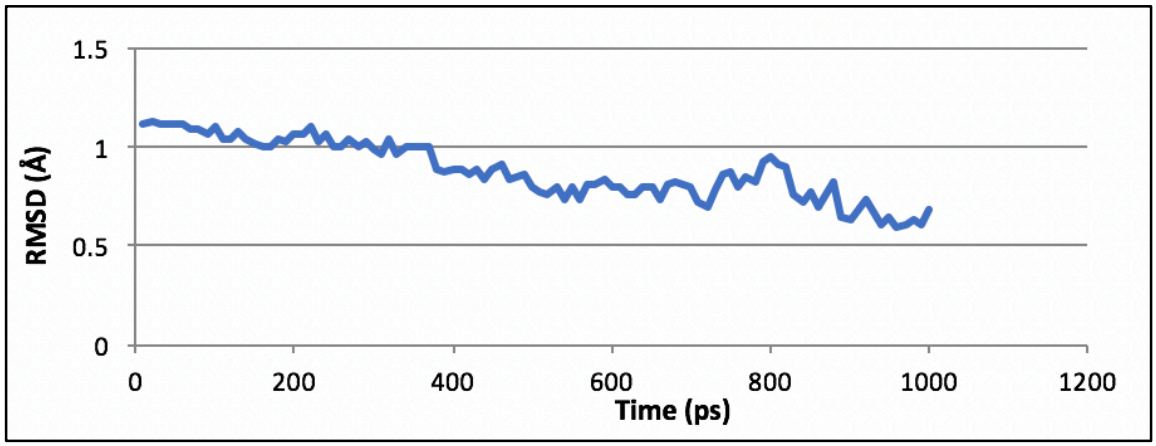


Figure 4.4. Calculated RMSD for MjThi4 active site with neutral His176.

Compared to the X-ray structure, after simulation, the distances of atoms bound to Fe(II) increased and remained between 2.0 Å and 2.5 Å long (Figure 4.6.A). In the X-ray structure obtained, and as it was proposed experimentally, distances ranged from between 2.0 Å to 2.2 Å (Figure 4.5.B).

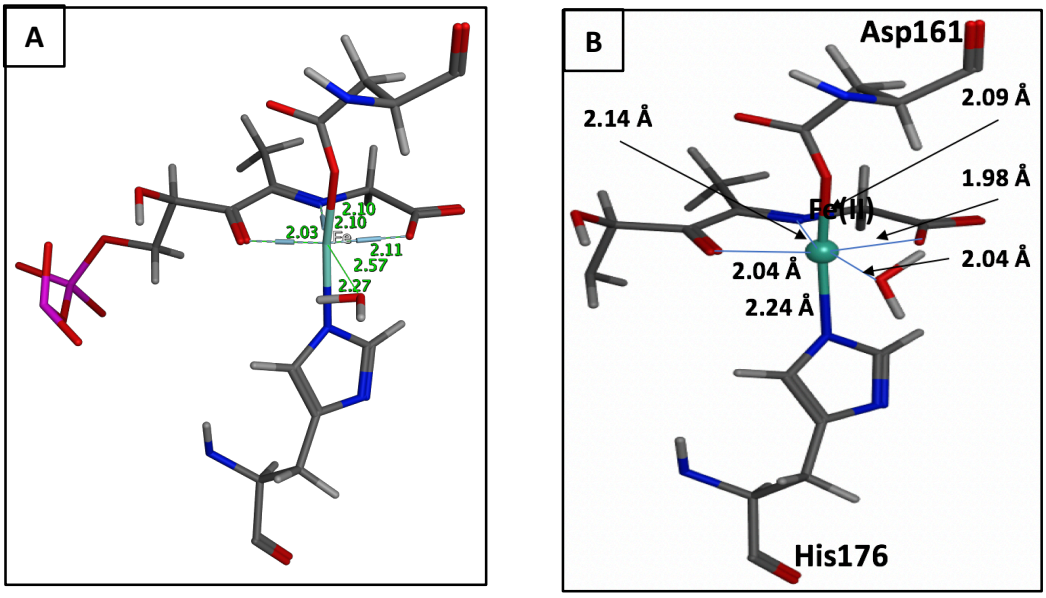


Figure 4.5. Schematic structures of the MjThi4 active site including Fe(II), glycine, His176, Asp161, Arg236, Glu198 and a water molecule **A)** After 1 ns MD simulation **B)** experimental X-ray structure.

QM-cluster results. A proposed mechanism for thiazole formation was examined computationally using the QM-cluster only approach. all intermediates were optimized based on the proposed mechanism with some modifications we obtained, beginning with the reactant complex **(RC)** as Fe(II) is bonded to six atoms, oxygen and nitrogen of substrate, oxygen of Glycine carboxylate, nitrogen of His176 imidazole ring, oxygen of Asp161 carboxylate, and a water molecule. After the addition of HS^- in the active site, the carbonyl oxygen of the substrate was rotated to be displaced by sulfide. A proton is then transferred to the carbonyl oxygen from the free sulfide, producing a water molecule. In Figure 4.6A and Figure 4.6B, full optimized structures of **RC**, **I1**, **I2** and **I3** are compared using B3LYP and B3LYP*.

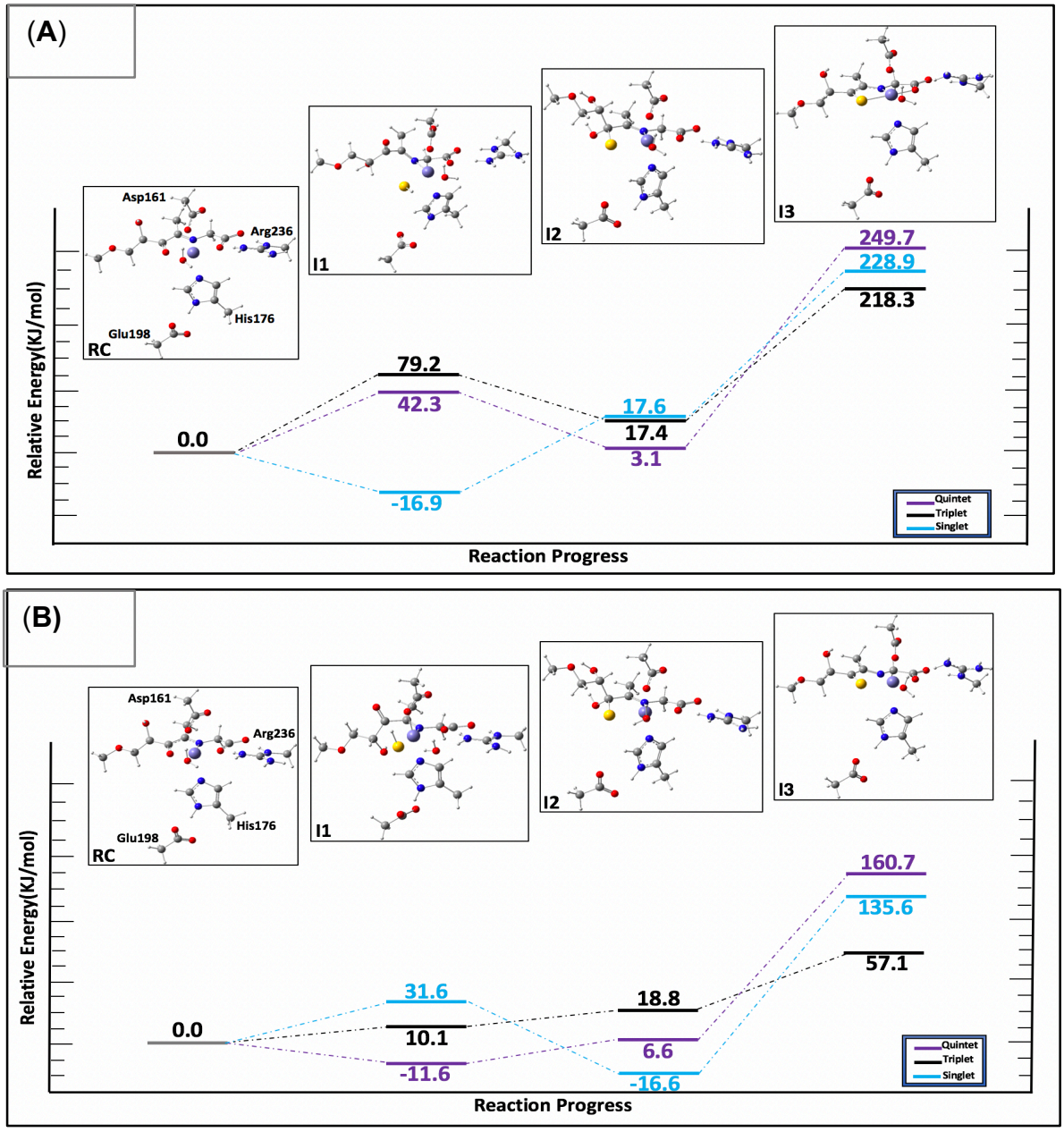


Figure 4.6. Calculated relative energy surface obtained using the methods (**A**) B3LYP and (**B**) B3LYP*, with the 6-31G(d) basis set for both. Three different spin states: quintet, triplet, and singlet states were analyzed as represented by purple, black, and blue respectively.

For **I1** using B3LYP, singlet showed the lowest energy with -16.9 kJ/mol. In contrast, the singlet state using B3LYP* was the lowest in energy at -16.6 kJ/mol for **I2**. On the other hand, **I3** was the highest energy intermediate in both methods. With B3LYP, the quintet state was the highest and followed by singlet and then triplet with 249.7 kJ/mol, 228.9 kJ/mol and 218.3 kJ/mol respectively. When B3LYP* was used, the energy of **I3** was lower by approximately 90 kJ/mol in the quintet and the singlet states, and by about 160 kJ/mol for the triplet.

Furthermore, the potential energy surface was obtained using single point energy calculations for the optimized structures using B3LYP (Figure 4.7) as outlined in the methods section.

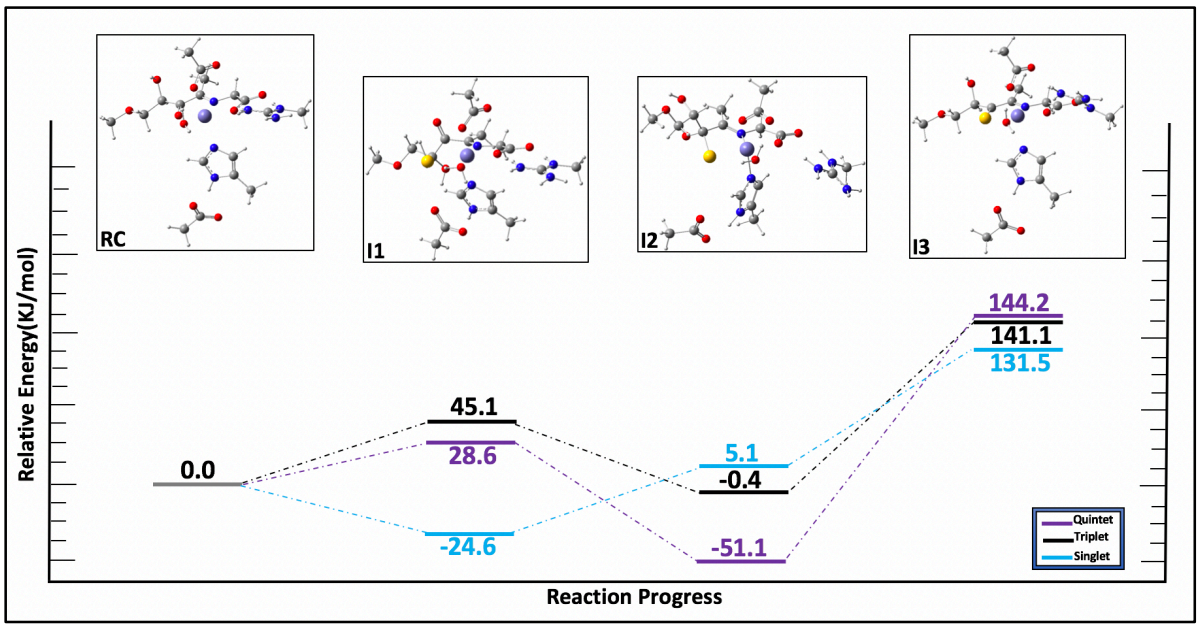


Figure 4.7. The potential energy surface of the sulfur attachment as calculated at B3LYP/6-311+G(2df,p). The various spin states of Fe(II) are shown in purple for the quintet, black for the triplet, and blue for the singlet.

As shown in Figure 4.7, compared to B3LYP/6-31G(d) and B3LYP*/6-31G(d), the overall energy is lower for all intermediates, except in some spin states. For **I3** the triplet state showed an increase in energy by 84 kJ/mol compared to B3LYP* and decreased by 77.2 kJ/mol with respect to B3LYP. Meanwhile, the singlet and quintet states were lower in energy. In the case of **I2**, all spin states appeared to be lower in energy compared to the optimization methods, with exception of the singlet state as it was higher in energy compared to B3LYP* by 21.7 kJ/mol. In **I1**, the energy of the triplet and quintet states were higher than B3LYP* and lower than B3LYP; however, the singlet was lower in energy in both B3LYP* and B3LYP.

In addition, we considered the effect of the active site's surrounding protein environment as outlined previously in the methods. For these calculations, the triplet and quintet states were considered.

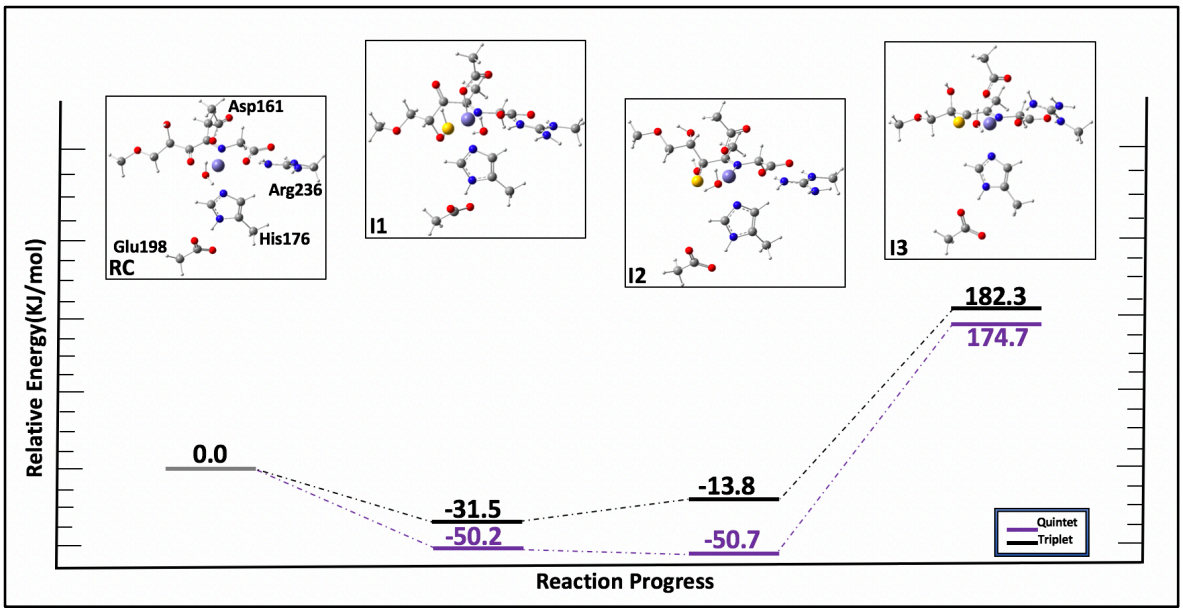


Figure 4.8. Schematic illustration of the potential energy surface obtained at IEFPCM-B3LYP/6-31G(d) level of theory. The triplet state is shown in black while the quintet state is in purple.

As we can see in Figure 4.8, compared to calculations conducted without considering the effect of protein environment, it is evident that there is an overall increase in energy compared to B3LYP* where it is decreased compared to B3LYP.

In the hybrid QM/MM approach, the two layers were segregated as outlined in the Methods section. Notably, this model was obtained from the X-ray structure directly with some modifications done on the active site such as addition of the free sulfide and missing hydrogens.

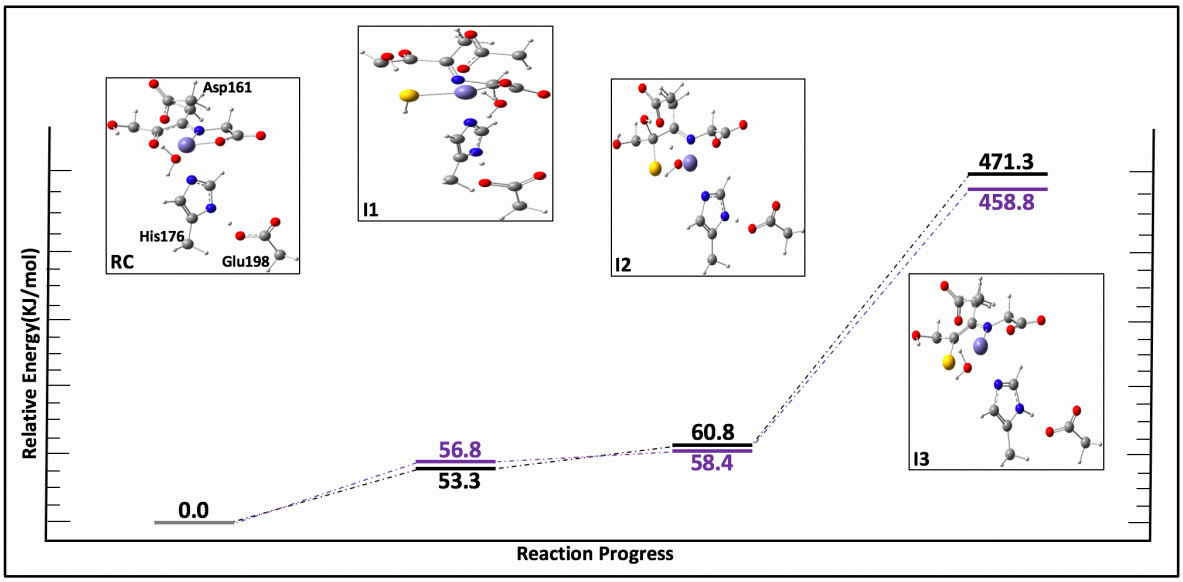


Figure 4.9. Schematic illustration of the potential energy surface obtained for the first half of the reaction calculated at the ONIOM(B3LYP/631G(d): AMBER96) level of theory. The triplet state is shown in black while the quintet is shown in purple.

As we can see that quintet state was the lower in energy in **I2** and **I3,** but not in **I1** where the triplet was slightly higher by 3.5 kJ/mol. Overall, **I1** was the lowest

in energy whereas **I3** was the highest, as opposed to the results obtained from QM-cluster calculations where **I2** was the lowest energy intermediate.

4.3B Second Half-Reaction.

MD results. The proposed mechanism for thiazole formation suggests a protonated His176.[16] For that reason, we considered doing MD for both cases; one with neutral His176 and one with protonated His176. For the simulations, we manually deleted Fe(II) from the active site and re-solvated the model. In the protonated and neutral His, a representative conformation was chosen after 1 ns molecular dynamic simulation based on RMSD plots on selected residues (Figure 4.10).

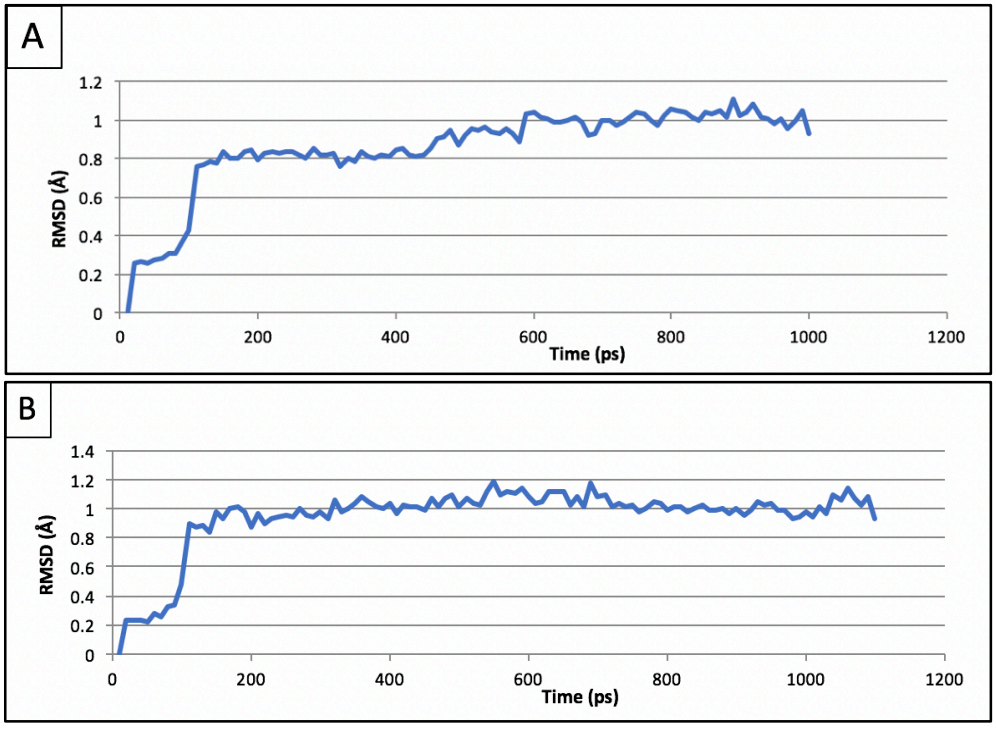


Figure 4.10. RMSD of the active site with **A**) protonated His176 **B**) neutral His176.

As noted in the case of protonated His176, the RMSD stabilized at about 1 Å. In the neutral case, the RMSD was at 1.1 Å after around 500 ps to 700 ps and started to decrease to 1 Å, rising near the end to 1.2 Å

The two models were analyzed and the effects of the protonation state of His176 over the 1 ns MD simulation were observed using ten representative structures (Figure 4.11).

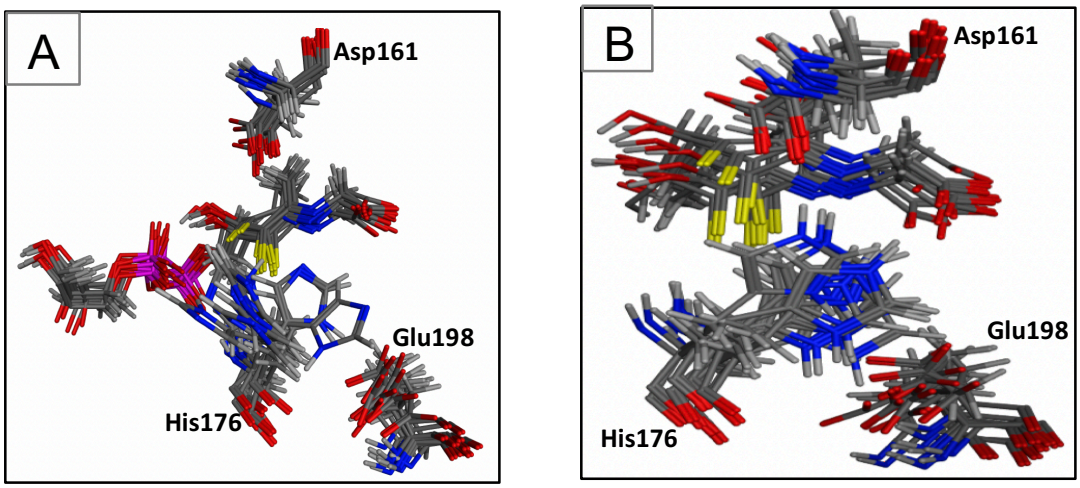


Figure 4.11. Overlay of ten representative conformations of the model with **A**) neutral His176 and **B**) protonated His176.

As we can see, a fluctuation occurs on the neutral His176 configuration and Glu198 appeared to move away from the active site in addition to Asp161. Meanwhile when His176 is protonated, positionally it is more stable over the course of the simulation due to its ability to form a hydrogen bond to the adjacent side-chain carboxylate of Glu198. This hydrogen bond, which more consistently is formed in the MD simulation involving protonated His176, conversely also helps retains Glu198 close to the active site compared to when His176 is neutral.

We then performed a solvent analysis on the active site of each representative conformation in both models, neutral and protonated His176 (Figure 4.12).

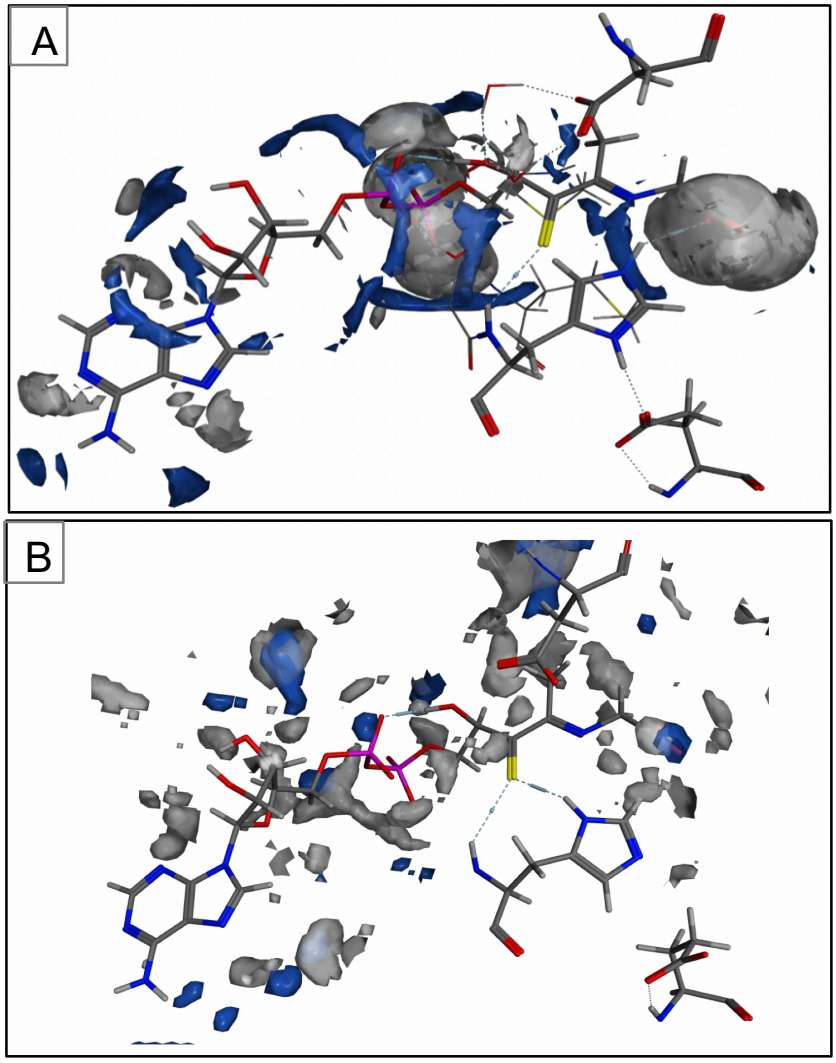


Figure 4.12. Water analysis of the active site with (**A**) protonated His176 (**B**) neutral His176 as gray indicating hydrogens and blue indicating oxygens.

In the case of protonated His176, we observe a richer presence of water protons. This facilitates proton transfer to the active site.

QM results. The optimized structures of the second half of the reaction were performed a QM-cluster only approach at the B3LYP/6-31G(d) level of theory. Once again, the two protonation states of His176 were considered. With the His176-H^+ structure, two different models were studied: one with 68 atoms and another with 74 atoms. We can clearly see and as it was reported in many studies that the larger model is more energetically favorable than the small one. Notably, triplet state was adopted for the second half of the reaction because it gave us the most reasonable conformations in the first half compared to quintet and singlet.

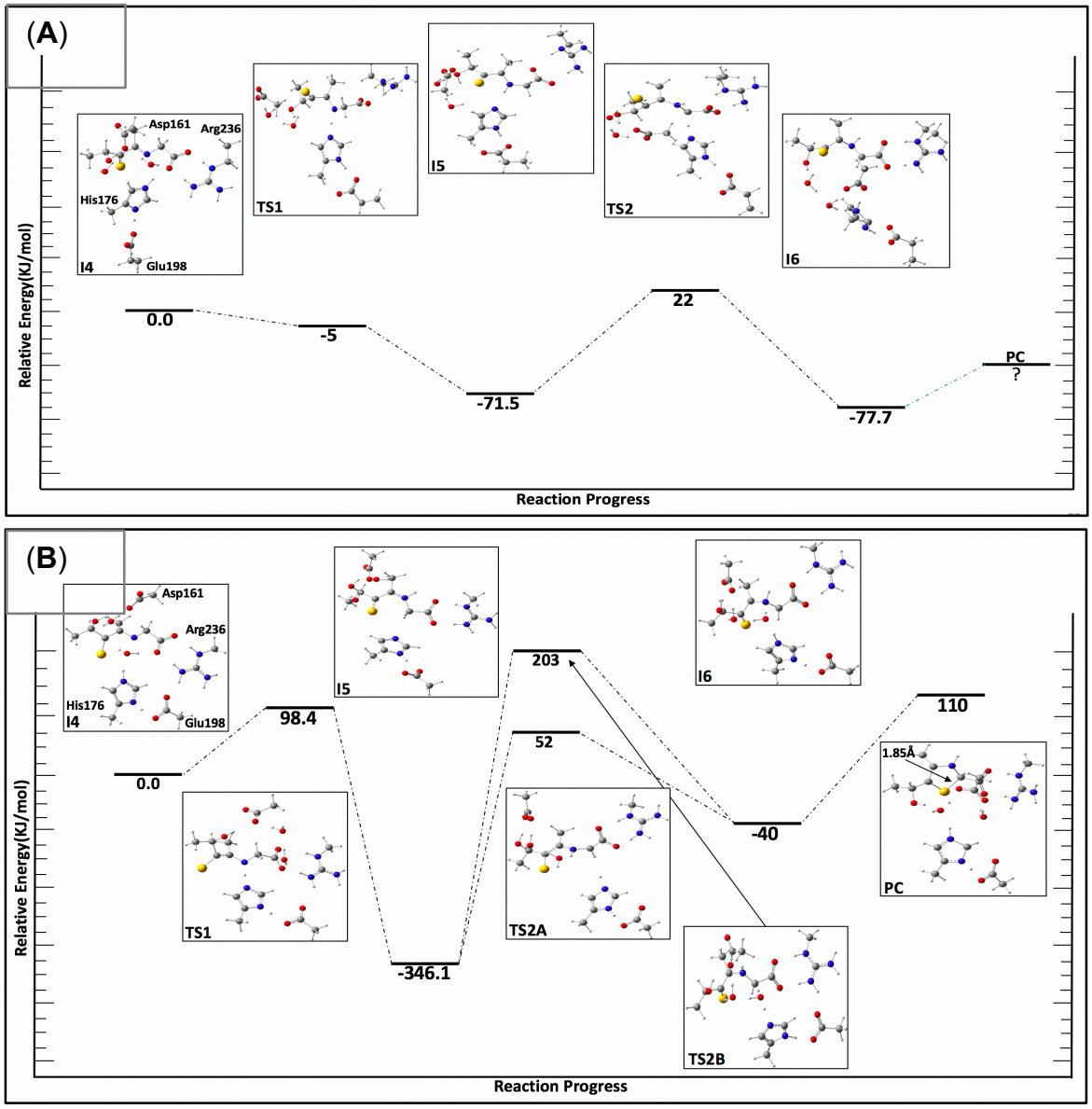


Figure 4.13. Reaction coordinate and optimized intermediates for the second half of the reaction with (**A**) protonated His176 with a 74-atom model and (**B**) protonated His176 with a 68-atom model.

On both surfaces, relative energy was calculated with respect to **I4**. In Figure 4.13A, calculations of **PC** are ongoing. The addition of the 6 atoms in the large model has a significant impact on the energy for all intermediates. Overall, the

model with 68 atoms is less energetically favorable than the larger model. However, **I5** was higher in energy in the larger model at -71.5 kJ/mol compared to -346.1 kJ/mol in the 68 atoms model. Two different pathways for proton transfer from the glycine backbone carbon to His176 were evaluated, corresponding to **TS2**. In Figure 4.13 **A,** a direct transfer to His176 is preferred in the larger model by 181 kJ/mol. In Figure 4.13 **B**, a water molecule facilitating the transfer to His176 is predicted to be stabilized by 151 kJ/mol in the small model. Notably, since the energy of **TS2** in the larger model is expectable, we did not investigate the other pathway using the large model where water facilitate the transfer of the proton to His176.

We have also examined the possibility of proton transfer from protonated His176 to sulfur instead of a direct transfer to the glycine nitrogen with the 68-atom model (Figure 4.14).

Following this pathway, our calculations show that proton transfer bridging through the sulfur atom are less favorable, with intermediates lying at 69.6 kJ/mol, 123.8 kJ/mol, and -27.3 kJ/mol for **I5**, **TS2,** and **I6** respectively. Comparing to the values obtained in Figure 4.13 A. Furthermore, we considered yet another pathway for proton transfer to nitrogen through sulfur. Instead, we used Asp161 as a mediator, forming a proton relay between sulfur and nitrogen (Figure 4.15).

Compared to **TS1** in Figure 4.13B that is a direct transfer of the proton to the glycine nitrogen, **TS3** in Figure 4.15 in which Asp worked as a mediator, showed 33.4 kJ/mol less in energy.

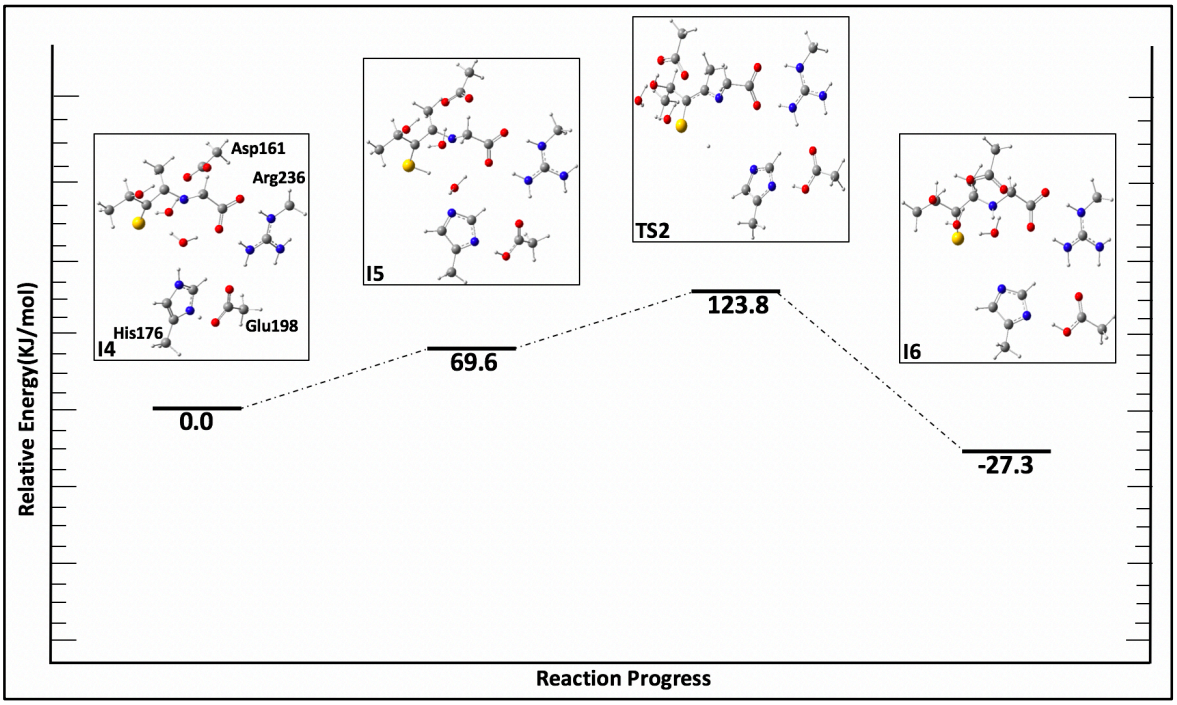


Figure 4.14. Potential energy surface of proton transfer from His176 to the glycine nitrogen bridging through a sulfur atom. Energies are calculated relative to **I4**.

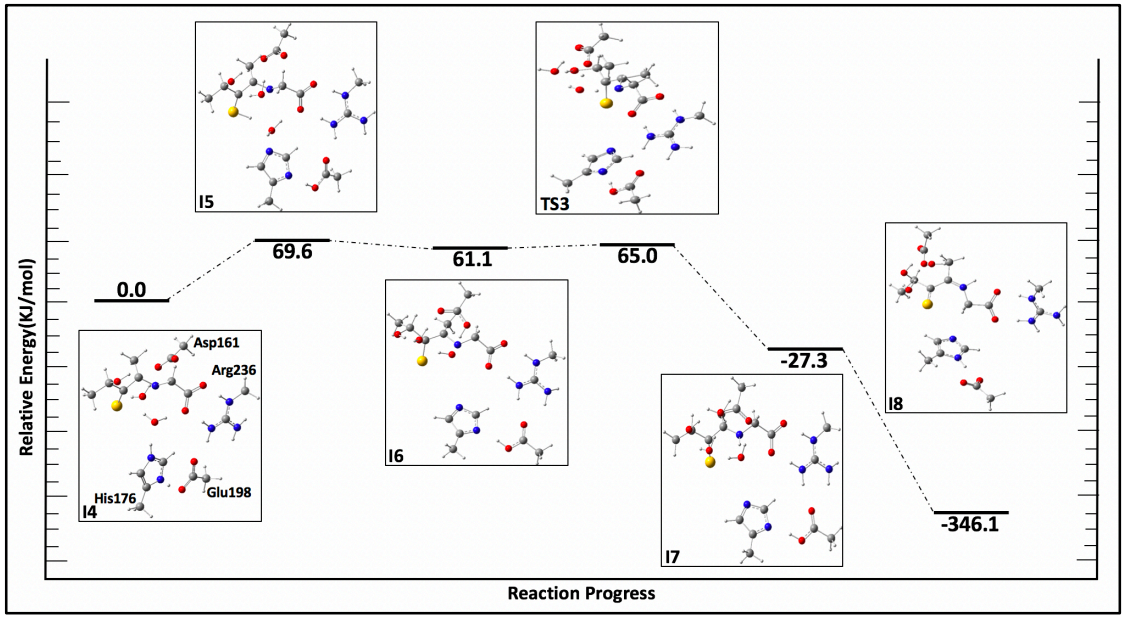


Figure 4.15. Optimized structures of the proton transfer relay from His176 to sulfur then to nitrogen using Asp161 as a mediator.

For the neutral form of His176, the relative energy was calculated for a 67-atom model (Figure 4.16).

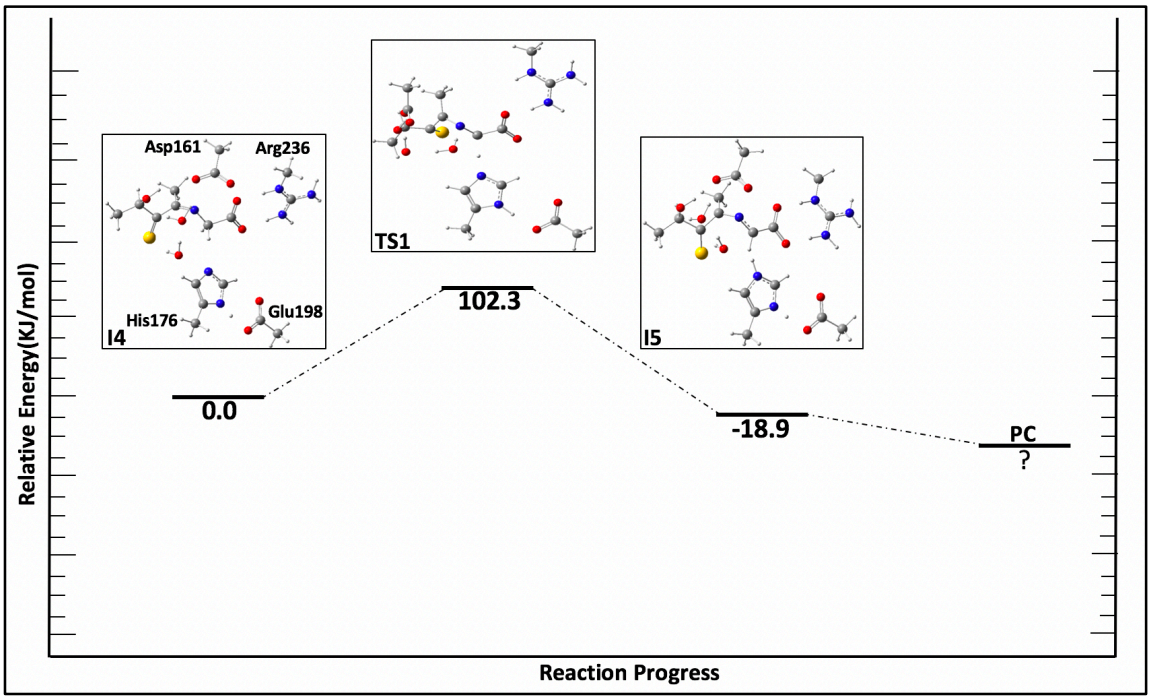


Figure 4.16. Potential energy surface showcasing optimized structures of the second half of the reaction with neutral His176. Calculations on **PC** are in progress.

The reaction proceeds through **TS1** for proton transfer to His176 from the glycine carbon where the intermediate lies -18.9 kJ/mol lower in energy than **I4**. It is important to mention that unlike the steps in the protonated His176, neutral His176 model lacks a proton in the Glycine nitrogen.

QM/MM results. As mentioned in the methods section, for the QM/MM approach, a variety of models were used to examine the mechanism, each differing in the number of atoms and residues included in the QM region. Figure 4.17 two different

models were investigated for the thiazole ring to be formed starting with a protonated His176 form. For all figures, for clarity, the MM layer was omitted.

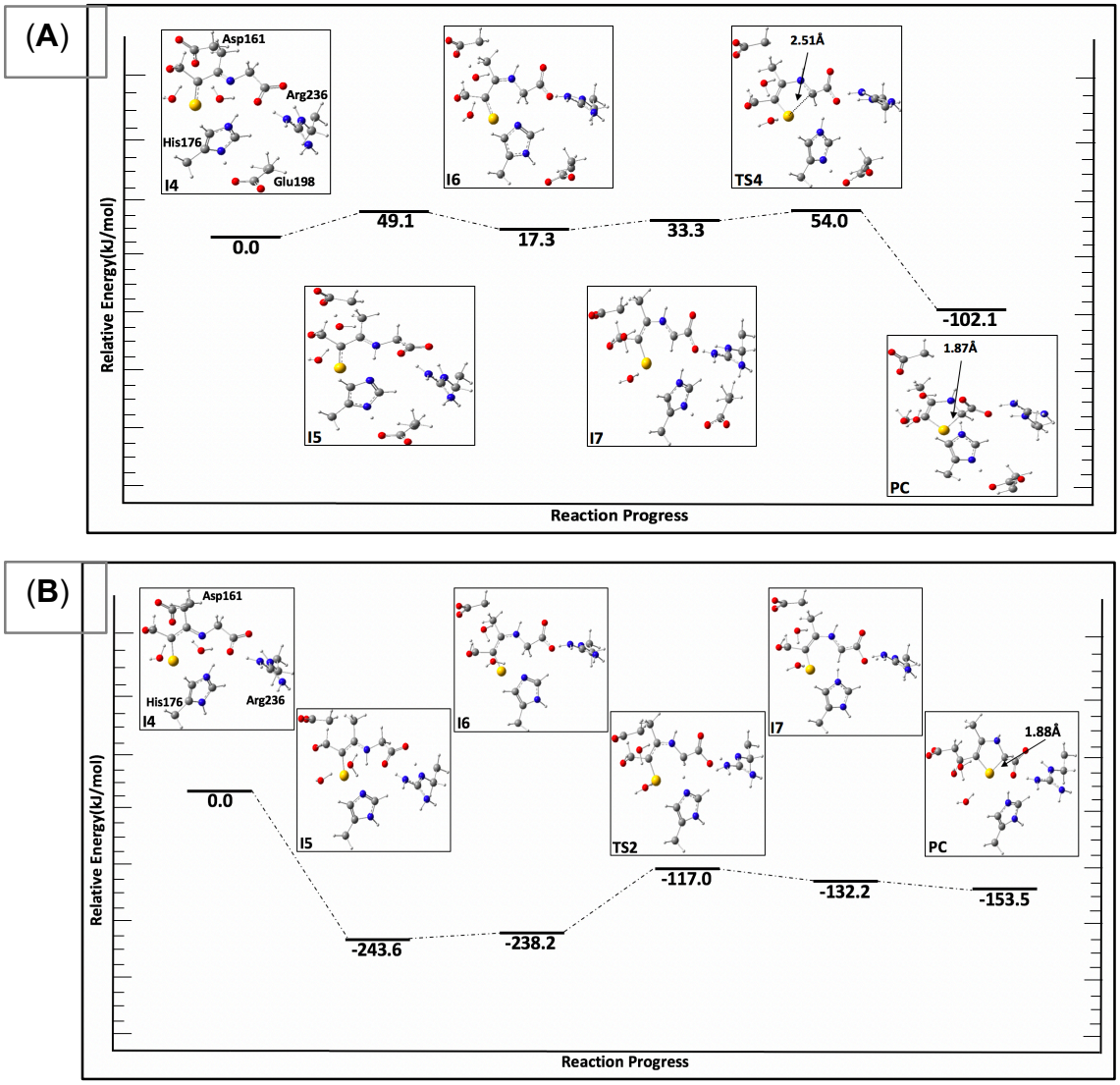


Figure 4.17. Potential energy surfaces of the QM/MM models. **A)** His176, Glycine, Asp161, Arg236, Glu198 and two water molecules are in the QM region while **B)** includes the same residues except for Glu198.

By comparison, the optimized structures of the intermediates were much lower in energy when Glu198 is not included in the QM region. Intermediates starting from **I4** to **I7** have huge difference in energy between the two models, calculated to be over -100 kJ/mol; however, the difference between the product complex is only -51.4 kJ/mol. The carbon-sulfur bond in the thiazole is 1.87 Å and 1.88 Å in models **A** and **B** respectively, which indicates Glu198 has no significant effect on the bond distance.

We also examined the potential pathway of proton transfer from protonated His176 to nitrogen through sulfur (Figure 4.18) as we have done previously with the QM-cluster models.

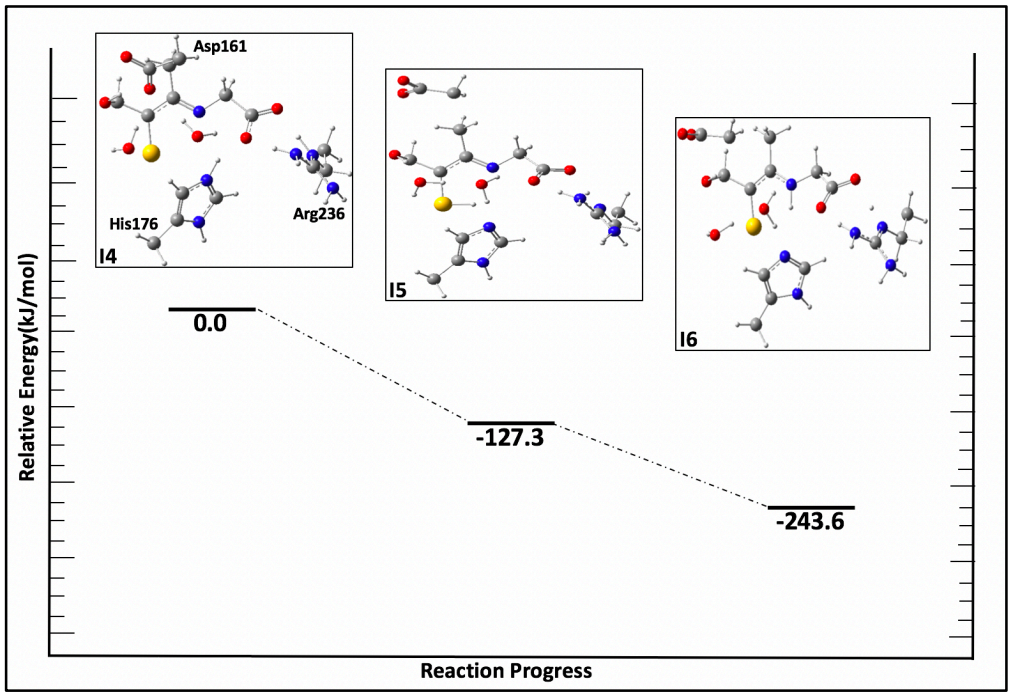


Figure 4.18. Reaction surface and optimized structures obtained for proton transfer to sulfur and then to Glycine-nitrogen including residues illustrated above in Figure 4.17**b**.

Furthermore, we investigated the possibility of the neutral His176 adopting the same residues in QM region as in model Figure 4.17**A** above.

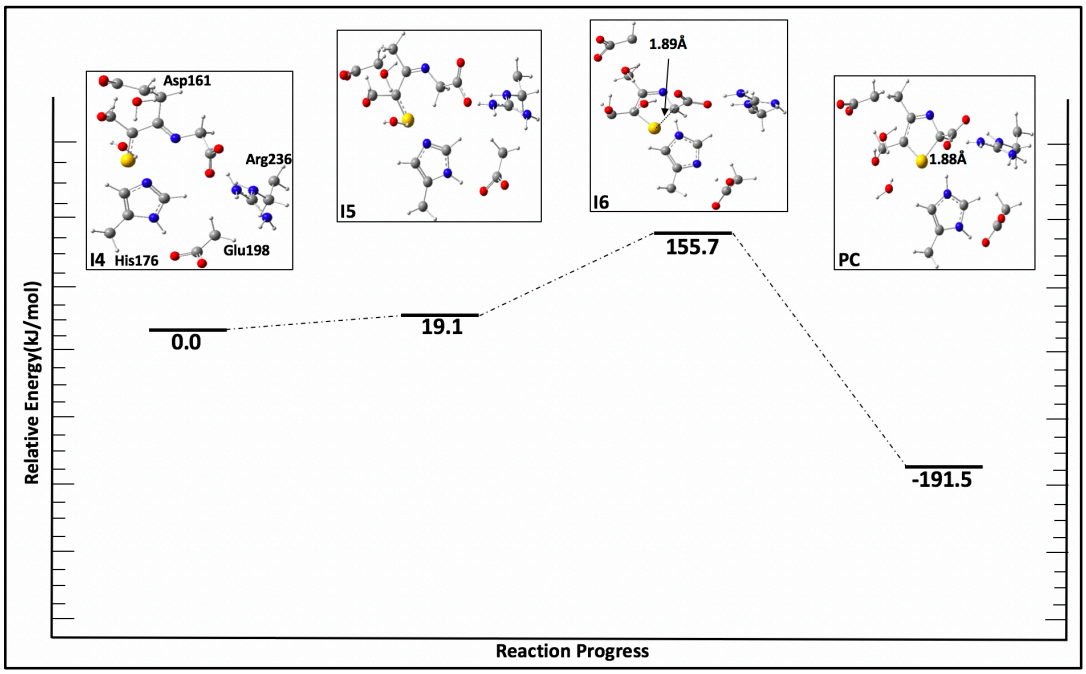


Figure 4.19. Schematic illustration of the second half of the reaction with neutral His176.

As it can be clearly seen, **I6** has the highest energy with respect to **I4**. It is also noted that the His176 proton has adjusted to be in close proximity to Glu198 at 1.75 Å and 1.01 Å respectively. The carbon-sulfur bond in **I6** was 0.01 Å longer than in **PC** at 1.89 Å.

The QM/MM calculation was used to investigate the effect of the surrounding environment on the active site. On the active site residues for almost all QM/MM structures, geometry of the residues has been distorted by the protein

environment. Furthermore, the free energy was also effected by the environment surrounding.

4.4 Conclusions:

In this work, we computationally examined the experimentally proposed mechanism for thiazole formation, as well as our own suggestions for a reasonable reaction mechanism. The investigation was performed using Quantum Mechanics (QM) only approach and Quantum Mechanics/Molecular Mechanics (QM/MM) methods with snapshots obtained from MD for the second half reaction only as we started the examination of the first half with the crystal structure. Notably, we have made some changes and modifications to the proposed mechanism to match the crystal structure of the protein used throughout the study. In the MD simulation, we have made various changes to examine the crystal structure that was presented experimentally. In comparison with the X-ray structure, a 1 ns molecular dynamic simulation resulted in longer bond distances to Fe(II). In our studies, we utilized two density functional methods, B3LYP and B3LYP*, with different basis sets to examine the effect of the method on the energy barrier.

Experimentally, proton transfer to the glycine carboxylate has been suggested to occur after Fe(II) exits the active site. However, as our investigation of the environment around the carboxylate group has shown, it is unlikely for a proton to be obtained in this situation. It is also the case with sulfur as a bridging intermediate. Contrary to what has been experimentally suggested, our simulations have shown that for electrophilic attack to form the thiazole ring, sulfur must be deprotonated.

4.5 References:

1. Song, Z.; Pan, J.; Xie, L.; Gong, G.; Han, S.; Zhang, W.; Hu, Y., Expression, Purification, and Activity of ActhiS, a Thiazole Biosynthesis Enzyme from Acremonium Chrysogenum. *Biochem. (Mosc).* **2017,** *82*, 852-860.

2. Rodionov, D. A.; Leyn, S. A.; Li, X.; Rodionova, I. A., A Novel Transcriptional Regulator Related to Thiamine Phosphate Synthase Controls Thiamine Metabolism Genes in Archaea. *J. Bacteriol.* **2017,** *199*.

3. Liu, H. M.; Sang, S. M.; Lu, Y.; Wang, Z. F.; Yu, X.; Zhong, C. J., Thiamine Metabolism is Critical for Regulating Correlated Growth of Dendrite Arbors and Neuronal Somata. *Sci. Rep.* **2017,** *7*, 12.

4. Mahan, L. K.; Escott-Stump, S., *Krause's Food, Nutrition, & Diet Therapy*. W.B. Saunders: 2000.

5. Piro, A.; Tagarelli, G.; Lagonia, P.; Tagarelli, A.; Quattrone, A., Casimir Funk: His Discovery of the Vitamins and Their Deficiency Disorders. *Ann. Nutr. Metab.* **2010,** *57*, 85-88.

6. Maltz, A., Casimer Funk, Nonconformist Nomenclature, and Networks Surrounding the Discovery of Vitamins. *J. Nutr.* **2013,** *143*, 1013-1020.

7. Northrop-Clewes, C. A.; Thurnham, D. I., The Discovery and Characterization of Riboflavin. *Ann. Nutr. Metab.* **2012,** *61*, 224-230.

8. Hohmann, S.; Meacock, P. A., Thiamin Metabolism and Thiamin Diphosphate-Dependent Enzymes in the Yeast Saccharomyces Cerevisiae: Genetic Regulation. *Biochimica et Biophysica Acta (BBA) - Protein: Struct., and Mol. Enzymol.* **1998,** *1385*, 201-219.

9. Brahman, P. K.; Dar, R. A.; Pitre, K. S., Adsorptive Stripping Voltammetric Study of Vitamin B-1 at Multi-Walled Carbon Nanotube Paste Electrode. *Arab. J. Chem.* **2016,** *9*, S1889-S1896.

10. Pourcel, L.; Moulin, M.; Fitzpatrick, T. B., Examining Strategies to Facilitate Vitamin B-1 Biofortification of Plants by Genetic Engineering. *Front. Plant Sci.* **2013,** *4*, 8.

11. Wolak, N.; Zawrotniak, M.; Gogol, M.; Kozik, A.; Rapala-Kozik, M., Vitamins B-1, B-2, B-3 and B-9 - Occurrence, Biosynthesis Pathways and Functions in Human Nutrition. *Mini-Rev. Med. Chem.* **2017,** *17*, 1075-1111.

12. Kamarudin, A. N.; Lai, K. S.; Lamasudin, D. U.; Idris, A. S.; Balia Yusof, Z. N., Enhancement of Thiamine Biosynthesis in Oil Palm Seedlings by Colonization of Endophytic Fungus Hendersonia Toruloidea. *Frontiers. Plant Sci.* **2017,** *8*.

13. Eser, B. E.; Zhang, X.; Chanani, P. K.; Begley, T. P.; Ealick, S. E., From Suicide Enzyme to Catalyst: The Iron-Dependent Sulfide Transfer in Methanococcus Jannaschii Thiamin Thiazole Biosynthesis. *J. Am. Chem. Soc.* **2016,** *138*, 3639-3642.

14. Rapala-Kozik, M., Vitamin B-1 (Thiamine): A Cofactor for Enzymes Involved in the Main Metabolic Pathways and an Environmental Stress Protectant. In *Biosynthesis of Vitamins in Plants: Vitamins a, B1, B2, B3, B5, Pt A: Vitamins a, B1, B2, B3, B5*, Rebeille, F.; Douce, R., Eds. Academic Press Ltd-Elsevier Science Ltd: London, 2011; Vol. 58, pp 37-91.

15. Raschke, M.; Bürkle, L.; Müller, N.; Nunes-Nesi, A.; Fernie, A. R.; Arigoni, D.; Amrhein, N.; Fitzpatrick, T. B., Vitamin B1 Biosynthesis in Plants Requires the Essential Iron–Sulfur Cluster Protein, THIC. *Proc. Natil. Acad. Sci.* **2007,** *104*, 19637-19642.

16. Zhang, X.; Eser, B. E.; Chanani, P. K.; Begley, T. P.; Ealick, S. E., Structural Basis for Iron-Mediated Sulfur Transfer in Archael and Yeast Thiazole Synthases. *Biochem.* **2016,** *55*, 1826-1838.

17. Begley, T. P.; Ealick, S. E.; McLafferty, F. W., Thiamin Biosynthesis: Still Yielding Fascinating Biological Chemistry. *Biochem. Soc. Trans.* **2012,** *40*, 555-560.

18. Chatterjee, A.; Abeydeera, N. D.; Bale, S.; Pai, P. J.; Dorrestein, P. C.; Russell, D. H.; Ealick, S. E.; Begley, T. P., Saccharomyces Cerevisiae THI4p is a Suicide Thiamine Thiazole Synthase. *Nature.* **2011,** *478*, 542-546.

19. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *Jo. Am. Chem. Soc.* **1984,** *106*, 765-784.

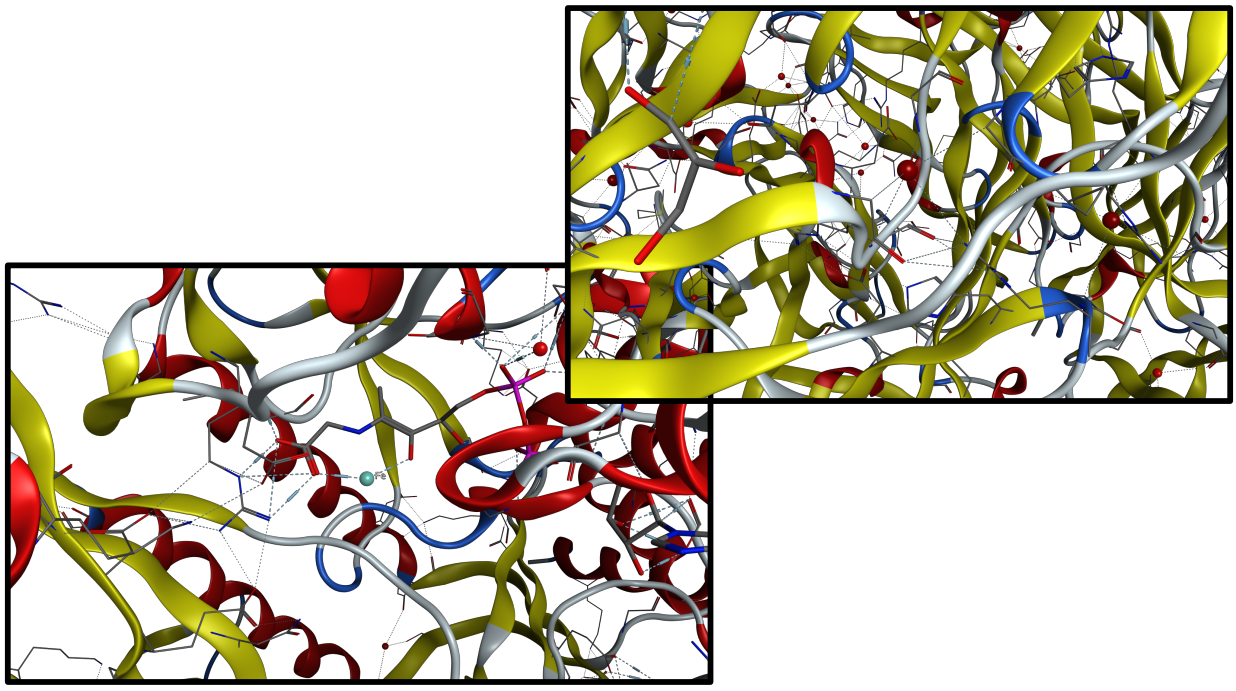
20. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996,** *118*, 2309-2309.

21. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo,

C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz de Montellano, P. R.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision E.01; Gaussian Inc.: Pittsburgh, PA, 2013.

22. Bushnell, E. A. C.; Gherib, R.; Gauld, J. W., Insights into the Catalytic Mechanism of Coral Allene Oxide Synthase: A Dispersion Corrected Density Functional Theory Study. *The Journal of Physical Chemistry B* **2013,** *117* (22), 6701-6710.

CHAPTER 5: Conclusions



5.1 Conclusions

In this thesis multi-scale computational approaches have been employed to examine two important biochemical processes in which sulfur plays a central role. Specifically, molecular dynamics (MD), QM-cluster, and QM/MM methods have been used to examine (i) proposed and possible mechanisms for formation of the sulfilimine bond, as facilitated by the haypohalous acid HOBR, in Collagen IV; and (ii) substrate binding in the iron-dependent thiazole synthase, water distribution in its bound active site, and proposed and possible mechanisms by which the enzyme catalyzes the biochemically important thiazole ring.

Chapter 3. A mechanism in which Sulfilimine was formed was investigated using QM only approach and ONIOM QM/MM methods. MD simulation was also performed to explore the possible reaction and distance between the Lys211 amine nitrogen and Met93 sulfur that calculated to be 3.56 Å after 15 ns simulations. And for further investigation, cluster analysis was done and a representative structure was chosen as the two residues Lys121 and Met93 are in good position for further interaction to occur. Next, we examined the two possible pathways that was being proposed experimentally; Formation of halosulfonium in which one water molecule was needed for the reaction to proceed. The second pathway was the formation of haloamine that needed two water molecules for a proton transfer to the BrOH oxygen. For the QM/MM calculation, formation of halosulfonium and haloamine were calculated to be lower in energy than RC` by -55.7 kJ/mol and -67.0 kJ/mol respectively. We concluded that the haloamine is more energetically favored than halosulfonium with respect to RC`. Further study on the mechanism is required to complete our atomistic-level understanding of the

mechanism of sulfilimine bond formation as well as the experimentally observed formation of the sulfoxide minor product.

Chapter 4. Detailed computational investigations were carried out on the thiamine thiazole synthetases using a proposed iron dependent reaction. MD simulations was first performed on the x-ray structure for 1 ns, and for the second half reaction, a representative structure was chosen for QM/MM optimization. Compared to the experimental data, the bond distances for atoms ligated to iron was slightly longer after the simulation. As it was proposed, proton transfers to the OH in the substrate forming water. For that we did a water density analysis on the active site that confirms the availability of protons around the active site. For the QM only approach and QM/MM calculations, two different models were used to investigate the mechanism; One with neutral His176 and a second one with protonated His176 as well as different models differing in number of atoms and residues included. We also examined different pathways for the proton transfer from the nitrogen of the His176 imidazole to the nitrogen in the substrate; one with a direct transfer and the second one is from His176 nitrogen to sulfur in substrate and then to the nitrogen. Based on our study and in contrast to the experimental proposed of proton transfer to the oxygen of the substrate carboxylate group and Sulphur, it was unlikely to happen based on the analysis we have made on the active site after iron exits. More investigations and study have to be done to explore and understand the synthesis of the thiamine.

In terms of the sulfilimine, we successfully examined the two pathways that was proposed for the crosslink to form. Investigations to be continued to model the last products of the mechanism proposed earlier in Chapter 3, sulfilimine crosslink and the sulfoxide. Gaining insights into this essential crosslink enables us to better

understand its critical role in collagen IV and potentially the effect and/or impact of diseases arising from such damaging reactive oxygen species as HOX (X = Cl or Br) acids.

Further study on the THI4 and possible investigation of other metals in the reaction will bring insights into understanding the mechanism by which the thiazole is formed. As we addressed previously in Chapter 4, we can also study the effect of mutating residues in the active site. Such studies will also provide greater insights into each of their roles within the mechanism.

Vita Auctoris

Name:	Taqred Alnakhli
Place of Birth:	Taif, Saudi Arabia
Year of Birth:	1990
Education:	Taiba University, B.Sc., Medinah, Saudi Arabia, 2012
	University of Windsor, M.Sc., Windsor, ON, 2018