Title
Design and Evolution of Metalloenzymes Through Multi-Scale Approaches

Author
Valdez, Crystal Edie

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Design and Evolution of Metalloenzymes Through Multi-Scale Approaches

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Crystal Edie Valdez

2015

ABSTRACT OF THE DISSERTATION

Design and Evolution of Metalloenzymes Through

Multi-scale Approaches

by

Crystal Edie Valdez

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2015

Professor Anastassia N. Alexandrova, Chair

Natural metalloenzymes are often the most proficient catalysts in terms of their activity, selectivity, and ability to operate at mild conditions. However, metalloenzymes are occasionally surprising in their choice of catalytic metals, and in their responses to metal substitution. Indeed, from the isolated standpoint of producing the best catalyst, a chemist designing from first principles would likely choose a different metal. Due to competing evolutionary pressures, many natural enzymes may not have evolved to be ideal catalysts and can be improved for the isolated purpose of catalysis *in vitro* when the competing factors are removed. To improve and, in due course, design metalloenzymes, extensive sampling and proper treatment of the electronic structure of the bound metal(s), is required, while seamlessly merging the required techniques to assess energies and entropies, or their changes, for the entire system. Approaching these challenges with a multi-scale approach, the Alexandrova group has developed an accurate and efficient quantum mechanical/molecular mechanics (QM/MM) hybrid dynamics method to model metalloproteins called quantum mechanics/discrete molecular

mechanics (QM/DMD). QM/DMD operates through an iterative scheme between QM and MM machineries. DMD is a flavor of molecular dynamics (MD) that approximates the continuous interaction potentials in classical MD with square-well potentials, course-graining the potentials and overall reducing the number of calculations needed. Due to these discretized potentials, DMD is driven by collision events rather than physical forces as in traditional MM and MD. Therefore, the user saves a tremendous amount of time with DMD by solving ballistic equations of motions rather than Newtonian equations of motions.

This fast and efficient hybrid dynamics tools has allowed us to investigate various metal-dependent phenomena in natural metalloenzymes such as: 1) exploring Nature's curious choices for specific metals using two amide hydrolases that contain different metals as a case study, 2) examining protein conformational responses to substrate binding and metal replacement as showcased by the role of a flexible loop β-lactamase in binding antibiotic substrates and 3) investigating how the species of the metal dictates the reaction mechanism in a pair acireductone dioxygenases (ARD/ARD').

Extending outside the realm of naturally occurring enzymes, our tools have the ability to span across formidable challenges such as metalloenzyme design, where stabilization of a transition state of the catalyzed reaction in the specific binding pocket around the metal needs to be achieved. QM/DMD was used in the redesign of a well-studied Zn^{2+} peptidase, carboxypeptidase A (CPA), an enzyme involved in the breakdown of proteins, with a slight preference for bulky hydrophobic groups. More specifically, the enzyme and substrate system were modified to create specific-specific binding and subsequent experiments proved the mutant to be catalytically active. Additionally, another tool called *Eris*-QM/DMD was formulated to better gauge the effect of mutation on protein structure during the design process. *Eris* is a stand-alone package that evaluates protein stability upon mutagenesis. Coupling the software to QM/DMD gives us the distinct advantage of accounting for the effect of the metal during protein alternations.

With this diverse set of tools, our future ambitious goals are to install catalytically potent non-physiological metals into proteins. While nature is limited to operating with bio-available elements, some metals such as Ir, Pd, Sc, and Rh, which have been shown to be excellent catalysts, even surpassing physiological metals. If the catalytic activity of these non-physiological metals can be combined with the superb selectivity and mild operational conditions characteristic of proteins, new proficient enzymes may emerge. Another advantage to enzymatic catalysis, done either *in vitro*, or *in vivo*, is that it can be cheaper, “greener”, and more efficient than synthetic catalysis. An early endeavor in this frontier of metalloenzyme design involves installing Pd^{2+} into an existing protein scaffold, specifically for intramolecular hydroarylation of C-C triple bonds to form coumarins.

The dissertation of Crystal Edie Valdez is approved.

Kendall N. Houk

Giovanni Zocchi

Anastassia N. Alexandrova, Committee Chair

University of California, Los Angeles

2015

To my dear family and friends

A mi querida familia y amigos

TABLE OF CONTENTS

LIST OF FIGURES

1. CHAPTER 1 *Mysteries of Metals in Metalloenzymes*

2. CHAPTER 2 *Computational Treatment of Metalloproteins*

3. CHAPTER 3 *Why Urease is a di-Nickel Enzyme, whereas the CcrA β-Lactamase is a di-Zinc Enzyme*

LIST OF SCHEMES

LIST OF TABLES

5. Chapter 5 *Metal-dependent activity of Fe and Ni acireductone dioxygenases: how two electrons reroute the catalytic pathway*

6. Chapter 6 *Co^{2+} Acireductone Dioxygenase: Fe^{2+} Mechanism, Ni^{2+} Mechanism, or Something Else?*

7. Chapter 7 ***Computational Redesign of Carboxypeptidase A and Experimental Confirmation***

ACKNOWLEDGEMENTS

My foremost and deepest gratitude and respect goes to my selfless, sacrificing, loving and spirited family. My mother, Victoria, and father, Edward, I thank you both endlessly for believing in me and my abilities to pursue my dreams. The tireless drive you both possess in anything you do continues to inspire me every day. I love you both with all of my heart and soul. The surprise camping trips were always cool too. To my sister, Carolyn, I thank you with every atom in my body for being the resilient foundation for my academic and personal growth. A few highlighted moments I treasure forever are: shark bites, scientific discussions at awkward family parties and the times you would prepare stress-relieving lavender tea for me. To my husband, Manasa Chandra: my foundation, my best friend, my love, my present and my future. My life changed when you came into the picture and only continues to grow more and more awesome. You are my inspiration and I struggle to find the exact words to describe my love and respect for you. All of my accomplishments are a manifestation of your love and support. Besos. To my extended family and friends in Los Angeles, the Antelope Valley, the Bay Area and all around the world, thank you for the constant love from near and far, family gatherings with home-made food, endless laughter and memories, and all the warm words of wisdom.

To my academic advisor, Professor Anastassia N. Alexandrova, I thank you tremendously and continuously for the invaluable guidance and remarkable opportunities you have provided me. I am constantly amazed by your scientific rigor, creativity, and excitement and deeply respect your work ethic and passion for life. I am thankful you took me on with patience and confidence in my scientific ability during your assistant professorship, all of which has been a priceless asset to my growth as a scientist and academic professional. I hope to continue doing science together, with an occasional ballet or yoga class sprinkled in.

I am deeply grateful to all the wonderful scientific collaborators who have taken the time and resources to train and mentor me. Some of these individuals include: Professor Nikolay

Dokholyan, who graciously hosted me at UNC Chapel Hill for ten days, training me on protein modeling techniques while making me feel part of the Dokh Lab family. I thank my committee members, Professor Bill Gelbart, Professor Ken Houk and Professor Giovanni Zocchi, for taking the time to serve on my committee and providing valuable mentorship along the way. I thank Hong Nyugen, student of Professor Joseph Loo, for believing this little computational chemist could do some experiments. Thank you to all of my wonderful professors who have been inspirational and monumental in my growth as a budding scientist and academic - Professor Ben Schwartz, Professor Alejandro Briseño, Professor Lüis Campos, Professor and Associate Dean Joe Watson, Professor and Associate Dean Carlos Grijavla, Professor and Vice Provost/Dean Robin Garrell, Professor Alex Levine and Professor Michael Maroney. To these professors, I am humbled and motivated by not only your commitment to research excellence but to your commitment to being good citizens, citizens that foster the development and well being of the global community.

To my colleagues and dear friends I met while at UCLA: Manuel Sparta, thank you for taking me under your wing and mentoring me to become a patient, inquisitive and thorough computational scientist. Forever your Padawan. To other AA lab members - Anjan Nandula, Tony Smith, Jin Zhang, Sean Nedd, Jonny Dadras, Mioy Huynh, Michael Nechay, Mai-Anh Ha, PJ Robinson, Elisa Jimenez-Izal and Nathan Gallup. Thank you all for your support and love, I could not have survived this crazy thing called graduate school without you. To my awesome colleagues - Steven Alexander Lopez for life chats over coffee breaks. #rockstars. Geeta Vadehra, for good times and wine dates. Alexandra Mendoza, my UCLA sister and happy hour date. To my UCLA salsa family, just to name a few - Kevin Keys (my salsa dancing biomathematician buddy!), Misho Galbo, I-Wen Chang, Rachel Hirst, Francisco Toro and Christelle Guette. Thank you all for the dances that melted my stress away. To my Organization for Cultural Diversity in Science (OCDS) and Alliance in Science and Engineering (ADSE) family

– Ashay Patel, Tejas Shah, Ashley Comfort, Charles Frazier and Diana Azurdia. Thank you guys for the “diversity” of thought and science in my life, I cherished and continue to cherish it every moment of it.

As a firm believer in the ancient African proverb “It takes a whole village to raise a child”, I have learned it takes a whole community to raise a scientist. I bow my chin to my chest, with my heart lifted and my hands in prayer pose, in gratitude for this community. Namaste

This dissertation is a collected of previous publications listed below:

Chapter 1 is a modified version of the publication “Mysteries of metals in metalloenzymes”, Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N. *Acc. Chem. Res*. **2014**, *47*, 3110-3117. This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV), the DARPA Young Faculty Award N66001-11-1-4138 and Alfred P. Sloan Research Fellowship (ANA). This was an invited feature article.

Chapter 2 is a modified version of the publication **“**Computational treatment of metalloproteins”, Nechay, M. R.; Valdez, C. E.; Alexandrova, A. N. *J. Phys. Chem. B.,* **2015***, 119,* 5945-5956*.”* Nechay, M. R. and Valdez, C. E. contributed equally to this work. This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV), the UCLA dissertation year fellowship (CEV), the DARPA Young Faculty Award N66001-11-1-4138 and Alfred P. Sloan Research Fellowship (ANA). This was an invited feature article that was featured on the cover of the journal.

Chapter 3 is a modified version of the publication “Why Urease is a di-Nickel Enzyme, whereas the CcrA beta-Lactamase is a di-Zinc Enzyme*”,* Valdez, C. E.; Alexandrova, A. N. *J. Phys. Chem. B.,* **2012,** *116*, 10649-10656. This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV) and the DARPA Young Faculty Award N66001-11-1-4138 (ANA).

Chapter 4 is a modified version of the publication "*The Role of the Flexible L43-S54 Protein Loop of the CcrA Metallo-beta-lactamase in Binding Structurally Dissimilar beta-Lactam Antibiotics",* Valdez, C. E.; Sparta, M.; Alexandrova, A. N. *J. Chem. Theor. Comput.*, **2012,** *9*, 730-737. This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV) and the DARPA Young Faculty Award N66001-11-1-4138 (ANA).

Chapter 5 is a modified version of the publication "Metal-dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway", Sparta, M.; Valdez, C. E.; Alexandrova, A. N. **2013,** *J. Mol. Biol.*, *245*, 3007-3018. Nechay, M. R. and Valdez, C. E. contributed equally to this work. This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV) and the DARPA Young Faculty Award N66001-11-1-4138 (ANA). This article was featured on the cover of the journal.

Chapter 6 is a modified version of the publication "Valdez, C. E.; Gallup, N. M.; Alexandrova, A. N. *Co2+ Acireductone Dioxygenase: Fe2+ Mechanism, Ni2+ Mechanism, or Something Else?* **2014***, Chem. Phys. Lett.*, *604*, 77-82*."* This work as supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) #2011115747 (CEV), the DARPA Young Faculty Award N66001-11-1-4138 and Alfred P. Sloan Research Fellowship (ANA).

Chapter 7 is a modified version of the manuscript "Computational redesign of the Carboxypeptidase A (CPA) and Experimental Confirmation", Valdez, C. E.; Nguyen, H.; Morgenstern, A.; Eberhart, M. E.; Loo, J. A.; Alexandrova, A. N. *In preparation.*

Chapter 8 is a modified version of the manuscript *"*Eris-QM/DMD: A metalloprotein design tool", Valdez, C. E.; Shirvanyants, D.; Dokholyan, N. V.; Alexandrova, A. N. *In preparation.*

VITA

2005-2009	Bachelor of Science, Biochemistry California Polytechnic State University, San Luis Obispo, CA (Cal Poly)
2006	William Frost Research Fellowship Recipient, Cal Poly
2009	NSF Research Experience for Undergraduate (REU) Program, University of Oregon
2010	Physical Chemistry graduate student at UCLA
2010-2011	Teaching Assistant, University of California, Los Angeles (UCLA)
2011-2015	NSF Graduate Research Fellowship Program (GRFP), UCLA
2011-2012	Extreme Science and Engineering Discovery Environment (XSEDE) Scholars Program
2012	NOBCChE West Regional Meeting Student Travel Grant
2012	Carl Storm Underrepresented Minority (CSURM) Fellowship
2013	UCLA Graduate Division Research Travel Grant
2013	63rd Lindau Nobel Laureates Meeting UCLA and NSF Delegate
2014	UCLA Dissertation Year Fellowship
2014	UCLA Graduate Division AGEP Professional Development Award
2015	Edward A. Bouchet Honor Society Inductee
2015	Graduate Research Advocacy Day Nominee (1 of 4 students selected at UCLA)

PUBLICATIONS

Nechay, M. R. [*]; **Valdez, C. E.** [*]; Alexandrova, A. N. "Computational treatment of metalloproteins.*" J. Phys. Chem. B.* **2015**, *119,* 5945-5956. **Invited feature article**. **Featured on Journal Cover**. [*]*Authors contributed equally.*

Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N. "*Mysteries of metals in metalloenzymes." Acc. Chem. Res*. **2014**, *47*, 3110-3117. **Invited Article.**

Valdez, C. E.; Gallup, N. M.; Alexandrova, A. N. "*Co2+ Acireductone Dioxygenase: Fe2+ Mechanism, Ni2+ Mechanism, or Something Else?" Chem. Phys. Lett.* **2014***, 604*, 77-82*.*

Sparta, M.[*]; **Valdez, C. E.**[*]; Alexandrova, A. N. "Metal-dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway*." J. Mol. Biol.* **2013**, *245*, 3007-3018. **Featured on Journal Cover.** [*]*Authors contributed equally.*

Valdez, C. E.; Sparta, M.; Alexandrova, A. N. "*The Role of the Flexible L43-S54 Protein Loop of the CcrA Metallo-beta-lactamase in Binding Structurally Dissimilar beta-Lactam Antibiotics." J. Chem. Theor. Comput.* **2012**, *9*, 730-737*.*

Valdez, C. E.; Alexandrova, A. N. "*Why Urease is a di-Nickel Enzyme, whereas the CcrA beta-Lactamase is a di-Zinc Enzyme." J. Phys. Chem. B.* **2012**, *116*, 10649-10656. **Featured in September 22, 2014 issue of C&EN**

SELECTED PRESENTATIONS

Valdez, C. E.; Alexandrova, A. N. **"**Understanding metalloenzyme structure and function for the design of artificial enzymes." 12th Annual Yale Bouchet Conference on Diversity and Graduate Education. New Haven, CT. 2015. **Invited talk**

Valdez, C. E; Alexandrova, A. N. "The curious case of ARD: metal-dependent functionalities of metalloenzymes", 2014 SACNAS National Conference. Los Angeles, CA, 2014. Oral presentation.

Valdez, C. E.; Alexandrova, A. N. "Evolution of metalloenzymes through multi-scale dynamics modeling", UCLA Research Showcase, 243th National Meeting of the American Chemical Society (ACS), San Francisco, CA, 2014. Poster presentation.

Valdez, C. E.;Alexandrova, A. N. "How the electronic structure of metals contribute, and impede, metalloenzyme structure and function." 2nd International Conference on Chemical Bonding, Lihue, HI, USA, 2014. – **Featured in September 22, 2014 issue of C&EN. Invited talk**

Valdez, C. E.; Alexandrova, A. N. "Evolution of metalloenzymes through multi-scale dynamics modeling", Gordon Research Seminars: Bioinorganic Chemistry (GRS), Ventura, CA, 2014. Poster presentation.

Valdez, C. E.; Sparta, M.; Alexandrova, A. N. "Elucidation of Nature's Choice for Catalytic Metals: A Comparative Study of di-Nickel Urease and di-Zinc beta-lactamase", Gordon Research Conference (GRC) in Computational Chemistry, West Dover, VT, 2012.

Valdez, C. E.; Sparta, M.; Alexandrova, A. N. "The Role of Beta-lactamase Loop in Binding Antibiotics", NOBCChE West Regional Meeting, The Scripps Research Institute, La Jolla, CA, 2012. Oral Presentation

CHAPTER 1

Mysteries of Metals in Metalloenzymes

1.1. INTRODUCTION

From the isolated standpoint of catalysis, the general principles that should govern the construction of metal-containing active sites in enzymes are fairly clear. For example, a metal playing the role of a Lewis acid needs to have an appropriate charge and coordination geometry, and the quality of its interactions with a substrate can be described via the HSAB theory. Redox centers must have appropriate reduction potentials, and orbital populations that result in minimal structural rearrangements upon charge transfer (in order for charge transfer to be efficient, in accord with Marcus theory). However, understanding the evolutionary optimization of enzymes goes beyond following these chemistry principles. Metalloenzymes exist in a cellular environment where they have to remain stable, properly structured, and have the ability to find and bind the needed metals. Metals are predominantly toxic, scarce, or too reactive to be readily available without an elaborate metal-storage and delivery mechanism. Cells are dependent on metal trafficking pathways that bypass metal detoxification steps and allow metals to be utilized by some enzyme in the cell.[2,3] Metals can also compete for binding sites with other cations of comparable size, making concentration and selection regulation necessary for cell life. The cell must overcome challenges imposed by the relative stability of complexes formed by divalent transition metals, also known as the Irving-Williams series. [4,5] One way to overcome is through localized folding within the cell, passing the enzyme through compartmentalized locations where specific metal ions exist.[6] Additionally, catalytic processes in the cell are tuned to proceed at a physiologically convenient rate, i.e. not too fast and not too slow. Evolution takes care of all aforementioned aspects at once, thus imposing multiple constraints on enzymes as catalysts. Phenomenologically, this means enzymes should be improvable for catalysis *in vitro*, and understanding them requires a holistic approach at the cellular or organismal level. Resultantly, so far, nature is the only true holder of the key to enzymology. Examining natural metalloenzymes through the looking glass of a chemist, we sometimes find surprises. In this Account, we aim to highlight a few cases of perplexing choices

for metals in metalloenzymes and unexpected effects of metal replacement on the activity of enzymes, to illustrate this conundrum. For some systems, the explanations can be found in the structures and catalytic mechanisms, but some cannot be explained without the involvement of the bigger, alas poorly understood picture. Our goal herein is to provide not a full review of the field, but hopefully a diverse set of provocations.

1.2. STRANGE OR SUBOPTIMAL CHOICES FOR CATALYTIC METALS: METAL-DEPENDENT HYDROLASES

Metal-dependent hydrolases, such as peptidases, phosphatases, amidases, and ATPases, catalyze hydrolysis. They contain one or two metal centers that activate the nucleophile (typically H_2O or OH^-) and/or the substrate for the key step in the reaction, nucleophilic attack. The metals play the simple role of a Lewis acid. Many hydrolases use Zn^{2+} as the metal of choice, but some are inhibited by Zn^{2+}, and activated by Mg^{2+}, Mn^{2+}, Co^{2+}, Ni^{2+}, or Ca^{2+}. Furthermore, sometimes enzymes hydrolyze very similar substrates but by different metals. From the standpoint of bioavailability, the recruitment of redox-active metals (e.g. iron, whose relative abundance in the Earth's crust is much higher than in living organisms) as Lewis acids seems wasteful. On the other hand, it is possible that Zn^{2+} is so frequently used only because of its easy availability.[7] The natural question arises, have enzymes evolved to utilize not the most catalytically optimal metal, but the most readily available one? The degree to which proteins are selective to these divalent metals is also intriguing.

One interesting monometallic hydrolase where the trade-off between activity and availability can be observed is Histone Deacetylase 8 (HDAC8). It participates in the regulation of gene expression. HDACs and their complements, histone acetyltransferases (HATs), modify the lysine residues of histones by removing or adding an acetyl moiety to the ε amino group (Fig. 1.1).[8,9] Histones are the primary protein component of chromatin, the DNA and protein "package" that fits inside a cell nucleus. Acetylated histone lysines are associated with an open

chromatin structure that permits DNA transcription, while deacetylated lysines are associated with tightly-packed DNA that is resistant to transcription.[10] Thus, deacetylation is correlated with gene silencing.

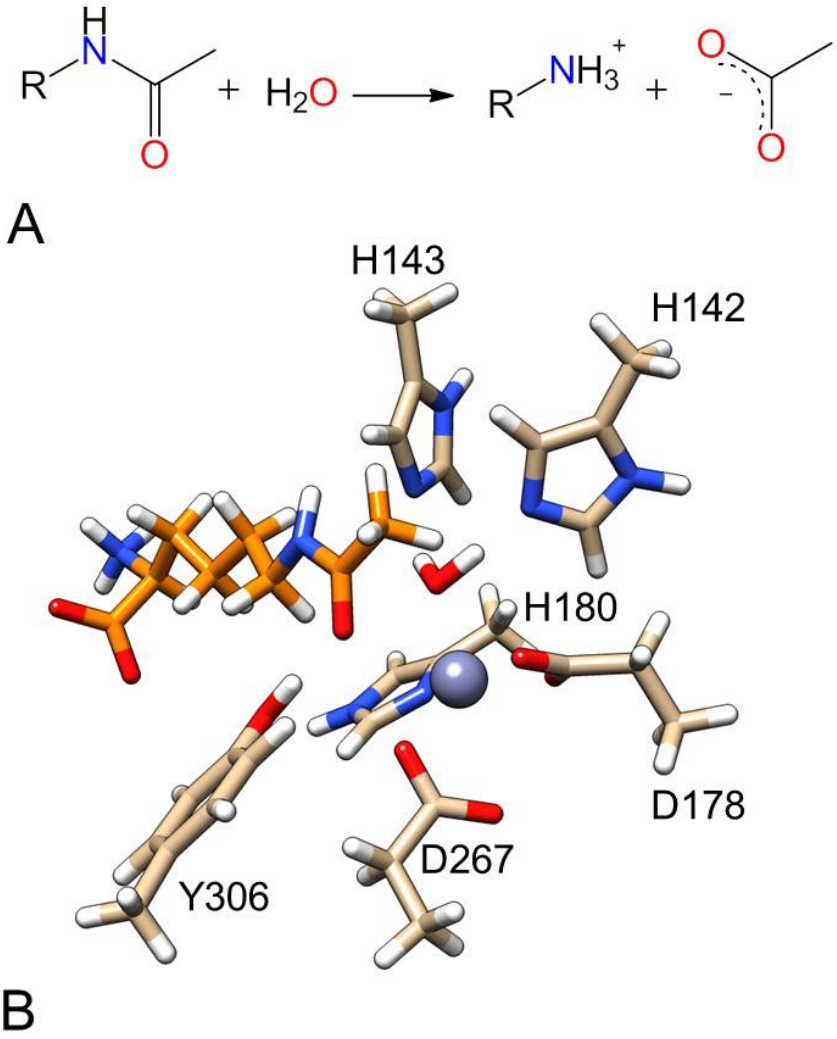


Figure 1.1. Reaction catalyzed by HDACs (A), including HDAC8. (B) Structure of HDAC8 active site, based on crystal structure with PDB accession code 2V5W. The acetylated lysine substrate is shown in orange. The protein coordinates a metal ion via an Asp_2His binding pocket.

HDAC8, the best-characterized HDAC, is generally classified as a zinc-dependent enzyme even though its Asp_2His metal binding site is considered unusual for zinc (Fig. 1B).[11] Indeed, experimental studies[11] demonstrate that HDAC8 has a 10^6-fold greater affinity for Zn^{2+}

over Fe^{2+}. However, the Fe^{2+}-containing enzyme is significantly more catalytically active, with a k_{cat}/K_M value almost three times that of the Zn^{2+} variant (2300 ± 160 M^{-1} s^{-1} for Fe^{2+} vs. 800 ± 50 $M^{-}1$ s^{-1} for Zn^{2+}).[12] Depending on experimental conditions, HDAC8 either does not bind Fe^{3+} or suffers a >85% loss in activity when Fe^{2+} is oxidized to Fe^{3+}.[11,12] The question of which metal the enzyme contains *in vivo* remains contentious, as typical methods used to extract and purify the enzyme could oxidize Fe^{2+}, resulting in its replacement with Zn^{2+}.[11] Additionally, despite the greater affinity for Zn^{2+}, typical intracellular concentrations of Fe^{2+} are considerably higher (0.2 - 6 μM for Fe^{2+} vs. 5 - 400 pM for Zn^{2+}), which could allow iron to out-compete zinc in the absence of non-thermodynamic considerations (e.g. metallochaperones).[11] Of further interest is the even higher catalytic activity of Co^{2+} (k_{cat}/K_M = 7500 ± 300 M^{-1} s^{-1}, a 9-fold increase over the zinc-containing enzyme).[12] With the exception of vitamin B_{12} (cobalamin) cofactors, cobalt-containing enzymes are exceedingly rare, and the intracellular concentration of cobalt is low.[13] The fact that HDAC8 has evolved to use iron and/or zinc may represent a compromise with metal bioavailability. Another intriguing possibility is that evolutionary pressures may select for a more limited catalytic activity. An overly active HDAC could be too proficient at silencing genes and hiding them from transcription factors. HDACs co-evolved with their complement HATs, and cells must maintain a balance between their respective activities for optimal function.

Bimetallic hydrolases also exhibit unusual metal-dependent functionalities. Here, we focus on bimetallic amidases, ureases and β-lactamases. Urease is present in plants, bacteria and yeast, and catalyzes the hydrolysis of urea.[14] Its function is widely encompassed in medical and agricultural settings, not only in nitrogen mineralization but also as the main virulence factor in many pathogens, such as *Helicobacter pylori*.[15] Urease was the first enzyme characterized to contain Ni in its active site, a puzzling discovery[16] since the job of a Lewis acid can be done easily by a non-redox metal.[7,17]

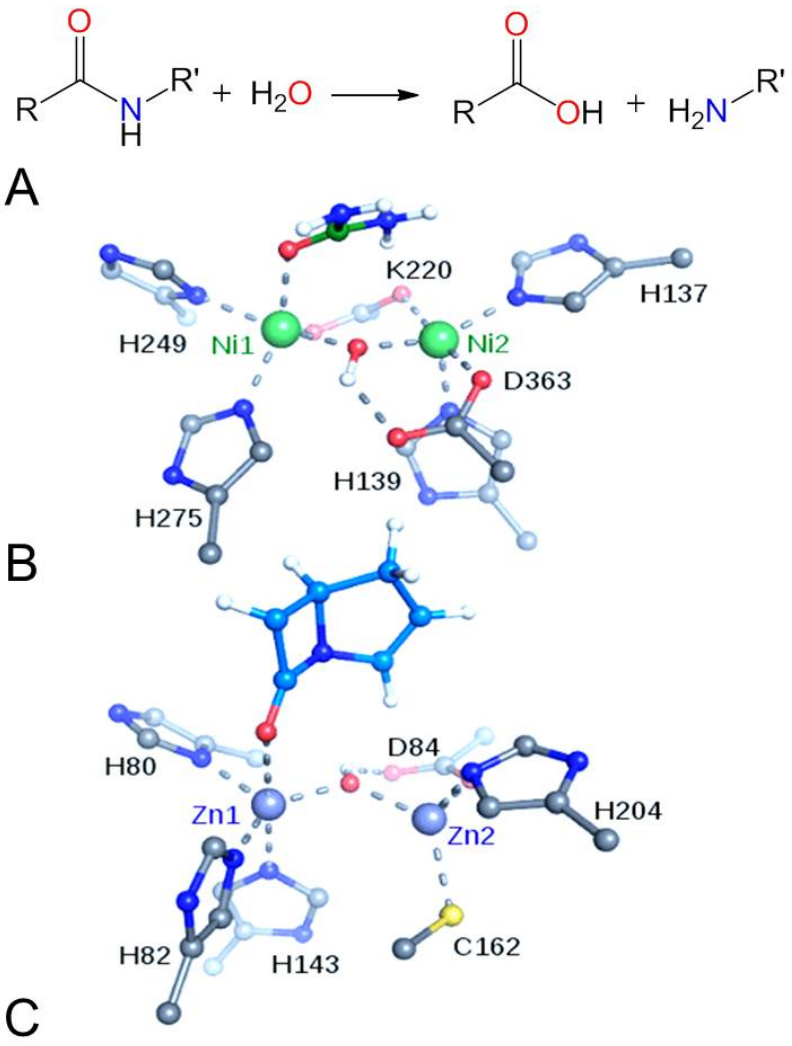


Figure 1.2. Amide hydrolysis (A) catalyzed by metal-containing amidases. The active sites of (B) urease, and (C) metallo-β-lactamase with their respective substrates share similar ligand coordination environment, yet contain different metals.

In comparison, metallo-β-lactamase is an amidase that employs two Zn^{2+} ions. Metallo-β-lactamases hydrolyze β-lactam antibiotics, such as penicillin, and are the primary cause of antibiotic resistance in bacteria.[18,19] Thus, urease and metallo-β-lactamase both hydrolyze amides, according to the general scheme shown in Fig. 2A, and both contain a di-metal active site with a very similar arrangement of ligands (Fig. 2B,C). What is the reason behind the selection of Zn^{2+} or Ni^{2+}?

Considering the natural substrates catalyzed by the respective enzymes, urea has a resonance stabilized structure that makes it intrinsically more difficult to hydrolyze compared to the strained four-membered ring found in β-lactam antibiotics. In a purely computational study,[20] we employed urease to hydrolyze β-lactam and found that the reaction proceeds easily. Conversely, β-lactamase is unable to hydrolyze resonance-stabilized urea. From a structural standpoint, the active site of urease is probably suboptimal for binding β-lactam drugs. Therefore, we substituted the two Zn^{2+} ions in β-lactamase with Ni^{2+} ions. The Ni-substituted β-lactamase is predicted to catalyze the hydrolysis of the β-lactam ring with an efficiency surpassing that of natural di-Zn β-lactamase.[20] This suggests that the use of Zn in β-lactamase does not provide maximal possible efficiency for the enzyme. Zinc gives a performance that is satisfactory for biological purposes but could be improved via substitution with Ni. The use of Zn instead of Ni in β-lactamase could be dictated by other factors in the cell, such as the toxicity or lower availability of Ni. Computationally derived mechanisms have the ability to elucidate these metal-dependent features of metalloenzyme catalysis.

Another interesting example is a urease found in *Heliobacter mustelae*, a gastric pathogen in ferrets that employs two Fe^{2+} ions in the active site.[21,22] (Fig. 1.3) Although the Fe-containing urease is less active then the native Ni-containing urease, the activity is adequate for the survival of the pathogen in the ferret's low-nickel gastric environment. This could represent an instance of evolutionary adaption of the enzyme for its specific niche, surviving a low pH and low Ni environment. Carter *et al.*[23] show this Fe-containing urease could be substituted with Ni under specific low Fe level and high Ni level *in vitro* conditions. The Fe-containing urease differs from the nickel urease by 92 residues, which are hypothesized to map to a region near the active site that influences metal loading.

Through these examples, we infer that hydrolases exemplify the interplay between most proficient catalysis and other evolutionary pressures. Metal use in enzymes is not straightforward, and therefore not always optimal considering catalytic properties alone. Taking

this message one step further, we can suggest that some natural enzymes could be improved for catalysis *in vitro* through metal replacement.

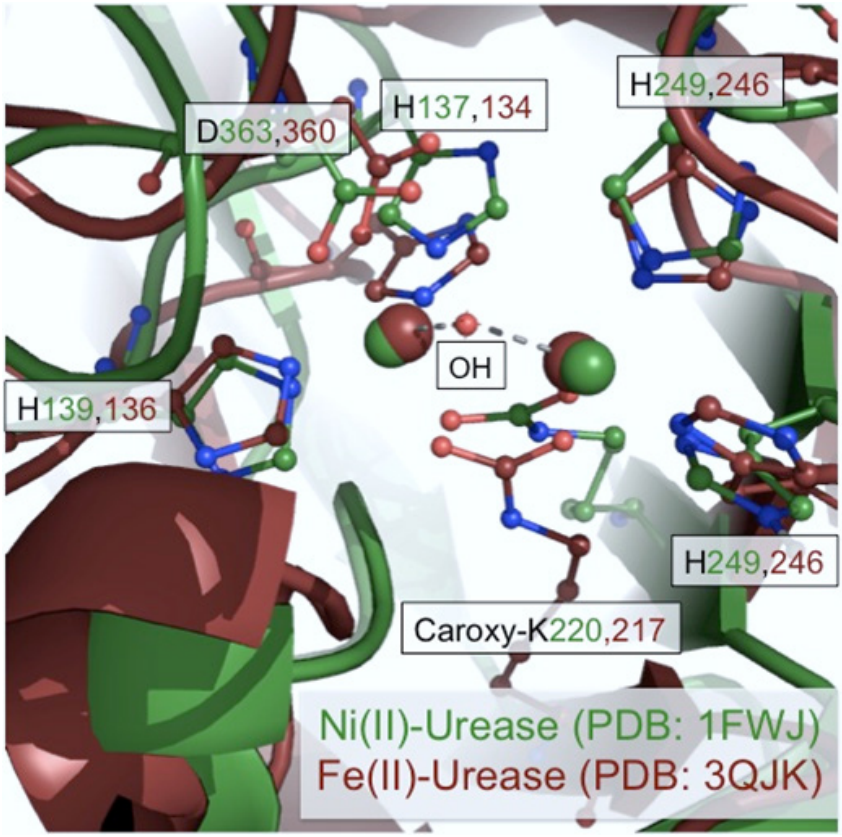


Figure 1.3. Overlay of Ni2+ and Fe3+-containing ureases reveals closely identical active site ligand environments.

1.3. ENZYMES THAT USE SEEMINGLY UNSUITABLE REDOX METALS

In view of the sensitivity of enzymes as simple as hydrolases to the identity of the catalytic metal ion, it would seem unreasonable to expect redox enzymes to swap metals: a specific reduction potential is required for each reaction. Nevertheless, there are cases when metal-replacement works, specifically in situations where the needed metal is, or was, scarce. Through a variety of strategies, enzymes tune the reduction potential of one metal to behave

more like another, with a varying degree of success. Few such mechanisms are currently understood, and likely many of them are still undiscovered.

1.3.1. MN IN PLACE OF FE IN RIBONUCLEOTIDE REDUCTASES (RNRS)

Most organisms, including humans, depend on Fe for a variety of enzyme-facilitated processes, from electron transport to cell replication. However, in the presence of large amounts of H_2O_2, i.e. in conditions of extreme oxidative stress, Fe^{2+} can be oxidized and removed from enzymes through Fenton chemistry.[24] Most organisms have developed peroxide scavenging enzymes, catalases and peroxidases that quickly remove H_2O_2 from the cells.[25] However, it was recently discovered that some organisms have a different means of survival in conditions of oxidative stress: they employ Mn instead of Fe in certain crucial enzymes.[26,27] One prominent enzyme in which this Mn-Fe trick was observed is bacterial RNR, specifically binuclear Fe_2^{2+} RNRs. These redox enzymes catalyze the conversion of ribonucleotides into deoxyribonucleotides (Fig. 4A),[28,29] and thus are needed for the replication of DNA.

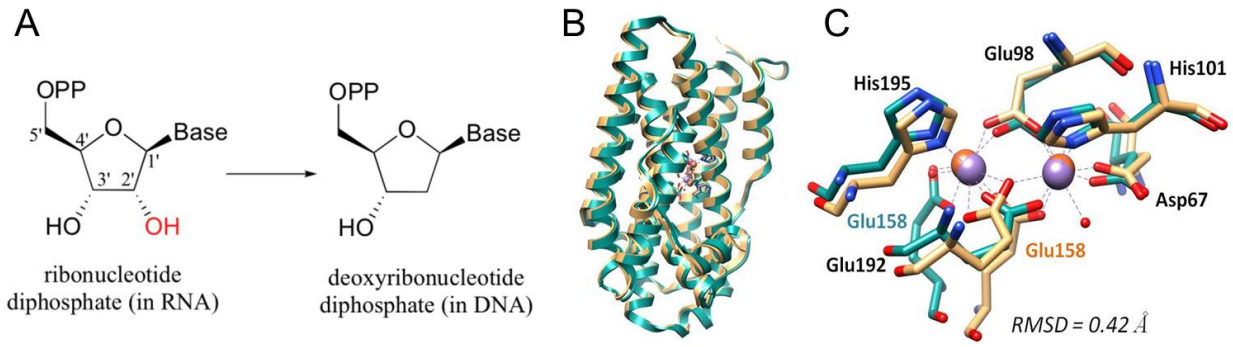


Figure 1.4. RNR reaction (A). DiFe RNR with the native Fe and substituted with Mn: (B) secondary structure overlays very closely, (C) but there are shifts in the immediate coordination of the metals, especially the residue Glu158.[30]

There is an interesting Mn-vs-Fe chemistry that takes place in the b-subunit of this protein, shown in Fig. 4B,C. The β-subunit is responsible for the production of the tyrosyl radical, the oxidant of the Cys residue located 35 Å away from in the a-subunit, where the

reaction in Fig. 4A takes place. Some bacteria, such as *Streptococcus sanguinus*, and *E.Coli*, have the NrdAB RNR, with a partially understood mechanism of action. Additionally, these organisms have a homologue of NrdAB, called NrdEF, whose function has long remained unclear. A series of recent works[27,31] revealed NrdEF becomes active in conditions of oxidative stress. NrdEF preferentially binds Fe in the presence of excess Fe, but in this form its activity is insufficient to drive DNA replication.[27] In low Fe conditions, it binds Mn, and gains a 5-fold increase in activity. In contrast, NrdAB is an incompetent catalyst when in the Mn-form. The site in Fig. 4C can function as di-Mn, or Fe-Mn.[32,33]

What are the factors that make one RNR active and the other inactive with Mn as a cofactor? Electronic differences between the two metals and the subtle differences in the geometries of the Fe and Mn active sites (Fig. 4B,C) may lead to differences in the reduction potential.[27,31] Overall, structures of NrdEF with Fe or Mn overlay very closely, but the active site has an appreciable RMSD. It is indeed known that reduction potential is a very sensitive quantity, altered dramatically even by residues in the second coordination sphere and weak H-bonds at the active site.[28,34] Additionally, long-range electrostatics play a significant role. This is well-known in modeling, where proper electrostatic embedding is required to reproduce the reduction potential.[35,36] Structural differences may also alter the electron conduction. Hence, there are a handful of unexplored possibilities to explain the Fe-Mn conundrum of RNR, but no conclusive explanation available to date.

1.3.2. MN AND NI INSTEAD OF CU IN SUPEROXIDE DISMUTASE (SOD)

The presence of Mn and Ni instead of Cu in superoxide dismutase (SOD) is another example of successful replacement of a redox metal (Fig. 5). SODs are vital antioxidants in organisms that slow-down or prevent oxidative damage, inflammation, aging, and cancer.[37-39]

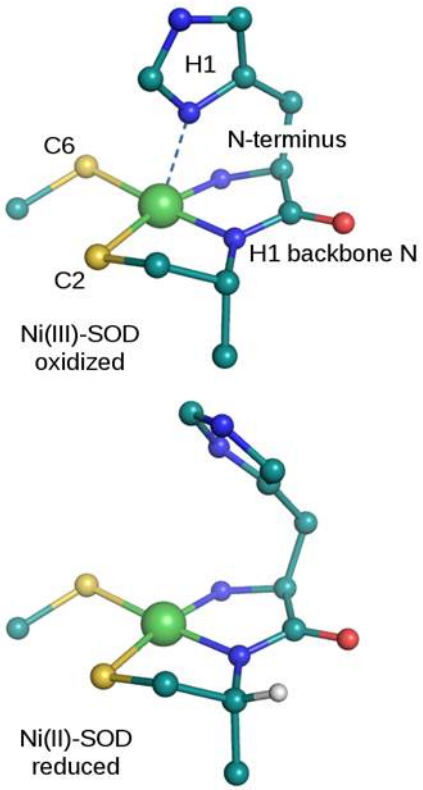


Figure 1.5. Two coordination modes of Ni in Ni-SOD, where the cycling redox state of Ni from Ni(II) to Ni(III) modulates the coordination of the N1 of H1.

They catalyze the disproportionation of the superoxide radical to yield O_2 and H_2O_2. SODs accomplish this through a two-step, ping-pong mechanism via a redox-active metal center, typically Fe or Mn, or a Cu/Zn pair, or with the unique case of Ni.[40] While our focus will be on Ni-SOD, we note a common link between Fe-Mn in RNRs and SODs. Mn can replace Fe in the role of redox metal in some SODs to avoid the deleterious effects of Fenton chemistry.[41] Interestingly as well, in certain bacteria, nickel levels enhance NiSOD expression while

simultaneously repressing FeSOD expression, suggesting the evolution of NiSOD when Fe, or perhaps even Mn, concentrations were low.[42]

Ni-SODs exist only in *Streptomyces* and cyanobacteria. Unlike all other metals utilized in SOD, Ni^{2+} does not catalyze superoxide dismutation in aqueous solution due to an improper redox potential (a calculated +2.26 V when the optimum reduction potential is 0.36 V).[43] The active site of Ni-SOD is strikingly different from those of the other SODs.[44] The Ni ion, coordinated by a "Ni-hook" motif,[45] has a square-pyramidal coordination geometry when Ni is in its oxidized form (Ni^{3+}). It has four equatorial ligands -- two thiolates from Cys2 and Cys4, a deprotonated amide of the Cys2 backbone, the N-terminal group of His1 -- and one axial amidazole from His1. Upon reduction, Ni^{2+} loses the His1 amidazole ligand and becomes square planar, making Ni-SOD the only observed SOD with a coordination number that changes as a function of metal oxidation state.[45,46] The flexible coordination geometry activates Ni and confers the proper redox potential needed to functional as a SOD.

1.3.3. EXCHANGE OF MO WITH W IN DMSO REDUCTASE (DMSO-R)

In contrast to the specific coordination environment needed for Ni-SOD, exchanging Mo^{4+} with W^{4+} in the enzyme DMSO reductase (DMSO-R) for redox chemistry purposes sometimes yields an equally viable and active enzyme (Fig. 6). Mo is used by all forms of life.[47,48] W is very close to Mo in coordination preferences, electronic properties, and ionic radii: 79, 75, and 73 pm for Mo, and 80, 76, 74 pm for W, for the three oxidation states of VI, V, and VI, respectively. Organisms can be grouped according to those that prefer W, those that prefer Mo, and a significantly smaller group that can (to some extent) use both.[49] Most Mo-containing enzymes are mononuclear, and their activity involves transferring an oxygen atom to/from the substrate directly bound to the metal.[50] In Mo-DMSOR, found in *Rhodabacter capsulatus*, the direct replacement with W is efficient, and results in an active enzyme very close in structure to the Mo form (Fig. 6).[1] W-DMSOR was shown to be significantly more active than Mo-DMSO for

the reduction of DMSO, but it does not oxidize DMS, unlike Mo-DMSOR. The difference stems from the dissimilarity in the reduction potentials: W^{4+} is a stronger reductant and a weaker oxidant than Mo^{4+}. Redox behavior and the relative bio-availability of the metals (W is more bio-available than Mo)[51] [52] are suggested to be major factors in determining whether an organism uses Mo or W in enzymes.

The rare organisms that can accommodate both Mo and W provide the link. The thermophilic methanogen *Methanobacterium wolfei* grows one of two isozymes determined by its environment.[53] In addition, growing *E. coli,* which usually uses Mo, on media containing W in the absence of Mo has been shown to generate an active enzyme with higher catalytic activity and increased thermal stability.[54]

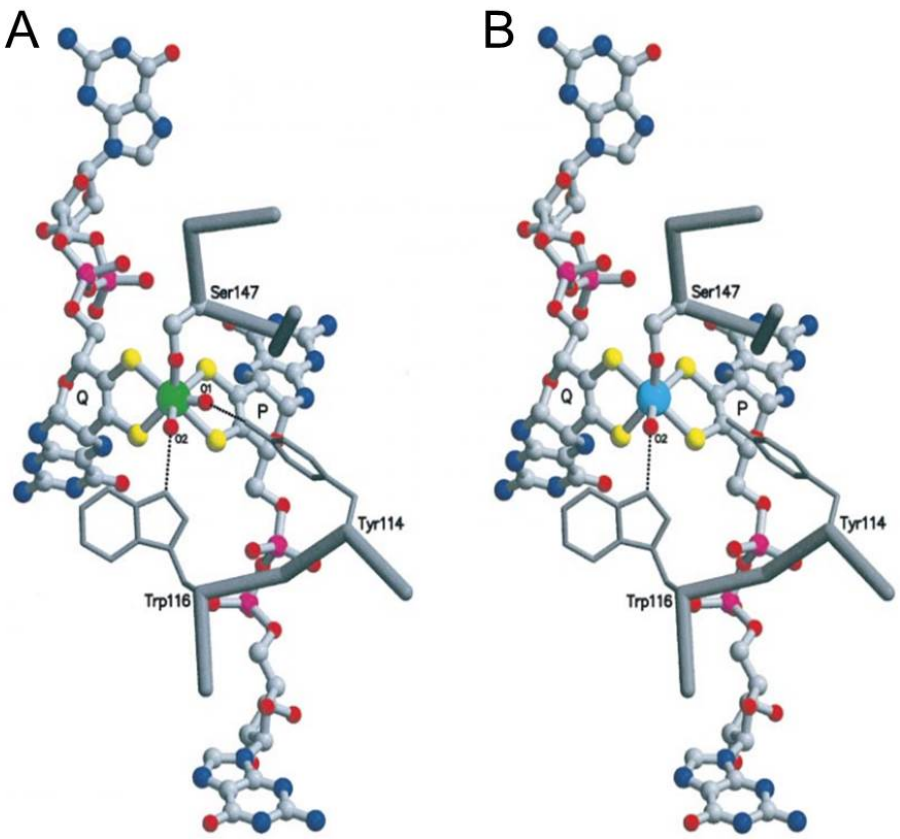


Figure 1.6. The structure of DMSO-R with (A) Mo and (B) W. The RMSD between the two structures is 0.06 Å.[1]

1.4. CONSERVATIVE METAL REPLACEMENTS CAN DEACTIVATE In contrast to dramatic redox exercises, simple template metal ions in enzymes are sometimes chosen perfectly for their purpose and cannot be replaced even by seemingly interchangeable metals. At first glance, the group 2 metals, Mg^{2+} and Ca^{2+}, are very similar electronically. Both metals interact with amino acids primarily by electrostatics, leaving the biggest difference between these two to be ionic radius - Mg^{2+} being smaller than Ca^{2+}, and closer in size to common biologically-used metals. Calcium's size, propensity for a higher coordination number,[55] and other related geometric strains[56] could be the sources of Mg^{2+} enzyme inhibition by Ca^{2+}. Interestingly, in some cases, enzymes preferentially bind Ca^{2+} over Mg^{2+}.[57-59] An example is the EF-hand motif, common to many calcium signaling and transport proteins, which undergoes conformational change to an inactive form when octahedrally-coordinated Mg^{2+} is present.[57] These conformational changes can lead to a dramatic cascading effect, such as along the signaling transduction pathway.[58] We discuss catechol-O-methyltransferase to illustrate our point.

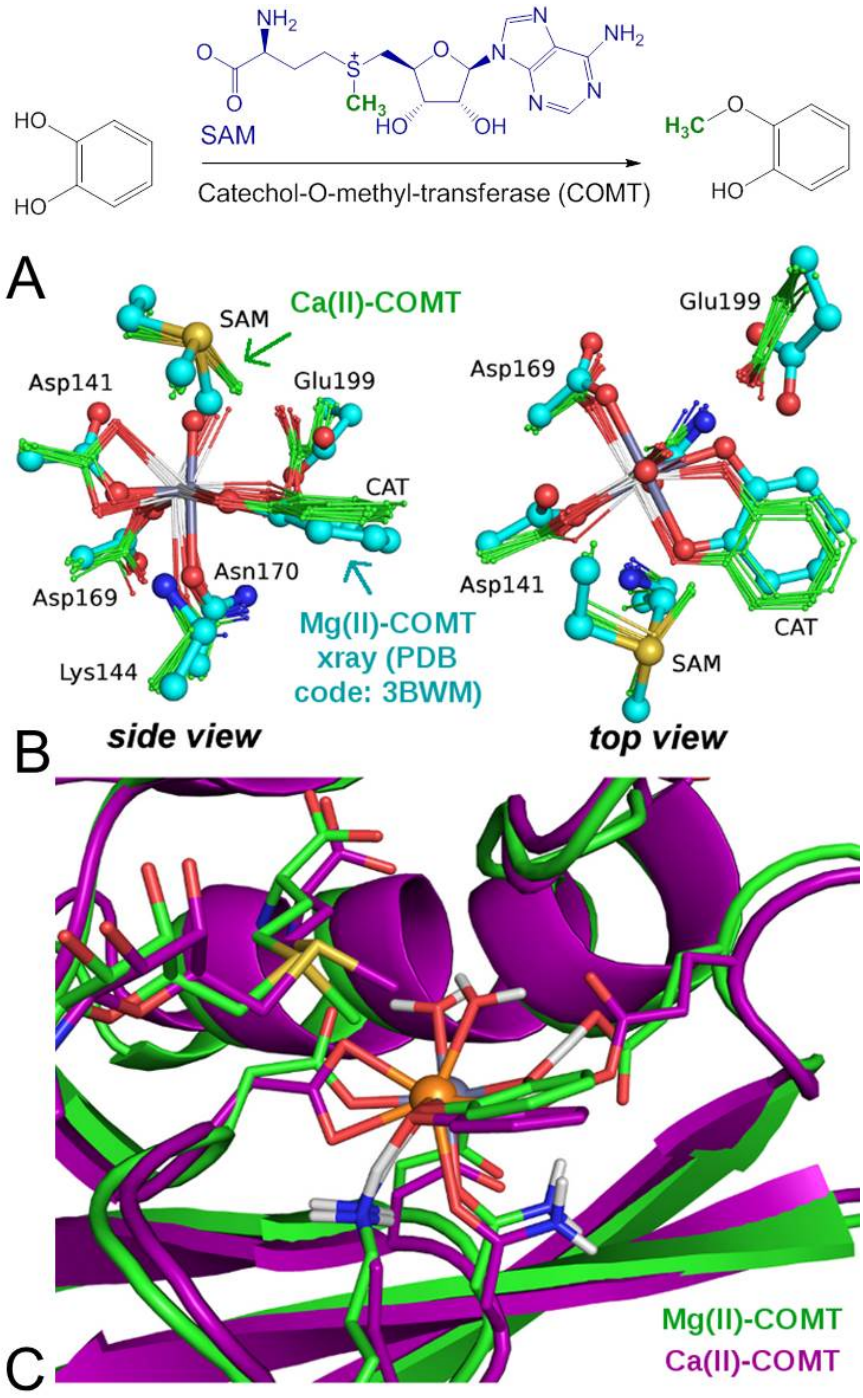


Figure 1.7. (A) Reaction catalyzed by COMT. (B) Comparing simulated Ca^{2+}-COMT structures (thin sticks, small spheres) with X-ray structure of Mg^{2+}-COMT (bold sticks, large spheres) reveals the shift of the methyl group on SAM with respected to catechol (CAT). (C) Representative simulation snapshots show overall structure adjustment when Mg^{2+} is replaced with Ca^{2+}.

Catechol-O-Methyltransferase (COMT) is a Mg^{2+} metalloenzyme involved in the regulation of catecholamine neurotransmitters in the brain and other organs (Fig. 7A).[60,61] The enzyme catalyzes the transfer of the methyl group from the cofactor S-adenosyl-L-methionine (SAM) to catechol. The role of the metal in COMT is to bind and position the catechol substrate in the correct reactive orientation towards SAM for the methylation step.[62-64] Native Mg^{2+} in COMT can be replaced with Co^{2+}, Mn^{2+}, Zn^{2+}, Cd^{2+}, Fe^{2+}, Fe^{3+}, Ni^{2+}, and Sn^{2+}.[65] However, metal replacement leads to varying changes in the activity and structure of the enzyme. The replacement of Mg^{2+} with Ca^{2+} leads to the inhibition of COMT. Fe^{2+} creates an only slightly weaker catalyst compared to Mg^{2+}, but Fe^{3+} is an inhibitor. We addressed the structure and function of these metal variants of COMT through an extensive computational experiment.[66] For Ca^{2+}-COMT, the reaction energetics are predicted to be unfavorable due to restructuring of the active site. The larger cation, Ca^{2+}, coordinates one additional ligand as compared to Mg^{2+}, sits deeper in to the binding cavity and distorts the reacting parts of catechol and SAM out of alignment for methyl transfer (Fig. 7B,C). Thus, inhibition is a purely geometric effect in this case. In contrast, the inhibitory effect of Fe^{3+} comes solely from the electronic properties of the metal, specifically its high electrophilicity.

1.5. METAL REPLACEMENT CAN ALSO REROUTE

Besides boosting or killing the native activity of enzymes, metal replacement can have much more sophisticated effects on enzyme function. One such example is acireductione dioxygenases (ARD). ARD is the only enzyme that utilizes the same protein scaffold to catalyze two different oxidation reactions, depending exclusively on whether Fe^{2+} or Ni^{2+} is bound.[67-69] Fe-dependent ARD' oxidizes the substrate 1,2-dihydroxy-3-keto-5-(methylthio)pentene (acireductone) into two products: the α-keto acid precursor of methionine and formate. Fe-ARD' is part of the methionine salvage pathway in cells.[69] Ni-dependent ARD instead oxidizes acireductone into three products, methylthiopropionate, CO, and formate, and provides a shunt

into the methionine salvage cycle.[69] Interestingly, the protein has micromolar affinity for both metals (Kd < 0.4 μM for Fe and Kd < 0.1 μM for Ni), readily allowing interconversion between the two. In contrast to other dioxygenases, such as non-heme iron dioxygenases that require the metal to activate the dioxygen for oxidation,[70,71] ARD binds the substrate directly.[72] The mechanistic difference between Fe and Ni for acireductione oxidation has been speculated to arise from coordination of the substrate to the metal center. However, recent experiments and computational studies challenge this view and propose a new explanation.[70,73,74] We found that the two metals bind the substrate in the same way, exposing the same sites to the nucleophilic attack by O_2 (Fig. 8). Instead, an additional intermediate forms in the Fe-ARD' mechanistic pathway that is not observed in Ni-ARD.[74] This extra intermediate is also found in biomimetic complexes mimicking ARD.[73] The ability of Fe but not Ni to stabilize an additional intermediate comes from the redox flexibility of Fe^{2+}, allowing for the flow of electrons from the residues to the bound O_2. The more electron rich Ni^{2+} does not allow for this.[74]

Another interesting question is that of the functionality of intermediate Co-ARD. We showed computationally that, for Co^{2+}-substituted ARD, both Fe and Ni routes are energetically accessible.[75] Experimentally, ARD has been observed to confer Ni-ARD activity with Mn^{2+} and Co^{2+}, whereas incorporation of Mg^{2+} confers Fe-ARD' activity.[68,72]

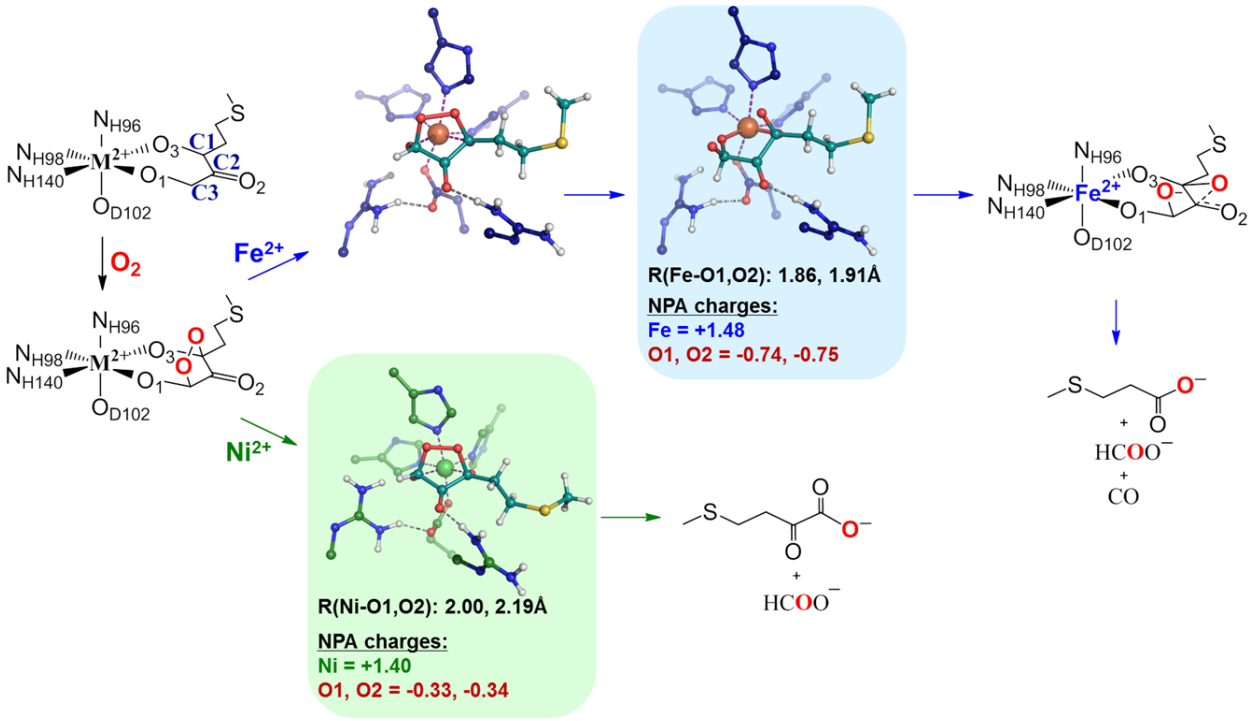


Figure 1.8. ARD/ARD' reaction mechanisms, illustrating how the redox flexibility of Fe^{2+} allows for the electron transfer to the bound dioxygen and it dissociation, leading to a Fe-specific intermediate and a different reaction mechanism.

1.6. CONCLUSIONS AND FUTURE OUTLOOK

By highlighting a variety of specific examples, this Account illustrates how the selection of a catalytic metal in the process of metalloenzyme evolution is based on factors other than optimizing catalysis alone. Some metal selections may be dictated by metal availability, as described in numerous examples of hydrolases. Alternate metal use, though surprising at first sight, is in fact meaningful from the standpoint of catalysis. Very similar metals, such as Mg^{2+} and Ca^{2+}, might have very different catalytic properties due to metal-induced large-scale structural effects. In other enzymes, such as the ARD/ARD' system, the metal steers the catalyzed reaction in different ways due exclusively to the electronic effects. In some cases, radical metal substitution is tolerated by an enzyme, e.g. redox active enzymes that adapt to use a metal with a very different reduction potential by tuning it through the nature of the

ligands, strain, and other structural factors. For instance, some bacteria evolved to be able to survive in Fe^{2+} starvation by utilizing Mn^{2+} instead. SOD normally uses Cu or Fe to recycle superoxide, but some SODs can use Ni or Mn by forcing the needed reduction potential on these metals through their unusual ligand environments. These evolutionary moves may have been forced at some point by scarcity of the more commonly used metals, or may have resulted from random mutations that enabled the organisms to expand their habitable environments. In summary, the discussion presented here further emphasizes that natural enzymes are not necessarily perfect catalysts, but rather optimized within a broader evolutionary context. The enzyme design community aims to prepare maximally efficient quasi-natural enzymes for catalysis, with few examples of enzymes used by pharmaceutical manufacturers.[76] Metalloenzyme design specifically is unprecedented and provides a largely unexplored frontier for chemists, where roles of metals is extremely important but characterization can be problematic. These natural limitations provide an exciting opportunity for research into the improvement of many catalytic properties *in vitro.*

1.7. REFERENCES

1 Stewart, L. J.; Bailey, S.; Bennett, B.; Charnock, J. M.; Garner, C. D.; McAlpine, A. S.: Dimethylsulfoxide reductase: an enzyme capable of catalysis with either molybdenum or tungsten at the active site. *J. Mol. Biol.* 2000, *299*, 593-600.

2 Banci, L.; Rosato, A.: Structural genomics of proteins involved in copper homeostasis. *Acc. Chem. Res.* 2003, *36*, 215-221.

3 Luk, E.; Jensen, L. T.; Culotta, V. C.: The many highways for intracellular trafficking of metals. *J. Biol. Inorg. Chem.* 2003, *8*, 803-809.

4 Irving, H.; Williams, R. J. P.: Order of stability of metal complexes. . *Nature* 1948, *162*, 746–747.

5 Fraústo da Silva, J. J. R.; Williams, R. J. P.: *The Biological Chemistry of the Elements* Oxford Univ. Press, 2001.

6 Waldron, K. J.; Robinson, N. J.: How do bacterial cells ensure that metalloproteins get the correct metal. *Nat. Rev. Microbiol.* 2009, *7*, 25-35.

7 Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. S.: *Biological Inorganic Chemistry: Structure and Reactivity*. ; University Science Books, 2007.

8 Hodawadekar, S. C.; Marmorstein, R.: Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism, and implications for effector design. *Oncogene* 2007, *26*, 5528-5540.

9 Haberland, M.; Montgomery, R. L.; Olson, E. N.: The many roles of histone deacetylases in development and physiology: implications for disease and therapy. . *Nat. Rev. Genet.* 2009, *10*, 32-42.

10 Gallinari, P.; di Marco, S.; Jones, P.; Pallaoro, M.; Steinkuhler, C.: HDACs, histone deacetylation, and gene transcription: from molecular biology to cancer therapeutics. *Cell Res.* 2007, *17*, 195-211.

11 Dowling, D. P.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W.: Structures of metal-substituted human histone deacetylase 8 provide mechanistic inferences on biological function. *Biochemistry* 2010, *49*, 5048-5056.

12 Gantt, S. L.; Gattis, S. G.; Fierke, C. A.: Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion. *Biochemistry* 2006, *45*, 6170-6178.

13 Kobayashi, M.; Shimizu, S.: Cobalt proteins. *Eur. J. Biochem.* 1999, *261*, 1-9.

14 Blakeley, R. L.; Hinds, J. A.; Kunze, H. E.; Webb, E. C.; Zerner, B.: Jack bean urease (EC 3.5.1.5). Demonstration of a carbamoyl-transfer reaction and inhibition by hydroxamic acids. . *Biochemistry* 1969, *8*, 1991-2000.

15 Montecucco, C.; Rappuoli, R.: Living dangerously: how Helicobacter pylori survives in the human stomach. *Nat. Rev. Mol. Cell Biol.* 2001, *2*, 457-466.

16 Dixon, N. E.; Gazzola, C.; Blackeley, R.; Zerner, B.: Jack bean urease (EC 3.5.1.5). A metalloenzyme. A simple biological role for nickel? . *J. Am. Chem. Soc.* 1975, *97*, 4131-4132

17 Maroney, A. J.; Ciurli, S.: Nonredox Nickel Enzymes. . *Chem. Rev.* 2014, *114*, 4206–4228.

18 Page, M. I.: *The Chemistry of β-Lactams*; Blackie: London, 1992.

19 Estiu, G.; Suárez, D.; Merz, K. M.: Quantum mechanical and molecular dynamics simulations of ureases and Zn β-lactamases. *J. Comput. Chem.* 2006, *27*, 1240– 1262.

20 Valdez, C. E.; Alexandrova, A. N.: Why Urease Is a Di-Nickel Enzyme whereas the CcrA β-Lactamase Is a Di-Zinc Enzyme. *J. Phys. Chem. B* 2012, *116*, 10649–10656.

21 Fox, J.G.; Correa, P.; Taylor, N.S.; Lee, A.; Otto, G.; Murphy, J.C.; Rose, R. Helicobacter mustelae-associated gastritis in ferrets: an animal model of Helicobacter pylori gastritis in humans. *Gastroenterology* 1990, *99*, 352–361.

22 Solnick, J.; Schauer, D. B.: Emergence of diverse Helicobacter species in the pathogenesis of gastric and enterohepatic diseases. . *Clin. Microbiol. Rev.* 2001, *14*, 59–97.

23 Carter, E. L.; Tronrud, D. E.; Taber, S. R.; Karplus, P. A.; Hausinger, R. P.: Iron-containing urease in a pathogenic bacterium. . *Proc. Natl. Acad. Sci.* 2011, *108*, 13095-13099.

24 Fenton, H. J. H.: Oxidation of tartaric acid in presence of iron. . *J. Chem. Soc., Trans.* 1984, *65*, 899–911.

25 Loew, O.: A new enzyme of general occurrence in organisms. . *Science* 1900, *11*, 701–702.

26 Sobota, J. M.; Imlay, J. A.: Iron enzyme ribulose-5-phosphate 3-epimerase in Escherichia coli is rapidly damaged by hydrogen peroxide but can be protected by manganese. . *Proc. Natl. Acad. Sci.* 2011, *108*, 5402-5407.

27 Martin, J. E.; Imlay, J. A.: The alternative aerobic ribonucleotide reductase of Escherichia coli, NdrEF, is a manganese-dependent enzyme that enables cell replication during periods of iron starvation. *Mol. Microbiol.* 2011, *80*, 319-334.

28 Reichard, P.: From RNA to DNA, why so many ribonucleotide reductases? *Science* 1993, *260*, 1773-1777.

29 Stubbe, J.: Ribonucleotide reductases in the twenty-first century. *Proc. Natl. Acad. Sci.* 1998, *95*, 2723-2724.

30 Boal, A. K.; Cotruvo, J. A., Jr.; Stubbe, J.; Rosentzweig, A. C.: Structural Basis for Activation of Class Ib Ribonucleotide Reductase. *Science* 2010, *329*, 1526-1530.

31 Makhlynets, O.; Boal, A. K.; Rhodes, D. V.; Kitten, T.; Rosenzweig, A. C.; Stubbe, J.: Streptococcus sanguinis Class Ib Ribonucleotide Reductase. High activity with both iron and manganese cofactors and structural insights. *J. Biol. Chem.* 2014, *28*, 6259-6272.

32 Boal, A. K.; Cotruvo, J. A.; Stubbe, J.; Rosenzweig, A. C.: The Dimanganese(II) Site of Baccilus subtilis Class Ib Ribonucleotide Reductase. *Biochemistry* 2012, *51*, 3861-3871.

33 Dassama, L. M. K.; Boal, A. K.; Krebs, C.; Rosenzweig, A. C.; Bollinger, J. M., Jr: Evidence that the b subunit of chlamydia trachomatic ribonucleotide reductase is active with the manganase ion of its manganese(IV)/iroon(III) cofactor in site 1. *J. Am. Chem. Soc.* 2012, *134*, 2520-2523.

34 Cammack, R.: Iron-sulfur clusters in enzymes: themes and variations. *Adv. Inorg. Chem.* 1992, *38*, 281–322.

35 Gamiz-Hernandez, A. P.; Galstyan, A. S.; Knapp, E. W.: Understanding rubredoxin redox potentials: role of H-bonds on model complexes. *J. Chem. Theory Comput.* 2009, *5*, 2898–2908.

36 Gamiz-Hernandez, A. P.; Kieseritzky, G.; Knapp, E. W.: Rubredoxin function: redox behavior from electrostatics. *J. Chem. Theory Comput.* 2011, *7*, 742–752.

37 Beyer, W.; Imlay, J.; Fridovich, I.: *Prog. Nucleic Acid Res. Mol. Biol.* 1991, *40*, 221-253.

38 Wallace, D. C.: *Science* 1992, *256*, 628-632.

39 Sheng, Y.; Abreu, I. A.; Cabelli, D. E.; Maroney, M. J.; Miller, A.; Teixeria, M.; Valentine, J. S.: Superoxide Dismutases and Superoxide Reductases. . *Chem. Rev.* 2014, *114*, 3854–3918.

40 Abreu, I. A.; Cabelli, D. E.: Superoxide dismutases: A review of the metal-associated mechanistic variations. *Biochim. Biophys. Acta* 2010, *1804*, 263-274.

41 Aguirre, J. D.; Culotta, V. C.: Battles with iron: manganese in oxidative stress protection. *J. Biol. Chem.* 2012, *287*, 13451-13548.

42 Kim, F. J.; Kim, H. P.; Hah, Y. C.; Roe, J. H.: Differential expression of superoxide dismutases containing Ni and Fe/Zn in Streptomyces coelicolor. *Eur. J. Biochem.* 1996, *241*, 178-185.

43 Cabelli, D. E.; Riley, D.; Rodriguez, J. A.; Valentine, J. S.; Zhu, H.: In *Biomimetic Oxidations Catalyzed by Transition Metal Complexes*; Meunier, B., Ed.; Imperial College Press: London, 2000.

44 Fiedler, A. T.; Bryngelson, P. A.; Maroney, M. J.; Brunold, T. C.: Spectroscopic and Computational Studies of Ni Superoxide Dismutase: Electronic Structure Contributions to Enzymatic Function. *J. Am. Chem. Soc.* 2005, *127*, 5449-5462.

45 Barondeau, D. P.; Kassmann, C. J.; Bruns, C. K.; Tainer, J. A.; Getzoff, E. D.: Nickel Superoxide Dismutase Structure and Mechanism. *Biochemistry* 2004, *43*, 8038-8047.

46 Wuerges, J.; Lee, J.; Yim, Y.; Yim, H.; Kang, S.; Carugo, K. D.: Crystal structure of nickel containing superoxide dismutase reveals another type of active site. *Proc. Natl. Acad. Sci.* 2004, *101*, 8569-8574.

47 Grimaldi, S.; Schoepp-Cothenet, B.; Ceccaldi, P.; Guigliarelli, B.; Magalon, A.: The prokaryotic Mo/W-bisPGD enzymes family: a catalytic workhorse in bioenergetic. . *Biochimica et Biophysica Acta* 2013, *1827*, 1048–1085.

48 Schink, B.: Fermentation of acetylene by an obligate anaerobe, Pelobacter acetylenicus sp. nov. *Arch. Microbiol.* 1985, *142*, 295–301.

49 L'vov, N. P.; Nosikov, A. N.; Antipov, A. N.: Tungsten-containing enzymes. Biochemistry. *Biokhimiia* 2002, *67*, 196–200.

50 Johnson, M. K.; Rees, D. C.; Adams, M. W. W.: Tungstoenzymes. *Chem. Rev.* 1996, *96*, 2817-2839.

51 Courcy, B.; De Pedersen, L. G.; Parisel, O.; Gresh, N.; Silvi, B.; Pilme, J.: Understanding Selectivity of Hard and Soft Metal Cations within Biological Systems Using the Subvalence Concept. 1. Application to Blood Coagulation: Direct Cation-Protein Electronic Effects versus Indirect. *J. Chem. Theor. Comput.* 2010, *6*, 1048-1063.

52 Williams, R. J.: The natural selection of the chemical elements. *Cel. Mol. Life Sci.* 1997, *53*, 816–829.

53 Schmitz, R. A.; Albracht, S. P. J.; Thauer, R. K.: *Eur. J. Biochem.* 1992, *209*, 1013-1018.

54 Buc, J.; Santini, C. L.; Giordani, R.; Czjzek, M.; Wu, L. F.; Giordano, G.: *Mol. Microbiol.* 1999, *32*, 159-168.

55 Peeraer, Y.; Rabijns, A.; Collet, J. F.; Van Schaftingen, E.; De Ranter, C.: How calcium inhibits the magnesium-dependent enzyme human phosphoserine phosphatase? . *Eur. J. Biochem.* 2004, *271*, 3421-3427.

56 Yang, L.; Liao, R.-Z.; Ding, W.-J.; Liu, K.; Yu, J.-G.; Liu, R.-Z.: Why calcium inhibits magnesium-dependent enzyme phosphoserine phosphatase? A theoretical study. . *Theor. Chem. Acc.* 2012, *131*, 1275.

57 Dudev, T.; im, C.: Principles governing Mg, Ca, and Zn binding and selectivity in proteins. . *Chem. Rev.* 2003, *103*, 773-788.

58 Gifford, J. L.; Walsh, M. P.; Vogel, H. J.: Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. . *Biochem. J.* 2007, *405*, 199-221.

59 De Courcy, B.; Pedersen, L. G.; Parisel, O.; Gresh, N.; Silvi, B.; Pilme, J.: Understanding selectivity of hard and soft metal cations within biological systems using the subvalence concept. I. Application to blood coagulation: direct cation-protein electronic effects vs. indirect interactions through water networks. *J. Chem. Theor. Comput.* 2010, *12*, 1048-1063.

60 Männistö, P. T.; Kaakkola, S.: Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. . *Pharmacol. Rev.* 1999, *51*, 593–628.

61 Guldberg, H. C.; Marsden, C. A.: Catechol-O-methyl transferase: pharmacological aspects and physiological role. *Pharmacol. Rev.* 1975, *27*, 135-140.

62 Lotta, T.; Vidgren, J.; Tilgmann, C.; Ulmanen, I.; Melen, K.: Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and

description of the thermolabile variant of the enzyme. *Biochemistry* 1995, *34*, 4202–4210.

63 Rutherford, K.; Trong, I. L.; Stenkamp, R.; Parson, W.: Crystal structures of human 108V and 108M catechol O-methyltransferase. *J. Mol. Biol.* 2008, *380*, 120–130.

64 Vidgren, J.; Svensson, L. A.; Liljas, A.: Crystal structure of catechol O-methyltransferase. *Nature* 1994, *368*, 354–358.

65 Axelrod, J.; Tomchick, R.: Enzymatic O-methylation of epinephrine and other catechols. *J. Biol. Chem.* 1958, *233*, 702–705.

66 Sparta, M.; Alexandrova, A. N.: How metal substitution affects the enzymatic activity of catechol-O-methyl transferse. *PLoS ONE* 2012, *7*, e47172.

67 Martin-Diaconescu, V.; Maroney, M. J.: In *Comprehensive Inorganic Chemistry II*; Reedijk, J., Poppelmeier, K. R., Eds.; Elsevier: Oxford, 2013.

68 Dai, Y.; Wensink, P. C.; Abeles, R. H.: One protein, two enzymes. *J. Biol. Chem.* 1999, *274*, 1193–1195

69 Pochapsky, T. C.; Ju, T.; Dang, M.: Nickel in acireductone dioxygenase. In *Nickel and Its Surprising Impact in Nature*; Sigel, A., Sigel, H., Sigel, R. K. O., Eds.; John Wiley & Sons, Ltd.: Chichester, 2007; Vol. 2.

70 Ye, S.; Riplinger, C.; Hansen, A.; Krebs, C.; Bollinger, J. M. J.; Neese, F.: Electronic structure analysis of the oxygen-activation mechanism by FeII- and α-ketoglutarate (αKG)-dependent dioxygenases. . *Chem. Eur. J.* 2012, *18*, 6555–6567.

71 Christian, J. G.; Ye, S.; Neese, F.: Oxygen activation in extradiol catecholate dioxygenases—a density functional study. . *Chem. Sci.* 2012, *3*, 1600–1611.

72 Dai, Y.; Pochapsky, T. C.; Abeles, H. R.: Mechanistic studies of two dioxygenases in the methionine salvage pathway of Klebsiella pneumonia. *Biochemistry* 2001, *40*, 6379–6387.

73 Allpress, C. J.; Grubel, K.; Szajna-fuller, E.; Arif, M. A.; Berreau, L.: Regioselective aliphatic carbon–carbon bond cleavage by a model system of relevance to iron-containing acireductone dioxygenase. *J. Am. Chem. Soc.* 2013, *135*, 659–668.

74 Sparta, M.; Valdez, C. E.; Alexandrova, A. N.: Metal-Dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway. *J. Mol. Biol.* 2013, *245*, 3007-3018.

75 Valdez, C. E.; Gallup, N. M.; Alexandrova, A. N.: Co^{2+} Acireductone Dioxygenase: Fe^{2+} Mechanism, Ni^{2+} Mechanism, or Something Else? *Chem. Phys. Lett.* 2014, *604*, 77-82.

76 Thayer, A. M.: Biocatalysts. *C&E News* 2012, *90*, 13.

CHAPTER 2

Computational Treatment of Metalloproteins

2.1. INTRODUCTION

Metalloproteins present many challenges when it comes to computational modeling.[1,2] The strong Coulombic forces of metal cations and charged amino acids induce considerable perturbations to protein tertiary structure. For example, installation or removal of metal cations can yield large conformational changes.[3] In addition, metalloproteins commonly perform electron and proton transfers - a mechanism beyond the most common ways to describe protein dynamics today, i.e., classically. Metalloproteins are large dynamic molecules whose elements operate synergistically and on different time scales during enzymatic catalysis. The challenges in computationally modeling metalloproteins stem from the need to simultaneously describe the complex metal coordination site(s) and efficiently sample the protein backbone. Motions of the protein backbone and smaller metal site(s) are coupled, so the remaining challenge is balancing on a time scale relevant protein machinery: picoseconds, milliseconds, and beyond. Because of the dynamic nature of proteins, sampling and assessment of free energies or their changes are critical, and it is not a simple task. Many dream applications that stand to benefit from modeling *in silico* include engineering metalloproteins and/or their substrates to enrich catalysis and inhibition. Together with traditional mechanistic studies (especially for proteins that perform electron and proton transfer), these applications rely on an adequate multi-scale methodology. Computational treatment of metalloproteins is in a crucial developmental stage, and we believe some of the most exciting applications are just over the horizon, with more prospects to follow.

Sampling of protein structure typically employs classical force field based methods: molecular dynamics (MD) and Monte Carlo statistical mechanics (MC). Much work has been done in developing more efficient sampling of proteins via MD[4,5] and MC via the Metropolis-Hastings algorithm.[6,7] Both methods lend themselves well to parallelization, and serve faithfully in areas of computational biophysics such as drug design.[8] However, in modeling metalloproteins, the classical force fields usually are parameterized to handle only a specific

metal coordination, which then has no freedom to change (by acquiring or losing ligands as in a catalyzed reaction or in response to reduction/oxidation of the metal) in the course of a simulation. Such changes would not be possible within the harmonic oscillator-like description. This limitation is also characteristic of the Empirical Valence Bond (EVB) approach.[2] Merz et. al. are currently developing parameters for 24 divalent metal ions, however they are currently limited to calculating hydration free energies of metal-water clusters within the constraints of a harmonic oscillator-like description of the metal-ligand bonds.[9] On the other hand, fragmentation-based techniques - methods that carefully partition a large system of interest into smaller fragments suitable for *ab initio* QM calculations, with the intent to give the same energy and properties of the complete large system - show promise in expediting the computational time required with massive parallelism for QM calculations.[10] The literature is rich in such techniques but some notable examples for biological systems of interest in this paper such as fragment molecular orbital (FMO) methods. A full review on these methods can be found in other works[11, 12,13] However, the computational power needed for these methods is far from realistic for metalloproteins containing hundreds of residues.

Hybrid methodologies which bridge the quantum mechanical description of the metal-ligand interactions with classical sampling of protein structure are very promising. There is a rich history behind these methods. The quantum mechanical/molecular mechanical (QM/MM) approach was first introduced by Levitt and Warshel[14] to calculate the energies of intermediate states in enzymatic reactions. QM/MM methods have undergone remarkable advancement, and now it is possible to study reaction pathways in large systems, such as solvated enzymes.[15,16] Car-Parrinello MD (CPMD)[17,18] techniques account for electronic and nuclear motion in an approximate manner and have substantial success. Presently, CPMD can describe the time evolution of molecular systems with ~100 heavy atoms for ~100 ps without relying on a force field. Nonetheless, hybrid molecular mechanics/CPMD (MM/CPMD)

schemes,[17,19] that do employ a force field for part of the simulation, have been successfully applied to several different metal-containing systems. A few examples include determination of the catalytic role of Zn^{2+} and Mg^{2+} in metalloenzymes,[20,21,22] ligand-DNA interactions with transition-metal-dependent anticancer drugs,[23,24] and properties of electron transport proteins.[25] Fragmentation-based methods have been recently extended into the QM/MM formalism, in particular for simple polymers[26], biopolymers[27,28] and zeolites.[28] Approaches that contain intergrated QM with MM, such as ONIOM,[29] which compute entire system properties at a lower level of theory and add higher level components to chosen areas. For large metal-containing proteins, these methods are not developed enough yet.

A primary issue with even the most efficient QM/MM methods is still speed. Conformational sampling is expensive, restricting the use of QM/MM methods in docking drug candidates, protein structure design, and specific metal-dependent functions. Towards reducing the computational cost, semi-empirical methods such as PM7 and PDDG have taken the place of QM in QM/MM in pioneering work by Gao and Truhlar,[30] and in continued expansion.[31,32,33] However, the reliability of semi-empirical methods is known to be more limited. For sensitive quantities, such as reduction potentials of metals in proteins, one needs to advance beyond semi-empirics into a quality quantum mechanical description.[34,35,36] A powerful approach, specific for the description for electron transfer in metalloproteins, was designed by Voth,[37] which skips the QM part all together. Voth instead coupled a coarse-grained version of the classical mechanical description of the protein to quantum dynamics describing electron transfer. This simulation yields electron transfer between FeS clusters [FeFe]-hydrogenase. Despite its indisputable beauty, the method cannot describe the FeS clusters themselves and address such questions as what electronic states are involved in accepting and releasing electrons, and how the cluster-coordination environment impacts those results. In contrast to pure electron transfer, which employs cooperation from experimental optical spectroscopy

methods, computation is necessary in studying Proton Coupled Electron Transfer (PCET) due to experimental limitations in capturing proton translocation. PCET is commonly considered a suitable reaction mechanism for proteins and metalloproteins such as cytochrome oxidase. Many theoretical studies on PCET have been done by Hammes-Schiffer,[38,39] including capturing reaction mechanisms and reproducing reaction rates in the metalloenzyme soybean lipoxygenase, via vibronically non-adiabatic formulation for PCET reactions in solution and proteins.[40]

In this feature article we discuss selected methodologies on the forefront of this field and how sampling, entropy, and description of the metal centers are currently addressed. In particular, we believe that our QM/DMD method (DMD standing for Discrete Molecular Dynamics[41]) is among the most efficient for sampling of metalloprotein structure.[42] Further, we will present specialized techniques for applications such as metalloprotein design, which demands a robust sampling strategy. We illustrate our discussion with computational studies that showcase the power of our methods, and a look forward to where we envision they could be successfully applied.

2.2. METHODS FOR METALLOPROTEIN TREATMENT

2.2.1. QUANTUM MECHANICS/DISCRETE MOLECULAR DYNAMICS (QM/DMD)

QM/DMD is a metalloprotein sampling engine.[42] It operates through an iterative scheme between QM and DMD machineries (Figure 2.1). DMD is a flavor of MD that approximates the continuous interaction potentials in classical MD with square-well potentials, (Figure 2.1A) course-graining the potentials and overall reducing the number of calculations needed. Due to these *discretized* potentials, DMD is driven by collision events rather than physical forces as in traditional MM and MD. Therefore, the user saves a tremendous amount of time with DMD by solving ballistic equations of motions rather than Newtonian equations of motions (Figure 2.1A).

Complete details can be found in earlier works.[43] QM/DMD operates in the following scheme: the simulation begins with a DMD simulation of the entire protein keeping the metal and atoms directly bound to the metal frozen, and a few other constraints possibly being included (Figure 2.1A). This saves one from the need to parameterize the classical force field for the metal-ligand interactions. Following DMD, a structure is selected from the trajectory representative of the ensemble. A larger, chemically meaningful QM region is extracted from the protein for optimization at the QM level. Most of our work employs Density Functional Theory (DFT) for the QM management area due to the size and transition metal species under consideration. However, any *ab initio* QM formalism, whether it be density of wave-function based methods, can be chosen for the purely QM region. This region can be something normally used for a QM mechanistic study on an enzyme using a cluster model (Figure 2.1A). During the relaxation, the structural changes in the protein predicted by DMD can influence the metal coordination. The relaxed part is then reinstalled back into the protein, the QM-DMD boundary shrinks back to going right around the metal center or centers, and the small region inside the boundary is again fixed. The simulation continues with DMD. During the DMD stage, the updated geometry of the QM region has a chance to influence the rest of the protein. The simulation proceeds in the analogous iterative manner to convergence (illustrated in Figure 2.1B). The described "breathing" QM-DMD boundary is a simple solution for the communication between the two simulation machineries. Then last but not the least, DMD is highly suitable for being incorporated into the hybrid method; DMD does not require calculations of forces, and operates on discontinuous potentials, and as a result is insensitive facing a discontinuity of the potential at the QM-DMD boundary. To the best of our knowledge, QM/DMD provides record metalloprotein sampling speeds for simulations done on CPU.

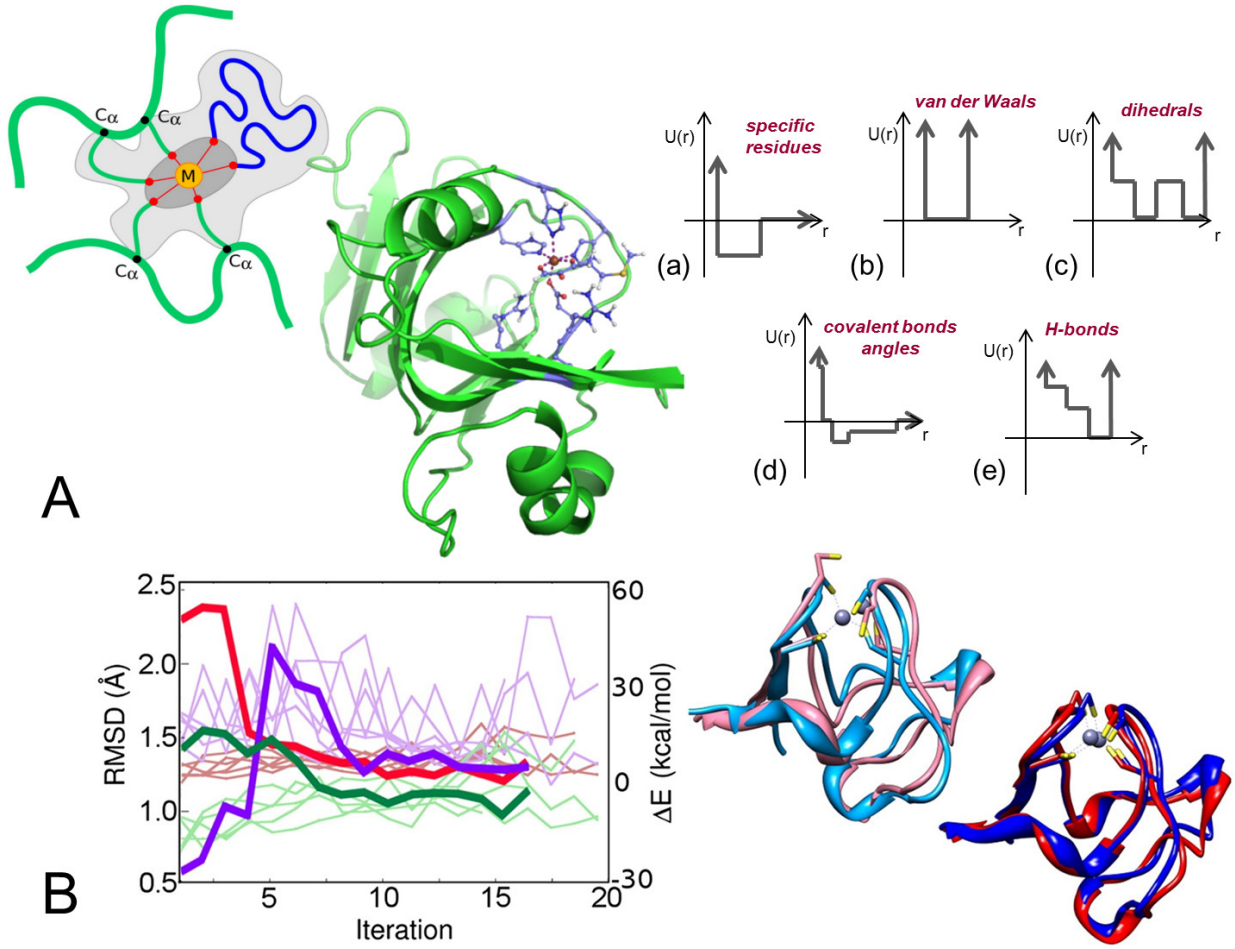


Figure 2.1. QM/DMD. (A) The unique feature of QM/DMD is the "breathing" QM-DMD boundary: the light grey area on the scheme is managed in both the QM and DMD regions. The alpha carbons of the "breathing" residues are held frozen during the pure QM phase, and the atoms directly coordinated to the metal (red circles) are frozen during the pure DMD phase. The dark grey region is managed exclusively by QM. A real example of the separation into the DMD-only and QM-DMD regions is shown by the green/purple protein (the ARD system). A few step function potentials in DMD are shown. Each is defined with an example in parenthesis: (a) Hard-shell interaction potentials (hard-sphere radius, attractive potential well), (b) single-infinite square well (covalent bonds), (c) dihedrals (peptide bonds), (d) Discretized van der Waals (solvation non-bonded) and (e) hydrogen-bonding auxiliary distance potential function. (B) The graph shows a representative QM/DMD simulation with converged data such as RMSD (Å) from the x-ray structure (light green lines), the QM energies (pink lines) and the DMD energies (light purple lines). The thick lines illustrate the fast return of a distorted wild type structure of rubredoxin to equilibrium. The structures compare the overlay of the x-ray structure (light blue) and distorted structure (pink) with a representative QM/DMD equilibrated structures starting from x-ray (dark blue) and the distorted structure (red). Adapted with permission from Reference 42.

QM/DMD has the ability to recapitulate native protein structure from native and distorted ones,[42,44] and provide finer structural details of the active site at the level needed for, for example, subsequent mechanistic studies.[45] Large-scale motions of protein parts are also captured efficiently.[46] For the equilibrated QM/DMD structures, one can apply most rigorous QM methods to obtain such sensitive properties as changes in the reduction potentials of the metal, or barriers of catalyzed reactions.[42] Naturally, since sampling is done on the entire protein while metal coordination is purely in the QM management, events such as ligand attachment[44] or detachment[47] are easily captured. This is often important in mechanistic studies. We are particularly excited about the demonstrated ability of QM/DMD to predict how the structure of a metalloprotein would react to the changes in protein sequence or the nature of the bound metal,[42,44,45,48] removal of one of the metal cations,[47] or binding a metal to a protein that originally did not contain a metal. These capabilities make QM/DMD a good platform for metalloprotein design, as is currently being tested. (manuscript in preparation) Undoubtedly, metalloprotein design is a very interesting and exciting goal pursued in the field.[49,50,51,52,53,54] It is also known that in the design of specific, buried binding pockets, as opposed to surface- or interface-exposed binding sites, sampling the protein backbone dramatically expands the design repertoire.[55,56] Hence, the idea is to utilize QM/DMD sampling in design, enabling a true design of buried active sites containing transition metals, perhaps for the first time.

2.2.2. APPROACH TO METALLOENZYME DESIGN: GENERAL OUTLINE

The design process we aim for is based on the “inside-out” protocol developed by Baker *et al*.[57] The original idea was to design an arrangement of amino acids around the rate-limiting transition state of the catalyzed reaction such that it would be stabilized selectively from the reactants and then incorporate this design into a pocket of an existing protein scaffold. For metalloenzymes, this process acquires an additional dimension: the design of the electronic

structure of the metal center or centers. Indeed, the electronic structure of the metal defines catalysis, and it is what makes metalloenzymes such great catalysts, often accomplishing the most dramatic reactions in a single step. Design here means tuning Lewis acidity, reduction potential, or order and population of d-atomic orbitals (d-AOs) on the metal, determined by its ligand environment. Thus, *ab initio* calculations of the catalytic transition metal complex are at the root of the computational design process. It is then needed to find proteins that can arrange for such metal coordination, which may be nontrivial since only a few (natural) amino acids and the backbone N and C=O groups can be the ligands. After that, the stabilization of the transition state of interest should contain continuous QM treatment of the metal for its optimal description, and sampling of the protein backbone. Hence, we put together a set of tools enabling every step in this process.

2.2.2.1. EREBUS

Erebus is one of such tools we found necessary to develop to assist in the design process. To increase the chance of a designed protein to fold into a correct structure, it is safest to use an existing robust protein scaffold that already contains a metal. If desired, it generally can be arranged for the metal to be replaced with another one *in vitro*, or *de novo* synthesis of the protein can be done with the desired metal being supplied. Mutations in the active site are typically kept to a minimum in the design, again to preserve the native fold. *Erebus*[58] is a data mining tool for searching through the Protein Data Bank (PDB) for substructures in proteins. For the design application, *Erebus* can look for arrangements of the metal, its ligands, and possibly other amino acids near the metal, that are close to what is desired. The idea is that one can use *ab initio* calculations to construct a potentially catalytic metal-amino acid residues complex and then look for this arrangement in the full set of available protein structures. For example, the calculated complex shown in Figure 2.2A was found in the

natural protein in Figure 2.2B. The *Erebus* search is based on sub-graph isomorphism algorithm with a user-defined allowed structural uncertainty, as illustrated in Fig. 2.2C. It browses through the entire PDB in an efficient manner (on the order of under 1 hour for a single search). Depending on the allowed uncertainty, many matches can be uncovered.

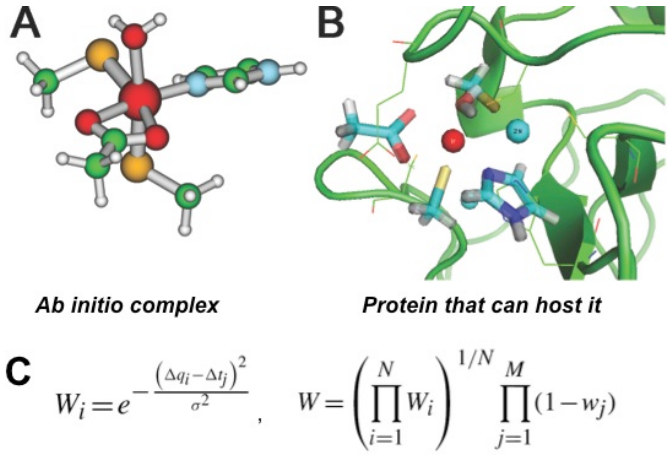


Figure 2.2. Illustration of *Erebus*. (A) A metal-ligand complex calculated using DFT. Hypothetically, this complex is desired for catalysis inside a protein pocket. (B) The active site of di-Zn β-lactamase found by *Erebus* to contain the Asp, two Cys, His residues, and a water molecule in the geometry close to the one desired from (A). Notice that it is already a Zn-binding site. (C) W_i represents the accuracy of a match to the input scaffold, more specifically W_i represents how much the position of every heavy atom *i* matches that in the designed structure, subject to user-defined uncertainty σ, and the resultant overall weight of the structure, *W*, used by *Erebus* to judge the overall quality of the match.

2.2.2.2. MULTI-SCALE DESIGN TOOL

The last necessary tool in our arsenal should assess the change in protein free energy upon mutagenesis, to simultaneously optimize for the transition state binding and protein stability. These evaluations are also important beyond protein design, for a wide variety of research areas concerning structure/function relationships of proteins. However, assessing the effect of mutagenesis on protein stability is difficult since it requires addressing the complex nature of many interactions in a secondary structure. With metalloproteins, the additional

complexity arises because every bit of repacking in the binding site can have a serious influence on the delicate electronic structure of the metal through both bonded (shifting ligands) and non-bonded (polarization, electrostatics) interactions. As much as nature fine-tunes the metal to have just the right Lewis acidity, or just the right reduction potential for the reaction it catalyzes,[59] we too must be mindful in a similar way when designing.

Intuition can provide initial guesses for useful mutations to introduce into a protein. However, they only take us so far, where evaluating whether one or two seemingly innocent mutations could cause the entire protein to denature becomes impossible. Additionally, changes beyond a few catalytic residues might be beneficial for protein stability and overall packing of the active site, and usually human intuition is completely inept at this task. Computation is required.

One of the most popular and powerful design tools is classical force field-based *Rosetta*.[60,61] It samples through the discrete rotameric space of amino acids as well as their chemical nature, in an effort to repack and redesign binding sites for catalysis/stability and has been successful in several cases.[62,63,64,65] A competing tool, *Eris*, is a protein stability prediction software that utilizes the slightly-coarse grained force field, *Medusa*.[66] *Eris* measures protein stability when a mutation is introduced by calculating the ΔΔG of the mutation, where $\Delta\Delta G=\Delta G_{mutant}-\Delta G_{wild\text{-}type}$.[55,67] It is beneficial that *Eris* utilizes some form of protein backbone flexibility, as it was shown to be helpful in design,[55,67] and is expected to be of a major importance for metalloproteins. However, none of the existing methods in their current forms capture the structural impact of metals ions on the protein, and in turn allow the protein to impact the electronic structure of the metal.

We have combined our QM/DMD sampling software[42] with *Eris*, creating a recipe for gauging metalloprotein stability induced by mutations called *Eris*-QM/DMD. The *Eris*-QM/DMD method begins with a QM/DMD simulation of the system the user is interested in mutating.

After each iteration of QM/DMD, the protein structure is mutated to the desired residue(s), undergoes an additional sampling step and the ΔΔG of the mutation is evaluated. This method has demonstrated proof-of-principle native sequence recovery for two structurally different proteins, where each protein is mutated in several ways, and *Eris*-QM/DMD is able to recapitulate all native sequences of each protein except for one mutation, which is undergoing experimental validation. (manuscript in preparation)

2.2.2.3. ENTROPY EVALUATION

Entropy evaluation is a separate and complicated issue. QM/DMD sampling predicts equilibrium protein structure and calculations of energies. Entropic effects are either ignored (for example, in certain mechanistic studies where it is assumed that entropy does not change significantly as the system progresses from reactants to the transition states to products), or is included in an empirical way (as in *Eris* and *Eris-QM/DMD*). However, a quantitative evaluation of configurational entropy is desired, especially for sensitive applications such as drug design, where every fraction of a kcal/mol in the affinity matters greatly. Entropy evaluation presents a fundamental hurdle in the field - protein structures are routinely solved as static crystal structures and fewer experimental studies are able to extract a structural ensemble. In the case of substrate binding, it is commonly accepted that there is correlation between ΔH and $T\Delta S$, however, the relevant free energy in protein and substrate binding is not correlated with either of these individual descriptors (Figure 2.3).[68]

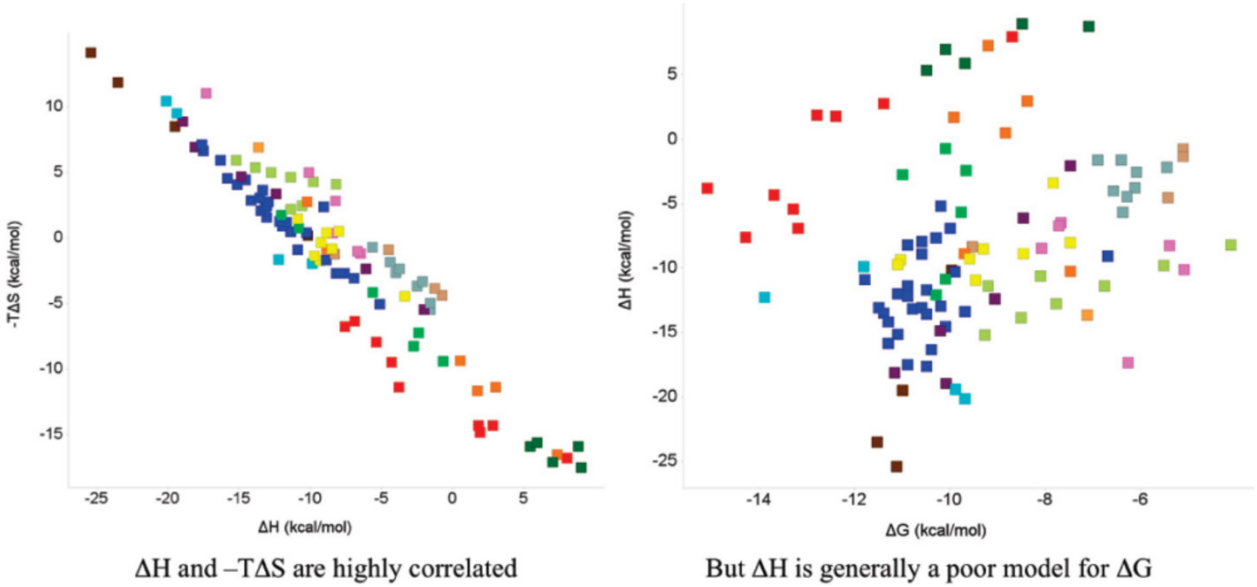


Figure 2.3. Analysis of the experimental binding thermodynamics for approximately 100 protein-ligand complexes. Reprinted with permission from Reference 68.

Computational methods allow us to follow precise positions of not only tertiary and secondary structures but atomic positions, and these motions are coupled with corresponding energies. Provided we have a crystal structure or another good "guess" of a starting point, these motions are accessible. (However, getting to that starting point *in silico* remains an active field in itself[69]). Normal mode analysis, for example, probes local curvature near a stationary point via the Hessian and by using quantum mechanical modeling one can deduce a related entropy component to sum across all modes. Karplus and Kushick laid down groundwork for quasi-harmonic analysis using internal coordinates[70] which probes the global extent of configurational space accessible to a system at a given temperature. By evaluating a mass-weighted covariance matrix from MD simulation and diagonalizing, the resulting quasi-harmonic frequencies can be used with the QM expression for entropy of a harmonic oscillator. This approach was later reformulated by Andricioaei and Karplus to work in Cartesian coordinates.[71] Similar methods exist, such as in calculating an "upper" limit to entropy by use of convergence

matrices.[72] These methods are common in simulation packages together with MD trajectories, however, are often more useful for qualitative results than quantitative comparisons of entropy.

Consequently, free energy is left as the direct target of computation, rather than handling energy and entropy separately. The need for properly handling contributors including roles of disorder, fluctuations, protein dynamics, and multiple pathways in reaction dynamics is crucial and solutions are developing rapidly. Free Energy Perturbation (FEP) is one technique which can handle direct calculation of free energy changes, however, is very expensive especially if done in QM/MM settings. Umbrella sampling can be used to include entropic effects in 1-D free energy profiles and 2-D energy landscapes, and Gao, Truhlar, and others use ensemble averaging[73] to include entropic effects in free energy computation.[74,75,76,77,78,79,80,81]

As discussed, DMD efficiently samples configurational space at the atomic level. Bolstered by the efficient and discrete nature of molecular dynamics, Schofield et al. quantitatively measured the free energy and entropy of folding via probabilities of structures defined by their sharp non-covalent bond networks.[82] Basically, the function of the potential in DMD allows for a very straightforward definition of a conformational state: every well is either populated or it is not. Then, entropy evaluation becomes accessible. Schofield showed this on a very simple system (short polypeptide) and using a simpler DMD potential. Extending this idea to larger systems of interest and a more realist form of DMD is the subject of an ongoing implementation based on QM/DMD, in our lab. Indeed, the sampling of DMD combined with any quantitative method to extract free energy or entropy would provide the missing link between binding free energies and already-robust energy calculations.

With these tools, difficult questions of interest to the metalloenyzme community have been answered and our efforts to tackle more difficult and multi-faceted problems, such as the design of artificial metalloenzymes and inhibitors of them, continue to evolve. We now describe

a few successful applications of our methods, and in the end we will outline what is still missing or being under construction, and what future applications are becoming within reach.

2.3. EXAMPLES OF APPLICATIONS

2.3.1. LARGE-SCALE MOTION IMPORTANT FOR METALLOENZYME CATALYSIS

The need for adequate sampling for the protein structure is occasionally unnecessary, as the large, structural equilibration does not influence active site chemistry. However, in many instances backbone motion induces conformational changes around or far away from the active site. These structural motions could close or open channels for substrates to flow through or arrange residues at positions needed for catalysis. One such example of the latter is acireductione dioxygenases (ARD).

ARD is an enzyme that catalyzes two different oxidation reactions, depending solely on whether Fe^{2+} or Ni^{2+} is bound to the protein.[83,84,85] Fe-dependent ARD' that recycles methionine in the methionine salvage pathway by oxidizing the substrate, 1,2-dihydroxy-3-keto-5-(methylthio)pentene (acireductone), into two products: the α-keto acid precursor of methionine and formate. Ni-dependent ARD instead oxidizes acireductone into three products, methylthiopropionate, CO, and formate, and provides a shunt out of the methionine salvage cycle.[2] Interestingly, interconversion between the Fe and Ni forms of ARD is relatively simple as the protein has micromolar affinity for both metals (K_d < 0.4 μM for Fe and K_d < 0.1 μM for Ni). The long-standing hypothesis for why Fe and Ni catalyze different oxidation reactions is due exclusively to the coordination mode of the substrate to the metal center (Figure 2.4). It was proposed that a large conformation change in a nearby protein loop facilitates this different binding together with the metal replacement.[86] However, it is unclear why two divalent metals of similar radii would bind the substrate in dramatically different ways, and especially also cause

large conformational changes in the protein. Indeed, this view has been challenged by recent experiments and our QM/DMD computational studies, and a new mechanism was proposed.[37,69]

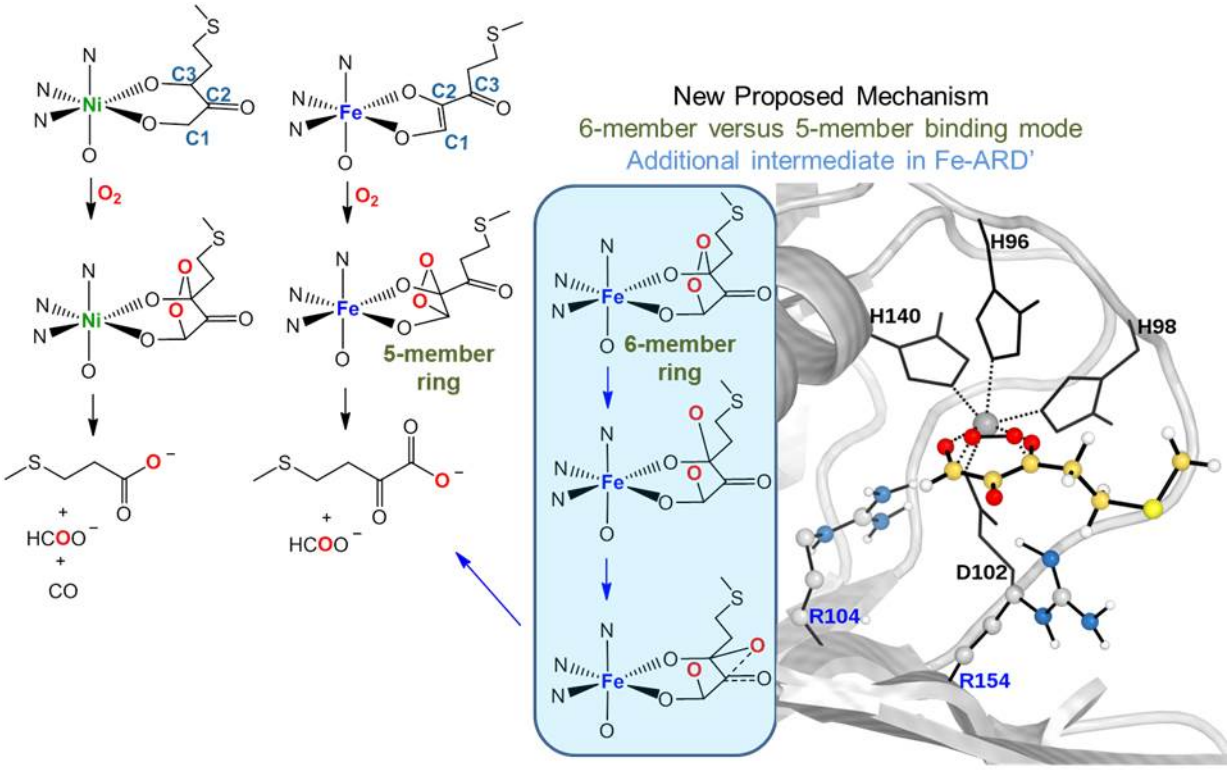


Figure 2.4. The old hypothesis on product differentiation of Fe-ARD' and Ni-ARD was disproved through our QM/DMD simulations coupled with DFT mechanistic studies and shown to stem purely from electronic structure of the metal, not the binding mode of the substrate to the metal. However, the residues R104 and R154 stabilize the substrate in the reactive orientation.

Through QM/DMD simulations, we found both Fe and Ni ARD bind the substrate in the same orientation, via O1 and O3 of the substrate (Figure 2.4). The protein pocket stabilizes this specific orientation by two Arg residues, R104 and R154, which forms a hydrogen bond to the doubly deprotonated substrate bound to the metal in this way. The X-ray structure does not contain a bound substrate or its analog, and therefore, upon docking the substrate, these hydrogen bonds were not present. DMD sampling during the simulations elucidated the role of these residues and demonstrated the importance of sampling in metalloproteins systems such

as this one. However, since substrate orientation is identical in the two systems, coordination mode to the metal center can no longer be the reason for product differentiation in Fe and Ni. QM mechanistic studies revealed an additional intermediate that forms in the Fe-ARD' mechanistic pathway not observed in Ni-ARD. This extra intermediate is also found in biomimetic complexes mimicking ARD.[87] The ability of Fe but not Ni to stabilize an additional intermediate comes from the redox flexibility of Fe^{2+}, allowing for the flow of electrons from the residues to the substrate and the bound O_2. Thus, the anti-bonding p* orbital in dioxygen gets populated, and O_2 dissociates. The O atoms easily migrate around, and one accessible epoxy-like transition state produces the mentioned intermediate. The more electron-rich Ni^{2+} does not allow for O_2 dissociation. With ARD, sampling was required to predict the correct substrate binding pose, and dismiss the old mechanistic hypothesis. The protein structure remained consistent upon metal replacement.

However, sometimes, a simple metal replacement can induce large-scale protein motions and repacking, as showcased by catechol-O-methyltransferase (COMT). COMT is a Mg^{2+}-dependent enzyme involved in the biology of pain.[88,89] COMT catalyzes the transfer of the methyl group from the cofactor S-adenosyl-L-methionine (SAM) to catechol subsequently, regulating the amount of catecholamine neurotransmitters in the brain and other organs (Figure 2.5). The metal in COMT binds and positions the catechol substrate in the correct reactive orientation towards SAM for the methylation step.[90,91,92] Native COMT contains Mg^{2+}, which however can be replaced with Co^{2+}, Mn^{2+}, Zn^{2+}, Cd^{2+}, Fe^{2+}, Fe^{3+}, Ni^{2+}, and Sn^{2+}.[93] Metal replacement leads to varying changes in the activity and structure of the enzyme. Surprisingly, the replacement of Mg^{2+} with Ca^{2+} leads to the inhibition of COMT, while replacement of Mg^{2+} with Fe^{2+} creates an only slightly weaker catalyst compared to Mg^{2+}. Fe^{3+} is also a complete inhibitor. The structure and function of these three metal variants of COMT were investigated computationally.[44] Through the course of the simulation, the larger cation, Ca^{2+}, is shown to

coordinate one additional ligand as compared to Mg^{2+} (Figure 2.5). Ca^{2+} also sits deeper into the binding cavity and distorts the reacting parts of catechol and SAM out of alignment for methyl transfer (Figure 2.5). These structural rearrangements with Ca^{2+} are predicted to produce disfavored reaction energetics. Thus, inhibition is a purely geometric effect in this case. Importantly, without sampling, and using just a cluster model, this geometric change is not observed: the Ca^{2+} containing site incorrectly looks identical to the Mg^{2+} containing site.[44] Predictably, the inhibitory effect of Fe^{3+} comes solely from the electronic properties of the metal, specifically its high electrophilicity.

COMT also is an illustration of how ligand capture can be important in addressing a mechanistic difference. This would not be done without the QM treatment of the metal, accompanied by adjustment of the portion of the backbone, which are the key features of QM/DMD. There are further, more dramatic, examples of this sort.

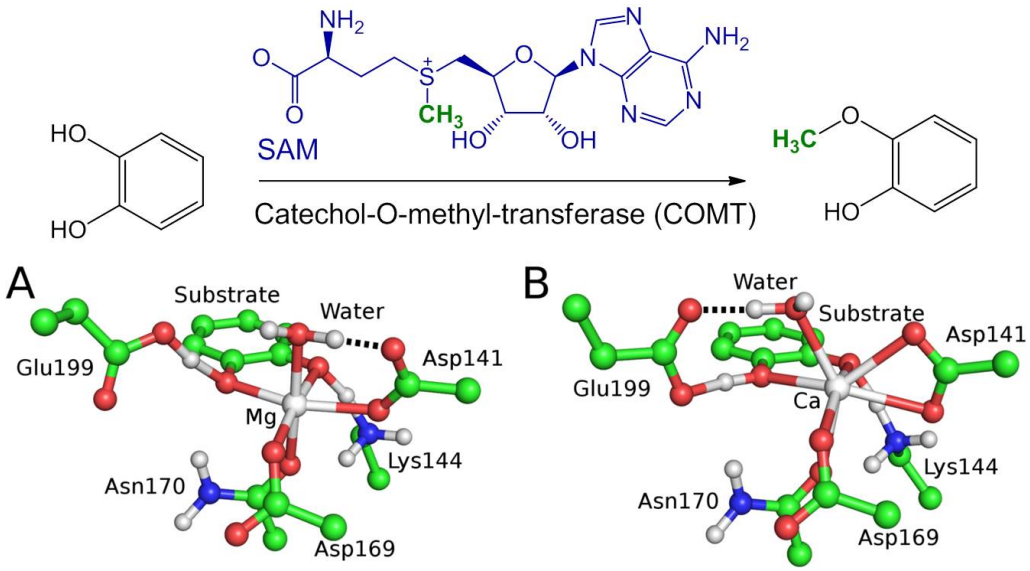
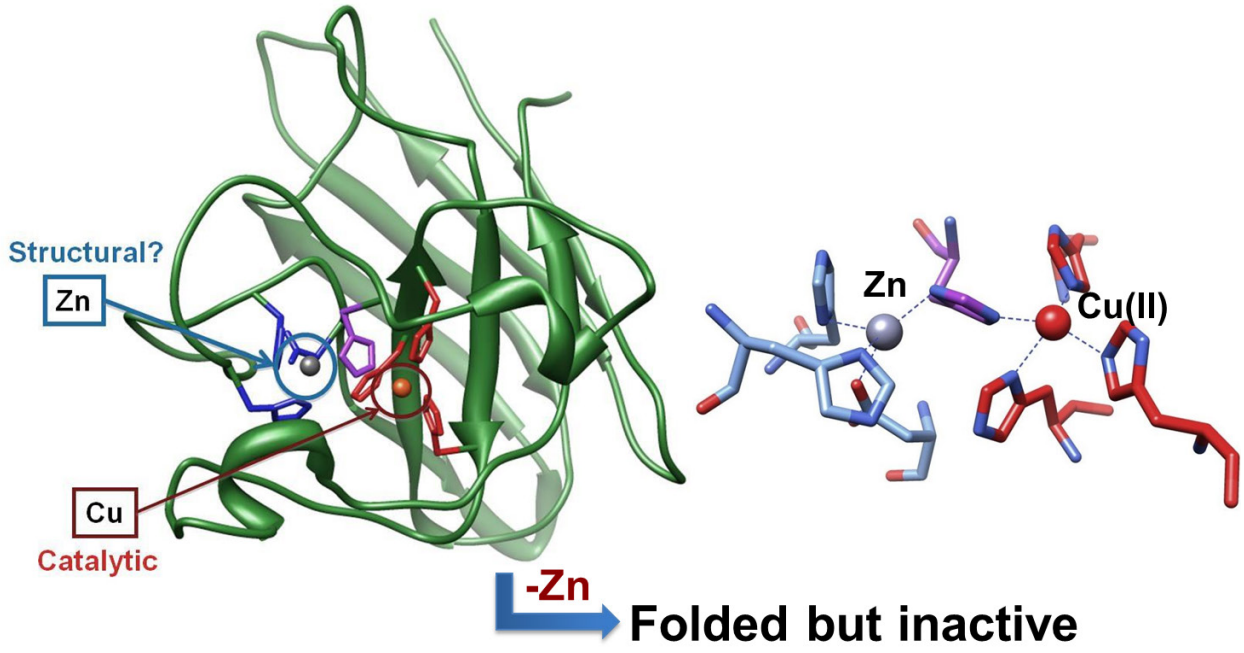


Figure 2.5. Catechol-O-methyl-transferase (COMT) catalyzes the methyl transfer from the cofactor SAM to a catechol motif found in neurotransmitters. (A) Native COMT binds Mg^{2+}, which positions the substrate in the proper orientation toward SAM, for methyl transfer. However, when Ca^{2+} (B) is bound, the active site distorts, putting reacting parts out of alignment, and thus leaving the enzyme inactive. Reprinted in part with permission from Reference 42.

2.3.2. CHANGES IN THE COORDINATION GEOMETRY OF THE METAL

Intriguingly, the class of superoxide dismutase (SOD) enzymes exhibits a wide range of redox activities modulated by ligand detachment or attachment. This review will highlight two specific ones, the first one being Cu,Zn dependent superoxide dismutase in humans (SOD1). SODs catalyze the dismutation/disproportionation of superoxide (Figure 2.6), a dangerous species linked to aging and other oxidation stress processes in organisms. The loss of SOD1 function and subsequent aggregation is known to lead to the neurodegenerative disease Amyotrophic Lateral Sclerosis (ALS) that affects the motor neurons of afflicted patients.[94,95,96,97,98] The metal-dependent nature of SOD is two-fold: catalysis occurs at the Cu site, whereas the Zn site is believed to serve an important structural role for the whole protein. Experiments show that monomers without Zn are precursors to the SOD aggregates[99] while Cu-less SOD1 does not show aggregation and retains its structure.[100] Without the presence of Zn, Cu is not able to catalyze the dismutation reaction.



$$M^{(n+1)+} - SOD + O_2^- \rightarrow M^{n+} - SOD + O_2$$

$$M^{n+} - SOD + O_2^- + 2H^+ \rightarrow M^{(n+1)+} - SOD + H_2O_2$$

Figure 2.6. In the Cu,Zn dependent superoxide dismutase (SOD), Cu plays the role of the catalytic metal however, Zn, although it plays a purely structural role, plays vital role in keeping the adequate structure for Cu-mediated dismutation.

A combined computational and experimental study addressed the effect of Zn removal on the protein structure, electronic structure of the Cu site, and overall catalytic function of SOD.[47] The results show that Zn plays a structural role in SOD1 and directly influences the catalysis, enabling proper coordination and reduction potential of the Cu site. Removal of Zn causes the elimination of the catalytic activity of SOD1 even without protein unfolding and aggregation. It also makes the Cu center prone to deactivation due to immediate reduction to Cu(I) in the resting state. QM/DMD sampling was required to obtain the proper, folded but inactive structure of SOD1. In it, Cu lost one ligand.

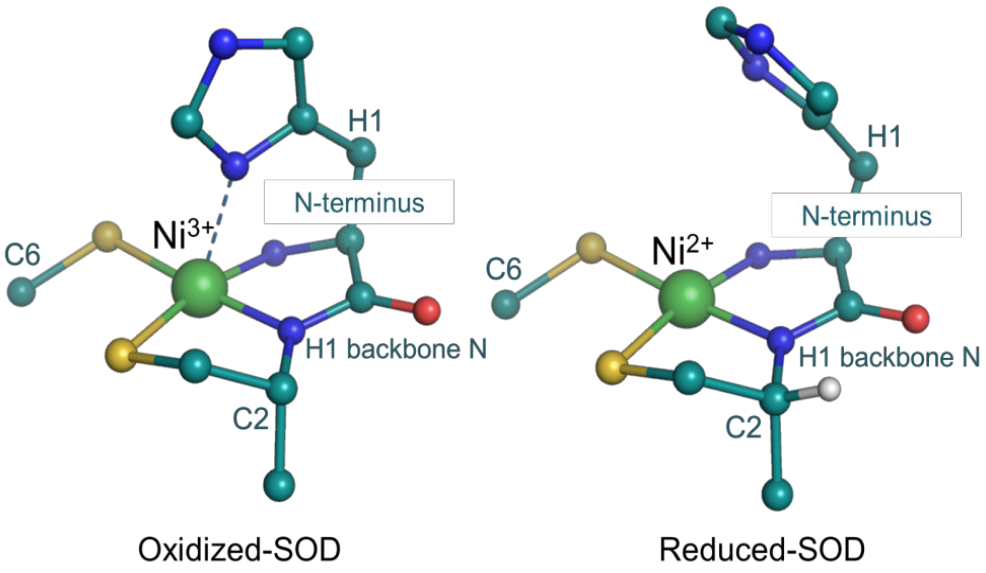


Figure 2.7. The redox reaction in Ni-SOD is modulated through the changing coordination environment of the histidine (H1) residue.

Ni-SODs exist only in *Streptomyces* and cyanobacteria. Unlike all other metals used in SOD, Ni^{2+} does not catalyze superoxide dismutation in aqueous solution due to an improper redox potential (a calculated +2.26 V when the optimum reduction potential is 0.36 V).[101] The active site of Ni-SOD is strikingly different from those of the other SODs.[102] The Ni ion, coordinated by a "Ni-hook" motif,[103] has a square-pyramidal coordination geometry when Ni is in its oxidized form (Ni^{3+}). It has four equatorial ligands: two thiolates from Cys2 and Cys4, a deprotonated amide of the Cys2 backbone, and the N-terminal group of His1; and one axial imidazole from His1. (Figure 2.7) Upon reduction, Ni^{2+} loses the His1 ligand and becomes square planar, making Ni-SOD the only observed SOD with a coordination number that changes as a function of metal oxidation state.[43,104,105,106,107] The flexible coordination geometry activates Ni and confers the proper reduction potential needed to function as a SOD. To address such a structural effect, an extensive sampling within a quantum-classical formalism would be essential.

2.3.3. METALLOENZYMES DESIGN

Metalloenzyme design is the hallmark challenge, where all difficulties revealed need to be addressed. We make the first attempts to design specific buried binding pockets in metalloproteins for desired catalysis. In all cases, QM/DMD sampling for structure prediction, followed by QM mechanistic studies in the core of the effort. In a more conservative approach, the substrate of a known enzyme gets modified, and the binding pocket then gets redesigned slightly to accommodate it. These works involve mostly intuitive mutations done by hand, sampling, and mechanistic investigation, followed by experimental validations. We also use the full machinery described in the section Methods, including *Erebus*, *Eris-QM/DMD*, *QM/DMD*, and *QM*, to design artificial enzymes catalyzing non-natural reactions of interest, using non-physiological metals in the active sites. These efforts are also in the stage of experimental testing of several promising designs. Details of the approach and results will be coming in future publications.

2.4. FUTURE ASPIRATIONS

Perhaps one of the most practical and significant applications of an efficient metalloprotein simulation engine would be computational design of drugs/inhibitors. In this case, it is necessary to predict the binding free energy of a substrate to a protein, or relative free energies of binding of one substrate with respect to another. The field of computational drug design has seen tremendous growth over the past decades.[8] By far, the most widely used approaches hinge on continuously evolving scoring functions,[108] classical MD, Monte Carlo, sampling via normal mode analysis, among other non-QM treatments of large molecules. However, with just these classically-based methods, metalloenzyme targets remain largely inaccessible. In order to assess the induced fit, i.e., the conformational response of the protein to the bound ligand, the QM, or parameterized MM treatment of the metal is paramount.

While there are experimental methods that help bridge this gap, for example, fragment-based lead design (FBLD) which has been used to identify the first building blocks of inhibitors to several metalloenzymes,[109] effective computational approaches would advantageously complement metalloenzyme inhibitor design. Insight into preferred ligand poses without a co-crystal and "growing" the best targets within the active site of a metalloenzyme for optimum affinity and protein dynamics is currently not commonly available *in silico*. Leveraging these methods would lessen the need to synthesize ligand candidates every step of the way (as in FBLD).

We are starting to see QM/MM methods used in screening small libraries, including so-called "On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking" which reveals significantly improved success rate for a zinc-binding protein data set via QM's comprehensive description of the charge-transfer effects and QM-refinement of binding poses.[110] Indeed, QM offers precise relaxation of ligands into protein active sites and measurement of electronic interactions, from the basic potentials (vdW, electrostatics, H-bonds, ligand strain, etc.) that scoring functions already perform well with, to the unparamaterized metals. Even if interactions and poses in the QM active site are well defined, a method to perform extensive sampling and estimate other key effects including entropy and desolvation is required to help complete the thermodynamic picture. This is no easy task, as we've seen for many years drug design was stuck with rigid models of proteins, showing appreciable effectiveness, however clearly inadequate for the more interesting cases.[111,112]

While efficient drug design is still an open question with many pending answers, only a few of which have been discussed here, a couple things have become abundantly clear. One barrier to any current approach in drug design is efficient sampling of design space, which quickly becomes expensive when targeting larger systems, and has recently been incorporated into many docking and scoring methods primarily by sampling of the larger protein movements

followed by rigid substructure searches. Methods such as DMD show promise in quickly and accurately sampling the design environment, not only of the pure protein but with a potential ligand. When coupled with accurate QM estimation of binding site energies, new opportunities in drug design arise, and being explored in our lab.

To conclude, the computational treatment of metalloproteins has been accelerating, and in this article we have introduced many current challenges and developments, especially within our own work. Ongoing improvements to current methodologies and expansion to new methods continue to push the frontier of both structural and mechanistic studies of metalloproteins. This frontier spans exciting directions such as artificial metalloenzymes and metalloenzymes inhibitor design where the efforts in our lab are pushing these boundaries.

2.5. REFERENCES

1 Chung, L. W.; Li, X.; Morokuma, K., Modeling Enzymatic Reactions in Metalloenzymes and Photobiology by Quantum Mechanics (QM) and Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations. In *Quantum Biochemistry*, Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2010; pp 85-130.

2 Warshel, A., *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. Wiley: New York, NY, 1997.

3 Ding, F.; Dokholyan, N. V., Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (50), 19696-19701.

4 Adcock, S. A.; McCammon, J. A., Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006,** *106* (5), 1589-1615.

5 Ben-Nun, M.; Martinez, T. J., Ab initio quantum molecular dynamics. *Adv. Chem. Phys.* **2002,** *121*, 439-512.

6 Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953,** *21* (6), 1087-1092.

7 Hastings, W. K., Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970,** *57* (1), 97-109.

8 Jorgensen, W. L., The Many Roles of Computation in Drug Discovery. *Science* **2004,** *303* (5665), 1813-1818.

9 Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M., Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013,** *9* (6), 2733-2748.

10 Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V., Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2011,** *112* (1), 632-672.

11 Fedorov, D. G.; Kitaura, K., Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2007,** *111* (30), 6904-6914.

12 Dmitri, G. F.; Kazuo, K., Introduction. In *The Fragment Molecular Orbital Method*, CRC Press: Boca Raton, FL, 2009; pp 1-3.

13 Massa, L., Book Review. *Int. J. Quant. Chem.* **2011,** *111* (12), 3251-3251.

14 Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227-249.

15 Hu, H.; Yang, W., Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008,** *59*, 573-601.

16 Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A., Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annu. Rev. Phys. Chem.* **2011,** *62* (1), 41-64.

17 Car, R.; Parrinello, M., Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985,** *55* (22), 2471-2474.

18 Marx, D.; Hutter, J., Ab initio molecular dynamics: Theory and implementation. In *Modern Methods and Algorithms of Quantum Chemistry*, Jülich, 2000; Vol. 1, pp 301-449.

19 Laio, A.; VandeVondele, J.; Rothlisberger, U., A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* **2002,** *116* (16), 6941-6947.

20 Dal Peraro, M.; Vila, A. J.; Carloni, P.; Klein, M. L., Role of zinc content on the catalytic efficiency of B1 metallo β-lactamases. *J. Am. Chem. Soc.* **2007,** *129* (10), 2808-2816.

21 De Vivo, M.; Ensing, B.; Klein, M. L., Computational study of phosphatase activity in soluble epoxide hydrolase: high efficiency through a water bridge mediated proton shuttle. *J. Am. Chem. Soc.* **2005,** *127* (32), 11226-11227.

22 De Vivo, M.; Ensing, B.; Dal Peraro, M.; Gomez, G. A.; Christianson, D. W.; Klein, M. L., Proton shuttles and phosphatase activity in soluble epoxide hydrolase. *J. Am. Chem. Soc.* **2007,** *129* (2), 387-394.

23 Magistrato, A.; Ruggerone, P.; Spiegel, K.; Carloni, P.; Reedijk, J., Binding of novel azole-bridged dinuclear platinum(II) anticancer drugs to DNA: insights from hybrid QM/MM molecular dynamics simulations. *J. Phys. Chem. B* **2006,** *110* (8), 3604-3613.

24 Spiegel, K.; Rothlisberger, U.; Carloni, P., Cisplatin Binding to DNA Oligomers from Hybrid Car-Parrinello/Molecular Dynamics Simulations. *J. Phys. Chem. B* **2004,** *108* (8), 2699-2707.

25 Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U., Role of protein frame and solvent for the redox properties of azurin from Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (52), 19641-19646.

26 Li, H.; Li, W.; Li, S.; Ma, J., Fragmentation-Based QM/MM Simulations: Length Dependence of Chain Dynamics and Hydrogen Bonding of Polyethylene Oxide and Polyethylene in Aqueous Solutions. *J. Phys. Chem. B* **2008,** *112* (23), 7061-7070.

27 He, X.; Wang, B.; Merz, K. M., Protein NMR Chemical Shift Calculations Based on the Automated Fragmentation QM/MM Approach. *J. Phys. Chem. B* **2009,** *113* (30), 10380-10388.

28 Guo, W. W., A.; Xu, X., XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010,** *498*, 203-208.

29 Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K., ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.* **1996,** *100* (50), 19357-19363.

30 Gao, J.; Truhlar, D. G., Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002,** *53*, 467-505.

31 Guimarães, C. R. W.; Repasky, M. P.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L., Contributions of Conformational Compression and Preferential Transition State Stabilization to the Rate Enhancement by Chorismate Mutase. *J. Am. Chem. Soc.* **2003,** *125* (23), 6892-6899.

32 Tubert-Brohman, I.; Acevedo, O.; Jorgensen, W. L., Elucidation of Hydrolysis Mechanisms for Fatty Acid Amide Hydrolase and Its Lys142Ala Variant via QM/MM Simulations. *J. Am. Chem. Soc.* **2006,** *128* (51), 16904-16913.

33 Alexandrova, A. N.; Jorgensen, W. L., Origin of the Activity Drop with the E50D Variant of Catalytic Antibody 34E4 for Kemp Elimination. *J. Phys. Chem. B* **2008,** *113* (2), 497-504.

34 Heimdal, J.; Kaukonen, M.; Srnec, M.; Rulisek, L.; Ryde, U., Reduction potentials and acidity constants of Mn superoxide dismutase calculated by QM/MM free-energy methods. *ChemPhysChem* **2011,** *12* (17), 3337-3347.

35 Gámiz-Hernández, A. P.; Galstyan, A. S.; Knapp, E.-W., Understanding Rubredoxin Redox Potentials: Role of H-Bonds on Model Complexes. *J. Chem. Theory Comput.* **2009,** *5* (10), 2898-2908.

36 Gámiz-Hernández, A. P.; Kieseritzky, G.; Ishikita, H.; Knapp, E. W., Rubredoxin Function: Redox Behavior from Electrostatics. *J. Chem. Theory Comput.* **2011,** *7* (3), 742-752.

37 McCullagh, M.; Voth, G. A., Unraveling the Role of the Protein Environment for [FeFe]-Hydrogenase: A New Application of Coarse-Graining. *J. Phys. Chem. B* **2013,** *117* (15), 4062-4071.

38 Hammes-Schiffer, S., Introduction: Proton-Coupled Electron Transfer. *Chem. Rev.* **2010,** *110* (12), 6937-6938.

39 Hammes-Schiffer, S.; Stuchebrukhov, A. A., Theory of Coupled Electron and Proton Transfer Reactions. *Chem. Rev.* **2010,** *110* (12), 6939-6960.

40 Hatcher, E.; Soudackov, A. V.; Hammes-Schiffer, S., Proton-Coupled Electron Transfer in Soybean Lipoxygenase: Dynamical Behavior and Temperature Dependence of Kinetic Isotope Effects. *J. Am. Chem. Soc.* **2006,** *129* (1), 187-196.

41 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V., Ab initio folding of proteins using all-atom discrete molecular dynamics. *Structure* **2008,** *16* (7), 1010-1018.

42 Sparta, M.; Shirvanyants, D.; Ding, F.; Dokholyan, Nikolay V.; Alexandrova, Anastassia N., Hybrid Dynamics Simulation Engine for Metalloproteins. *Biophys. J.* **2012,** *103* (4), 767-776.

43 Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I., Discrete molecular dynamics studies of the folding of a protein-like model. *Fold Des.* **1998,** *3* (6), 577-587.

44 Sparta, M.; Alexandrova, A. N., How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. *PLoS ONE* **2012,** *7* (10), e47172.

45 Valdez, C. E.; Alexandrova, A. N., Why Urease Is a Di-Nickel Enzyme whereas the CcrA β-Lactamase Is a Di-Zinc Enzyme. *J. Phys. Chem. B* **2012,** *116* (35), 10649-10656.

46 Valdez, C. E.; Sparta, M.; Alexandrova, A. N., The Role of the Flexible L43-S54 Protein Loop in the CcrA Metallo-β-lactamase in Binding Structurally Dissimilar β-Lactam Antibiotics. *J. Chem. Theory Comput.* **2012,** *9* (1), 730-737.

47 Nedd, S.; Redler, R. L.; Proctor, E. A.; Dokholyan, N. V.; Alexandrova, A. N., Cu,Zn-Superoxide Dismutase without Zn Is Folded but Catalytically Inactive. *J. Mol. Biol.* **2014,** *426* (24), 4112-4124.

48 Sparta, M.; Valdez, C. E.; Alexandrova, A. N., Metal-Dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway. *J. Mol. Biol.* **2013,** *425* (16), 3007-3018.

49 Calhoun, J. R.; Nastri, F.; Maglio, O.; Pavone, V.; Lombardi, A.; DeGrado, W. F., Artificial diiron proteins: From structure to function. *Biopolymers* **2005,** *80* (2-3), 264-278.

50 Bell, C. B.; Calhoun, J. R.; Bobyr, E.; Wei, P.-p.; Hedman, B.; Hodgson, K. O.; DeGrado, W. F.; Solomon, E. I., Spectroscopic Definition of the Biferrous and Biferric Sites in de Novo Designed Four-Helix Bundle DFsc Peptides: Implications for O_2 Reactivity of Binuclear Non-Heme Iron Enzymes. *Biochemistry* **2008,** *48* (1), 59-73.

51 Kaplan, J.; DeGrado, W. F., De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (32), 11566-11570.

52 Fry, H. C.; Lehmann, A.; Sinks, L. E.; Asselberghs, I.; Tronin, A.; Krishnan, V.; Blasie, J. K.; Clays, K.; DeGrado, W. F.; Saven, J. G.; Therien, M. J., Computational de Novo Design and Characterization of a Protein That Selectively Binds a Highly Hyperpolarizable Abiological Chromophore. *J. Am. Chem. Soc.* **2013,** *135* (37), 13914-13926.

53 Der, B. S.; Jha, R. K.; Lewis, S. M.; Thompson, P. M.; Guntas, G.; Kuhlman, B., Combined computational design of a zinc-binding site and a protein–protein interaction: One open zinc coordination site was not a robust hotspot for de novo ubiquitin binding. *Proteins: Struct. Funct. Bioinform.* **2013,** *81* (7), 1245-1255.

54 Pabo, C. O.; Peisach, E.; Grant, R. A., Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem* **2001,** *70* (1), 313-340.

55 Yin, S.; Ding, F.; Dokholyan, N. V., Modeling backbone flexibility improves protein stability estimation. *Structure* **2007,** *15* (12), 1567-1576.

56 Liao, R.-Z.; Thiel, W., On the Effect of Varying Constraints in the Quantum Mechanics Only Modeling of Enzymatic Reactions: The Case of Acetylene Hydratase. *J. Phys. Chem. B* **2013,** *117* (15), 3954-3961.

57 Han, K. E.; Bystroff, C.; Baker, D., Three-dimensional structures and contexts associated with recurrent amino acid sequence patterns. *Protein Sci.* **1997,** *6* (7), 1587-1590.

58 Shirvanyants, D.; Alexandrova, A. N.; Dokholyan, N. V., Rigid substructure search. *Bioinformatics* **2011,** *27* (9), 1327-1329.

59 Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N., Mysteries of Metals in Metalloenzymes. *Acc. Chem. Res.* **2014,** *47* (10), 3110-3117.

60 Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D., Accurate computer-based design of a new backbone conformation in the second turn of protein L. *J. Mol. Biol.* **2002,** *315* (3), 471-477.

61 Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D., Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003,** *302* (5649), 1364-1368.

62 DiMaio, F.; Terwilliger, T. C.; Read, R. J.; Wlodawer, A.; Oberdorfer, G.; Wagner, U.; Valkov, E.; Alon, A.; Fass, D.; Axelrod, H. L.; Das, D.; Vorobiev, S. M.; Iwai, H.; Pokkuluri, P. R.; Baker, D., Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **2011,** *473* (7348), 540-543.

63 Huang, P.-S.; Oberdorfer, G.; Xu, C.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D., High thermodynamic stability of parametrically designed helical bundles. *Science* **2014,** *346* (6208), 481-485.

64 Dahiyat, B. I.; Mayo, S. L., De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997,** *278* (5335), 82-87.

65 Kuhlman, B.; Baker, D., Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **2000,** *97* (19), 10383-10388.

66 Ding, F.; Dokholyan, N. V., Emergence of Protein Fold Families through Rational Design. *PLoS Comp. Biol.* **2006,** *2* (7), e85.

67 Yin, S.; Ding, F.; Dokholyan, N. V., Eris: an automated estimator of protein stability. *Nat. Methods* **2007,** *4* (6), 466-467.

68 Reynolds, C. H.; Holloway, M. K., Thermodynamics of Ligand Binding and Efficiency. *ACS Med. Chem. Lett.* **2011,** *2* (6), 433-437.

69 Dill, K. A.; MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* **2012,** *338* (6110), 1042-1046.

70 Karplus, M.; Kushick, J. N., Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981,** *14* (2), 325-332.

71 Andricioaei, I.; Karplus, M., On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001,** *115* (14), 6289-6292.

72 Schlitter, J., Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993,** *215* (6), 617-621.

73 Masgrau, L.; Truhlar, D. G., The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* **2015,** *48* (2), 431-438.

74 Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G., Free Energy Surface, Reaction Paths, and Kinetic Isotope Effect of Short-Chain Acyl-CoA Dehydrogenase. *J. Phys. Chem. B* **2003,** *107* (35), 9567-9578.

75 Pang, J.; Pu, J.; Gao, J.; Truhlar, D. G.; Allemann, R. K., Hydride transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. *J. Am. Chem. Soc.* **2006,** *128* (24), 8015-8023.

76 Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W., Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface. *J. Chem. Phys.* **2008,** *128* (3), 034105.

77 Ruiz-Pernía, J. J.; Garcia-Viloca, M.; Bhattacharyya, S.; Gao, J.; Truhlar, D. G.; Tuñón, I., Critical Role of Substrate Conformational Change in the Proton Transfer Process Catalyzed by 4-Oxalocrotonate Tautomerase. *J. Am. Chem. Soc.* **2009,** *131* (7), 2687-2698.

78 Senn, H. M.; Kästner, J.; Breidung, J.; Thiel, W., Finite-temperature effects in enzymatic reactions: insights from QM/MM free-energy simulations. *Can. J. Chem.* **2009,** *87* (10), 1322-1337.

79 Kanaan, N.; Ferrer, S.; Martí, S.; Garcia-Viloca, M.; Kohen, A.; Moliner, V., Temperature Dependence of the Kinetic Isotope Effects in Thymidylate Synthase. A Theoretical Study. *J. Am. Chem. Soc.* **2011,** *133* (17), 6692-6702.

80 Riniker, S.; Christ, C. D.; Hansen, H. S.; Hünenberger, P. H.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. F., Calculation of Relative Free Energies for Ligand-Protein

Binding, Solvation, and Conformational Transitions Using the GROMOS Software. *J. Phys. Chem. B* **2011,** *115* (46), 13570-13577.

81 Huang, W.; Lin, Z.; van Gunsteren, W. F., Use of Enveloping Distribution Sampling to Evaluate Important Characteristics of Biomolecular Force Fields. *J. Phys. Chem. B* **2014,** *118* (24), 6424-6430.

82 Movahed, H. B.; van Zon, R.; Schofield, J., Free energy landscape of protein-like chains with discontinuous potentials. *J. Chem. Phys.* **2012,** *136* (24), 1-13.

83 Martin-Diaconescu, V.; Maroney, M. J., *Comprehensive Inorganic Chemistry II*. Elsevier: Oxford, 2013.

84 Dai, Y.; Wensink, P. C.; Abeles, R. H., One protein, two enzymes. *J. Biol. Chem.* **1999,** *274* (3), 1193-1195.

85 Pochapsky, T. C.; Ju, T.; Dang, M.; Beaulieu, R.; Pagani, G. M.; OuYang, B., Nickel in Acireductone Dioxygenase. In *Nickel and Its Surprising Impact in Nature*, John Wiley & Sons, Ltd: Hoboken, NJ, 2007; pp 473-500.

86 Ju, T.; Goldsmith, R. B.; Chai, S. C.; Maroney, M. J.; Pochapsky, S. S.; Pochapsky, T. C., One Protein, Two Enzymes Revisited: A Structural Entropy Switch Interconverts the Two Isoforms of Acireductone Dioxygenase. *J. Mol. Biol.* **2006,** *363* (4), 823-834.

87 Allpress, C. J.; Grubel, K.; Szajna-Fuller, E.; Arif, A. M.; Berreau, L. M., Regioselective Aliphatic Carbon–Carbon Bond Cleavage by a Model System of Relevance to Iron-Containing Acireductone Dioxygenase. *J. Am. Chem. Soc.* **2012,** *135* (2), 659-668.

88 Männistö, P. T.; Kaakkola, S., Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors. *Pharmacol. Rev.* **1999,** *51* (4), 593-628.

89 Guldberg, H. C.; Marsden, C. A., Catechol-O-Methyl Transferase: Pharmacological Aspects and Physiological Role. *Pharmacol. Rev.* **1975,** *27* (2), 135-206.

90 Lotta, T.; Vidgren, J.; Tilgmann, C.; Ulmanen, I.; Melen, K.; Julkunen, I.; Taskinen, J., Kinetics of Human Soluble and Membrane-Bound Catechol O-Methyltransferase: A Revised Mechanism and Description of the Thermolabile Variant of the Enzyme. *Biochemistry* **1995,** *34* (13), 4202-4210.

91 Vidgren, J.; Svensson, L. A.; Liljas, A., Crystal structure of catechol O-methyltransferase. *Nature* **1994,** *368* (6469), 354-358.

92 Rutherford, K.; Le Trong, I.; Stenkamp, R. E.; Parson, W. W., Crystal structures of human 108V and 108M catechol O-methyltransferase. *J. Mol. Biol.* **2008,** *380* (1), 120-130.

93 Axelrod, J.; Tomchick, R., Enzymatic O-Methylation of Epinephrine and Other Catechols. *J. Biol. Chem.* **1958,** *233* (3), 702-705.

94 Khare, S. D.; Dokholyan, N. V., Common dynamical signatures of familial amyotrophic lateral sclerosis-associated structurally diverse Cu, Zn superoxide dismutase mutants. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (9), 3147-3152.

95 Pardo, C. A.; Xu, Z.; Borchelt, D. R.; Price, D. L.; Sisodia, S. S.; Cleveland, D. W., Superoxide dismutase is an abundant component in cell bodies, dendrites, and axons of motor neurons and in a subset of other neurons. *Proc. Natl. Acad. Sci. USA* **1995,** *92* (4), 954-958.

96 Valentine, J. S.; Hart, P. J., Misfolded CuZnSOD and amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. USA* **2003,** *100* (7), 3617-3622.

97 Wilcox, K. C.; Zhou, L.; Jordon, J. K.; Huang, Y.; Yu, Y.; Redler, R. L.; Chen, X.; Caplow, M.; Dokholyan, N. V., Modifications of superoxide dismutase (SOD1) in human erythrocytes: a possible role in amyotrophic lateral sclerosis. *J. Biol. Chem.* **2009,** *284* (20), 13940-13947.

98 Sheng, Y.; Abreu, I. A.; Cabelli, D. E.; Maroney, M. J.; Miller, A.-F.; Teixeira, M.; Valentine, J. S., Superoxide Dismutases and Superoxide Reductases. *Chem. Rev.* **2014,** *114* (7), 3854-3918.

99 Lyons, T. J.; Liu, H.; Goto, J. J.; Nersissian, A.; Roe, J. A.; Graden, J. A.; Café, C.; Ellerby, L. M.; Bredesen, D. E.; Gralla, E. B.; Valentine, J. S., Mutations in copper-zinc superoxide dismutase that cause amyotrophic lateral sclerosis alter the zinc binding site and the redox behavior of the protein. *Proc. Natl. Acad. Sci. USA* **1996,** *93* (22), 12240-12244.

100 Strange, R. W.; Antonyuk, S.; Hough, M. A.; Doucette, P. A.; Rodriguez, J. A.; Hart, P. J.; Hayward, L. J.; Valentine, J. S.; Hasnain, S. S., The structure of holo and metal-deficient wild-type human Cu, Zn superoxide dismutase and its relevance to familial amyotrophic lateral sclerosis. *J. Mol. Biol.* **2003,** *328* (4), 877-891.

101 Meunier, B., *Biomimetic Oxidations Catalyzed by Transition Metal Complexes*. Imperial College Press: London, 2000.

102 Fiedler, A. T.; Bryngelson, P. A.; Maroney, M. J.; Brunold, T. C., Spectroscopic and Computational Studies of Ni Superoxide Dismutase: Electronic Structure Contributions to Enzymatic Function. *J. Am. Chem. Soc.* **2005,** *127* (15), 5449-5462.

103 Barondeau, D. P.; Kassmann, C. J.; Bruns, C. K.; Tainer, J. A.; Getzoff, E. D., Nickel Superoxide Dismutase Structure and Mechanism. *Biochemistry* **2004,** *43* (25), 8038-8047.

104 Dokholyan, N. V., Studies of folding and misfolding using simplified models. *Curr. Opin. Struct. Biol.* **2006,** *16* (1), 79-85.

105 Ding, F.; Guo, W.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J. E., Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. *J. Mol. Biol.* **2005,** *350* (5), 1035-1050.

106 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V., Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* **2008,** *16* (7), 1010-1018.

107 Wuerges, J.; Lee, J.-W.; Yim, Y.-I.; Yim, H.-S.; Kang, S.-O.; Carugo, K. D., Crystal structure of nickel-containing superoxide dismutase reveals another type of active site. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (23), 8569-8574.

108 Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004,** *3* (11), 935-949.

109 Jacobsen, J. A.; Fullagar, J. L.; Miller, M. T.; Cohen, S. M., Identifying Chelators for Metalloprotein Inhibitors Using a Fragment-Based Approach. *J. Med. Chem.* **2010,** *54* (2), 591-602.

110 Chaskar, P.; Zoete, V.; Röhrig, U. F., Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *J. Chem. Inf. Model.* **2014,** *54* (11), 3137-3152.

111 Alvarez-Garcia, D.; Barril, X., Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J. Chem. Theory Comput.* **2014,** *10* (6), 2608-2614.

112 Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A., Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem.* **2014,** *186* (0), 31-45.

CHAPTER 3

Why Urease is a di-Nickel Enzyme,

whereas the CcrA β-Lactamase is a di-Zinc Enzyme

3.1. INTRODUCTION

Ureases (EC 3.5.1.5) and metallo-β-lactamases (EC 3.5.2.6) are metallo-enzymes that hydrolyze amides: urea and β-lactam antibiotics, respectively. Urease[1-3] is present in plants, bacteria and yeast, and converts urea into ammonia and carbon dioxide. Conversion of urea is the last step of nitrogen mineralization in nature. The proficiency of ureases is the highest known among hydrolases.[2,4] It was also suggested that elimination of ammonia from urea may be catalyzed by urease, and the energetics are fairly competitive.[5] Remarkably, urea on its own decomposes exclusively via deamination route rather than hydrolysis.[4,6] Understanding the functionality of ureases is extremely important, especially in the development of urease inhibitors. Inhibition would reduce the environmental pollution, increase nitrogen uptake in plants, and help the treatment of infections due to ureolytic bacteria.

Ureases are unique among hydrolases in having Ni cations at the active site. They are also Ni-selective in vivo, so no Zn-substituted urease can be studied.[7] The presence of Ni in this hydrolase puzzled scientists ever since its discovery.[8] Ni (d^8) is more toxic than Zn (d^{10}), which seems to be the evolutionary force opposing the use of Ni. It was speculated that Ni was required to coordinate simultaneously urea and a water molecule for the reaction, which would cause one of the Ni centers to adopt a coordination close to octahedral, typical for Ni, but atypical for Zn. Thus, a role of the second Ni as a cofactor was speculated. However, it was also shown that water is unlikely to remain in the active site upon the substrate binding.[9] Hence, the choice for Ni in urease remains elusive. At the same time, the catalytic potency of Ni is very high and the proficiency of urease is perhaps attributable to it. From this prospective, we wonder why it is that no other hydrolases use Ni instead of Zn.

β-lactamases are also amide hydrolases, responsible for antibiotic resistance in pathogenic bacteria.[10-13] They render β-lactam antibiotics ineffective by hydrolyzing the lactam ring. With antibiotic resistance being a major medicinal problem, inhibition of b-lactamases is an

active area of research. Among several kinds of β-lactamases, metallo-b-lactamases are the most proficient and hence most problematic enzymes. Metallo-β-lactamases contain two Zn cations in the active site, in contrast to ureases.

Since Ni seems to be a more potent metal, why is it that Zn is chosen in β-lactamases, as well as in many other hydrolases? Is this choice based on maximizing the potency of the enzyme, for not so obvious electronic or structural reasons? Or is this choice an indication of some other evolutionary factors present in cells? Are ureases and β-lactamases are "perfect" catalysts? Surely the evolutionary choice could be balanced by a number of biological factors other than enzyme efficiency, such as availability of given metals in given organisms, metal toxicity, or the presence of enzyme-building machinery that delivers metals to the active sites. If the catalytic performance is partially sacrificed in this overall evolutionary process, then it would suggest that enzymes could be further improved as catalysts, at least for the use in vitro.

In order to answer the above questions, analysis of the stages of the catalytic mechanism, where the metal plays the crucial role, is needed. There are several proposed mechanisms for both enzymes, with especially vigorous debated points being the group playing the role of the acid in the reaction, the role of the second metal center, the role of the protein flap above the active site, and the overall order of events in the catalytic cycles. However, it is generally agreed upon that one of the metal ions coordinates the substrate via its carbonyl oxygen atom.[2,14,15,16] The carbonyl group thus gets polarized, and the carbonyl carbon becomes more susceptible to the nucleophilic attack. This pose places the substrate near the nucleophilic OH group bridging the metals (WB). For β-lactamases, the nucleophilic attack by WB is thought to be the first, and, according to some studies, the rate-determining step of hydrolysis.[17-19] Even though a controversy exists about the mechanism of catalytic hydrolysis for ureases,[20] the nucleophilic attack being the first step is still one of the most viable reported possibilities.[21,22] It is additionally supported by the tight binding of inhibitors mimicking the tetrahedral transition state

to the urease active site.[23,24] The nucleophilic attack is also the step where the metal plays a major role: coordinating and polarizing the substrate and the nucleophilic water molecule. In this study, we focus on just the attack. We hypothesize that the mechanism and energetics of this step should shed light on the reason for Nature's preference for Zn in β-lactamase and Ni in urease. Our study thus addresses the specific choices for Zn1 and Ni1 in the active sites (metal centers that coordinate the carbonyl carbon), but not necessarily Zn2 and Ni2, although they will appear to play a role too, as will be shown shortly.

In this study, we do not address the substrate binding, and therefore do not assess k_{cat}/K_m (i.e. a typical measure of enzymatic performance). The reason is that both enzymes perform at the diffusion limit, and have solvent-exposed active sites. Thus, the actual event of the substrate binding is not hard. Furthermore, considering our task of realizing the reasons for the metal choices in the two enzymes from the point of view of catalysis, starting from the enzyme-substrate complexes seems reasonable.

This study utilizes a combined statistical mechanical and quantum mechanical approach, for the following reason: As was pointed out,[2,11] the active site and the protein structure at large mutually impact each other in the studied hydrolases, and the overall induced fit upon substrate arrival is expected. As a result, neither pure quantum mechanical studies on cluster models, nor classical molecular dynamics simulations of the studied hydrolases are entirely valid, since they produce conflicting ideas about the catalytic mechanisms.[11] Instead, the best approach to study these enzymes would be a combination of quantum mechanics (QM) and classical molecular mechanical (MM) molecular dynamics (MD). However, full QM/MM MD simulations covering nanoseconds of dynamics would be prohibitively expensive, and to date, no QM/MM dynamics studies have been reported for the studied hydrolases. We recently developed a method capable of performing this task, called QM/DMD.[25] Its features include affordable nanosecond time-scale sampling via discrete molecular dynamics (DMD), and QM treatment of the metal

containing active site. QM/DMD utilizes a combination of classical dynamics and gradient following relaxation at the QM level, the novel “breathing” QM-DMD boundary, clustering, and annealing. The method was demonstrated to be affordable, balanced, and reliable for modeling of natural and altered metalloproteins.[25] In the present study, we employ QM/DMD to find the preferred docking position of the β-lactam antibiotic model in the CcrA active site, with fully flexible protein. After the docking poses were identified, the mechanistic study is done quantum mechanically, on the cluster models.

3.2. THEORETICAL METHODS AND STRUCTURES

3.2.1. INITIAL PROTEIN STRUCTURES AND QM/DMD SIMULATIONS FOR B-LACTAMASE

For urease, the 2.2 Å crystal structure (PDB ID code: 1FWJ)[15] was used to dock urea and prepare the active site model. The 1.85 Å crystal structure of the CcrA β-lactamase (PDB ID code: 1ZNB)[26] was used for the study of this protein. Despite all our efforts, we failed in our search for the bound reactive complex of the chosen truncated model of the β-lactam antibiotic (shown in Figure 3.1A) and the protein coming directly from the crystal structure. The substrate dissociated away from the active site during geometry optimizations. This showed the need for a dynamical docking procedure. We performed it using our newly implemented QM/DMD method.[25]

The specific details of the method are explained elsewhere.[25] In short, QM/DMD is an new mixed quantum mechanical / molecular mechanical method, which features several unique tactics that make it affordable, and yet reliable on a variety of scales, from large (> 10 Å) to small (< 1 Å). In QM/DMD simulations, the protein is partitioned into three regions: (1) the QM-only region, consisting of the metals and their immediate coordination sphere, and moved only at the QM level, (2) the larger QM-DMD region, constituting a larger active site (cluster model), and moved by either QM or DMD, depending on the stage of the simulation, and (3) the DMD-

only region, which includes the rest of the protein, moved exclusively classically (Figure 3.1A). QM/DMD simulations proceed in an iterative fashion.

They start with a DMD simulation on the entire protein (i.e. on DMD-only, and QM-DMD regions), with frozen metal centers and their immediate coordination environment (QM-only region). A set of additional constrains, as derived from the QM calculations, were imposed to direct the sampling procedure. In particular, for each pair of atoms connected by green lines in Figure 3.1B, the sampling of the interatomic distance was limited to the interval defined by the QM optimized value ±0.01Å.

In this way, we avoid the need for parameterization of the classical force field for the metals. The beginning of the DMD simulation is done in conjunction with annealing, used for better sampling, and then the temperature is kept low for the final 10,000 DMD time units (0.5 ns), when the data is collected. This final ensemble of structures is clustered according to a geometric similarity, as follows. The Kabsch[27] RMSD is computed for all pairwise snapshot structures to quantify their geometric similarity. Based on the RMSD matrix, a hierarchical clustering algorithm[28] is used to organize all the snapshots in distinct clusters. The user can specify either the number of clusters, or the maximum relative distance (hence the amplitude) within each cluster. For each cluster, a specific representative structure is selected, and one can either choose the structure in the subset closest to the centroid, or the structure with the lowest DMD energy. In this study, we requested the number of clusters to be 5, and both the structure closest to the centroid, and the one with the lowest energy were used as representatives. Thus, there are 10 structures resultant from the DMD stage of the simulation.

The QM-DMD regions are then cut from those 10 structures. The atoms at the QM-DMD boundary are capped with hydrogen atoms and frozen in Cartesian space to maintain the positions dictated by the protein backbone. Single point QM calculations on these structures are performed at the TPSS[29]/def2-SVP[30] level with an additional empirical dispersion correction,[31]

using *TURBOMOLE* v.6.3.[32] A single structure is then chosen based on the combined scoring index (*SI*) that depends on both the QM and DMD energies, and their differences from the minimum energies found for the given set of structures:

$$SI^{i} = n(E^{i}_{DMD} - E^{min}_{DMD}) + (1-n)(E^{i}_{QM} - E^{min}_{QM}).$$

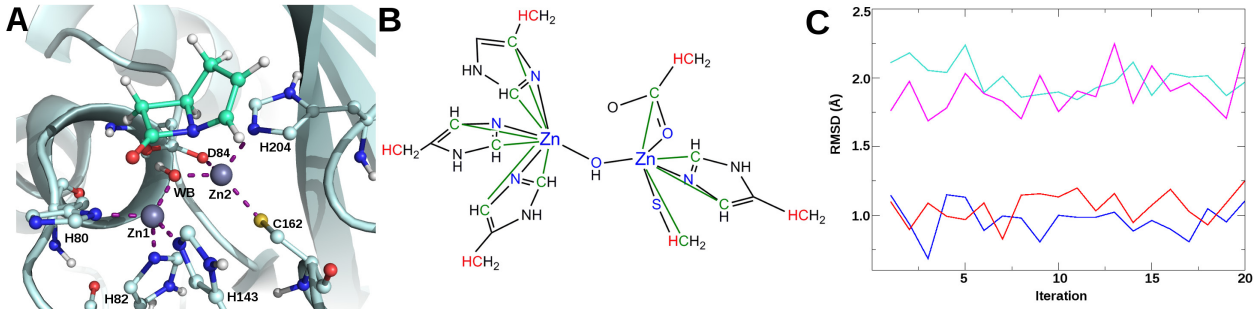


Figure 3.1. (A) A representative QM/DMD equilibrated structure of β-lactamase with the bound substrate. QM/DMD partitioning is demonstrated. The metal centers and atoms directly coordinated to them are moved only during the QM stage of the simulation. All other atoms shown as sticks can be moved by both QM and DMD. The rest of the protein shown as ribbons is moved exclusively classically by DMD. (B) The green lines show additional constrains that are imposed during the DMD simulation, to retain the chemistry determined at the QM level of theory, the red atoms are frozen during QM calculations, and the blue color marks the atoms in the *QM-only* domain. (C) QM/DMD convergence is demonstrated by the RMSD values for the protein backbone, and the all-atom RMSD of the active site, as a function of iteration number. The teal and blue lines correspond to the protein backbone RMSD and all atom RMSD of the active site, respectively, for the bidentate binding. Similarly, magenta and red correspond to the monodentate binding.

In this work, *n* is set to 0.5. This structure is then partially optimized at the QM level (for 200 cycles of optimization, after which the optimization is interrupted to save time), and after that uncapped and reinstalled into the protein. The QM-DMD boundary shrinks again, back to the immediate coordination of the metals, and the simulation continues with the second DMD stage, when the geometric information from the active site can propagate to the rest of the protein. The alternating DMD-QM cycle continues for 20 iterations (when the convergence was reached in the present study). The convergence is judged on the basis of the QM and DMD

energies of the system, and the RMSD values for the backbone of the protein and the active site (Figure 3.1C). Two separate QM/DMD simulations were run for the same system: one with monodentate initial binding pose, and one with a bidentate binding pose. Both binding poses retained the initial coordination, indicating that both are minima on the QM/DMD potential energy surface. The result is in agreement with earlier mechanistic studies on cluster models,[20g] which means that the role of the protein in the structure of the substrate-active site complex is minor. The bidentate pose yields the structure in the reactive orientation and overall lower QM and DMD energies.

From the equilibrated protein with the substrate coordinated in the bidentate fashion, a single structure with the lowest combined QM and DMD energy was chosen for the mechanistic study. That representative structure yielded by QM/DMD is shown in Figure 3.1A. QM/DMD was able to bind the carbonyl O atom of the substrate to the Zn1 center, with a bond length R(O-Zn1) of 3.7 Å. The nucleophilic O of WB was located 3.7 Å from the carbonyl C. The binding mode was monodentate.

The Ni-substituted β-lactamase was obtained by replacing Zn with Ni and optimizing the complex with the substrate bound, exclusively at the QM level. The overall structure did not change significantly, and docking was straightforward. Additionally, we studied a complex of β-lactamase with urea. Urea, being a small molecule, fits in the active site of β-lactamase without problems. No QM/DMD sampling was required for these two systems.

3.2.2. ACTIVE SITE COMPLEXES

For the mechanistic study, all active sites amino acid residues were truncated at the beta carbons and capped with hydrogens. These carbon atoms were fixed, preserving the backbone structure of the complexes. The urease active site (Figure 3.2A) contains two pseudo-octahedral Ni atoms: Ni1 is coordinated to residues H249, H275, a carbamylated lysine (K220),

a water molecule (W1) and bridging water (WB); Ni2 is coordinated to H137, H139, D363, a water molecule (W2) and the bridging water (WB) and carbamylated lysine (K220) coordinated to Ni1. As based on the crystallographic of the DAP inhibited urease (PDB code 3UBP),[23] the bidendate binding nature of urea displaces W1 and W2 coordinated to the nickels, a binding pose suggested to mimic a transition state.[9] Benini et al.[23,24] proposed hydrolysis of urea to occur with nucleophilic attack of bridging hydroxide, with necessary bidentate ligation of urea to the two electrophilic nickel ions, due to the close proximity of urea to the bridging hydroxide. Hence, W1 and W2 were removed from the complex (Figure 3.2A).

We tested different spin states of the di-nickel complexes, and found that the high-spin quintet state is lowest in energy (triplet was ~9 kcal/mol higher in energy). This is in agreement with previous theoretical and experimental reports.[20] The active site of ureases is not EPR detectable. However, magnetic measurements on jack bean urease have been performed, and showed that the active site is mostly high spin, with some spin heterogeneity being required for the best fit with experimental data.[20a-f] Thus, the ideal theoretical study would address the mixed-spin state by multireference calculations. However, this is prohibitively expensive for the bimetallic active site of ureases. Hence, the adopted broken-symmetry approach appears to be the best, and it showed the high spin complex being of the lowest energy. Additionally, in a recent theoretical study on the model complex, the quintet also was found to be the lowest energy state.[20g] Thus, we further considered only the quintet.

All spectroscopic and structural data point at the fact that both Ni centers are in the formal oxidation state of +2 in ureases. Hence, the charge on the complex was +1, and no other charge states were considered.

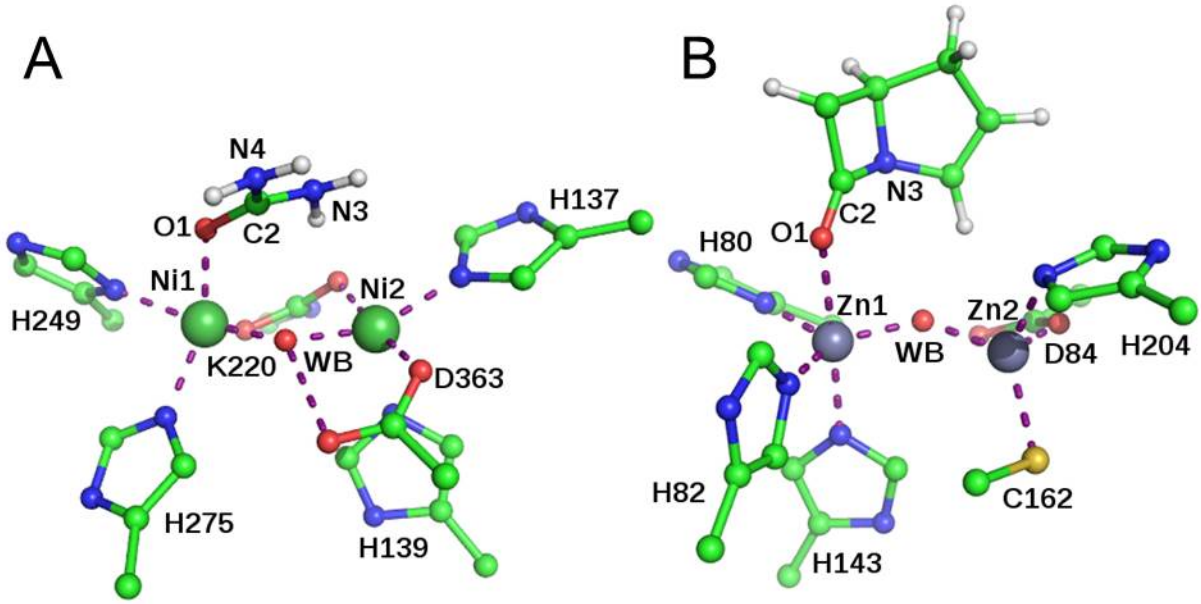


Figure 3.2. Truncated active sites of (A) urease, and (B) β-lactamase, used in cluster mechanism studies. The C atoms of amino acids located far most from the metal centers are protonated to satisfy all valencies, and fixed at the positions dictated by the rest of the protein.

β-lactamase active site (Figure 3.2B) contains a tetrahedrally coordinated Zn1 with three histidine residues (H80, H82, and H143) and bridging water (WB), and a bypyramidal Zn2 coordinated by D84, C162, and H204, the bridging water coordinated to Zn1 (WB), and an additional water molecule (W2). W2 is speculated to be readily displaced by the β-lactam ring nitrogen, facilitating the hydrolysis, as suggested by Merz et. al.[33] Thus, W2 was removed. The protonation state of D103 in CcrA was questioned in the past, and it was linked to the stability of the Zn-OH-Zn bridge. Results of classical MD simulations suggested that protonated D103 yields a looser bridge and higher activity of the enzyme.[33,34] However, it was shown that both forms have similar activities predicted computationally, and so in this study, D103 was chosen to be deprotonated, in agreement with Merz and coworkers.[20g] The spin of the active site is zero, and the total charge is +1. From our theoretical assessment, the active site of the Ni-substituted β-lactamase was a quintet state, charged +1 as well.

All cluster studies were done at the TPSS/def2-SVP and B3LYP[35]/def2-SV(P) level with empirical dispersion correction, using *TURBOMOLE*. Geometry optimizations of the reactants, transition states (TSs) and intermediates were carried out. All TSs and minima were confirmed by frequency calculations. Each TS had only one imaginary frequency, corresponding to the normal mode coinciding with the reaction coordinate.

3.3. RESULTS AND DISCUSSION

In order to explore the role of the metals in the nucleophilic attack on the carbonyl group in hydrolysis, five enzyme/substrate systems were considered: urease/urea, urease/β-lactam, β-lactamase/urea, β-lactamase/β-lactam, and di-Ni substituted β-lactamase/β-lactam. Our starting reasoning reads as follows: β-lactam is likely easier to hydrolyze than urea, due to the high strain in the four-membered ring, and the diminished resonance stabilization as compared to that in urea. At the same time, Ni is likely a more catalytically potent metal than Zn in this reaction. Therefore, it seems like the nucleophilic attack in urease should be more easily accomplished on β-lactam than on urea, if the structure of the active site of urease otherwise permits. On the other hand, hydrolysis of both substrates should be more easily done by urease than by β-lactamase. Finally, urea should be less eagerly, if at all, hydrolyzed by β-lactamase. Additionally, there is a possibility that the overall structure of the urease active site is suboptimal for β-lactam binding. Hence, we also consider Ni-substituted β-lactamase, with the aim to combine the seemingly more potent metal with the protein structure optimal for binding β-lactam, to complete the picture. In this study, we are effectively on the hunt for the facts that would break the elementary logics outlined in this paragraph, which would thereby reveal some hidden evolutionary reasons for the metal selection for catalysis performed by the two studied hydrolases.

3.3.1. HYDROLYSIS BY UREASE

Figure 3.3 and Table 3.1 show the reaction profiles and associated energetic for the nucleophilic attack on two substrates catalyzed by urease. For both substrates, the binding starts from monodentate coordination of the carbonyl O to Ni1, and then evolves into the bidentate coordination where the amide N of the substrate attaches to Ni2. The energy goes down upon bidentate coordination for both substrates. The nucleophilic attack by O_{WB} then takes place. The barrier to this process is 14.3 kcal/mol for urea, in a reasonable agreement with previous studies.[36] As expected, the barrier toward the nucleophilic attack on β-lactam is smaller than that on urea by ca. 6.8 kcal/mol. Both systems end up in the tetrahedral intermediates, which would undergo further decomposition through protonation of the amide N2 atoms and cleavage of the C2-N3 bonds. This subsequent step is more complicated, and details of various possible reaction routes are extensively described, and debated in the literature. Here, we do not consider reaction mechanisms any further, as justified in the Introduction. Our goal is to reveal the preference for Ni and Zn in these two hydrolases, and this should be seen from the nucleophilic attack step alone.

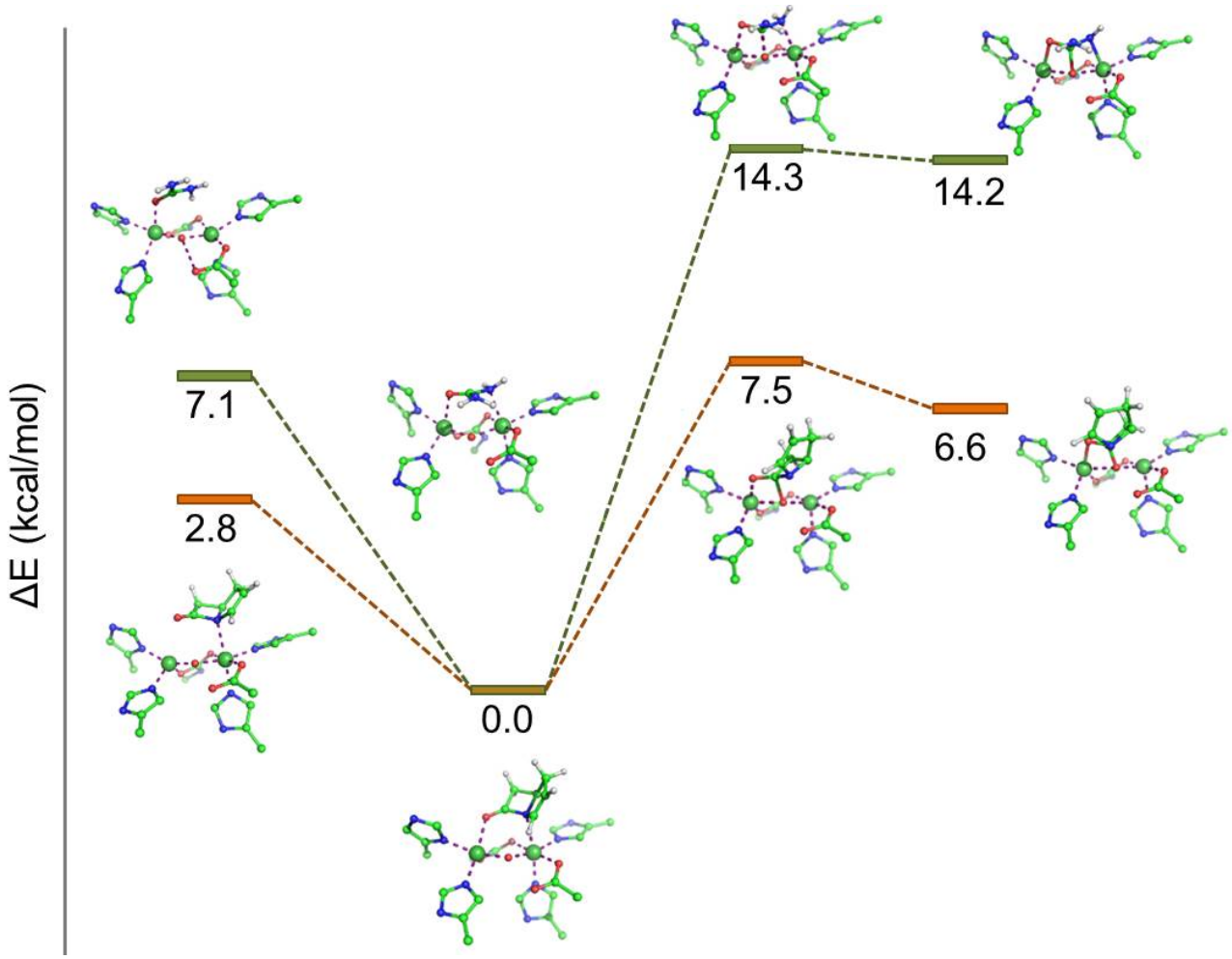


Figure 3.3. Reaction profile of nucleophilic attack catalyzed by urease with urea (green line) and β-lactam antibiotic model (orange line), in reference to bidentate coordination of substrates to nickel atoms. Most of hydrogen atoms are omitted for clarity. All energies are ZPE-corrected.

	Urease/urea				**Urease/β-lactam**			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
TPSS/ def2-SVP	7.1	0.0	14.3	14.2	2.8	0.0	7.5	6.6
B3LYP/ def2-SV(P)	5.9	0.0	17.8	17.7	1.8	0.0	12.2	9.6

Table 3.1. Relative free energy (kcal/mol) for the Urease systems (see Figure 3.3)

	Urea				β-lactam			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
Ni1-O1	2.11 (2.11)	2.31 (2.31)	2.06 (2.06)	2.03 (2.03)	3.01 (2.94)	2.22 (2.26)	2.09 (2.07)	2.08 (1.96)
O1-C2	1.26 (1.26)	1.25 (1.25)	1.31 (1.30)	1.32 (1.32)	1.21 (1.20)	1.22 (1.21)	1.28 (1.27)	1.40 (1.33)
C2-O_{WB}	2.81 (2.09)	2.57 (2.57)	1.69 (1.71)	1.61 (1.61)	2.95 (2.98)	2.97 (3.00)	1.78 (1.82)	1.59 (1.51)
C2-N3	1.36 (1.36)	1.40 (1.40)	1.47 (1.47)	1.48 (1.48)	1.44 (1.43)	1.41 (1.41)	1.44 (1.43)	1.47 (1.50)
N3-Ni2	3.08 (3.08)	2.25 (2.25)	2.13 (2.16)	2.12 (2.12)	2.43 (2.45)	2.46 (2.49)	3.77 (3.88)	3.77 (3.98)
Ni1-O_{WB}	1.96 (1.96) 2.15 [a] 1.991[b]	1.91 (1.91)	2.08 (2.06)	2.11 (2.11)	1.87 (1.88)	1.91 (1.91)	2.08 (2.07)	2.01 (2.18)
O_{WB}-Ni2	1.93 (1.93) 1.98[a] 2.071[b]	1.97 (1.97)	2.05 (2.06)	2.06 (2.06)	1.97 (1.98)	1.99 (1.99)	2.01 (2.02)	2.16 (2.05)
Ni1-Ni2	3.51 (3.51) 3.59[a] 3.557[b]	3.43 (3.43)	3.63 (3.64)	3.66 (3.66)	3.44 (3.48)	3.44 (3.46)	3.63 (3.63)	3.63 (3.65)

Table 3.2. Interatomic distances for the nucleophilic attack catalyzed by urease, for urea and β-lactam, in Å, with TPSS/def2-SVP and B3LYP/def2-SV(P) (in parenthesis). All atoms are numbered as in Figure 3.2. Structural data compared to [a] X-ray structure of native Urease (PDB 1FWJ) and [b] cluster calculations from ref. 35 where both do not contain substrate.

Table 3.2 illustrates the characteristic geometric parameters in the active site, and their evolution from the reactants, to TS, then to the tetrahedral intermediate. As the substrates binds to urease, and go into the bidentate coordination, the Ni1-O1 becomes longer (2.1-2.3 Å), allowing ideal positioning for Ni2-N3 coordination. This tight coordination to Ni1 also persists throughout the nucleophilic attack step. Upon binding of the substrates, the carbonyl C-O bond distances (R(O1-C2)) elongate slightly, by 0.02-0.04 Å, as compared to those in the unbound substrates, and become significantly longer once the attack is accomplished. Binding to Ni1 and

activation of carbonyl are very similar between the two substrates. Therefore, the smaller barrier to the reaction in the case of β-lactam is due solely to the properties of the β-lactam molecule (high strain, reduced p-resonance). The C2 and O_{WB} atoms are found to be in reactive orientations in the starting complexes, and naturally R(C2-O_{WB}) becomes shorter in the course of the reaction. For urea, the TS happens relatively late, at smaller R(C2-O_{WB}), which correlates with the higher barrier to the reaction, and the tetrahedral intermediate corresponding to a very shallow and high energy minimum on the potential energy surface. For β-lactam, the TS is early, at longer R(C2-O_{WB}), is lower in energy, and the intermediate is more stable.

3.3.2. Hydrolysis by β-lactamase

Now, let us consider hydrolysis of the two substrates by the CcrA β-lactamase, in its natural di-Zn form, and in the di-Ni substituted form. Figure 3.4 shows the reaction profiles for the nucleophilic attack occurring in these systems. In this case, the system still evolves from monodentate to bidentate coordination, which in turn precedes the attack. However, the bidentate complexes are higher in energy than the monodentate ones for the natural di-Zn enzyme, possibly due to the lower affinity of Zn to the electrophilic groups of the substrates, as compared to Ni. The TS for urea hydrolysis is high in energy (19.6 kcal/mol), and definitely much higher than when this reaction is catalyzed by urease, as expected. For β-lactam, the reaction barrier is 15.7 kcal/mol. Apparently, β-lactamase is unable to outdo urease in facilitating the nucleophilic attack on β-lactam, which at least in part is attributable to the higher potency of Ni1. However, one must also remember that the active site of urease is probably suboptimal for binding the TS of large β-lactam containing molecule. In order to account for this latter effect, we considered Ni-substituted β-lactamase. Replacing Zn with Ni in the β-lactamase active site leads to the reduction of the barrier to the attack. As anticipated, di-Ni substituted β-lactamase is the strongest catalyst for the nucleophilic attack on β-lactam. Furthermore, it is known that the

nucleophilic attack is also the rate determining step of enzymatic hydrolysis of β-lactam.[18-20] the product of the nucleophilic attack is lower than that for the native β-lactamase. Therefore, our results suggest that having Ni in the active site of β-lactamase would be better for both the nucleophilic attack, and overall enzyme performance than having Zn.

To support the qualitative observations presented in the previous paragraph, geometric parameters characteristic of the nucleophilic attack on the two substrates catalyzed by β-lactamases are shown in Tables 3 and 4. Overall, the parameters of the active site and the substrate are similar to those observed in the case of urease. The coordination to Zn1 is looser, as judged from the longer R(Zn1-O1). However, the carbonyl group still gets elongated and thus activated. This activation is stronger for β-lactam than for urea. The C2 and O_{WB} atoms arrange into reactive mutual orientations at bidentate coordination. R(C2-O_{WB}) in the bidentate complex for β-lactam is longer than that for urea. Not surprisingly, urea is a smaller molecule then the lactam antibiotic model and when arranged in the bidentate mode coordination with a short N3-Zn2 bond, it comes in close proximity to WB. The TS for β-lactam happens later, in terms of R(C2-O_{WB}), which could be viewed as the reaction coordinate. According to our calculations, there does not exist a tetrahedral intermediate minimum for the β-lactamase/urea complex. Instead, when the normal mode of the imaginary frequency in the TS is followed toward the products of the nucleophilic attack, the substrate detaches from the active site, and the system evolves back to the unbound reactants. In fact, the very existence of the TS may be an artifact of the harmonic approximation in calculations of vibrational frequencies. Nevertheless, considering the energetic of the process, this result suggests that β-lactamase is unable to hydrolyze urea.

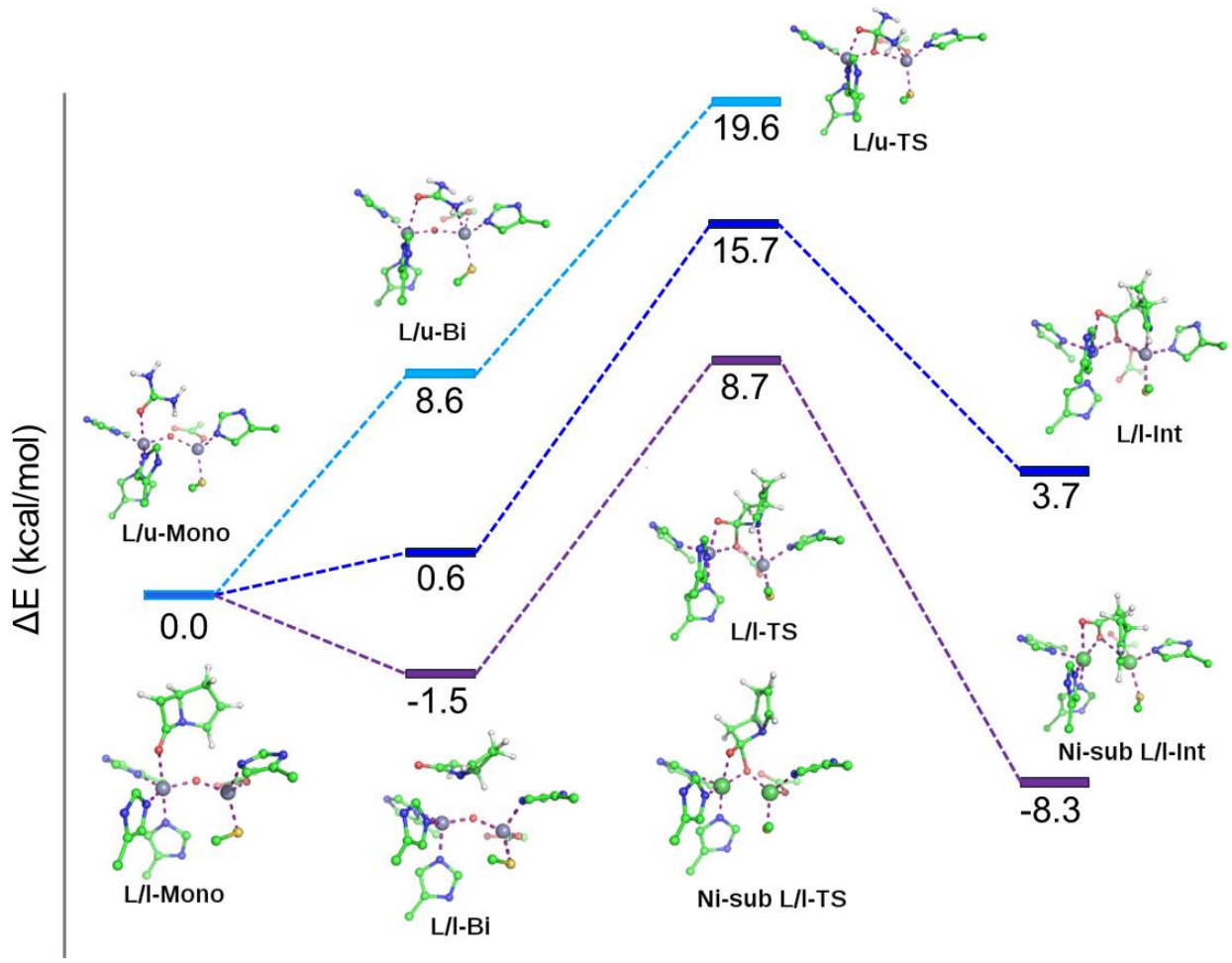


Figure 3.4. Reaction profile of nucleophilic attack catalyzed by β-lactamase with urea (light blue line), and β-lactam molecule (dark blue line), and Ni-substituted β-lactamase with β-lactam (purple line). Energies are shown relative to the monodentate complexes, and are ZPE-corrected. Structures for Ni-substituted lactamase are similar to di-Zn β-lactamase, and were omitted for clarity, except for the intermediate, which is different in the two enzymes. Most of hydrogen atoms are omitted for clarity.

Table 3.5 contains analogous data on the Ni-substituted β-lactamase. The properties of this active site appear to be similar to those of di-Zn β-lactamase. The only prominent differences include tighter binding of the substrate to Ni1, tighter Ni1-OH-Ni2 bridge, a slightly longer C2-O_{WB} distance, and a slightly shorter C2-N3 distance in the intermediate. Overall, this correlates with expected stronger electrophilic character of Ni versus Zn.

	β-Lactamase/urea				β-Lactamase/β-lactam			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
TPSS/def2-SVP	0.0	8.6	19.6	n/a[a]	0.0	0.6	15.7	3.7
B3LYP/def2-SV(P)	0.0	9.3	29.2		0.0	0.1	18.4	2.2

Table 3.3. Relative free energy (kcal/mol) for the β-lactamase systems (see Figure 3.4).

	Urea				β-lactam			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
Zn1-O1	2.47 (2.47)	2.68 (2.81)	2.13 (2.17)	n/a[a]	2.35 (2.39)	3.77 (3.75)	2.04 (2.01)	2.40 (2.48)
O1-C2	1.25 (1.25)	1.24 (1.23)	1.29 (1.28)		1.23 (1.22)	1.22 (1.21)	1.30 (1.30)	1.24 (1.23)
C2-O_{WB}	3.32 (3.32)	2.68 (2.74)	1.66 (1.66)		3.43 (3.45)	3.76 (3.75)	1.57 (1.53)	1.35 (1.33)
C2-N3	1.36 (1.36)	1.41 (1.40)	1.51 (1.50)		1.38 (1.38)	1.40 (1.40)	1.52 (1.50)	2.72 (2.84)
Zn1-O_{WB}	1.94 (1.94)	1.89 (1.89)	2.04 (2.02)		1.90 (1.90)	1.89 (1.89)	2.13 (2.13)	2.03 (2.01)
N3-Zn2	3.84 (3.84)	2.43 (2.50)	2.14 (2.15)		4.52 (4.53)	4.19 (4.14)	2.86 (2.95)	1.96 (1.91)
O_{WB}-Zn2	2.01 (2.01)	1.99 (1.99)	2.30 (2.29)		1.96 (1.96)	1.97 (1.97)	2.06 (2.08)	2.21 (2.31)
Zn1-Zn2	3.67 (3.67)	3.63 (3.63)	4.02 (3.97)		3.49 (3.51)	3.44 (3.46)	3.59 (3.65)	3.33 (3.74)

Table 3.4. Interatomic distances for the nucleophilic attack catalyzed by β-lactamase, for urea and β-lactam, in Å, with TPSS/def2-SVP and B3LYP/def2-SV(P) (in parenthesis). All atoms are numbered as in Figure 3.2.[a] See text for details.

	Monodentate	Bidentate	TS	Interm.
TPSS/def2-SVP	0.0	-1.5	8.7	-9.3
B3LYP/def2-SV(P)	0.0	-0.9	12.4	-4.8

Table 3.5. Relative free energy (kcal/mol) for the Ni-substituted β-lactamase systems (see Figure 3.4) where the thermal correction was added to the TPSS/def2-SVP data.

	Monodentate	Bidentate	TS	Interm.
Ni1-O1	2.30 (2.24)	3.69 (3.73)	2.11 (2.03)	2.36 (2.42)
O1-C2	1.23 (1.22)	1.21 (1.20)	1.27 (1.30)	1.25 (1.24)
C2-O_{WB}	3.20 (3.24)	3.66 (3.83)	1.80 (1.54)	1.32 (1.31)
C2-N3	1.38 (1.37)	1.44 (1.43)	1.44 (1.48)	2.58 (2.62)
Ni1-O_{WB}	1.92 (1.91)	1.91 (1.91)	1.96 (2.03)	2.14 (2.11)
N3-Ni2	4.23 (4.34)	3.66 (3.90)	3.49 (3.47)	1.99 (2.02)
O_{WB}-Ni2	1.97 (1.99)	1.96 (1.99)	1.98 (2.06)	2.52 (2.73)
Ni1-Ni2	2.90 (3.23)	2.89 (3.25)	2.76 (2.88)	3.03 (3.14)

Table 3.6. Interatomic distances for the nucleophilic attack on β-lactam catalyzed by Ni-substituted β-lactamase, in Å. All atoms are numbered as in Figure 3.2, except that Zn1 is now Ni1, and Zn2 is Ni2.

There is one critical difference to be observed between the ways that all three proteins interact with β-lactam. In natural and Ni-substituted β-lactamases, as soon as the attack on the carbonyl is complete, the C2-N3 bond splits spontaneously. This effect is not observed for urease, but prominent for both β-lactamases. In natural β-lactamase, it is additionally facilitated by the weakness of the Zn1-OH-Zn2 bridge. Notice that the O_{WB}-Zn2 and O_{WB}-Ni2 both increase significantly when the nucleophilic attack is complete. The Zn2-N3 distance in β-lactamase is comparable to the Ni2-N3 bond in the Ni-substituted enzyme. Recalling that the nucleophilic attack is the rate determining step in the hydrolysis of β-lactam,[17-19] the subtle energy difference between the tetrahedral intermediates in the two β-lactamases is probably unimportant for the overall performance of the enzymes.

On the other hand, apparently, the active site of urease is not suited for the splitting of the C2-N3 bond in β-lactam. At the same time, the C2-N3 bond is stronger in urea, and so no

active site can split it spontaneously. This agrees with the previous reports suggesting that the splitting of the C2-N3 bond and elimination of ammonia is the rate-determining step of enzymatic hydrolysis of urea.

To summarize this Section, we infer that having Ni in the active site of hydrolases is probably better for the overall enzymatic performance than having Zn. At least the nucleophilic attack on the substrate is much more strongly facilitated by Ni.

3.4. DISCUSSION AND CONCLUSIONS

This theoretical study explored the possible catalytic mechanistic reasons for the evolutionary choice of metals in two amide hydrolases: urease and β-lactamase. The active sites of these enzymes are similarly structured, except that the former uses two Ni cations, and the latter uses two Zn cations, to catalyze very similar reactions. Urease is unique among hydrolases to use Ni, and that puzzled the scientific community since the discovery of urease. However, from the inferred higher catalytic potency of Ni, we rather wondered why no more hydrolases use Ni, and why the seemingly less potent Zn cation is often used instead. We do not address all the factors present in the cell that may put pressure on the specific choice for the metal. Rather, we would like to know if the mechanism and rate of the catalyzed reactions are fully optimized with the given choice for the metals, or if instead they are compromised.

We considered the first step of hydrolysis, the nucleophilic attack, in which the metal center should play the key role of positioning and activating the carbonyl group of the substrates. We considered the hydrolysis of urea and a mimic of a β-lactam antibiotic in the active sites of urease, β-lactamase, and Ni-substituted β-lactamase. We found that β-lactam is most easily hydrolyzed by Ni-substituted β-lactamase, exhibiting the lowest barrier to the nucleophilic attack, and also the spontaneous splitting of the C2-N3 bond in the substrate, after the attack. The latter is also characteristic of di-Zn β-lactamase, but the barrier to the attack is

higher. The barrier in urease is smaller than that in natural β-lactamase, but the C2-N3 bond does not spontaneously open in the urease active site. On the other hand, hydrolysis of urea in di-Zn β-lactamase is unfeasible. Thus, hydrolysis of urea requires Ni, and cannot be done with Zn, but hydrolysis of β-lactam can benefit from the presence of Ni rather than Zn at the active site.

Nature's preference for Zn over Ni in the active site of β-lactamase may be dictated by other factors in the cell, such higher toxicity or lower availability of Ni. On the other hand, considering the already high performance of di-Zn β-lactamases in hydrolyzing β-lactam, it is possible that they are simply "good enough" as catalysts, while satisfy other requirements present in the cell, and so no further optimization by mental substitution was pursued. From the fundamental prospective of protein science, this suggests that natural enzymes could be further improved, especially if the intention is to use them for catalysis *in vitro*, rather than *in vivo*. This message is also of interest to the field of artificial enzyme design.

3.5. REFERENCES

1 Zambelli, B.; Musiani, F.; Benini, S.; Ciurli, S. *Acc. Chem. Res.* **2011**, *44,* 520-530.

2 Karplus, P. A.; Pearson, M. A.; Hausinger, R. P. *Acc. Chem. Res.* **1997**, *30,* 330-337.

3 Krajewska, B. *J. Mol. Catal. B* **2009**, *59*, 9-21.

4 Callahan, B. P.; Yuan, Y.; Wolfenden, R. *J. Am. Chem. Soc.* **2005**, *127,* 10828-10829.

5 Estiu, G.; Merz, K. M. *J. Am. Chem. Soc*. **2004**, *126,* 11832-11842.

6 Alexandrova, A. N.; Jorgensen, W. L. *J. Phys. Chem. B* **2007**, *111,* 720-730.

7 Zambelli, B.; Danielli, A.; Romagnoli, S.; Neyroz, P.; Ciurli, S.; Scarlato, V. *J. Mol. Biol.* **2008**, *383,* 1129–1143.

8 Dixon, N. E.; Gazzola, C.; Blakeley, R.; Zerner, B. *J. Am. Chem. Soc*. **1975**, *97,* 4131-4132.

9 Ciurli, S.; Benini, S.; Rypniewski, W. R.; Wilson, K. S.; Miletti, S.; Mangani, S. *Coord. Chem. Rev.* **1999**, *190-192,* 331-355.

10 Page, M. I., *The Chemistry of β-Lactams,* Ed; Blackie: London, 1992.

11 Estiu, G.; Suárez, D.; Merz, K. M. *J. Comp. Chem.* **2006**, *27,* 1240-1262.

12 Lipscomb, W. N.; Sträter, N. *Chem. Rev*. **1996**, *96,* 2375-2433.

13 T. Abraham, E. P.; Waley, S. G. *The β-Lactamase,* Hamilton-Miller, J. M. T.; Smith, J. T., Eds.; Academic Press: London, 1979. pp. 311-338.

14 Jabri, E.; Carr, M. B.; Hausinger, R. P.; Karplus, P. A. *Science* **1995**, *268,* 998-1004.

15 Pearson. M. A.; Michel, L. O.; Hausinger, R. P.; Karplus, P. A. *Biochemistry* **1997**, *36,* 8164-8172.

16 Park, H.; Brothers, E. N.; Merz, K. M. *J. Am. Chem. Soc.* **2005**, *127,* 4232-4241.

17 Krauss, M.; Gresh, N.; Antony, J. *J. Phys. Chem. B* **2003**, *107,* 1215-1229.

18 Estiu, G.; Suárez, D.; Merz, K. M. *J. Comput. Chem*. **2007**, *27,* 1240-1262.

19 Xu, D.; Guo, H.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129,* 10814-10822.

20 (a) Wilcox, D. E. Chem. Rev. 1996, 96, 2435-2458. (b) Clark, P. A.; Wilcox, D. E. Inorg.

21 Chem. 1989, 28, 1326-1333. (c) Dixon, N. E.; Gazzola, C.; Asher, C. J.; Lee, D. S. W.; Blakeley, R. L.; Zerner, B. Can. J. Biochem. 1980, 58, 474-480. (d) Blakeley, R. L.; Dixon, N. E.; Zerner, B. Biochim. Biophys. Acta 1983, 744, 219-229. (e) Clark, P. A.; Wilcox, D. E. Inorg. Chem. 1989, 28, 1326-1333. (f) Day, E. P.; Peterson, J.; Sendova, M. S.; Todd, M. J.; Hausinger, R. P. Inorg. Chem. 1993, 32, 634-638. (g) Suarez, D., Diaz, N., Merz, K. M. *J. Am. Chem. Soc*. **2003**, *125,* 15324-15337.

22 Musiani, F.; Arnold, E.; Casadio, R.; Ciurli, S. *J. Biol. Inor. Chem*. **2001**, *6,* 300-314.

23 Pearson, M. A.; Park, I.-S.; Schaller, R. A.; Michel, L. O.; Karplus, P. A.; Hausinger, R. P. *Biochemistry* **2000**, *39,* 8575-8584.

24 Benini, S.; Rypniewski, W. R.; Wilson, K. S.; Miletti, S.; Ciurli, S.; Mangani, S. *Struct. Fold. Des.* **1999**, *7,* 205–216.

25 Benine, S.; Rypniewski, W. R.; Wilson, K. S.; Miletti, S.; Ciurli, S.; Mangani, S. *Structure* **1999**, *7,* 205–216.

26 Sparta, M.; Ding, F.; Shirvanyants, D.; Dokholyan, N. V.; Alexandrova, A. N. *Biophys. J*. **2012**, 103*, 767-776.*

27 Concha, N.O.; Rasmussen, B.A.; Bush, K.; Herzberg, O. *Structure* **1996***, 4,* 823-836*.*

28 Kabsch, W. *Acta Crystallogr. A* **1976**, 32, 922-923.

29 Barton, G.J. (1993, 2002) "OC - A cluster analysis program", University of Dundee, Scotland, UK; http://www.compbio.dundee.ac.uk/downloads/oc.26. 29. (a) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992***, 45, 13244-13249.* (b) Tao J., Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91,* 146401-146404.

30 Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97,* 2571-2577*.*

31 Grimme, S. *J. Comput. Chem*. **2004**, *25,* 1463-1473.

32 Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162,* 165-

169.

33 Diaz, N.; Suárez, D.; Merz, K. M. *J. Am. Chem. Soc*. **2000**, *122,* 4197-4208.

34 Suárez, D.; Brothers, E. N.; Merz, K. M. *Biochemistry* **2002**, *41,* 6615-6630.

35 (a) R. G. Parr, W. Yang, Density-functional theory of atoms and molecules, Oxford Univ. Press, Oxford, 1989. (b) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. (c) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

36 Suarez, D.; Diaz, N.; Merz, K. M. *J. Am. Chem. Soc*. **2003**, *125*, 15324-15337.

CHAPTER 4

The Role of the Flexible L43-S54 Protein Loop in the CcrA Metallo-β-lactamase in Binding Structurally Dissimilar β-lactam Antibiotics

4.1. INTRODUCTION

Antibiotic resistant bacteria are a growing worldwide concern.[1] The production of β-lactamases (EC 3.5.2.6) is the most common cause of antibiotic resistance in clinically relevant bacteria. These enzymes disable the most potent antibiotics through the hydrolysis of the b-lactam ring. β-lactamases are classified according to their amino acid sequence into 4 major classes (A-D). The enzymes in Class A, C and D share a serine-based mechanism, whereas enzymes of class B depend on the presence of one or two zinc ions in the catalytic center and are known as metallo-β-lactamases.[2,3] For example, zinc-β-lactamases are responsible for antibiotic resistance in such pathogenic bacteria as *Bacteroides fragilis,*[4] *Bacillus cereus,*[5] *Stenotrophomonas meltophilia,*[6] *Pseudomonas and Klebsiella,*[7] *Bradyrhizobium japonicum*,[8] and *Pseudomonas aeruginosa*.[9] Among them, the di-Zn β-lactamase from *Bacteroides fragilis* (CcrA) considered in this study is one of the most versatile and dangerous,[10] due to its capability of hydrolyzing a broad spectrum of antibiotics at the diffusion-limited rate.[11]

One of the open questions regarding the functionality of CcrA is the mechanistic and structural role of residues 43-54 in its structure. These residues form a flexible hydrophobic β hairpin loop located above the active site in CcrA. It is speculated that the loop contributes to the promiscuity of CcrA and its ability to bind and hydrolyze antibiotics of diverse structures, by molding around the docked substrates. Several studies have reported on the flexibility of the b hairpin loop. Backbone heteronuclear NOE measurements indicate greater flexibility on the pico- to nano-second timescale when compared to the entire protein.[12] Furthermore, the loop shows greater flexibility in the free CcrA compared to the inhibitor bound complexes.[13] The crystal structure of CcrA bound to a tricyclic inhibitor exhibited obvious sandwiching of the inhibitor between the active site and flexible β-hairpin loop, specifically around the GWG region.[14] Mutagenic studies of the tryptophan in the GWG region confirmed the existence of large hydrophobic contributions to the non-specific antibiotic/loop interactions.[15] Merz and

coauthors investigated the role of the loop with molecular dynamics (MD), and observed the loop's mobility and adjustment in response to the bound substrate.[16] However, purely molecular mechanical MD was unable to predict chemically meaningful binding to the metal centers, and the substrate was located over 5 Å above the active site, i.e. in a nonreactive orientation. This showed that a quantum mechanical (QM) treatment of the active site is required to produce chemically meaningful docking poses. Xu et al. investigated the Michaelis complex of moxalactam in the active site of the L1 lactamase, by means of QM/MM MD, but could not report on the role of the loop due to the short time scale of their simulation.[17] In general, elucidating the role of the loop in catalysis is a challenge, both for theory and for experiment. Experimentally capturing the complexes of b-lactamases with their native substrates is virtually impossible, due to the high catalytic efficiency of these enzymes. In theoretical modeling, quality docking that would assess the role of the loop requires both the quantum mechanical treatment of the di-Zn active site and extensive sampling of the protein structure, including the backbone and the 43-54 loop in particular.

Here, we aim at elucidating the role of the 43-54 protein loop in binding three different substrates (imipenem, ampicillin, and cephalorodine) to the active site of CcrA. Each of these molecules is a representative for a different class of b-lactam antibiotics: carbapenems, penicilins, and cephalosporins, respectively. Comparing loop interactions with a broad spectrum of substrates will elucidate the role of the loop and may contribute to the design of new generation CcrA-resistant antibiotics and inhibitors. In this study, we used our newly developed QM/DMD methodology that allows for flexible dynamic docking to metalloenzymes, with the inclusion of the quantum mechanical treatment of the metal centers. We then refine the binding energies via high level *ab initio* calculations. We elucidate the role of the flexible protein loop in binding structurally diverse substrates.

4.2. THEORETICAL METHODS

4.2.1. SYSTEM PREPARATION

The crystal structure of CcrA (1.85 Å resolution PDB code, 1ZNB)[18] was used as a starting point in the present investigation. Although it crystalizes as a dimer, CcrA operates as a monomer and was simulated accordingly (monomer A was used in this study). Crystallographic water molecules were removed, except for the two of them directly coordinated to the active site. The active site of CcrA contains two non-equivalent zinc ions connected by a bridging hydroxyl group (W1). Three histidine residues (H99, H101, and H162) complete the tetrahedral coordination of the first zinc (Zn1), whereas the second metal (Zn2) is coordinated by D103, C181, and H223, and a water molecule (W2) arranged in a trigonal bipyramid. This arrangement allows Zn1 to accept an additional ligand, (i.e. going from tetrahedral to trigonal bipyramid). In fact, Merz et al. demonstrated, through QM/MM calculations, that the binding mode of nitrocefin involves the interaction between the carbonyl oxygen atom of the β-lactam ring and Zn1.[19] In such conformation, Zn1 polarizes the carbonyl group, and WB then acts as a nucleophile, attacking the carbonyl carbon atom, in the first step of hydrolysis. These ideas about substrate binding contribute to our initial guess for the CcrA-antibiotic complexes.

The loop located above the active site consists of the residues L43, A44, E45, I46, E47, G48, W49, G50, M51, V52, P53, and S54. The crystal structure did not contain residues G48 and W49, which therefore were manually built in. Upon the initial manual placement of the substrate molecules into the active site, the loop was not in contact with the substrates. Furthermore, the conformations of the side chains holding the Zn centers captured in the X-ray structure do not fully allow for the proper docking of the studied substrates in reactive orientations with respect to Zn1 and the nucleophilic hydroxyl group. In the course of QM/DMD simulations, the results of which are presented below, the substrates went deeper into the active site and the loop was found to accommodate this binding.

The chemistry of the system dictates the choice for the QM/DMD partitioning of the system in the simulations. In QM/DMD, the system is partitioned onto three regions: the QM-only region moved by QM only, QM-DMD region that can be moved by both QM and DMD, and the rest of the system moved only by DMD (Figure 4.1A). The QM-DMD boundary is fluxional, and it can go around the QM-only region, or around the QM-DMD region, depending on the stage in the simulation. The QM-only region includes the two Zn centers and atoms immediately attached to them, belonging both to the substrate and to the amino acid fingers. Furthermore, additional constraints were imposed to prevent any of the residues coordinating zinc to become too distorted and lose their coordination during DMD phases. A few distances between Zn cations and certain atoms in the active site were allowed to be sampled by DMD only within ±0.01 Å from the values predicted by QM. The specific atoms are in the histidine residues, the two carbons adjacent to N coordinating Zn, the gamma carbon of the aspartate, and the beta carbon of the cysteine (Figure 4.1B). The QM-only region is required in order to avoid any parameterization of the classical force field for the metals, and to have instead their fully quantum mechanical description. The fuller active site contains the amino acid ligands (up to Ca atoms) coordinating to the two Zn cations, W1, and the substrate constituted the QM-DMD region. The rest of the protein, including the 43-54 protein loop, was treated exclusively with DMD. We recently applied QM/DMD to study a closely related system, CcrA with a smaller mimic of the β-lactam antibiotic, and obtained results for structure and reaction energies in close agreement with available theoretical and experimental data.[20]

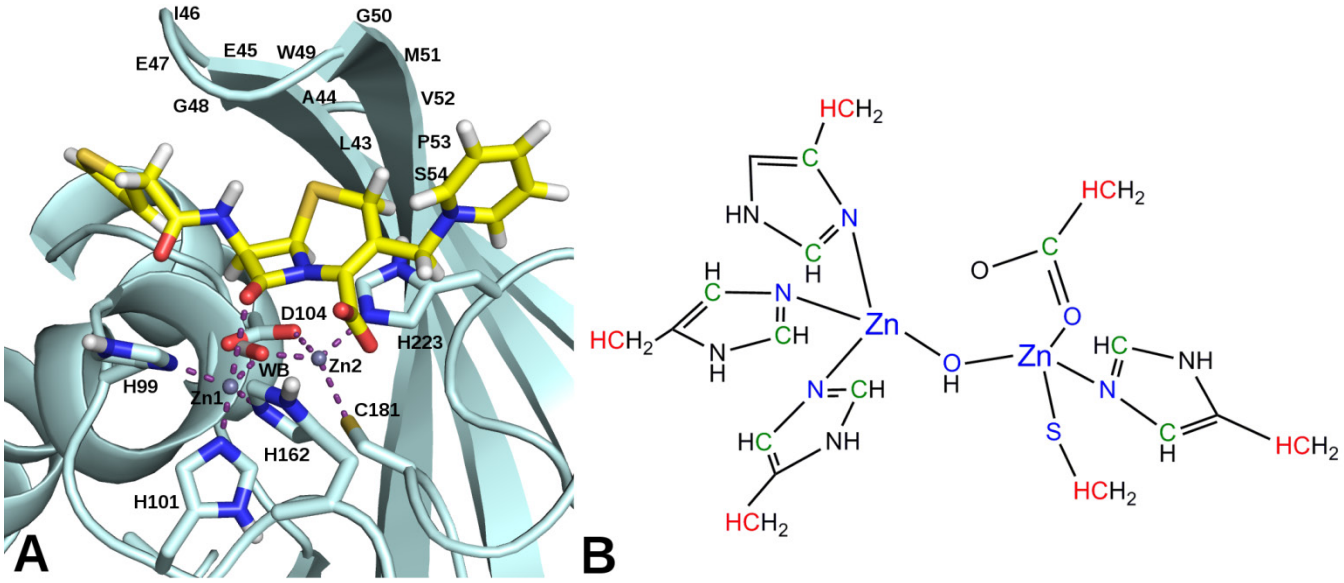


Figure 4.1. CcrA with loop, active site and docked antibiotic cephalorodine are shown. (A) In the QM/DMD simulations, the region shown in sticks constituted the QM-DMD region. (B) The green atoms show additional constrains that are imposed during the DMD simulation. To retain the chemistry determined at the QM level of theory, the red atoms are frozen during QM calculations, and the blue color marks the atoms in the QM-only domain.

4.2.2. QM/DMD SIMULATIONS

QM/DMD is our newly developed method that involves a QM description of the active site coupled with a discrete molecular dynamics (DMD) simulation of the entire system, including the protein backbone.[21] Briefly, DMD is a type of molecular dynamics based on solving equations of conservation of energy and momentum rather than Newtonian equations of motion.[22] It operates on a set of square well potentials, where atoms interact only as long as they are within a given distance from each other, defined by the square well. DMD is a fast and exceptionally successful method, capable of rapid sampling and simulations of protein folding and aggregation, and routinely covering time-scales on the order of tens of microseconds at least for small peptides (ca. 60 residues).[23] Density Functional Theory (DFT) was used for the QM part of simulations. The functional chosen is a pure meta-GGA functional TPSS[24] with Grimme's empirical dispersion correction[25] (TPSS-D) in conjunction with double zeta type basis sets[26] implemented in the Turbomole software package.[27] In QM calculations, the charge of the

active site is -1 for imipenem and cephalorodine, and -2 for ampicillin, and all the active sites are singlets as predicted through our preliminary calculations.

QM/DMD simulations are performed in an iterative fashion. Each run starts from DMD operating on most of the protein, including the backbone, except for the metal centers and atoms constituting their immediate coordination sphere (QM-only region). Simulations are done using the annealing temperature profile: the temperature is initially ramped up and gradually reduced for 3,500 time units (t.u.) of DMD with each time unit corresponding roughly to 50 fs. After that, the temperature is kept low for 10,000 t.u., and at that stage, data is collected. Annealing was found to help improve sampling and statistics of the resulting ensemble. The DMD-produced ensemble is clustered according to geometric similarity. In the present study, we used five clusters. Geometric centroids and lowest-energy structures are then selected from each cluster, hence a total of ten structures are selected to represent the outcome of the current DMD phase. The QM-DMD regions are then cut out of the protein, and capped with H atoms, to satisfy all dangling valencies. The single point QM energies of the active sites are calculated. The structures are then scored, based on both the QM energy of the QM-DMD region and the DMD energy of the entire system minus the metal centers, and the smallest values of the two found in the given set of ten structures, according to the formula:

$$SI^i = n(E^i_{DMD} - E^{min}_{DMD}) + (1-n)(E^i_{QM} - E^{min}_{QM})$$

In this work, n is set to 0.5. The structure having the smallest SI is chosen, and its active site is partially optimized quantum mechanically (for 200 steps), with the positions of the atoms bridging the QM region to the rest of the protein (Ca atoms) being fixed. Truncated optimizations were found to be adequate to reach sufficient convergence, and yet save time and increase the robustness of the algorithm. After that, the active site is uncapped and reinstalled in the protein. The QM-DMD boundary shrinks back to going around the QM-only region. The simulation continues with DMD, and at this point, the QM predicted changes in the active site have a

chance to propagate to the rest of the system, in particular to the position of the protein loop of interest. QM and DMD jobs alternate to convergence in the protein structure and QM and DMD energies. The use of the fluxional QM-DMD boundary permits extensive and algorithmically simple communication between the two regions. The advantage of using the gradient following geometry optimization at the QM level instead of *ab initio* dynamics is also the gain in computational speed. The QM/DMD method is computationally affordable, while reliable for modeling of metalloproteins at a variety of scales from large (>10 Å), to intermediate (active site structure), to small (electronic level). Its success was recently demonstrated on such examples as b-lactamases, urease, catechol-O-methyl transferase, and rubredoxin.[20,21,28] QM/DMD recapitulated native metalloenzyme structures, predicted structural response of the proteins to metal replacement and sequence mutations, recovered distorted proteins to the equilibrium, and generated ensembles of structures suitable for quantitative predictions of electronic properties and redox activities of metalloenzymes. In particular, QM/DMD can be used for mixed quantum-classical flexible docking. Details of the method are described elsewhere.[21]

QM/DMD simulations were run for the CcrA protein without substrate and for the complexes with each antibiotic in duplicates. Each simulation consisted of 20 iterations alternating between DMD and QM. This corresponded to approximately 10.5 ns of dynamics for each protein. The exact timing of simulations is not extractable, because the QM part of the simulations is not done as dynamics, and the timing covered by QM may vary significantly between iterations. Finally, a long DMD-only simulation was run on the equilibrated structure, with the QM-only region fixed, in order to accumulate more statistics on the position of the loop.

4.2.3. LOOP/ANTIBIOTIC INTERACTIONS

QM/DMD produced sets of structures, each resulting from the end of a single iteration, after the DMD stage. These structures (collected after the convergence within a simulation was

reached) were used to calculate the interaction energies between the substrate and the protein loop, at higher levels of *ab initio* theory. Structures were chosen to be collected after the DMD phase because the protein loop of interest was not part of the QM-DMD region, and so its position was adjusted only by DMD. From each structure, we extracted the full active site with substrate and residues I46-W49 from the loop, since these were the residues that contribute to the binding of the substrate directly, in the QM/DMD generated geometries. Maintaining the geometries, we calculated the interaction energies between the active site/substrate complex and the loop using the TPSS-D3 functional with triple zeta type basis sets[23] in both gas phase, and solvent using the Conductor-like Screening Model (COSMO)[29] with the dielectric constant set to 78.4 since the active site is very solvent exposed. Additionally, we performed B3LYP-D3[30] and Møller-Plesset perturbation theory calculations within the RI formalism, RI-MP2,[31] with triple zeta type basis sets for the interaction energies. Our systems, due to its size, and number of intermolecular interactions, are subject to basis superposition errors (BSSE). BSSE was accounted for using the counterpoise correction (CP) procedure as laid out by Boys and Bernardi.[32] Turbomole27[7] was used for all QM calculations.

4.3. RESULTS AND DISCUSSION

4.3.1. QM/DMD SIMULATIONS OF THREE STRUCTURALL DIFFERENCE ANTIBIOTICS

For all studied systems, the loop was brought in close contact with the substrate almost immediately, within 1-2 iterations. To quantify the convergence, the backbone and all heavy atom root-mean-square deviation (RMSD) of the loop as a function of the iteration number was calculated (one snapshot was used from each iteration). Because the initial PDB x-ray structure did not contain coordinates for the two residues in the flexible loop region, G48 and W49, and those residues had to be built by hand, the resultant initial structure could not be used as a reference in calculating RMSDs. Instead, the reference structure for each system was chosen

via the following approach: the RMSD was calculated between all 20 structures with each other. The structure that gave the smallest average RMSD was chosen as the reference structure. Figure 4.2 reports the variation of RMSD with respect to this chosen structure, along the simulation. For every system, there is first a gradual reduction of RMSD, as the simulation progresses, and then stabilization of the RMSD value around the 10^{th} iteration, indicating convergence. The relatively large fluctuations of RMSD in Figure 4.2 are typical for QM/DMD, [20,21,28] because the collected structures come from the ends of the DMD phases and are scored based on the DMD and single point QM energies. These structures thus represent the way the DMD treats the system (on the set of square-well potentials), where there is a subtle mismatch between the QM and DMD potential energy surfaces.

QM/DMD was capable of efficiently moving protein parts in conjunction with QM description of the active site; the loop position becomes more consistent and fixed upon antibiotic binding. In Table 4.1, RMSD values averaged over the final 10 iterations of QM/DMD (the stage when the simulations converged) are reported for each system, in duplicates. One may see that the loop mobility varies for each antibiotic. Upon binding of ampicillin and intramolecular H-bound cephalorodine (see below), the loop mobility is indeed reduced slightly upon antibiotic binding, by ~0.5 Å for the atoms of the loop backbone, as compared to the protein without a bound substrate. For imipenem and free cephalorodine, the RMSD values do not imply loop stabilization. The all-atom RMSD of the loop calculated in the same manner are also reported in Table 4.1. Notice also that for cephalorodine, two different structures are reported; one of them contains an intramolecular H-bond, as will be described in detail below.

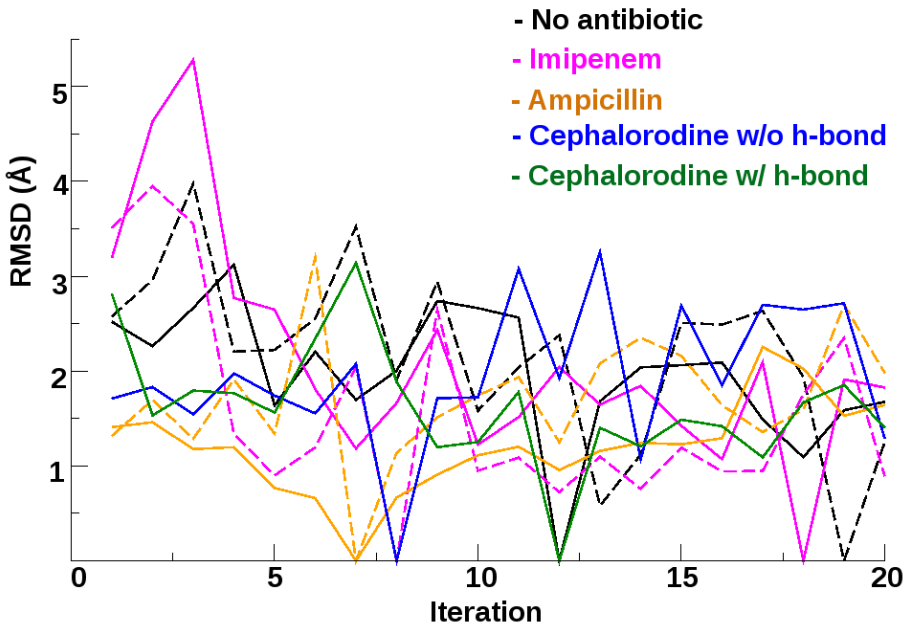


Figure 4.2. Shown are the all atom RMSDs of the active site as a function of iteration number of each antibiotic system, no antibiotic (black), imipenem (pink), ampicillin (orange), cephalospordine without constrained H-bond (blue) and cephalospordine with constrained H-bond (green). Duplicate runs are indicated by dotted lines. Each plot hits zero RMSD at the reference structure, which is different in every case (see text). The convergence of the simulations in terms of protein structure is apparent from leveling out of the plots after 10th iteration, or earlier. Results of just one of two independent simulations are shown for each system.

RMSD of loop backbone				
No Antibiotic	Imipenem	Ampicillin	Cephalorodine	Cephalorodine (H-bond)
1.2 ± 0.4 1.5 ± 0.5 (1.19 ± 0.29)[a]	1.2 ± 0.4 1.0 ± 0.3	1.1 ± 0.4 1.4 ± 0.3	1.8 ± 0.6	1.1 ± 0.4
All-atom RMSD of the loop				
1.9 ± 0.5 1.9 ± 0.7	1.7 ± 0.3 1.2 ± 0.5	1.4 ± 0.4 1.9 ± 0.5	2.3 ± 0.7	1.5 ± 0.2

Table 4.1. RMSD of loop with and without antibiotics, averaged over the final 10 iterations of QM/DMD, when the simulations are considered converged, in Å. [a]Data from reference 15.

4.3.2. OVERALL STRUCTURES OF CCRA-ANTIBIOTIC COMPLEXES PREDICTED BY QM/DMD

Our simulations predict that the CcrA β-lactamase indeed can tightly bind a wide spectrum of antibiotics, and the non-specific hydrophobic interactions provided by the protein loop above the active site contribute to this versatility of CcrA. The loop visibly adjusts its structure and position in response to the presence of different substrates, and it comes in a direct and tight contact with all three of them (Figure 4.3). The geometries of the loops adopted in each complex are distinctly different, and complementary to the bound substrates (Figure 4.3). At the atomic level, the only specific interaction found in the complexes is the one between the W49 side chain of the loop and the aromatic ring in the cephalorodine antibiotic. Specifics of this contact that helps substrate binding are explicated below. All other interactions between the loop and the substrates are hydrophobic and non-specific. Our results support the earlier proposal[13-15] that the loop molds around substrates of different shapes, upon substrate binding.

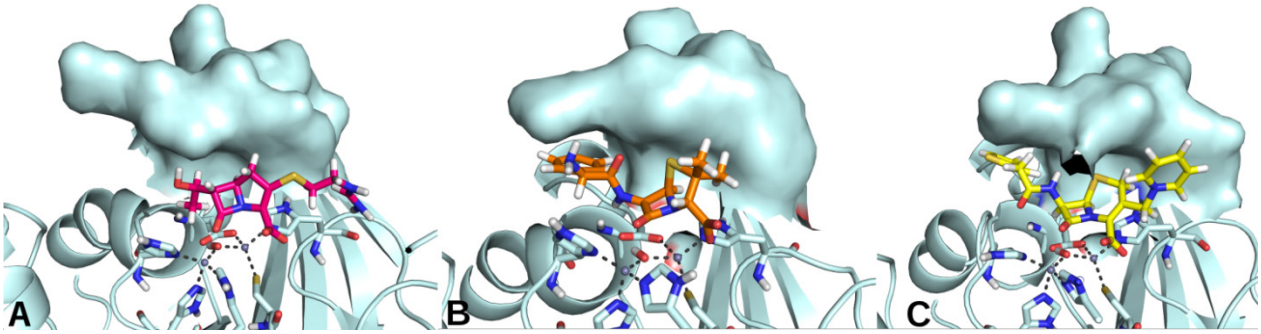


Figure 4.3. Representative snap-shots from QM/DMD simulations for imipenem (A) ampicillin (B) and cephalorodine (C) bound to CcrA show the distinctly different conformations of the residues in the hydrophobic β-hairpin loop adopted in response to binding different substrates. The loop in space filling mode is shown above the active site.

The structural parameters of the active site of CcrA not bound to any substrate were found to be close to those predicted by purely quantum mechanical calculations on small model complexes.[33] For the bound active site-substrate complexes, our simulations predict binding

that is consistent with quantum mechanical calculations on the active site model, but that is much tighter as compared to the structures produced by purely force field based molecular dynamics (Table 4.2 and Figure 4.4).[14] The carbonyl oxygen atom is located within ca. 2.5 Å from Zn1, and the carbonyl carbon atoms in within ca. 3.5 Å from the nucleophilic OH. These structural characteristics of the protein-substrate complexes are remarkably consistent within each QM/DMD ensemble, and also across the set of studied antibiotics. The average bond lengths of the lactam ring carbonyl and Zn1 of the three antibiotic systems are listed in Table 4.1. One important observation should be made about cephalorodine. In its complex with CcrA, two competitive configurations were identified. In one of them, the molecule formed an intramolecular H-bond between H1 and O3 (see Figure 4.3), and in the other one the molecule was extended. These structures are listed separately in Tables 4.1 and 4.2; detailed discussion is presented in the next section. All structures of the protein-antibiotic complexes are available upon request. They can be used as starting points for mechanistic studies.

	Imipenem	Ampicillin	Cephalorodine	Cephalorodine (H-bond)
Zn1-O1	2.9 ± 0.4 2.6 ± 0.1 (5.93 ± 0.40)[a]	2.7 ± 0.5 4.7 ± 0.3	3.3 ± 0.3	2.6 ± 0.2
C2-O_{WB}	3.1 ± 0.3 3.0 ± 0.1 (5.55 ± 0.32)[a]	3.0 ± 0.3 3.8 ± 0.3	2.8 ± 0.2	2.4 ± 0.3
Zn2-O2	5.4 ± 1.2 6.3 ± 1.2	4.4 ± 1.6 3.6 ± 0.9	9.8 ± 0.3	9.1 ± 0.3
Zn2-O3	6.1 ± 1.3 7.6 ± 1.3	5.5 ± 0.8 5.7 ± 0.7	10.9 ± 0.3	10.1 ± 0.4

Table 4.2. Listed are the important antibiotic/active site interatomic distances averaged over 20 structures extracted from the converged states of two independent QM/MD simulations for each system, in Å. Atoms are labeled as in Figure 4.4.[a] Data from reference 14.

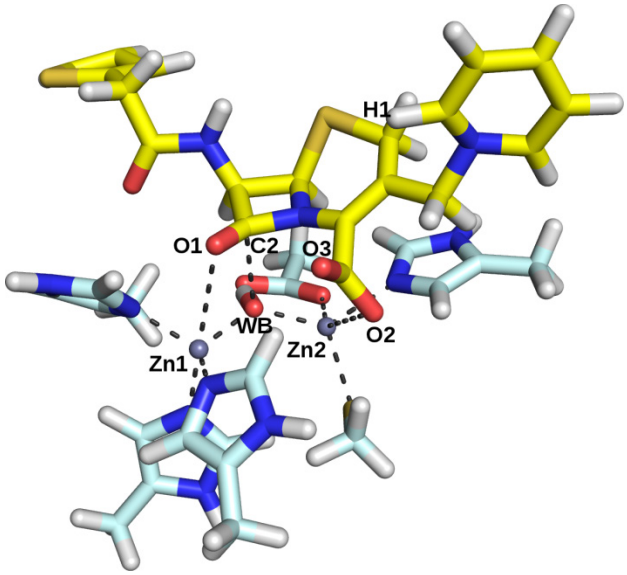


Figure 4.4. Illustrating important interatomic distances listed in Table 4.2 in the antibiotic/active site complexes (shown here with cephalorodine).

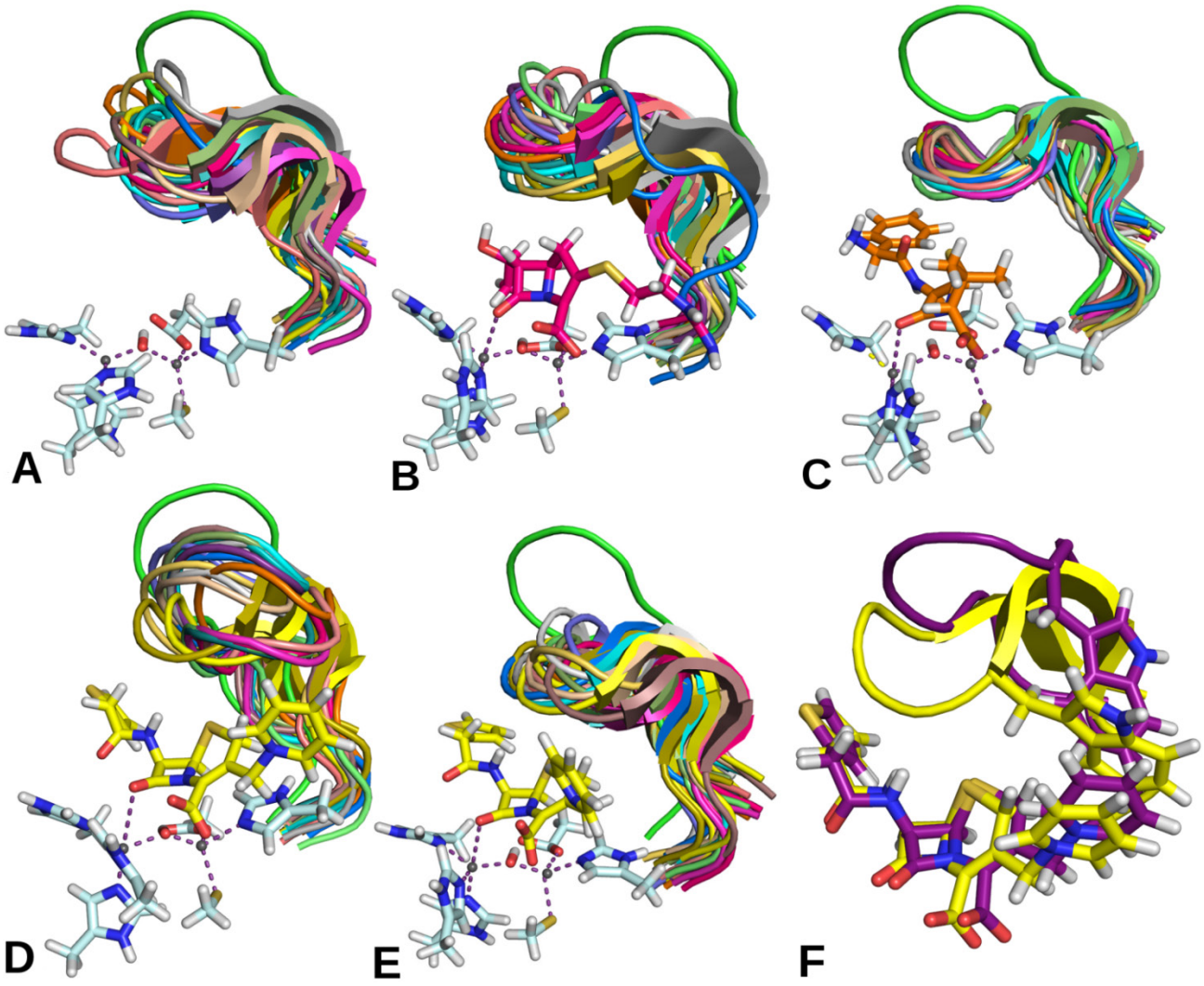


Figure 4.5. Overlaid images of the snapshots from the QM/DMD simulation illustrate the mobility and adjustment of the β-hairpin loop: (A) no antibiotic is bound, (B) imipenem, (C) ampicillin, (D) cephalorodine (simulation without a constraint), (E) cephalorodine (simulation with a constraint) bound to CcrA. (F) Two possible p-stacking orientations between W49 and the cephalorodine molecule, parallel and perpendicular, are seen in the simulation with (D). The larger mobility of the loop when no antibiotic is present at the active site can be seen. The snapshot in green is the initial conformation of the loop in each simulation.

4.3.3. POSITION AND MOBILITY OF THE LOOP IN THE CCRA ALONE, AND IN COMPLEXES WITH THREE ANTIBIOTICS

QM/DMD simulations done on the empty proteins indicate a significant mobility of the loop. In Figure 4.5, snapshots coming from QM/DMD simulations are overlaid. The snapshot shown in green is the initial structure, which is shown just for a reference, but really should not be considered. The rest of the ensemble is not well-clustered, and samples a large conformational space.

When antibiotics molecules bind to the active site, the preferred conformations and the mobility of the loop change. Different and unique loop conformations are predicted for each antibiotic up taken by CcrA. For ampicillin (Figure 4.5B and Table 4.1), the loop adopts a rigid conformation. The QM/DMD ensemble is tightly clustered. Clearly the loop rigidity is induced by the bound antibiotic. However, when imipenem is bound, the loop is not as well-structured and multiple conformations of the loop can be found, showing its flexibility (Figure 4.5C). The tail of the molecule extending toward the loop is smaller, hydrophilic and charged. Therefore, hydrophobic contacts with the loop are largely avoided by this part of the bound substrate. However, the loop molds around the rest of the substrate, as seen in Figure 4.3. The residue of the loop closest to imipenem is W49, but even it does not form any efficient interactions. Recent experiments with BlaB metallo-β-lactamase showed that both of these antibiotics are efficiently hydrolyzed by it, at a rate comparable with that of CcrA, and the smaller tail of imipenem did not have any effect on binding or turnover rate of the enzyme.[34] In agreement with this finding, the position of the substrate with respect to the Zn centers and the nucleophilic bridging water is consistent with those characteristic of other studied antibiotics (Table 4.2).

QM/DMD simulations yielded a more complicated story of the cephalorodine binding (Figure 4.5D-F). In this antibiotic, there appears to be a weak intramolecular H-bond between the carboxylic oxygen and the pyridine group hydrogen (H1-O3, in Figure 4.4). This H-bond is

not recognized by the force field implemented in DMD, because of the atypical H-bond donor involved. However, QM calculations predict this H-bond for the substrate bound to the active site. Whether it is QM or DMD that should be trusted is questionable, because despite the higher accuracy, QM calculations did not include the residues of the loop. As a result, our initial QM/DMD simulation on this complex failed to converge and showed swapping between two different conformations predicted by QM and DMD. At the same time, the pyridine group of substrate adopted two different conformations corresponding to perpendicular and parallel p-stacking interactions with the loop residue W49 (Figure 4.5D,F).

To further investigate this issue, we introduced an additional constraint on the distance between the carboxylic oxygen and the pyridine group C donating the H-bond in the DMD part of the simulation. Sampling of this distance was allowed within ±0.1Å from the value predicted by QM calculations. With this constraint, the simulation yielded a stable intermolecular H-bond, and an exclusively parallel p-stacking interaction with W49 (Figure 4.5E,F). Interestingly, constrained and unconstrained simulations yield similar QM energies, the differences within 4.5 kcal/mol, and converge to a similar DMD energy range. From this we infer that perhaps all three conformations of cephalorodine bound to CcrA may exist: with intramolecular H-bond and parallel p-stack with W49, without H-bond and with parallel p-stack, and without H-bond and with perpendicular p-stack. This versatility should contribute favorably to the binding process through configurational entropy and facilitate the hydrolysis of this molecule by CcrA. This result is somewhat paradoxical, since cephalorodine is the bulkiest of the studied antibiotic molecules. Our result agrees with an earlier report suggesting the interaction of cephalorodine with W49.[14]

In order to analyze the interactions of the loop with antibiotics, we assessed the mobility of the loop residues in each complex (Figure 4.6). It can be seen that, for each amino acid, the average all atom RMSD is the highest when no antibiotic is bound, in agreement with the reduced loop mobility upon locking onto the substrate. Also, naturally, the residues in the middle

of the loop are more mobile than the ones nears its ends. As mentioned above, the loop, when ampicillin is bound, adopts a slightly more rigid conformation and this is shown in the individual residue RMSDs. On the other hand, when imipenem is bound, the loop is more flexible and can adapt multiple conformations, especially the hydrophobic residues, W49-V52, closest to its hydrophilic tail. For cephalorodine, the simulation without the H-bond enforced yielded the two conformations of W49 while the one without the H-bonding did not. Comparing the RMSD of W49 of these two simulations, the RMSD for cephalorodine with the H-bond is definitely higher than the one without, by ~0.5Å. Overall, individual amino acid RMSDs differ relatively little between different antibiotics. These results are consistent with the bonding between the loop and the substrates being nonspecific in each case.

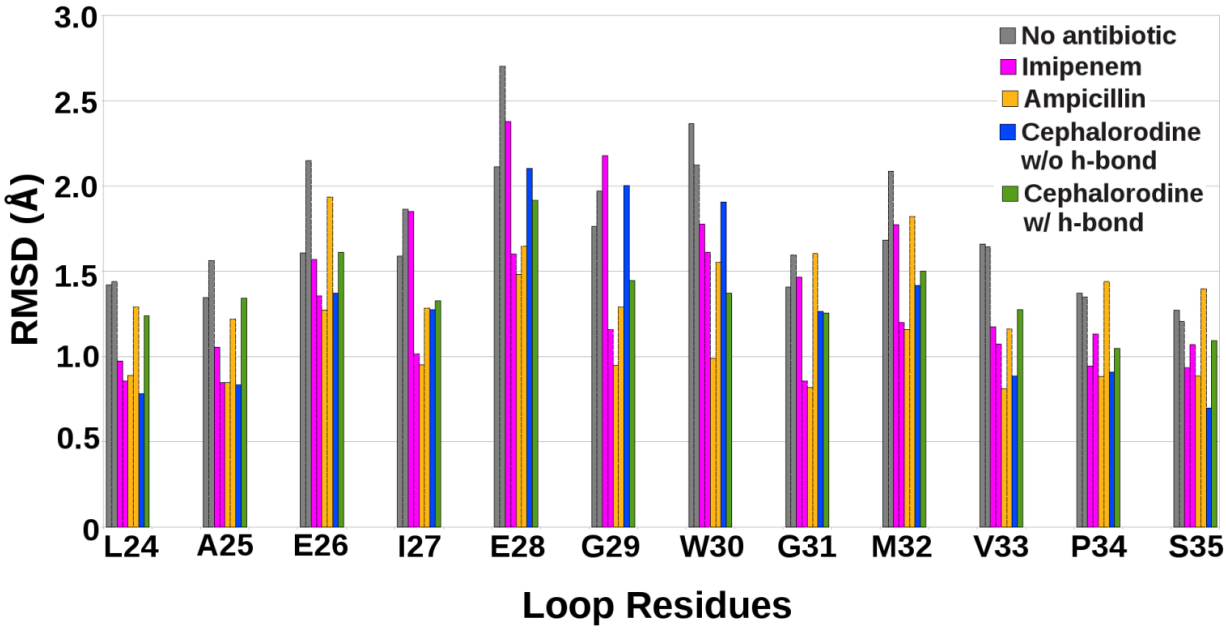


Figure 4.6. The average all atom RMSDs for each residue in the loop shows greater flexibility for the more solvent exposed, polar residues such as E28 and W49. The dotted line columns correspond to the duplicate simulation for the no antibiotic, imipenem and ampicillin systems. The reference structure used for calculating the RMSDs are the same ones used to calculate the RMSDs of the loop during the QM/DMD simulations.

4.3.4. LOOP-SUBSTRATE INTERACTION ENERGIES

The interaction between the β-hairpin loop and antibiotic was evaluated by comparing binding energies. Structures used to evaluate those interactions resulted from the DMD stages of QM/DMD simulations, when the position of the loop was the most realistic. The systems used for the binding energies calculations include the loop, antibiotic, and active site chopped as it is during the QM stage of QM/DMD. Of the 12 residues in the loop (L43-S54) only residues I46-W49 were used in these calculations, to decrease the system size. Those residues were found to have the closest contact with the substrates, and hence are expected to contribute the most to the substrate binding energies. Loop-substrate interaction energies were calculated as the difference between the energy of the entire complex, and the sum of the energies of the loop and the joint antibiotic-active site part. Table 4.3 lists all the BSSE-corrected binding energies with respective antibiotics. In this work, positive binding energies correspond to favorable and stabilizing interactions. As apparent from the values presented in Table 4.3, inclusion of solvation dramatically lowers the estimated loop-antibiotic affinities, and therefore is essential for the assessment. On the other hand, the large basis sets used in the final energy calculations limit the BSSE to ca. 1 kcal/mol.

As one would expect from the nature of the interaction between the antibiotics and the loop, the major component to the binding energy is the dispersion interactions. The contributions of dispersion to the binding energies are also included in all DFT calculations.

For all antibiotics, these values are similar, small and positive, indicating weakly attractive interactions. Our results suggest that, even though the loop above the active site plays a major role in binding structurally dissimilar substrates, the loop-substrate interaction energies are not different enough to be a defining factor in the efficiency of hydrolysis of these molecules by CcrA.

Binding Energies (B.E.)	Imipenem	Ampicillin	Cephalorodine	Cephalorodine (H-bond)
TPSS-D3/def2-TZVPP	48.6 ± 5.0 (51.1 ± 3.7)	16.7 ± 5.4 (16.3 ± 6.3)	56.6 ± 7.9	56.6 ± 4.7
TPSS/def2-TZVPP (no dispersion)	35.0 ± 3.7 (36.4 ± 4.8)	-6.7 ± 5.6 (-1.4 ± 4.8)	37.7 ± 5.2	33.9 ± 4.4
TPSS/def2-TZVPP (no dispersion) + CP	34.4 ± 3.7 (35.5 ± 4.8)	-7.1 ± 5.6 (-1.8 ± 4.8)	36.9 ± 5.2	32.6 ± 4.4
TPSS-D3/def2-TZVPP + solvent	5.9 ± 2.2 (4.0 ± 1.7)	4.3 ± 3.8 (3.2 ± 4.1)	0.9 ± 3.5	7.3 ± 2.6
TPSS-D3/def2-TZVPP + solvent + CP	5.2 ± 2.1 (3.2 ± 1.7)	3.9 ± 3.8 (2.9 ± 4.1)	-0.4 ± 3.5	6.1 ± 2.6
B3LYP-D3/def2-TZVP + solvent + CP	5.8 ± 2.1 (-7.8 ± 2.1)	4.0 ± 4.3 (6.9 ± 4.5)	0.0 ± 3.6	7.4 ± 3.2
MP2/def2-TZVPP	35.9 ± 7.0 (48.5 ± 9.6)	20.6 ± 5.4 (20.4 ± 8.1)	45.3 ± 11.2	51.8 ± 7.8

Table 4.3. Binding energies of all loop/antibiotics systems at the DFT level, in kcal/mol.

4.4. CONCLUSIONS

The CcrA b-lactamase enzyme is capable of hydrolyzing a large class of structurally different antibiotics. The presence of the flexible loop above the active site was long thought to contribute to the substrate binding and enzyme promiscuity. Through this work we were able to see that loop/antibiotic interactions play a role in stabilizing the antibiotic for hydrolysis. The work features our new method, QM/DMD, capable of performing efficient quality QM/MM docking to flexible active sites of metalloproteins. All antibiotic exhibit snag binding to the flexible loop, and greatly reduce its mobility. The positions of the loop are substrate-specific, indicating the loop adjustment and molding around bound substrates, in agreement with earlier proposals. For imipenem and ampicillin, no specific interactions between the loop and the substrates can be found, and the interaction is essentially hydrophobic. For cephalorodine, the aromatic ring of the substrate can p-stack with W49 of the loop, and orient itself in parallel or perpendicular orientation with respect to this residue. In all cases, the primary loop-substrate interaction

component is dispersion. This work contributes to mechanistic studies of CcrA, and potentially can help the design of inhibitors and more potent, CcrA-resistant antibiotics. Structures of all CcrA-antibiotic complexes are available upon request.

4.5 REFERENCES

1 Wang, Z.; Fast, W.; Valentine, A.M.,; Benkovic, S.J. *Curr. Opin. Chem. Biol*. **1999**, *3* (5), 614-622.

2 Ambler, R. P. *Phil. Trans. R. Soc. B* **1980**, *289,* 321–331.

3 Bush, K.; Jacoby, G.; Medeiros A. A. *Antimicrob. Agents Chemother*. **1995**, *39, 1211-1233.*

4 Cuchural, G.J.; Malamy, M.H.; Tally, F.P. *Antimicrob. Agents Chemother*. **1986**, *30,* 645–648.

5 Sabath, L.D.; Abraham, E.P. *Biochem. J.* **1966**, *98,* 11c–13c.

6 Ullah, J.H.; Walsh, T.R.; Taylor, I.A.; Emery, D.C.; Verma, C.S.; Gamblin, S.J.; Spencer, J. *J. Mol. Biol.* **1998**, *284,* 125–136.

7 Senda, K.; Arakawa, Y.; Ichiyama, S.; Nakashima, K.; Ito, H.; Ohsuka, S.; Shimokata, K.; Kato, N.; Ohta, M. *J. Clin. Microbiol*. **1996**, *34,* 2909–2913.

8 Docquier, J.-D.; Benvenuti, M.; Calderone, V.; Stoczko, M.; Menciassi, N.; Rossolini, G. M.; Mangani, S. *Antimicrob. Agents Chemother.* **2010**, *54*, 4343-4351.

9 Leiros, H. K.; Bora, P. S.; Bandsdal, B. O.; Edvardsen, K. S.; Spencer, J.; Walsh, T. R.; Samuelsen, O. *Antimicrob. Agents Chemother.* **2012**, *56*, 4341-4353.

10 Felici, A.; Amicosante, G.; Oratore, A.; Strom, R.; Ledent, P.; Joris, B.; Fanuel, L.; Frère, J. M. *Biochem. J.* **1993**, *291,* 151-155.

11 Carfi, A.; Duée, E.; Paul-Soto, R.; Galleni, M.; Frére, J. M.; Dideberg, O. *Acta Crystallogr. D* **1998**, *54,* 47–57.

12 Scrofani, S. D. B.; Chung, J.; Huntley, J. J. A.; Benkovic, S. J.; Wright, P. E.; Dyson, H. J. *Biochemistry* **1999**, *38*, 14507-14514.

13 Fitzgerald, P.M.D.; Wu, J. K.; Toney, J. H. *Biochemistry* **1998**, *37,* 6791-6800.

14 Payne, D. J.; Hueso-Rodrieguez, J. A.; Boyd, H.; Conca, N. O.; Janson, C. A.; Gilpin, M.;

Bateson, J. H.; Cheever, C.; Niconovich, N. L.; Pearson, S.; Rittenhouse, S.; Tew, D.; Diez, E.; Perez, P.; de la Fuente, J.; Rees, M.; Rivera-Sagredo, A. *Antimicrob. Agents Chemother.* **2002***, 46,* 1880-1886.

15 Huntly, J. J. A.; Fast, W.; Benkovic, S. J.; Wright, P. E.; Dyson, J*. Protein Sci.* **2003**, *12*, 1368-1375.

16 Suárez, D.; Días, N.; Merz, K.M. *J. Comput. Chem*. **2002**, *23,* 1587-1600.

17 Xu, D.; Cui, Q.; Guo, H. *J. Phys. Chem. A* **2007**, *111*, 5630–5636.

18 Concha, N.O.; Rasmussen, B.A.; Bush, K.; Herzberg, O. *Structure* **1996**, *4*, 823-836.

19 Park, H.; Brothers, E. N.; Merz, K. M. Jr. *J. Am. Chem. Soc.* **2005**, *127*, 4232-4241.

20 Valdez, C. E.; Alexandrova, A. N. *J. Phys. Chem. B*. **2012**, *116*, 10649-19656.

21 Sparta, M.; Ding, F.; Shirvanyants, D.; Dokholyan, N. V.; Alexandrova, A. N. *Biophys. J.* **2012**, *103*, 767-776.

22 (a) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Fold. Des* **1998**, *3*, 577-588. (b) Ding, F.; Guo, W.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J. E. *J. Mol. Biol.* **2005**, *350*, 1035–1050. (c) Dokholyan N. V. *Curr. Opin. Struct. Biol.* **2006**, *16,* 79-85. (d) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16,* 1010-1018.

23 (a) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Fold. Des.* **1998**, *3*, 577–587. (b) Dokholyan, N. V. *Curr. Opin. Struct. Biol*. **2006**, *16*, 79–85. (c) Ding, F.; Guo, W. H.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J. E. *J. Mol. Biol.* **2005**, *350*, 1035–1050. (d) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010-1018.

24 Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett*. **2003**, *91*, 146401-146401.

25 Grimme, S. *J. Comput. Chem*. **2006***, 27*, 1787–1799.

26 Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.

27 Turbomole v6.3 (2011), Program Package for ab initio Electronic Structure Calculations; University of Karlsruhe: Germany, available from http://www.turbomole.com.

28 Sparta, M.; Alexandrova, A. N. *PLoS ONE* **2012**, 7, e47172.

29 Klamt, A.; Schüürmann, G. J. *Chem. Soc. Perkin Trans.2* **1993**, *5,* 799–805.

30 (a) A. D. Becke*, J. Chem. Phys*. **1993***, 98, 5648-5652.* (b) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys*. **2010**, *132, 154104-15419*.

31 (a) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett*. **1988**, *153*, 503-506. (b) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett*. **1990**, *166*, 275-280. (c) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett*. **1990**, *166*, 281-289. (d) Weigend, F.; Häser, M. *Theor. Chem. Acc*. **1997**, 97*,* 331-340.

32 Boys, S. F.; Bernardi, F. *Mol. Phys*. **1970**, *19,* 553-566.

33 Días, N.; Suárez, D.; Merz, K. M. *J. Am. Chem. Soc*. **2000**, *122,* 4197-4208.

34 Vessillier, S.; Docquier, J.; Rival, S.; Frere, J.; Galleni, M.; Amicosante, G.; Rossolini, G.; Franceschini, N. *Antimicrob. Agents Chemother*. **2002**, *46* (6), 1921–1927.

CHAPTER 5

Metal-Dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway

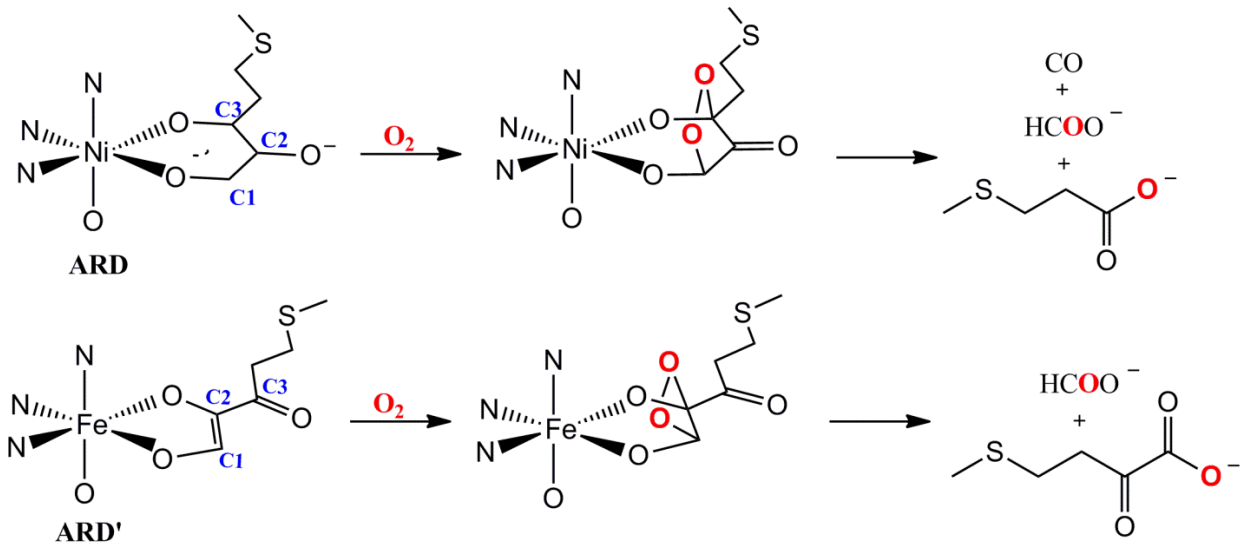
5.1. INTRODUCTION

The two acireductone dioxygenases (ARD and ARD') are found in the methionine salvage pathway in living organisms, including humans.[1,2] It is intriguing that these two enzymes, while sharing the same protein sequence and differing only by the nature of the bound metal, catalyze completely different reactions of oxidation of the same substrate, 1,2-dihydroxy-3-keto-5-(methylthio)pentene.[3,4] The Ni^{2+} dependent enzyme, ARD (EC 1.13.11.53) catalyzes the formation of formate, methylthiopropinate and carbon monoxide, whereas the Fe^{2+} dependent enzyme, ARD' (EC 1.13.11.54) promotes the formation of formate and 2-keto-4-methylthiobutyric acid (precursor of methionine).[4] Furthermore, it has been demonstrated that the two enzymes have the active sites with identical pseudo-octahedral coordination of the metal centers, by His96, His98, Glu102 and His140, with the remaining two positions being occupied by either water molecules or the substrate.[5,6] Finally, it has been shown that it is possible to reconstitute either of the enzymes by combining the apoprotein with a solution containing the proper metal.[7] Hence, although Fe^{2+} and Ni^{2+} have different binding constants, the protein does not preferentially bind either one of the cations.

Sequential reaction mechanisms have been proposed for both enzymes. Shown by kinetic analyses, a red λ_{max} shift that occurs upon substrate binding in anaerobic conditions to ARD and ARD' suggests that the substrate binds as a dianion to the enzymes.[6] The dianionic substrate is electron rich and susceptible to electrophilic attack by dioxygen, as confirmed by the non-enzymatic reaction studies. In the absence of metal ions, the oxidation of 1,2-dihydroxy-3-keto-5-(methylthio)pentene yields the same products as ARD' via a base-catalyzed oxidative cleavage, indicating that dioxygen activation is not needed for enzymatic activity.[6] Other studies have shown that non-heme iron proteins, in particular non-heme iron dioxygenases, require metal activation of dioxygen[8,9] whereas, in the presence of dioxygen alone both ARD and ARD'

do not bind oxygen, further proving the requirement of sequential binding of the substrate followed by dioxygen binding for catalysis.[6]

The exact mechanisms of the two reactions remain unknown. However, using isotopically labeled dioxygen and following the incorporation in the products, it has been postulated that the substrate may coordinate the metal center forming either a five- or six-membered ring, depending on the metal present. This would dictate different targets for the attack by dioxygen, with the subsequent cleavage of the C1-C2 bond (in ARD'), or the C1-C2 and C2-C3 bonds (in ARD) (Scheme 5.1). Isotopic studies have shown that dioxygen is incorporated into the products.[10,11] To the best of our knowledge, the catalytic mechanisms have never been investigated theoretically. On the other hand, the uncatalyzed reaction of the substrate with dioxygen has been investigated by Borowsky *et al.*[12] by means of density functional theory (DFT) calculations, and experimentally by Dai *et al.*[6] These studies concluded that the uncatalyzed reaction may occur via a radical mechanism. As shown Dai *et al.*[6] using a substrate modified with a radical trap, a mechanism involving long living radicals is not likely to occur during the catalyzed reaction.



Scheme 5.1. The old hypothesis for the reaction mechanisms of oxidation of acireductone catalyzed by ARD and ARD'. The suspected difference in the substrate binding is illustrated.

In this article, we report on the first theoretical study of the catalytic cycles promoted by ARD and ARD'. We first address the acireductone substrate binding to the two proteins and show that, contrary to the old belief, there is no difference in the binding mode of acireductone to ARD and ARD'. To show this, we use our new method for dynamics simulations of metalloproteins, QM/DMD.[13] We then tackle the catalytic mechanisms of ARD/ARD' using first-principle calculations, and show that the bifurcation of the mechanism occurs later in the reaction route. We propose a new catalytic mechanism for ARD', which includes an additional intermediate, as compared to the ARD pathway. This difference is due to just two extra electrons available in Ni^{2+} versus Fe^{2+}, and the resultant RedOx flexibility of Fe^{2+}. This allows for the rerouting of the reaction. As we were preparing this article, Allpress *et al*.[14] experimentally showed that in a bioinorganic complex mimicking ARD and ARD', the regioselectivity change of aliphatic carbon-carbon cleavage is also not dependent on the substrate binding mode, but rather dependent on the intermediate in the reaction pathway, an intermediate found in this study. Although their experiments used a non-native substrate, our results agree with the trends of their experiments.

5.2. THEORETICAL METHODS

The initial structures of ARD/ARD' were obtained from the Protein Data Bank (PDB) (ARD PDB code: 1ZRR[15], ARD' PDB code: 2HJI[16]). However, to examine specifically the metal-dependent effects on the mechanism, we only used one PDB structure, 2HJI, for the QM/DMD simulations using Fe and Ni. Since neither of the PDBs contained the coordinated substrate, 2-dihydroxy-3-keto-5-methylthiopentene, it was manually built into the protein. The coordination geometry chosen for each structure comes from experimental and theoretical predictions.[5,12]

5.2.1 QM/DMD SIMULATIONS

This study employs the recently developed QM/DMD[12] method, where DMD is short for discrete molecular dynamics.[17,18,19,20] QM/DMD is a hybrid method, a variant of QM/MM,[16-18] which efficiently captures metalloprotein dynamics on the order of tens of nanoseconds, in conjunction with the quantum mechanical description of the active site. It has been shown to perform exceptionally well for recapitulating and recovery of native protein structures down to the subtle structural details at the active sites,[13] electronic properties of the bound metals,[13,21] protein conformational responses to substrate binding and metal replacement,[21,22,23] and mechanistic studies of enzymatic reactions.[21,22] The main attractive feature of QM/DMD is its affordability as compared to that of other often equally capable QM/MM methods. The strength of the method is in part due to DMD,[16-19] a remarkably successful classical force-field-based method for sampling biological molecules and their complexes. No explicit solvent is used in DMD, but solvation is included in the averaged way into the force field in use.

DMD samples most of the protein, except for the immediate coordination of the metal. In this way, there is no need for the parameterization of the force field to describe the metal, which is then free to change the coordination environment in the course of the simulation. The QM part of the simulations consists of periodic relaxations following the *ab initio* gradients for nuclear motions of the larger active site that includes the metals, their ligands, and the substrate, and may include other important amino acids near the reactive center. The QM/DMD boundary is thus “breathing”, and goes around just the metal and its immediate coordination, or around the larger active site, depending on the stage of the simulation, as described below. The presence of the shared QM and DMD domain permits for the simple and efficient communication between the QM and DMD regions. We use QM/DMD to test the current hypothesis on the preferred binding modes of the substrate to ARD and ARD’.

5.2.1.1. QM/DMD PARTITIONING THE PROTEIN.

The coupling between QM and DMD in the QM/DMD methods is achieved via partitioning the whole system into three domains, and then using the "breathing" QM-DMD boundary that circumvents parts of the protein significantly different in size, depending on the stage of the simulation. Figure 5.1A shows how the protein is divided into three domains. The metal cation, Ni for ARD and Fe for ARD', and the portion of the active site immediately surrounding it constitutes the *QM-only* domain, which moves only during the QM phase of the simulation and is not allowed to be moved by classical DMD. The larger active site shown in purple in Figure 5.1A and zoomed in in Figure 5.1B is the shared *QM-DMD* domain, which can be moved by both QM and DMD, depending on the stage of the run. The *QM-DMD* domain is a minimalistic chemically meaningful cluster model of the active site. It includes the metal ion, His96, His98, His140, Glu102, doubly deprotonated substrate and two protonated Arg104 and Arg154 that stabilize the substrate through H-bonds. The truncation of the amino acids in the QM calculations occurred at the Cα-Cβ bond, Cδ-Cγ for arginines. Whenever a bond is truncated, the atom included in the QM region is saturated with a hydrogen atom positioned along the original bond at a distance equal to 0.7052*R(C-C/N). All saturating hydrogens and their bond partners were frozen during the QM optimization of the active site to retain the geometry imposed by the rest of the protein. Furthermore, additional constraints were imposed to prevent any of the residues coordinating the metal to become too distorted and lose their coordination during DMD phases. A few atoms near the metal centers in the active site were allowed to be sampled by DMD only within ±0.01 Å from the values predicted by QM. Those atoms are the two C atoms adjacent to N coordinating the metal in the His residues, and CG of Glu102. This set of constraints was developed and validated in our previous works.[13,20-22] The rest of the system, including most of the protein macromolecule and a large portion of ARD, constitutes the *DMD-only* domain.

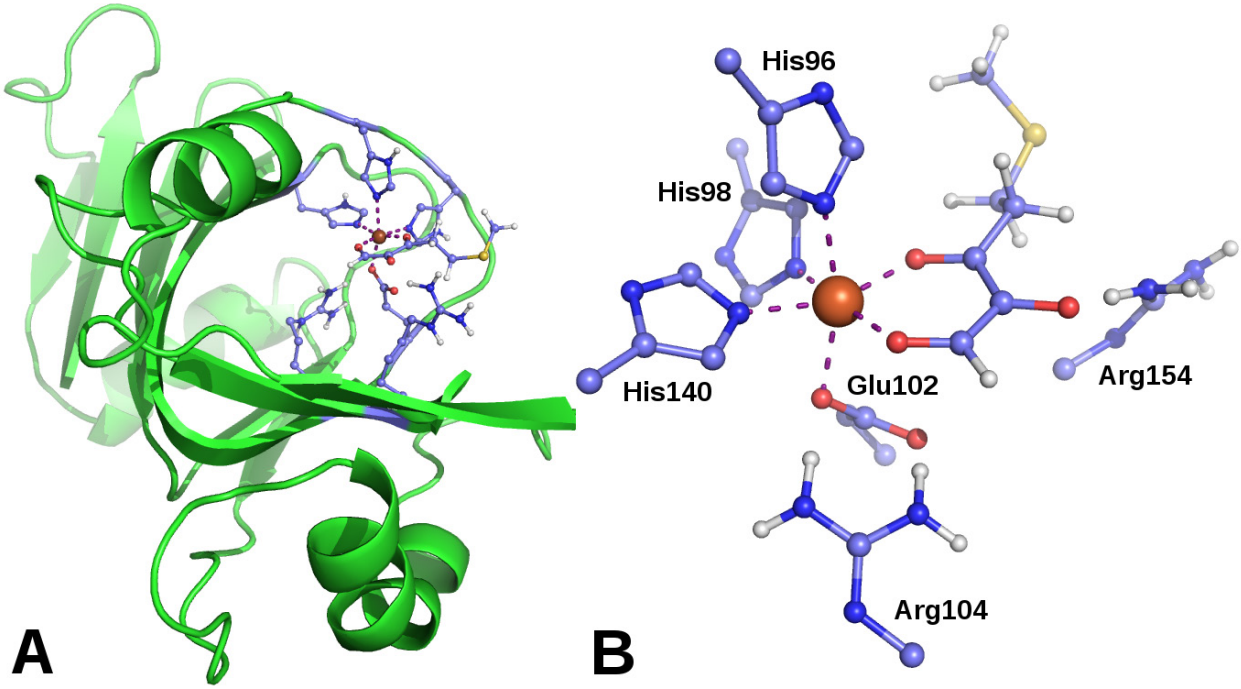


Figure 5.1. A) Schematic representation of the QM/DMD domains. The atoms directly coordinated to the metal are the *QM-only* domain, the purple area defines the *QM-DMD* domain, and the rest of the system constitutes the *DMD-only* domain. Protein backbone is shown as thick green lines. B) Schematic model of the system during QM calculations: atoms that are bordering with the *DMD-only* domain are frozen and their valences are saturated with hydrogen atoms. A portion of the substrate is shown bound to the metal center.

5.2.1.2. SIMULATION DETAILS

Each simulation starts from a short DMD run of 1,000 DMD time units (t.u.), where 1 t.u. roughly corresponds to 50 fs, at a temperature of T=0.10 kcal/(mol*K), with a high heat exchange rate of the protein with the bath for 10 $t.u.^{-1}$, using an Andersen thermostat.[24] This was found to remove clashes introduced by the pdb starting structures.[25] Then, the QM/DMD simulation on the prepared protein begins. It proceeds in an iterative fashion, where the DMD and QM machineries alternate. It starts with the DMD phase, which operates on most of the protein, including the backbone, except for the *QM-only* region. The temperature was ramped to 0.20 kcal/(mol*K), and then decreased stepwise (5 steps, consisting of 500 t.u. each) to anneal and equilibrate the structure. Annealing was found to help achieve better sampling and statistics

of the resulting ensemble. After equilibration, the temperature is kept low for 10,000 t.u., and at this stage, data are collected. The DMD-produced ensemble is clustered according to geometric similarity, based on the Kabcsh[26] RMSD for all pair-wise snapshot structures and by applying a hierarchical clustering algorithm.[27] For each cluster, both the structure closest to the centroid and the one with the lowest DMD energy were used as representatives for the QM phase.

Following each iteration, the *QM-DMD* domain is extracted from each of the centroid and low-energy structures, capped with hydrogens, and a single-point QM energy is calculated for it. For the QM calculations, DFT is used in the specific BP86[28,29,30,31,32] formulation of the exchange and correlation functional. A double ζ quality basis set (def2-SVP)[33] for H, C, N, O, and S and a triple ζ quality basis set (def2-TZVPP)[34] for the metals were used. Resolution of Identity (RI),[35] and Multipole Accelerated Resolution of Identity (MARI-J)[36] as implemented in *Turbomole*[37] were exploited to speed up the calculations. Empirical dispersion correction for DFT calculations was included in both the energy and gradient evaluations.[38] Solvent was included via the Conductor-like Screening Model (COSMO) continuum solvent with the dielectric constant set to 20.0.[39] After that, the structures are scored, based on both the QM energy of the *QM-DMD* region and the DMD energy.[13] A single structure is selected from each iteration, and it is partially optimized at the QM level with constraints fixing the points of attachment to the rest of the protein. Then, the active site is reinstalled into the protein, the QM/DMD boundary shrinks again to go around just the *QM-region*, and the simulation proceeds with the new DMD phase, when the structural information of the active site can be propagated to the rest of the protein. The overall QM/DMD simulation for each protein runs until convergence, and in this work 20 iterations were enough, corresponding to approximately 10.5 ns of dynamics. From each run, the structure with the lowest QM energy was selected for the mechanistic study.

5.2.2. METHODS USED IN THE MECHANISTIC STUDY

Staying consistent with the level of theory with the QM optimizations during the QM/DMD simulations, all stationary points along the reaction profile were calculated with the BP86 functional plus empirical dispersion and a double ζ quality basis set (def2-SVP) for H, C, N, O, and S and a triple ζ quality basis set (def2-TZVPP) for the iron and nickel. All calculations included the solvation via COSMO with the dielectric constant set to 20.0. The nature of each stationary point was confirmed with a frequency calculation: no imaginary frequencies confirmed the local minima, and one imaginary frequency with the normal mode going along the reaction coordinate confirmed TSs. Additionally, TPSSh[40]/def2-TZVPP and B3LYP[41]/ def2-TZVPP with empirical dispersion and implicit solvation were used to calculate the energies of the stationary points. The results reported in the main text are obtained with TPSSh, and those obtained with the BP86 and B3LYP functionals are given in the Appendix 9.2 (Table 9.2.1 and 9.2.2). All charges were computed using Natural Population Analyses (NPA)[42] at the TPSSh/def2-TZVPP level of theory.

5.3. RESULTS AND DISCUSSION

5.3.1. ARD AND ARD' BIND THE SUBSTRATE IN THE SAME WAY

It has been long speculated that the key difference between the mechanisms of ARD and ARD' is the binding modes of the substrate to the metals: the substrate would coordinate to Fe^{2+} with the O1 and O2, hence forming a five-membered ring, whereas to Ni^{2+}, it would bind with O1 and O3, producing a six-membered ring (see Scheme 5.1 for atom numeration). This difference in binding should indeed lead to the respective products. However, we challenge this speculation. The two proteins share the exact same sequence. The bond lengths to Fe^{2+} and Ni^{2+} could be slightly different, but this does not seem enough to change the binding pose so dramatically.

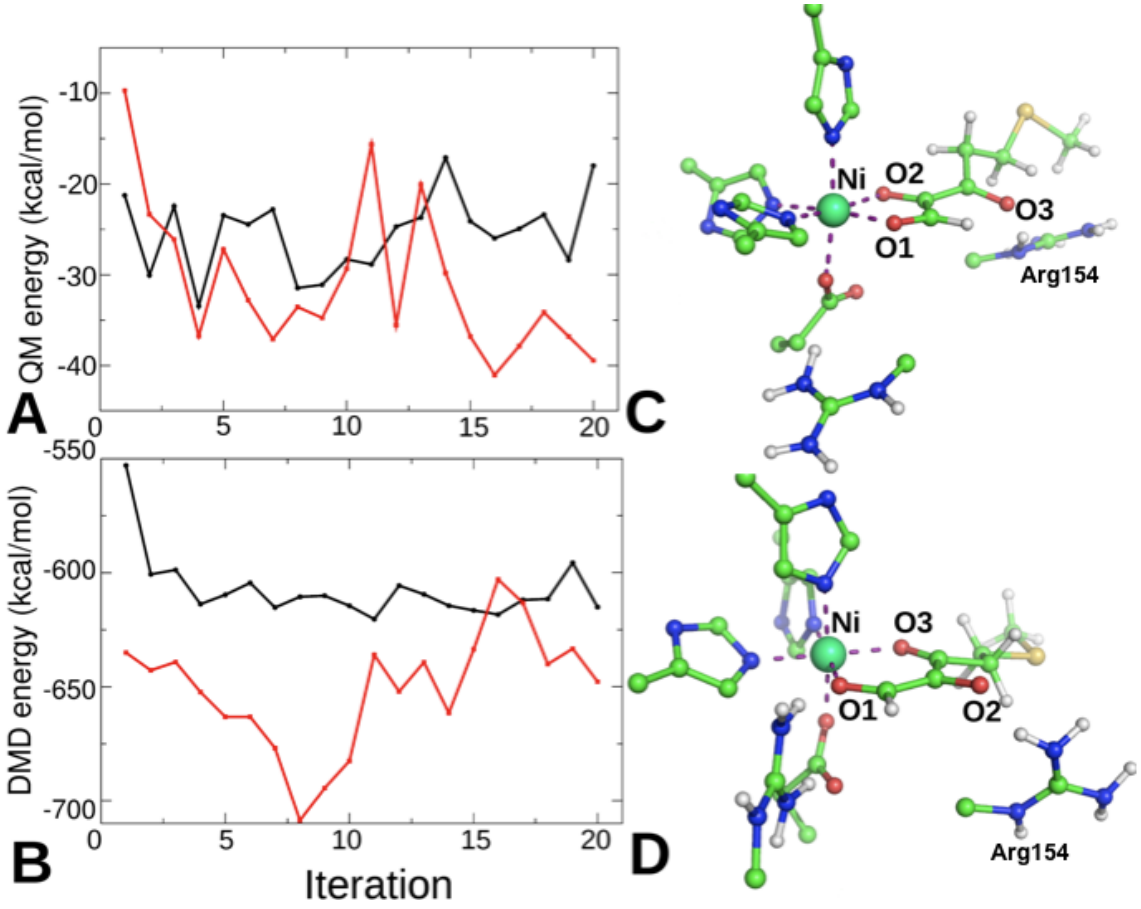


Figure 5.2. Results of QM/DMD simulations for the Ni^{2+} containing ARD coordinating the substrate in two different ways. (A, B) The QM and DMD energies, respectively, plotted as a function of the iteration number. The red lines correspond to the 6-membered ring coordination, and the black lines correspond to the 5-membered ring coordination. The preference for the 6-membered ring binding is apparent. (C, D) representative structures of the active site with the substrate coordinated as 5- and 6-membered rings, respectively.

To address whether or not there is a dependence of the binding mode on the nature of the metal, we performed QM/DMD simulations for both metal variants, and for two different binding modes, that is, a total of four simulations. The lowest energy spin states were found on the cluster models of the two active sites (triplet for Ni ARD, and quintet for Fe ARD'), and the QM/DMD simulations on the entire protein with these spin states were performed. In Figure 5.2 A and B we report the QM and DMD energies for the Ni^{2+} containing ARD coordinating the substrate as a six-membered ring (in red) and five-membered ring (in black), as a function of the

iteration number. Both energies show a clear and consistent preference for the six-membered ring coordination. Figure 5.2 C and D depict representative structures of the active site with the substrate coordinated in the two different modes, as resulted from QM/DMD simulations. Thus, for ARD, the old hypothesis about the binding mode is fully confirmed.

However, for ARD', there is also a clear preference for the six-membered ring coordination, as follows from the QM and DMD energies produced in QM/DMD simulations (Figure 5.3 A and B). From just the QM energies of the active site, the five-memebered ring coordination is discouraged by ca. 10 kcal/mol. This result contradicts the old hypothesis that Fe^{2+} promotes the 5-membered ring coordination.

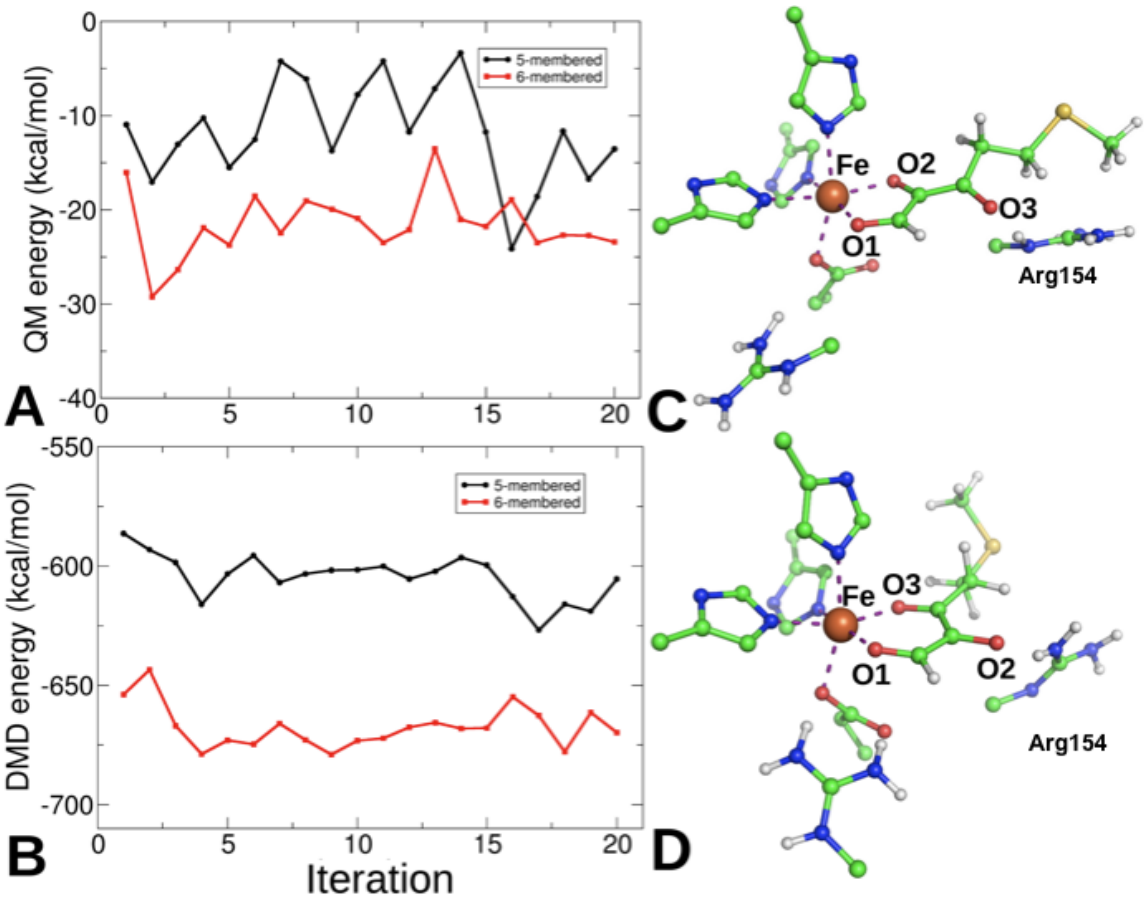


Figure 5.3. Results of QM/DMD simulations for the Fe^{2+} containing ARD' coordinating the substrate in two different ways. (A, B) The QM and DMD energies, respectively, plotted as a function of the iteration number. The red lines correspond to the six-membered ring coordination, and the black lines correspond to the five-membered ring coordination. The preference for the six-membered ring binding is again

apparent. (C,D) representative structures of the active site with the substrate coordinated as five- and six-membered rings, respectively.

In both ARD and ARD', coordination as a six-membered ring indeed makes sense from the structural point of view. Arg154 in the second coordination sphere of the metal forms a H-bond to the substrate O2 atom, stabilizing this coordination mode for both ARD and ARD' (Figure 5.3D and 5.4D). In the five-membered ring scenario, this interaction is impossible (Figure 5.2C and 5.3C). The overall structures of the binding pocket upon the substrate binding are also exceptionally consistent between the ARD and ARD' forms of the protein. Experimental data for the ARD and ARD' enzyme-substrate adducts are limited. However, similar UV-vis absorption of these adducts indicated that there may not be a difference in binding mode of ARD and ARD'.[14] We performed the time-dependent DFT calculations on the substrate coordinated to ARD and ARD' in the two studied orientations, and in solution (Appendix 9.2). It is noteworthy that in solution the five-membered ring-like conformation of the substrate is preferred by ca. 5 kcal/mol. Our results indicate that both coordination modes result in similar red shift in the UV-vis spectra, and so from the computational spectroscopy standpoint the two modes are also indistinguishable.

It was brought to our attention at the stage of the revision that, in the earlier work by Ju *et al.*,[16] classical molecular dynamics simulations were performed on the two metal variants of ARD, with the metal coordination kept fixed at the geometry observed in XAS experiments. Ten-picosecone dynamics simulations showed that Trp162 "guarding" the entryway to the binding site of ARD changed its conformation depending on the bound metal. It is "closed" in ARD interacting with the substrate and supposedly stabilizing the six-membetred ring binding, but "open" in the ARD' thereby facilitating the different binding mode. We performed extensive (ca. 100 ns) QM/DMD simulations on the Fe and Ni of the protein using the 1ZRR and 2HJI structures. The metals were treated quantum mechanically and their coordination was allowed

to move in the dynamics. We found very little difference in the position of the C-terminus of the protein and specifically Trip162 for the Ni and Fe forms. Results are given in the Appendix 9.2. Therefore, we expect that the possible weak dispersion interaction of the substrate with Tri162 to play no or very little role in defining the substrate binding. Instead, Arg154 appears to drive the preference for the six-membered ring coordination. In fact, if Arg154 is excluded from the shared QM-DMD domain in the simulations, the preference becomes less clear.

5.3.2. CATALYTIC MECHANISM FOR ARD AND ARD'

For mechanistic studies, the lowest QM energy structure was chosen from each of the QM/DMD simulations. We do not explicitly consider the barrier toward the dioxygen attack, since the difference between ARD and ARD' is rooted in the later steps along the reaction profile.

For ARD, the lowest energy spin state unequivocally is a triplet. However, other spin states were also considered. The dotted lines on Figure 5.4 connect stationary points along the reaction paths without the consideration of possible spin cross-over. The reaction proceeds as follows: dioxygen binds to C1 and C3 of the substrate forming a dioxygen adduct. The lowest energy state for this adduct is a triplet with other states being considerably higher in energy (Figure 5.4). In order to start from the triplet state of the initial complex and reach the triplet potential energy surface in the intermediate, the system has to undergo the spin cross-over, which is anticipated to be a facile process, considering the typically large magnitudes of the spin orbit coupling in the octahedral complexes of Ni^{2+}. The intermediate then has access to the low-barrier transition state in which dioxygen splits, and the substrate synchronously dissociates into the three products, carbon monoxide and two carboxylate products (Figure 5.4). Both the transition state and the products are triplets, with the singlet and the quintet being much higher in energy.

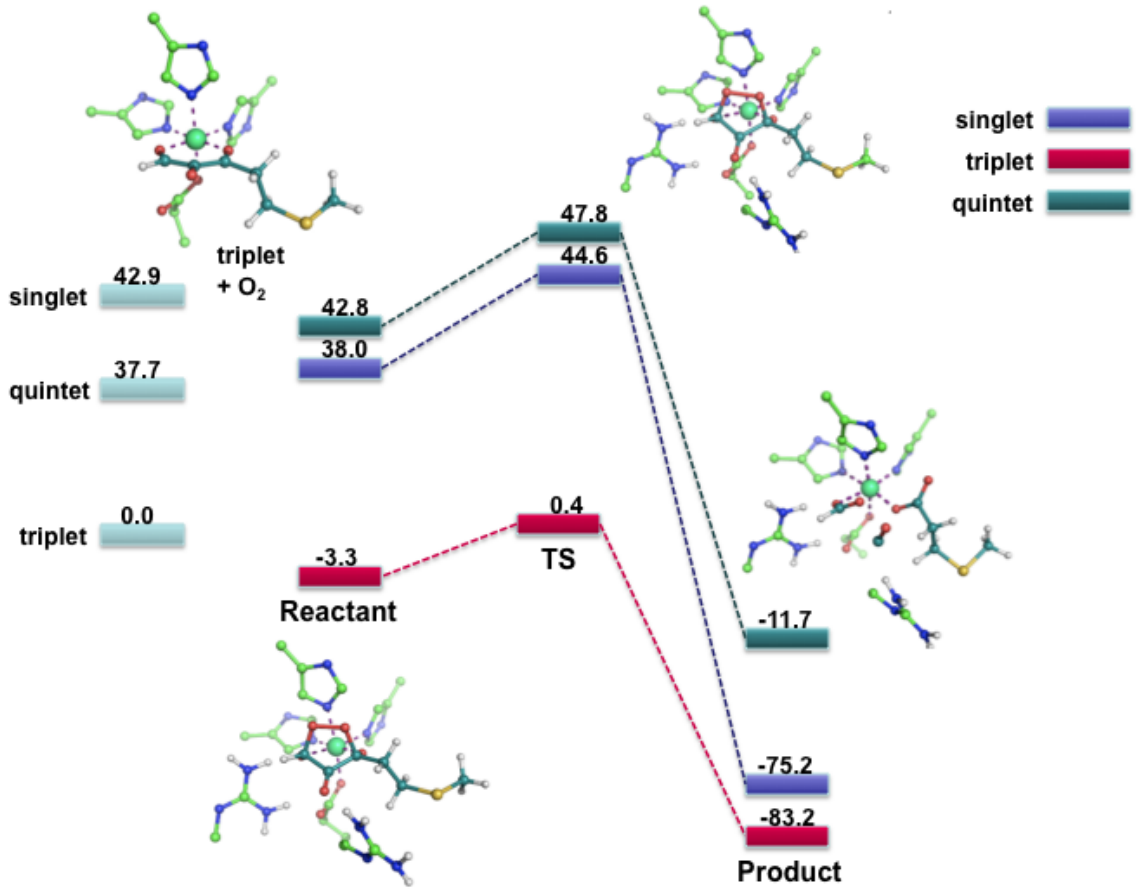


Figure 5.4. Reaction profile for the Ni^{2+} containing ARD. All numbers are calculated with TPSSh/def2-TZVPP.

For ARD', the mechanism contains an intermediate step that is not seen in ARD. Firstly, the starting complex has the splitting between the spin states that is much smaller than that in ARD. The quintet state is preferred by ca. 10 kcal/mol. The same dioxygen adduct is formed when dioxygen binds to C1 and C3 on the substrate. This intermediate is lower in energy than the reactants only on the quintet potential energy surface, with the singlet and the triplet being ca. 10 kcal/mol above it, and the septet being 20 kcal/mol higher. Getting to the singlet state from the initial quintet requires spin crossing, which is very typical for Fe^{2+}. The reaction proceeds with the splitting of dioxygen. However, unlike in ARD, on all but the singlet surface, the split leads to a new intermediate where the O-O bond is dissociated but with the reactants still being attached to the complex (Figure 5.5). The intermediate is 16-25 kcal/mol below the

reactants. On the singlet surface, the split of the dioxygen leads to the three ARD (Ni-like) products without a barrier. Since these products are not observed experimentally, we are forced to conclude that after the first intermediate the higher-energy singlet and also the septet surfaces are not reachable in this reaction. The reaction must proceed on the lower-energy quintet or triplet surfaces. Both of these states have an access to the low-energy epoxy-like transition state with oxygen attached to C3 transferring to C2. After this transition state, the system splits spontaneously to form the two ARD' products, an alpha-keto carboxylate and formate. Remarkably, in the recently published work on inorganic complexes mimicking ARD and ARD', it was suggested that the Fe form of this catalyst has to have an additional intermediate along the reaction path.[14] This is in a full agreement with our findings for ARD'.

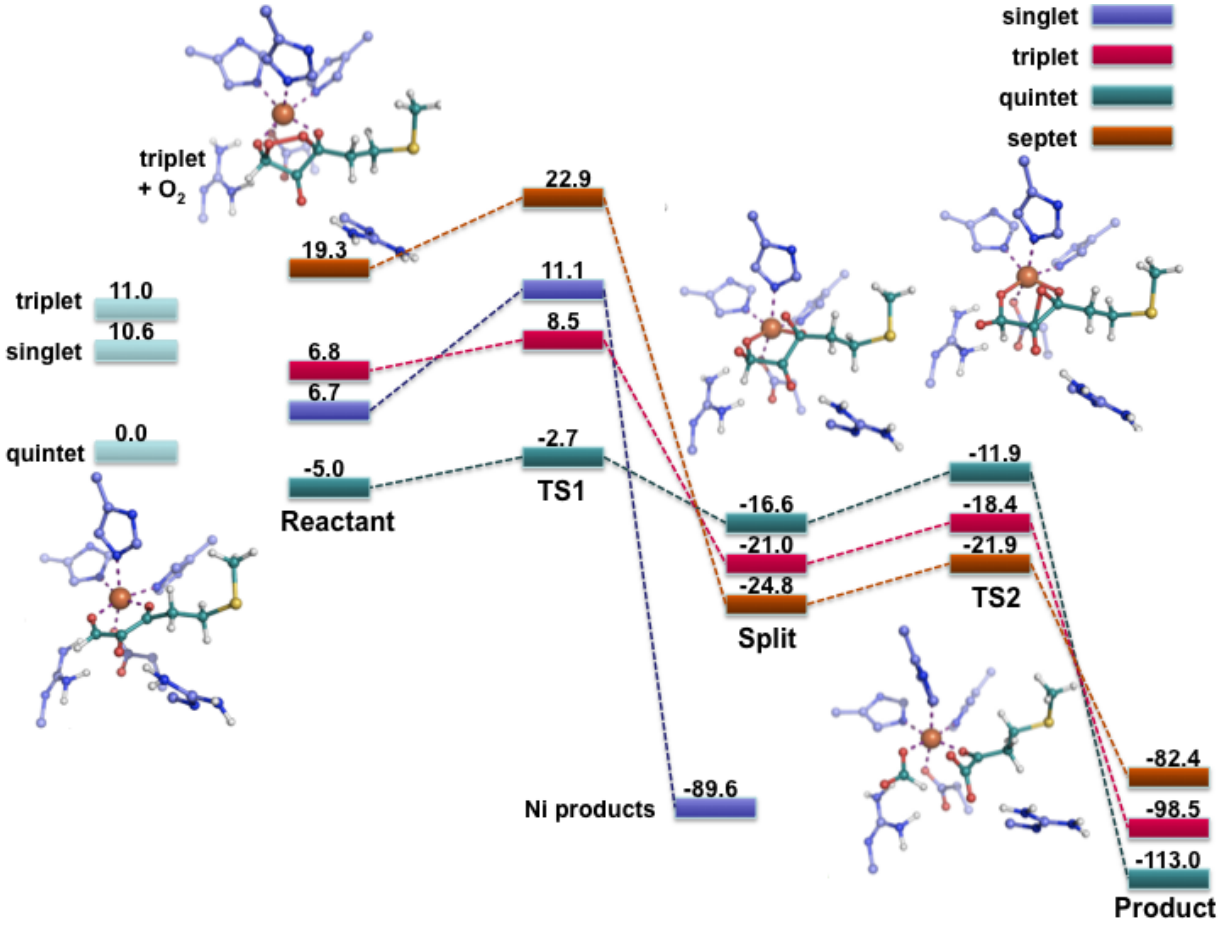


Figure 5.5. Reaction profile for the Fe^{2+} containing ARD'. All reported numbers are calculated with TPSSh/def2-TZVPP.

Considering the QM energy difference between the six-membered ring coordination mode and the five-membered ring coordination mode to ARD' being ca. 10 kcal/mol, we decided to still check the reaction mechanism for the five-membered ring coordination. In this process, dioxygen must attack the substrate at the C1 and C2 position, generating a strained four-membered cycle: C1-C2-O-O. All attempts to find the minimum on the potential energy surface for this structure failed. Dioxygen does not attach to the substrate in this fashion. Therefore, indeed, the old mechanism postulated for ARD' appears to be incorrect.

The key conclusion of this section is that the difference in reaction mechanisms between ARD and ARD' does not have to be dictated by the binding mode. Instead, the differentiation happens further along the reaction profile, during the splitting of dioxygen. This has to be caused by the electronic properties of the bound metal. (Figure 5.6) In the next section, we investigate this difference in detail.

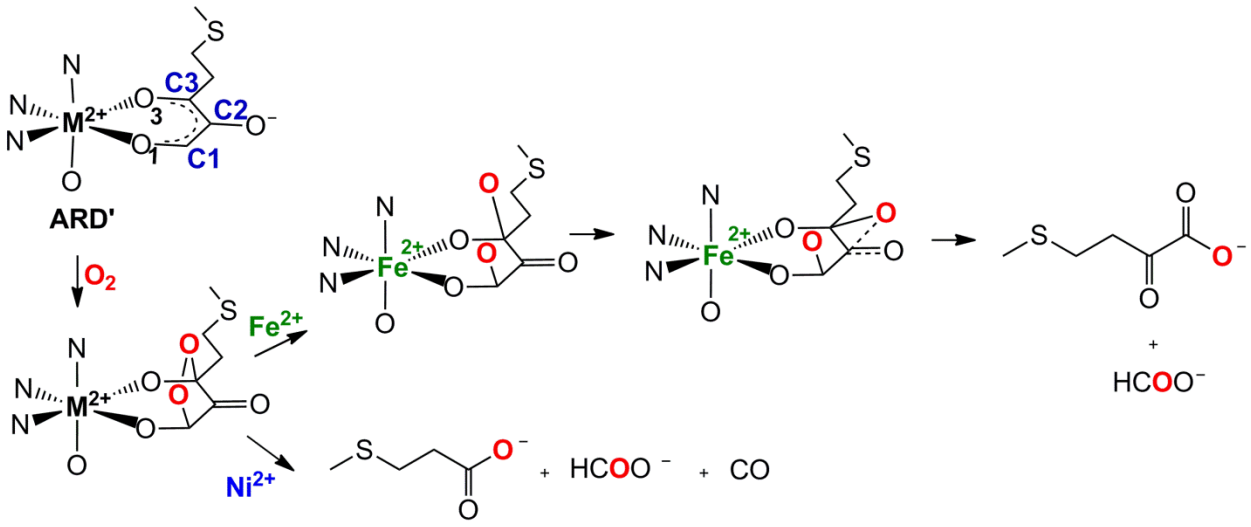


Figure 5.6. The new proposed mechanisms for ARD and ARD'.

5.3.3. WHY DOES ARD' HAVE EXTRA INTERMEDIATE, WHEREAS ARD DOES NOT?

The crucial bifurcation point in the reaction mechanism is the moment when dioxygen splits, either producing either the ARD-like products, or an additional intermediate (in the case of ARD'). The reason for this difference must be rooted in the electronic properties of the metal. To answer this question of why ARD' passed through a split dioxygen intermediate before proceeding to the products whereas ARD does not, we performed natural population analysis (NPA) calculations of the partial charges on all the stationary points of the lowest energy spin state for both proteins (Table 5.1 and 5.2).

	His96	**His98**	**Glu102**	**Arg104**	**His140**	**Arg154**	**Ni^{2+}**	**Substrate**	**O_2**
Reactant	0.09	0.09	-0.85	0.87	0.09	0.93	1.40	-0.97	-0.65
TS	0.09	0.09	-0.84	0.87	0.09	0.93	1.40	-0.95	-0.67
Product	0.07	0.10	-0.83	0.87	0.09	0.98	1.40	-0.22	-1.45

Table 5.1. NPA charges on the side chains, the Ni ion, the substrate, and the attached O_2 in ARD. Throughout the reaction mechanism, charges do not vary indicating the Ni atom does not donate or receive any electron density from the substrate. All charges are computed with TPSSh/def2-TZVPP.

For ARD, as the system progresses through to the transition state where dioxygen splits, there is no significant charge transfer occurring between the substrate, metal and side chains. At the final product formation stage, the charge redistribution takes place within the bound substrate with dioxygen attached (Table 5.1). Ni does not participate in any charge transfer to and from the substrate. Ni in ARD thus plays merely a Lewis acid role; it coordinates and polarizes the substrate, but it has no RedOx role in the reaction.

For ARD', the situation is different and more complicated (Table 5.2). Upon binding and splitting of dioxygen, the substrate gains the partial negative charge. NPA shows that the electrons are drawn from the protein residues coordinating the Fe ion. The charge on Fe itself

does not change. This can be seen from comparing all the charges from the reactants stationary point to the split intermediate: the positive charge on all the residues and the negative change on O_2 increase. Donation of electrons to the attached dioxygen populates its formerly p*-orbital, thus leading to its split. Then upon the formation of products, the substrate donates electrons back to Fe which transmits them to the side chains coordinating it (Table 5.2). The RedOx flexibility of Fe facilitates the process of O_2 splitting, which requires additional electrons. Ni, having a fuller *d*-shell of atomic orbitals is not nearly versatile. It is indeed not in vain that electron transport proteins typically contain Fe and Fe-S centers. What is unusual about ARD' is that Fe is coordinated by His residues, and not the usual Cys utilized in electron transporters, and yet Fe in ARD' pulls electrons out of His in the course of the reaction.

	His96	**His98**	**Glu102**	**Arg104**	**His140**	**Arg154**	**Fe^{2+}**	**Substrate**	**O_2**
Reactant	0.08	0.07	-0.82	0.88	0.06	0.94	1.46	-1.00	-0.66
TS1	0.08	0.07	-0.82	0.88	0.06	0.94	1.45	-0.98	-0.68
Split	0.13	0.15	-0.79	0.90	0.15	0.94	1.48	-0.66	-1.30
TS2	0.13	0.16	-0.79	0.90	0.16	0.93	1.50	-0.65	-1.34
Product	0.07	0.10	-0.81	0.90	0.09	0.92	1.47	-0.33	-1.41

Table 5.2. NPA charges on the side chains, Fe ion, the substrate and attached O_2 in ARD'. Upon splitting of the O_2, the attached O_2 draws the negative charge from the ligands coordinating Fe. All charges are computed with tpssh/def2-TZVPP.

The difference in how ARD and ARD' interact with the substrate can also be detected from the geometries of the bound complexes. In ARD the distances from the Ni cation to the O1 and O3 atoms of the substrate are 2.00 and 2.19 Å. In ARD' the coordination to the metal is much tighter: the same distances read 1.86 and 1.91 Å. The closer coordination to Fe^{2+} is again a manifestation of more "openings" in the set of *d*-AOs of the metal for s-donation from the substrate. Thus, stronger and more RedOx active interaction is observed. The electron donation

to dioxygen is responsible for the stabilization of the second intermediate along the reaction profile of ARD'. Therefore, the rerouting of the catalytic mechanism between ARD and ARD' is exclusively due to the differences in the electronic properties of the metal.

5.4. CONCLUSIONS

Oxidation of 1,2-Dihydroxy-3-keto-5-(methylthio)pentene by ARD and ARD' has important implications in biology. This enzyme either facilitates recycling of methionine in living cells, or exits this recycling pathway. Fe^{2+}-dependent ARD' produces the α-keto acid precursor of methionine and formate, whereas Ni^{2+}-dependent ARD instead produces methylthiopropionate, CO, and formate. The way by which the two proteins do it solely due to the nature of the bound metal has been a long-standing mystery. However, it has been speculated that the mechanistic difference stems from the differences in the binding of the substrate to ARD versus ARD'. In this work we show that this old hypothesis does not hold up. Both proteins bind the substrate in the same way: as a six-membered ring, via attaching O1 and O3 to the metal (as in Figure 5.2D). This was shown with the use of our powerful hybrid dynamics methods, QM/DMD. The subsequent mechanistic study revealed that the determination of the final products happens later in the reaction, after dioxygen attacks the bound substrate. Upon splitting of dioxygen, ARD immediately decomposes the reacting system into its three experimentally found products. Contrastingly, ARD' passes through an additional split dioxygen intermediate, and then proceeds through an epoxy-like transition state with a small activation energy to the two products. The ability of ARD' to stabilize an additional intermediate and thus produce the two products is due to the RedOx flexibility of the Fe^{2+} as compared to the more electron-rich Ni^{2+}. Fe^{2+} transmits electrons from the residues coordinating it to bound dioxygen and populating its formerly p*-orbital. This leads to dioxygen splitting in the second intermediate, and eventual access to the ARD' reaction route. It is remarkable how just

two additional electrons on the metal center in ARD *versus* ARD' cause a dramatic mechanistic difference. All findings are in agreement with the available spectroscopic and isotope labeling data. This is the first time the mechanisms of ARD and ARD' have been explicated.

5.5. REFERENCES

1 Myers, W. R., Wray, J. W., Fish, S. & Abeles, R. H. (1993) Purification and characterization of an enzyme involved in oxidative carbon-carbon bond cleavage reactions in the methionine salvage pathway of *Klebsiella pneumoniae*. J. Biol. Chem. 268, 24785-24791.

2 Oram S. W., Ai J., Pagani G. M., Hitchens M. R., Stern J. A., Eggener S., Pins M., Xiao W., Cai X., Haleem R., Jiang F., Pochapsky T. C., Hedstrom L., Wang Z. (2007) Expression and Function of the Human Androgen-Responsive Gene *ADI1* in Prostate Cancer. Neoplasia 9, 643–651.

3 Dai, Y., Wensink, P. C. & Abeles, R. H. (1999) One protein, two enzymes. J. Biol. Chem. 274, 1193-1195.

4 Pochapsky, T. C., Ju, T., Dang, M., Beaulieu, R., Pagani, G. M. & OuYang, B. (2007) In *Metal Ions in Life Sciences*; (Sigel, A., Sigel, H. & Sigel, R. K. O., Eds.) Vol. 2, pp 473-500 Wiley-VCH: Weinheim, Germany.

5 Al-Mjeni, F.; Ju, T.; Pochapsky, T. C.; Maroney, M. J. (2002) XAS investigation of the structure and function of Ni in acireductone dioxygenase. Biochemistry 41, 6761-6769.

6 Dai, Y., Pochapsky, C. T. & Abeles, H. R. (2001) Mechanistic studies of two dioxygenases in the methionine salvage pathway of *Klebsiella pneumoniae*. Biochemistry 40, 6379-6387.

7 Chai, S.; Ju, T.; Dang, M.; Goldsmith, R.; Maroney, M. J.; Pochapsky, T.C. (2008) Characterization of metal binding in the active sites of acireductone dioxygenase isoforms from Klebsiella ATCC 8724. Biochemistry 47, 2428 -2438.

8 Ye, S., Riplinger, C., Hansen, A., Krebs, C., Bollinger, J. M. Jr. & Neese, F. (2012) Electronic Structure Analysis of the Oxygen-Activation Mechanism by FeII- and α Ketoglutarate (αKG)-Dependent Dioxygenases. Chem. Eur J. 18, 6555-6567.

9 Christian, J. G., Ye, S. & Neese, F. (2012) Oxygen activation in extradiol catecholate dioxygenases – a density functional study. Chem. Sci. 3, 1600-1611.

10 Wray, W. J.; Abeles, H.R. J. Biol. Chem. (1993) Purification and characterization of an enzyme involved in oxidative carbon-carbon bond cleavage reactions in the methionine salvage pathway of *Klebsiella pneumoniae.* 268, 21466-21469.

11 Wray, W. J. & Abeles, H.R. (1995) The methionine salvage pathway in *Klebsiella pneumoniae* and rat liver. Identification and characterization of two novel dioxygenases. J. Biol. Chem. 270, 3147-3150.

12 Borowski, T., Bassan, A. & Siegbahn, P. E. M. (2006**)** DFT Study of the Uncatalyzed Dioxygenation of Acireductone. J. Mol. Struct. Theochem. 772, 89-92.

13 Sparta, M., Ding, F., Shirvanyants, D., Dokholyan, N. V. & Alexandrova, A. N. (2012) Hybrid dynamics simulation engine for metalloproteins. Biophys. J. 103, 767-776.

14 Allpress, C. J., Grubel, K., Szajna-fuller, E., Arif, M. A. & Berreau, L. (2013) Regioselective Aliphatic Carbon–Carbon Bond Cleavage by a Model System of Relevance to Iron-Containing Acireductone Dioxygenase. J. Chem. Soc. 135, 659-668.

15 Pochapsky, T.C., Pochapsky, S.S., Ju, T., Hoefler, C. & Liang, J. (2006) A refined model for the structure of acireductone dioxygenase from *Klebsiella ATCC 8724* incorporating residual dipolar couplings. J. Biomol. NMR. 34, 117-127.

16 Ju, T.; Goldsmith, R.; Chai, S.; Maroney, M.J., Pochapsky, S.S., Pochapsky, T.C. (2006) One protein, two enzymes revisited: A structural entropy switch interconverts the two isoforms of acireductone dioxygenase. J.Mol.Biol. 393, 823-834.

17 Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. (1998) Molecular dynamics studies of folding of a protein-like model. Fold. Des. 3, 577–587.

18 Dokholyan, N. V. (2006) Studies of folding and misfolding using simplified models. Curr. Opin. Struct. Biol. 16, 79–85.

19 Ding, F., Guo, W. H., Dokholyan, N. V., Shakhnovich, E. I. & Shea, J. E. (2005) Reconstruction of the src-SH3 Protein Domain Transition State Ensemble using Multiscale Molecular Dynamics Simulations. J. Mol. Biol. 350, 1035–1050.

20 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. (2008) Ab initio folding of proteins with all atom discrete molecular dynamics. Structure 16, 1010–1018.

21 Sparta, M.; Alexandrova, A. N. (2012) How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. PLoS ONE 7, e47172.

22 Valdez, C. E.; Alexandrova, A. N. (2012) Why Urease Is a Di-Nickel Enzyme whereas the CcrA β-Lactamase Is a Di-Zinc Enzyme. J. Phys. Chem. B 116, 10649-10656.

23 Valdez, C. E.; Sparta, M.; Alexandrova, A. N. (2012) The Role of the Flexible L43-S54 Protein Loop in the CcrA Metallo-β-lactamase in Binding Structurally Dissimilar β-Lactam Antibiotics. J. Chem. Theor. Comput. 9, 730-737.

24 Andersen, H. C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. J. Chem. Phys. 72, 2384-2393.

25 Ramachandran, S.; Kota, P.; Ding, F.; Dokholyan, N. V. (2011) Automated minimization of steric clashes in protein structures. Proteins 79, 261–270.

26 Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. Acta Crystallogr., Sect. A. 32, 922–923.

27 Barton, G. J. OC - A cluster analysis program, University of Dundee, Scotland, UK (1993, 2002), www.compbio.dundee.ac.uk/downloads/oc.

28 Dirac, P. A. M. (1929) Quantum mechanics of many-electron systems. Proc. Royal Soc. (London) A 123, 714.

29 Slater, J. C. (1951) A Simplification of the Hartree-Fock Method. Phys. Rev. 81, 385-390.

30 Vosko, S. H., Wilk, L. & Nusair, M. (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can. J. Phys. 80, 1200-1211.

31 Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys. Rev. A 38, 3098-3100.

32 Perdew, J. P. (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B 33, 8822-8824.

33 Schafer, A., Horn, H. & Ahlrichs, R. (1992) Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. J. Chem. Phys. 97, 2571-2577.

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

35 Arnim, M. V.; Ahlrichs, R. (1998) Performance of parallel TURBOMOLE for density functional calculations. J. Comput. Chem. 19, 1746–1757.

36 Sierka, M.; Hogekamp, A.; Ahlrichs, R. (2003) Fast evaluation of the coulomb potential for electron densities using multipole accelerated resolution of identity approximation. J. Chem. Phys. 118, 9136–9148.

37 Turbomole V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, Turbomole GmbH, since 2007, available from http://www.turbomole.com.

38 Grimme, S. (2004) Accurate description of van der waals complexes by density functional theory including empirical corrections. J. Comput. Chem. 25, 1463–1473.

39 Klamt, A. & Schüürmann, G. *J.* (1993) COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc. Perkin Trans.2 5, 799-805.

40 Staroverov, V., Scuseria, G. & Tao, J.; Perdew, J. (2003) Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. J. Chem. Phys. 119, 12129–12137.

41 (a) Becke, A. D. (1993) Density-Functional Thermochemistry. III. The Role of Exact Exchange. J. Chem. Phys. 98, 5648−5652. (b) Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. (2010) A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132, 154104−15419.

42 Reed, A. E., R. B. Weinstock & F. Weinhold. (1985) Natural population analysis. J. Chem. Phys. 83, 735-746.

CHAPTER 6

Co^{2+} Acireductone Dioxygenase: Fe^{2+} Mechanism, Ni^{2+} Mechanism, or Something Else?

6.1. INTRODUCTION

Metalloenzymes carry out a broad and diverse range of biological functions such as nitrogen fixation, metabolism, respiration, photosynthesis and oxygen transport.[1] The role of the metal varies greatly for each process and still remains a great mystery along with why Nature chooses specific metals for catalysis. Acireductone dioxygenases (ARD and ARD') is a monometallic enzyme involved in the methionine salvage pathway in living organisms, including humans.[2,3,4,5] ARD oxidizes the substrate 1,2-dihydroxy-3-keto-5-(methylthio)pentene (acireductone) using dioxygen. Interestingly, the nature of the bound metal is the exclusive determinant of whether the products are recycled or excreted in the methionine salvage pathway. Fe-ARD' (EC 1.13.11.54) oxidizes acireductone into two products, formate and 2-keto-4-methylthiobutyric acid, the precursor of methionine, and then continues along the metabolic pathway. Ni-ARD (EC 1.13.11.53) catalyzes the formation of formate, methylthiopropinate and carbon monoxide, which exit the pathway. Through an extensive quantum mechanical/molecular mechanical (QM/MM) and mechanistic density functional theory (DFT) study, the reaction pathway differentiation was found to be solely due to the electronic properties of the metal, disproving previous speculations that the binding mode of the substrate to the metal center dictated the direction of the reaction pathway (Figure 6.1).[2] Fe-ARD' has an additional intermediate on the reaction profile, where the dioxygen attached to the substrate is dissociated. For Ni-ARD, dioxygen is activated but remains bound. This contrasting behavior is the result of the greater RedOx flexibility of the Fe^{2+} cation: having more holes in the 3*d*-set of atomic orbitals, Fe^{2+} efficiently transmits electrons from its ligands to the bound substrate and the antibonding orbital of the attached dioxygen molecule, causing its dissociation. From the split-dioxygen intermediate, there is an access to the low-energy epoxy-like transition state, which leads to the Fe-ARD' products. Ni^{2+} has the 3d-set nearly full, it binds the substrate more weakly, and does not act as an efficient electron pump. Hence, the mechanistic difference.

Thus, the effect of just two extra electrons in the metal in ARD on the mechanism of reaction is dramatic. If to consider the nature of the metal as a means to tune the catalytic mechanism and performance of metallo-enzymes, ARD represents a curious case.

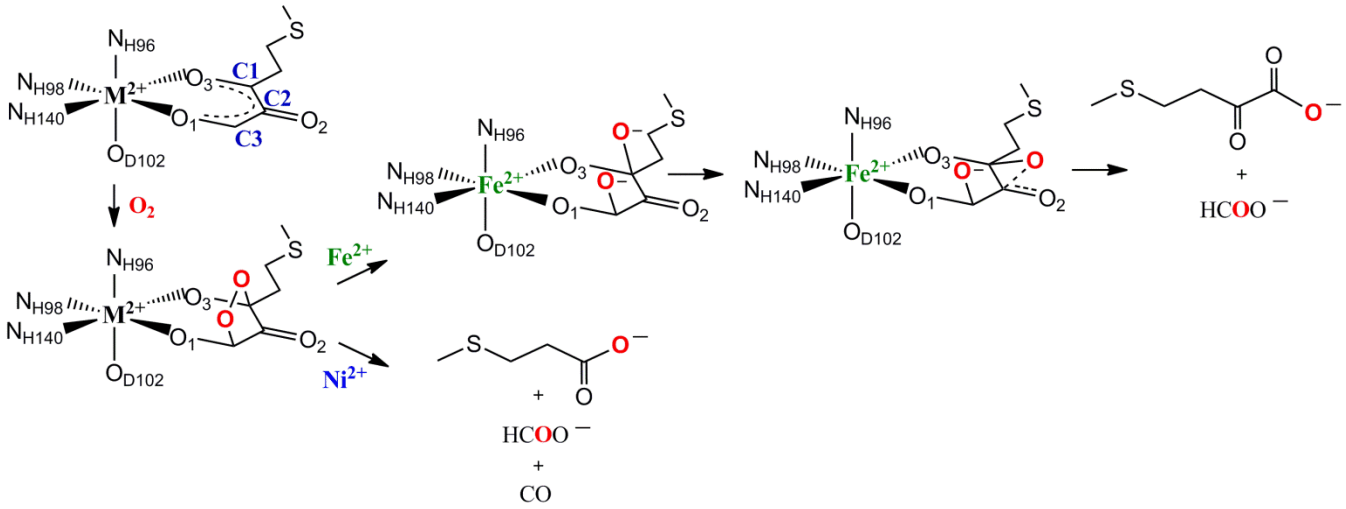


Figure 6.1. The mechanism of acireductone oxidation in ARD and ARD', where the additional "split-dioxygen" intermediate found in ARD' differentiates the two pathways.[2]

In this work we assess the catalytic mechanism accessible to Co^{2+} in the active site of ARD. Would Co-ARD follow the Fe-ARD' pathway, Ni-ARD pathway, or a completely different pathway? Co is a vital metal in biology,[6,7] found, for example, in cobalamin, commonly known as vitamin B12,[8] carbonic anhydrase,[9] and blue copper proteins.[10] In ARD, Co is not found naturally, but can be installed by ion exchange. Experimentally, when ARD is reconstituted with Co, it has been found to form primarily Ni products,[11,12] however the mechanistic details are not elucidated. In view of the exemplary sensitivity of the ARD enzyme to the electronic structure of the metal, we are interested to see how the Co-ARD mechanism unfolds and compares to the two native forms of the enzyme. This is a theoretical study employing mixed quantum-classical dynamics simulations for protein equilibration and identification of the substrate binding

geometries, and DFT calculations for the reaction mechanism, to fully explicate the mechanism of Co^{2+}-dependent ARD.

6.2. THEORETICAL METHODS

The initial structure of ARD was obtained from the Protein Data Base (PDB) (ARD' PDB code: 2HJI[13]) and used for this study, to fairly compare with our previous study.[2] Since the initial PDB does not contain the substrate, 2-dihydroxy-3-keto-5-methylthiopentene, coordinated to the metal, it was manually built into the protein. The coordination pose of the substrate to the metal, with O1 and O3 of the substrate deprotonated and positioned in a six-membered ring relative to the equatorial plane of the Co complex, was chosen based on experimental and theoretical predictions.[14,15,16] (Figure 6.2A) Previous QM/DMD simulations on the Fe-ARD' and Ni-ARD systems elucidated the need for adequate sampling of the backbone and substrate since the two residues, R104 and R154, were found to form vital hydrogen bonds with the substrate and active site.[2] R104 and R154 stabilize not only the doubled deprotonated substrate but also the six-membered ring orientation of the substrate to the metal.

This study utilizes our developed mixed quantum mechanical/discrete molecular dynamics[17,18,19,20] (QM/DMD)[21] method. QM/DMD is a hybrid method that efficiently captures metalloprotein dynamics with a quantum mechanical description of the active site on the order of tens to hundreds of nanoseconds. DMD differs from traditional molecular dynamics (MD) by the interaction potential functions, where the continuous potential functions are approximated with step functions of pairwise distances. This reduces DMD to event-drive MD simulations where a non-traditional time unit (t.u.) is introduced and defined as the shortest time between two consequent collisions in the system between two points. DMD has been utilized extensively on biomolecule systems such as the one in this study.[17,18,19,20] Examples of the capabilities of QM/DMD include comprehensive mechanistic studies of enzymatic reactions[22], unveiling the subtle structural details at metal-containing active sites[23] and proficient protein conformational

responses to metal replacement and substrate binding.[24] The affordability of QM/DMD is the key attractive quality of the method when compared to equally capable QM/MM methods.

Details of the QM/DMD method are described elsewhere.[25] To summarize them briefly, QM/DMD operates in an iterative scheme with three well defined regions – the QM-only, DMD-only and the QM-DMD boundary. First, DMD samples most of the protein, excluding the exclusively QM-only region - the metal and atoms directly coordinated to the metal (N from His, O from Glu and O1 and O3 of the substrate). With this constraint, no parameterization for the metal is needed in the force field since only QM can adjust the positions of the atoms relative to the metal. Following DMD, the extended QM region is extracted from the protein and goes through a geometry optimization, following the *ab initio* gradients for motions of the larger active site that includes the metals, ligands, the substrate, and may include other important amino acids near the reactive center. This extended QM region, the QM/DMD "breathing" boundary, allows the DMD and QM region to communicate with one another via the iterative scheme. (Appendix 9.3, Figure 9.3.1) The size of this boundary is adapted for each individual system and chosen carefully after preliminary simulations and observation of the behavior of the QM region with respect to other residues in the protein. The representative partitions, the QM-only, DMD-only and QM-DMD "breathing" boundary, are shown in Figure 6.2A. As applied in the previous ARD/ARD' study2, QM/DMD is used to sample the substrate bound to the Co. We also note that large fluctuations in QM and DMD energies are typical for complexes as large as the studied QM/DMD region.[22,23,24] Since Co is found in both low and high spin states in nature,[8,9,10] two sets of QM/DMD simulation were run for the two different spin states of Co.

6.2.1. QM/DMD PARTITIONING SCHEME

The QM/DMD partitioning is identical to previous work but in short goes as follows: The QM-only, DMD-only and QM-DMD boundary are all depicted in Figure 6.2. The QM-only domain

includes the Co^{2+} ion, NE2 of H96, H98 and H140, OE1 of E102 and the two deprotonated oxygens (O1 and O3) of the substrate. The QM-DMD boundary includes the remaining atoms of residues H96, H98, H140, E102, the substrate, and two arginines, R104 and R154, that stabilize the doubly deprotonated substrate and active site during the reaction. The truncated amino acids in the QM optimization are chopped at the Cα-Cβ bond, Cδ-Cγ for arginines, and saturated with hydrogens. The hydrogens are positioned along the original bond at a distance equal to 0.7052*R(C-C/N). All saturated hydrogens and their bond partners are frozen during the QM optimization to retain the protein scaffold geometry absent during this phase. During DMD, further constraints are imposed to prevent any residues coordinating the metal from distorting too much and losing their coordination with the metal. DMD is allowed to sample those constrained atoms only within ±0.01 Å from the values predicted by QM. These constraints have been shown to not hinder important sampling movement of the protein and are only present to prevent unrealistic distortions of the residues coordinated to the metal during the DMD phase.[21]

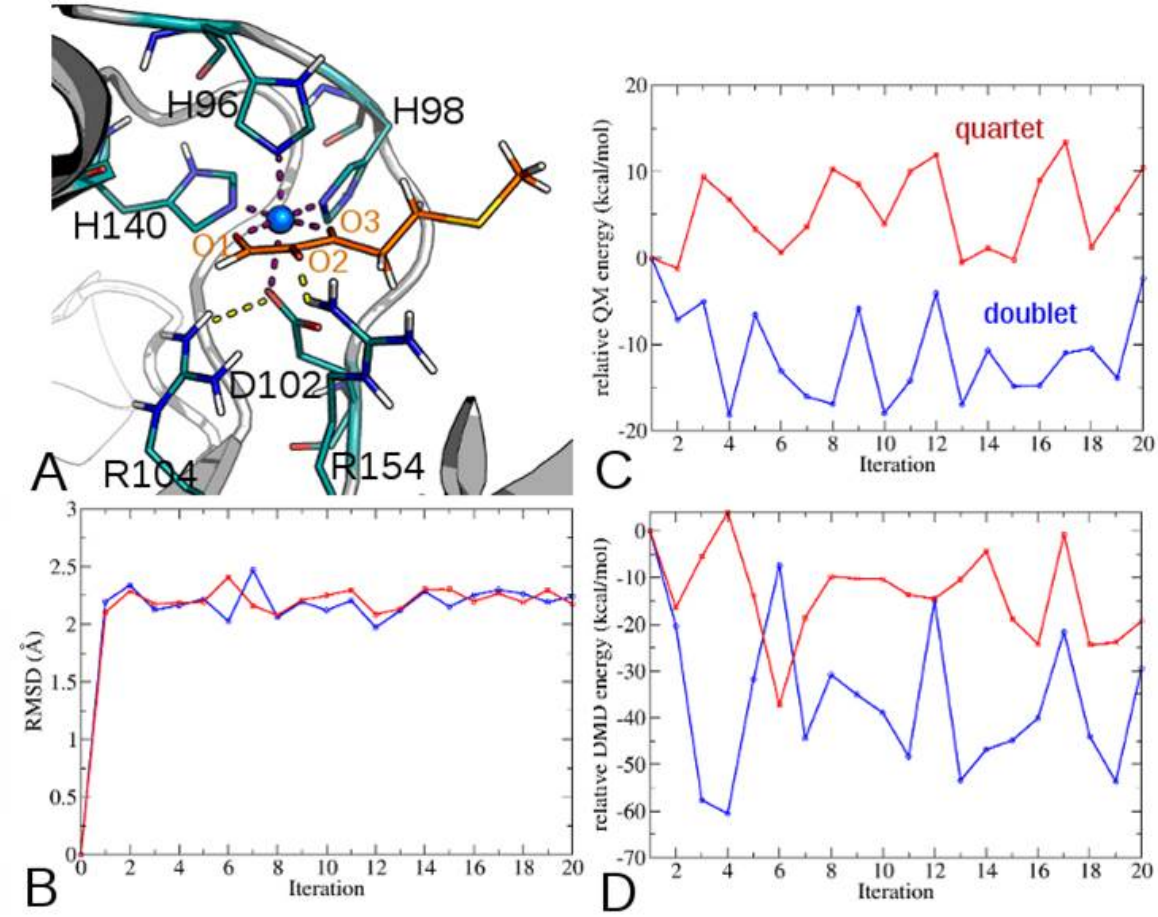


Figure 6.2. (A) The QM-DMD boundary in the protein is colored teal and includes residues H96, H98, D102 and H140 that directly coordinate the metal, the Co^{2+} metal, the substrate (orange) and R104 and R154, residues found to be important in stabilizing the dianionic substrate. Parts(B)-(D) who the convergence data from QM/DMD simulations for the doublet (blue lines) and quartet (red lines): (B) The backbone RMSD shows overall stabilization of the protein structure; (C) and (D) are the relative QM and DMD energies.

6.2.2. DETAILS OF THE QM/DMD SIMULATIONS

Each simulation begins with a short DMD run of 1,000 DMD time units (t.u.) (~ 50 fs) at a DMD temperature of 0.1 kcal/mol*K, with a high heat exchange rate of the protein with the bath for 10 $t.u.^{-1}$, using the Andersen thermostat.[26] This short DMD simulation with high heat exchange was found to remove clashes introduced from pdb starting structures. Following the removal of clashes, a simulated annealing procedure is performed, shown to be crucial for

avoiding local minimum traps during sampling and achieve better statistics of the resulting ensemble.[21] The annealing equilibration peaks at a temperature of 0.2 kcal/mol*K and, in 0.02 kcal/mol*K increments, cools down stepwise (5 steps, consisting of 500 t.u. each) to the production run simulation temperature with the equilibrated structure. The temperature is kept low for 10,000 t.u. during which the DMD trajectory is captured. The DMD ensemble is clustered according to geometric similarity based on the Kabcsh[27] RMSD for all pair-wise snapshot structures and by applying a hierarchical clustering algorithm.[28] For each cluster, both the structure closest to the centroid and the one with the lowest DMD energy are used as representatives for the QM phase.

Following DMD, the QM-DMD region is extracted from the centroid and lowest DMD energy structures and capped with hydrogens as discussed above. These QM ready structures undergo a single point calculation, the method of choice being DFT used in the BP86[29,30,31,32,33] formulation of the exchange and correlation functional. A double ζ quality basis set (def2-SVP)[34] for H, C, N, O, S, and a triple ζ quality basis set (def2-TZVPP)[35] for the metal were used. The use of mixed basis sets (MBS) is customary in DFT studies of systems containing transition metals[36] and we find it important for our systems due to the computational speed up without sacrificing any vital information. To speed up the calculations, the Resolution of Identity (RI),[37] and Multipole Accelerated Resolution of Identity (MARI-J)[38] as implemented in *Turbomole*[39] were exploited. Empirical dispersion correction for DFT calculations was included both in the energy and gradient evaluations.[40] Solvent was included via the Conductor-like Screening Model (COSMO) continuum solvent with the dielectric constant set to 20.0,[41] chosen as the best representation of the partially solvent-exposed active site of ARD and within the commonly used range of 4.0-20.0 used in computational studies of enzymes. [42,43] Next, after all QM single point calculations, each structure is scored based on both the QM and DMD energies, and a single structure is selected from each iteration. The selected active site structure is partially optimized

at the QM level, fixing the points of attachment during optimization to resemble the rest of the protein. Lastly, the active site is reinstalled into the protein, and the QM-DMD boundary shrinks back to the QM-only and the simulation continues to the next iteration (new DMD sampling phase). The overall QM/DMD scheme propagates until convergence is reached. The convergence criterion for our simulations is determined by the stabilization of the root-mean-squared-deviation (RMSD) of backbone Cα and QM and DMD energy components. These fluctuations are typical for complexes as large as the studied QM-DMD-region.[21,22,23] Monitoring these variables, these simulations converged quickly (Figure 6.2B-D). 20 iterations, roughly corresponding to 10.5 ns of dynamics, were found to be sufficient.

6.2.3. DFT MECHANISTIC STUDY

The QM/DMD simulations produce a set of well-equilibrated structures, of we which we chose the lowest QM energy structure to perform the mechanistic study. To continue on the same potential energy surface as the QM optimizations during QM/DMD, all stationary points along the reaction profile were calculated in an identical fashion: using the BP86 functional plus empirical dispersion correction, a double ζ quality basis set (def2-SVP) for H, C, N, O, S, a triple ζ quality basis set (def2-TZVPP) for cobalt and implicit solvation via COSMO with a dielectric of 20.0. BP86 is mentioned in a recent review[43] to give better energies for B_{12}-dependent enzymes that contain Co^{2+} when compared to B3LYP. In addition, TPSSh[44]/def2-TZVPP and B3LYP[45]/def2-TZVPP with empirical dispersion and implicit solvation were used to calculate the energies of the stationary points. Jensen et. al reported TPSSh is the most reliable choice of functionals for studying bioinorganic complexes.[46] The results reported in the main text are obtained with TPSSh, and those obtained with the BP86 and B3LYP functionals are given in the Supporting Information (Appendix 9.3, Table 9.3.1). All charges were computed using Natural Population Analyses (NPA)[47] at the TPSSh/def2-TZVPP level of theory. We recognize the

weakness of DFT when it comes to systems that exhibit spin cross over, however, with the size of our systems, we are limited to DFT and do believe we have captured the qualitative details of the system.

Mechanistic investigations were done on both low and high spin states of Co^{2+} structures generated from QM/DMD. Previous studies have shown that the low spin structures of Co^{2+} complexes are generally what exist in organometallic complexes and uniquely cobalamin, but high spin Co^{2+} exists for most natural metalloenzymes as shown by experiment and computation.[48] To include all possible spin combinations, the sextet adduct is calculated only in the mechanism and, as expected, too high in energy to be a viable catalytic option.

6.3. RESULTS AND DISCUSSION

6.3.1. CATALYTIC MECHANISM OF DOUBLET CO^{2+}

The lowest QM structure from the QM/DMD ensemble was selected and used in the mechanistic study. As with ARD/ARD', the barrier toward dioxygen attack is not considered because the differences between ARD/ARD' are rooted in later steps of the reaction profile.[2]

For doublet Co^{2+}, the mechanism proceeds as follows: dioxygen binds to C1 and C3 of the substrate (see Figure 6.1 for atom labeling) and forms the dioxygen adduct with a negligible stabilization of the complex (0.4 kcal/mol). Next, the dioxygen separates with a barrier of 3.7 kcal/mol immediately into the split intermediate found in Fe-ARD', where only the O-O bond is broken but the products are still coordinated to the metal. Following the split intermediate, O2 and C3 pass through an epoxy-like transition state of 4.6 kcal/mol and form the two products, formate and the α-keto acid. Doublet Co^{2+} was not observed to form Ni-ARD products upon the splitting of dioxygen (Figure 6.3).

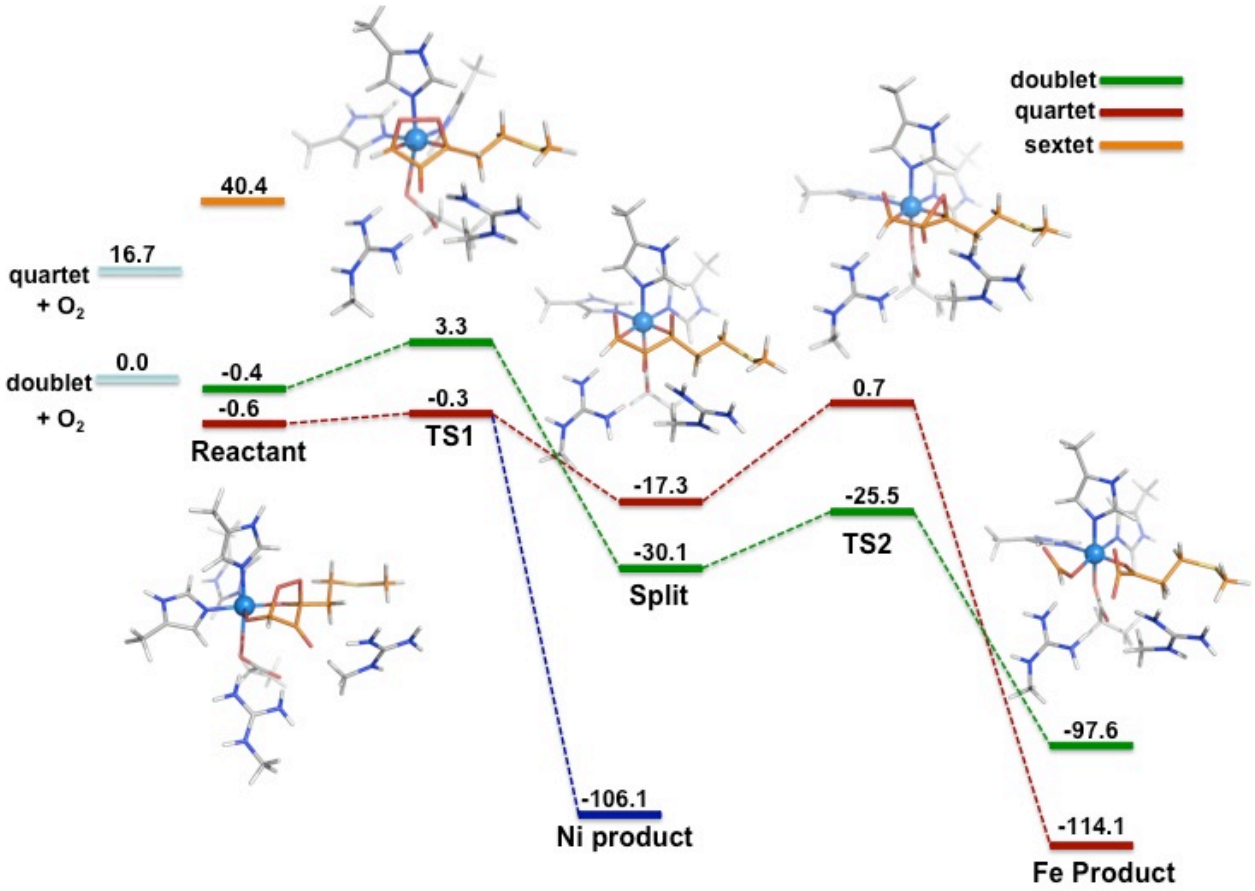


Figure 6.3. Catalytic mechanism of Co-ARD shows spin crossing at the dioxygen splitting and upon formation of the products (all relative energies reported in kcal/mol). A bifurcation in the PES is also found for the quartet system at TS1. All structures are optimized with BP86/def2-SVP (C,O,N,H) and def2-TZVPP (Co) with single point energies at TPSSh/def2-TZVPP.

6.3.2. CATALYTIC MECHANISM OF QUARTET CO^{2+}

The quartet complex without dioxygen is substantially higher in energy (16.7 kcal/mol) then the doublet complex. However, upon dioxygen binding and formation of the adduct, the doublet (-0.4 kcal/mol) and quartet adduct (-0.6 kcal/mol) are essentially degenerate (Figure 6.3). Continuing along the path, the dioxygen splitting transition state (TS1) proceeds with a negligible barrier (0.3 kcal/mol) into the Ni-ARD products. However, TS1 sits at an unexpected bifurcation in the PES, where the reaction has the choice to either proceed directly to the oxidation of the substrate to Ni-ARD products, or continue along the Fe-ARD' pathway.

Following the O-O bond split, we are able to capture the thermodynamically preferred Ni-ARD products. However, proceeding through to the split intermediate into the Fe-ARD' products is still energetically very favorable, through the epoxy-like TS2 with a barrier of 18.0 kcal/mol. The kinetics of this bifurcation and the resultant ratio of the forming products cannot be assessed without the use of on-the-fly dynamics, and this system is indeed interesting for the future dynamics study.

6.3.3. UNIQUE SPLIT DIOXYGEN INTERMEDIATE IN LOW AND HIGH SPIN Co^{2+}-ARD

The split dioxygen intermediate found in both low and high spin Co-ARD was first observed in Fe-ARD. To compare these intermediates to the Fe-ARD one, NPA calculations of partial charges are computed for all lowest energy stationary points for both spin states of Co (Table 6.1, listed in order of low spin Co-ARD, high spin Co-ARD and Fe-ARD'). If you follow the NPA charges of all the residues coordinating the metal, substrate and dioxygen throughout the duration of the mechanism, there are several charges to take note of: 1) Upon formation of the split dioxygen intermediate and into the epoxy-like transition state, the coordinating residues (H96, H98, D102 and H140) become more positively charged. The electrons withdrawn from these residues flow through the Co metal center and donate electron density to the π* orbital of dioxygen, stabilizing the split intermediate. 2) Although the metal facilitates the flow of electron density as a RedOx flexible center, it does not change oxidation state throughout the reaction profile.

Comparing bond distances of the two spin states of Co-ARD, overall low spin Co has shorter metal-ligand bond distances than high spin Co complexes. The low spin adduct has Co-O1 and Co-O3 bond distances of 1.93 and 1.91 Å, respectively. For high spin Co, the bond distances are 2.10 and 1.97 Å, respectively. The shorter bond distances of low spin Co adduct are comparable to the longer metal-ligand distances observed in Fe-ARD' (1.91 and 1.93 Å,

respectively) giving the low spin Co-ARD a stronger and more RedOx active interaction. The longer bond distances of high spin Co are similar to the ones found in Ni-ARD (2.19 and 2.00 Å, respectively). These longer bond distances could be a contributing factor in high spin Co-ARD bifurcation PES. Co can play the role of a Lewis acid, coordinating and polarizing the substrate without any RedOx activity, or Co can play an active role as a RedOx flexible metal and pass through the split intermediate into the Fe-ARD' products.

The two spin states considered here (starting from the complexes of the protein with the substrate and dioxygen bound) yield very accessible reaction profiles that are nearly-degenerate in the area of TS1, and go down-hill from there. Both Ni-ARD and Fe-ARD' products are predicted from our simulations for Co^{2+}-dependent ARD. Our calculations may not have recognized the high spin route as the exclusive option since we do not have the spin crossing profile and do not consider the dioxygen attachment to the substrate forming the adduct. However, we do see that the high spin Co profile does form Ni products very efficiently, which has been shown in experiments. The presence of spin cross-over between the doublet and quartet is evident but something we are not able to address because of prohibitively high computational cost. It is also possible that the Co-ARD mechanism is kinetically driven.

	H96	H98	D102	H140	R104	R154	Substrate	O_2	Co^{2+}
Reactant	**0.08**	**0.14**	**-0.82**	**0.17**	**0.85**	**0.96**	**-0.83**	**-0.65**	**1.09**
	0.05	***0.04***	***-0.85***	***0.08***	***0.90***	***0.99***	***-1.04***	***-0.65***	***1.48***
	0.08[a]	0.07	-0.82	0.06	0.88	0.94	-1.00	-0.66	1.46
TS1	**0.08**	**0.14**	**-0.82**	**0.17**	**0.85**	**0.97**	**-0.80**	**-0.68**	**1.09**
	0.06	***0.05***	***-0.86***	***0.08***	***0.90***	***0.99***	***-1.02***	***-0.65***	***1.46***
	0.08	0.07	-0.82	0.06	0.88	0.94	-0.98	-0.68	1.45
Split	**0.28**	**0.26**	**-0.67**	**0.26**	**0.92**	**0.94**	**-0.73**	**-1.36**	**1.10**
	0.11	***0.22***	***-0.78***	***0.22***	***0.82***	***0.95***	***-0.47***	***-1.47***	***1.40***
	0.13	0.15	-0.79	0.15	0.90	0.94	-0.66	-1.30	1.48
TS2	**0.27**	**0.26**	**-0.65**	**0.28**	**0.92**	**0.91**	**-0.75**	**-1.33**	**1.10**
	0.13	***0.15***	***-0.77***	***0.14***	***0.80***	***0.94***	***-0.51***	***-1.46***	***1.59***

	0.13	0.16	-0.79	0.16	0.90	0.93	-0.65	-1.34	1.50
Product (Fe)	**0.17**	**0.10**	**-0.78**	**0.23**	**0.91**	**0.92**	**-0.21**	**-1.41**	**1.08**
	0.09	***0.08***	***-0.83***	***0.11***	***0.86***	***0.92***	***-0.29***	***-1.41***	***1.47***
	0.07	*	-0.81	0.09	0.90	0.92	-0.33	-1.41	1.47
Product (Ni)	*	*	*	*	*	*	*	*	*
	0.09	***0.01***	***-0.82***	***0.08***	***0.88***	***0.91***	***-0.19***	***-1.45***	***1.50***

[a] Values are taken from reference 2

Table 6.1. Natural Population Analysis (NPA) charges are listed for residues H96, H98, D102, H140, the substrate, dioxygen and the metal for all stationary points along the reaction path for low spin Co-ARD (top row, bold), high spin Co-ARD (second row, bold and italized) and Fe-ARD' (third row). When comparing all the charges from the reactants' stationary point to the split intermediate, an overall trend of increasing positive charge on all the residues and the negative change on O_2 can be seen. Charges were taken from structures optimized at BP86/def2-SVP (H, C, O, N, H) and def2-TZVPP (Co) and singles points taken at TPSSh/def2-TZVPP

6.4. CONCLUSIONS

The pair of acireductone dixoygenases, ARD' and ARD, facilitate the recycling of methionine in cells when Fe^{2+} is present within the active site or exit the pathway entirely when Ni^{2+} is present. Our previous study explicated the mechanism behind this long-standing mystery, showing that the metal solely dictates the mechanism through the degree of charge transfer to dioxygen in the reaction, linked to the RedOx flexibility of the metal. This purely metal-dependent functionality led to exploring how Co, the element in between Fe and Ni, oxidizes the substrate, 1,2-dihydroxy-3-keto-5-(methylthio)pentene. Through mixed quantum-classical dynamics simulations using our QM/DMD method, and subsequent DFT mechanistic studies, we showed here that low spin Co^{2+} PES exclusively follows the Fe^{2+}-dependent pathways, producing α-keto acid precursor of methionine and formate, while the high spin Co^{2+} PES contains a bifurcation in the pathway that follows along both the Fe^{2+}-dependent pathway and Ni^{2+}-dependent pathway that produces methylthiopropionate, carbon monoxide and formate. Thus Co-ARD should be able to produce both sets of products, partially working as an

intermediate form between ARD and ARD'. The resultant product distribution could be assessed through dynamics studies of the complete reaction path, but this remains to be done in the future.

6.5. REFERENCES

1 Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, California, 1994.

2 Sparta, M.; Valdez, C. E.; Alexandrova, A. N. *J. Mol. Biol*. **2013**, *245*, 3007-3018.

3 Myers, W. R.; Wray, J. W.; Fish, S.; Abeles, R. H. *J. Biol. Chem*. **1993**, *268*, 24785-24791.

4 Oram S. W.; Ai J.; Pagani G. M.; Hitchens M. R.; Stern J. A.; Eggener S.; Pins M.; Xiao, W.; Cai X.; Haleem R.; Jiang F.; Pochapsky T. C.; Hedstrom L.; Wang Z. *Neoplasia*, **2007**, *9*, 643–651.

5 Maroney, M. J.; Ciurli, S. *Chem. Rev*. **2013**, DOI: 10.1021/cr4004488

6 Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valetine, J. S. *Biological Inorganic Chemistry: Structure and Reactivity*; University Science Books, 2007.

7 Carloni, P.; Alber, F.; Sebastiani, D.; Rothlisberger, U. *Advances in Density-functional based Modeling Techniques - Recent Extensions of the Car-Parrinello Approach. In Quantum Medicinal Chemistry*; Wiley-VCH: Weinheim, Germany, 2003; pp 5-36.

8 Eriksson, L. A. *Theoretical Biochemistry: Processes and Properties of Biological Systems*; Elsevier Science B. V., 2001.

9 Lindskog, S.; *Cobalt (II) in Metalloenzymes. A reporter of Structure-Function Relations*; Berlin, Germany, 1970, 153-196.

10 McMillian, D. R.; Rosenberg, R. C.; Gray, H. B. Proc. Nat. Acad. Sci. 1974, 71, 4760-4762

11 Dai, Y.; Wensink, P. C.; Abeles, R. H. *J. Bio. Chem*. **1999**, *274*, 1193-1195.

12 Dai, Y.; Pochapsky, T. C.; Abeles, R. H. *Biochemistry.* **2001**, *40*, 6379-6387.

13 Ju, T.; Goldsmith, R.; Chai, S.; Maroney, M.J.; Pochapsky, S.S.; Pochapsky, T.C. *J. Mol. Biol.* **2006,** *393*, 823-834.

14 F. Al-Mjeni, T. Ju, T.C. Pochapsky, M.J. Maroney. *Biochemistry,* **2002***, 41, 6761–6769.*

15 Pochapsky, T. C.; Ju, T.; Dang, M.; Beaulieu, R.; Pagani, G. M.; OuYang, B. In *Metal Ions in Life Sciences*; Wiley-VCH: Weinheim, Germany, 2007; Vol. 2, pp 473– 500

16 Wray, W. J.; Abeles, H.R. *J. Biol. Chem*. **1995**, *270*, 3147-3150.

17 Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Fold. Des*. **1998**, *3*, 577–587.

18 Dokholyan, N. V. *Curr. Opin. Struct. Biol.* **2006**, *16*, 79–85.

19 Ding, F.; Guo, W. H.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J. E. *J. Mol. Biol*. **2005**, *350*, 1035–1050.

20 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.

21 Sparta, M.; Ding, F.; Shirvanyants, D.; Dokholyan, N. V.; Alexandrova, A. N. *Biophys. J.* **2012**, *103*, 767-776.

22 Valdez, C. E.; Alexandrova, A. N. *J. Phys. Chem. B*. **2012**, *116*, 10649-10656.

23 Valdez, C. E.; Sparta, M.; Alexandrova, A. N. *J. Chem. Theory Comput*. **2012**, *9*, 730-737.

24 Sparta, M.; Alexandrova, A. N. *PLOS One*, **2012**, *7*, e47172

25 Sparta, M.; Ding, F.; Shirvanyants, D.; Dokholyan, N. V.; Alexandrova, A. N. *Biophys. J.* **2012**, *103*, 767-776.

26 Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384-2393.

27 Kabsch, W. *Acta Crystallogr. A.* **1976**, *32*, 922–923.

28 Barton, G. J. OC - A cluster analysis program, University of Dundee, Scotland, UK (1993, 2002), www.compbio.dundee.ac.uk/downloads/oc.

29 Dirac, P. A. M. *Proc. Royal Soc. (London) A*. **1929**, *123*, 714.

30 Slater, J. C. *Phys. Rev*. **1951**, *81*, 385-390.

31 Vosko, S. H.; Wilk, L.; Nusair, M.; *Can. J. Phys*. **1980**, *80*, 1200-1211.

32 Becke, A. D. *Phys. Rev. A*. **1998**, *38*, 3098-3100.

33 Perdew, J. P. *Phys. Rev. B*. **1986**, *33*, 8822-8824.

34 Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571-2577.

35 Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys*. **2005**, *7*, 3297–3305.

36 Yang, Y,; Weaver, M. N.; Merz, K. M. *J. Phys. Chem. A.* **2009**, *113*, 9843-9851.

37 Arnim, M. V.; Ahlrichs, R. *J. Comput. Chem*. **1998**, *19*, 1746–1757.

38 Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136–9148.

39 Turbomole V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, Turbomole GmbH, since 2007, available from http://www.turbomole.com.

40 Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

41 Klamt, A.; Schüürmann, G. J. *J. Chem. Soc. Perkin Trans. 2* **1993**, *5*, 799-805.

42 Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421-437.

43 Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.; Siegbahn, P.E.M. *Chem. Rev*. DOI:10.1021/cr400388t.

44 Staroverov, V.; Scuseria, G.; Tao, J.; Perdew, J. *J. Chem. Phys.* **2003**, *119*, 12129-12137.

45 (a) Becke, A. D. J. *Chem. Phys.* **1993**, *98*, 5648−5652. (b) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104−15419.

46 Jensen, K. P. *Inorg. Chem*. **2008**, *47*, 10357-10365.

47 Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys*. **1985**, *83*, 735-746.

48 Jensen, K. O.; Ryde, U. *J. Mol. Struc: THEOCHEM*, **2002**, *585*, 239-255.

CHAPTER 7

Computational Redesign of Carboxypeptidase A and Experimental Confirmation

7.1. INTRODUCTION

Nature provides well-folded protein scaffolds that catalyze some of the most difficult reactions known to humankind. Harnessing this catalytic fortitude is a goal within the enzyme design community, providing the opportunity to create “greener” ways to produce molecules of interest to pharmaceutical and industrial scale production lines.[1,2] Introducing slight modifications to naturally occurring proteins allows one to shift functionality by tuning properties such as substrate specificity, inducing stereospecific reactions,[3] altering protein stability, and even catalyzing non-native reactions.[4] In particular, computational redesign of natural enzymes is a powerful technique. Having the computational resources and capabilities to redesign natural enzymes through small adjustments to the active site or second coordination sphere residues, altering the function and/or catalytic ability without designing the protein from the ground up, worrying about the ability of the protein to fold and/or stay stable, proves to be a prevailing strategy to protein design. Many works demonstrate that the substitution of a single to just a few amino acids can alter the catalytic activity and specificity. The design and redesign of enzymes without catalytic metals is vast.[5,6,7,8,9,10,11,12] The substrate preference of a serine peptidase was altered to favor a new peptide substrate with a proline-glutamine (PQ) dipeptide motif over the native proline-arginine or proline-lysine (PR or PK) substrates.[13] The design efforts with metal-free enzymes has yielded many different *in silico* tools that have led to successful products, however these tools are insufficient for metal-containing enzymes.[14,15] When designing metalloenzymes, it is critical to properly treat the electronic structure of the metal, to correctly predict its coordination. It is also desirable to efficiently sample the protein backbone, since multiply-charged cations create strong polarization in the binding site, and restructuring might be significant.

There are efforts in creating metal-containing biocatalysts, though progress has been partly limited by available methodology. In a purely experimental effort, Edge, et. al. performed

mutagenesis on a zinc-containing peptidase, specificity targeting one residue in the binding pocket of protein and changing the polarity of the pocket.[16] Efforts in pure *de novo* protein design have been successful in not only creating artificial metalloenzymes[17] but also in altering and repurposing them. The function of a family of *de novo* proteins, di-iron oxo-proteins or "Due Ferri" proteins,[18] were altered through modifications of a loop far from the active site, mutations chosen to redirect the enzyme's intended function.[19] The addition of an extra metal-binding residue and adjusting the substrate cavity switches a *due ferri* protein originally intended to oxidize hydroquinones to catalyze the selective N-hydroxylation of arylamines.[20]

In this work, we probe some of our computational tools for the design of metal-dependent enzymes. The specific system to redesign is the natural enzyme, Carboxypeptidase A (CPA) (E.C. 3.4.17.1).[21,22,23] CPA is a Zn^{2+}-containing enzyme that has been extensively studied both computationally and experimentally, and provides a well-studied model system to test our tools. Our goal is to modify the substrate, and then predict mutations to the protein that lead to proper positioning of the substrate with respect to the catalytic Zn and important residues in the active site, and small reaction barrier.

CPA is an important exopeptidase produced in the pancreas and preferentially cleaves C-terminal alphatic and aromatic amino acids from dietary proteins.[24] Even though peptides are the intended substrates, CPA can also hydrolyze other unnatural substrates, including esters. Because CPA was crystallized early on, compared to other members of the zinc hydrolase superfamily, it served as a prototype for understanding catalysis and substrate binding of such enzymes.[25,26] Historically, much work on CPA was motivated by its sequence similarities to angiotensin-converting enzyme (ACE). ACE plays an important role in regulating blood pressure and its structure has only been available since 2003.[27] Therefore, much is known about the structure and function of CPA.

The active site of CPA consists of one Zn^{2+} ion coordinated by two histidines (His69 and

His196), one glutamate residue (Glu72), and one water molecule, a common motif for catalytic zinc sites.[28] Zinc plays a vital role in the enzymatic activity of CPA and other zinc-containing enzymes.[29] In CPA, the zinc ion acts as a Lewis acid, lowering the pKa of the bound water molecule therefore, rendering the water molecule a more potent nucleophile, and polarizing the carbonyl oxygen of the bound peptide substrate (Figure 7.1). Also, a higher resolution crystal structure of CPA suggests that the zinc-bound water is in its neutral state.[30] The C-terminal peptide specificity of CPA comes from the highly organized hydrogen bond network within the binding pocket that involves interactions between the terminal carboxylic acid, and residues Asn144, Arg145, Tyr248 and Arg127 (Figure 7.1B). Furthermore, the residue specificity comes from the hydrophobic pocket, also known as the S' subsite, that is responsible for the preference for aromatic and aliphatic side chains. Additionally, Arg127 not only interacts with the terminal carboxylic acid but also with the carbonyl oxygen of the peptide backbone coordinated to zinc, an interaction believed to contribute to catalysis and substrate binding.[31] During the course of the reaction, Arg127 stabilizes the oxyanion developed on the tetrahedral intermediate (Figure 7.1A).

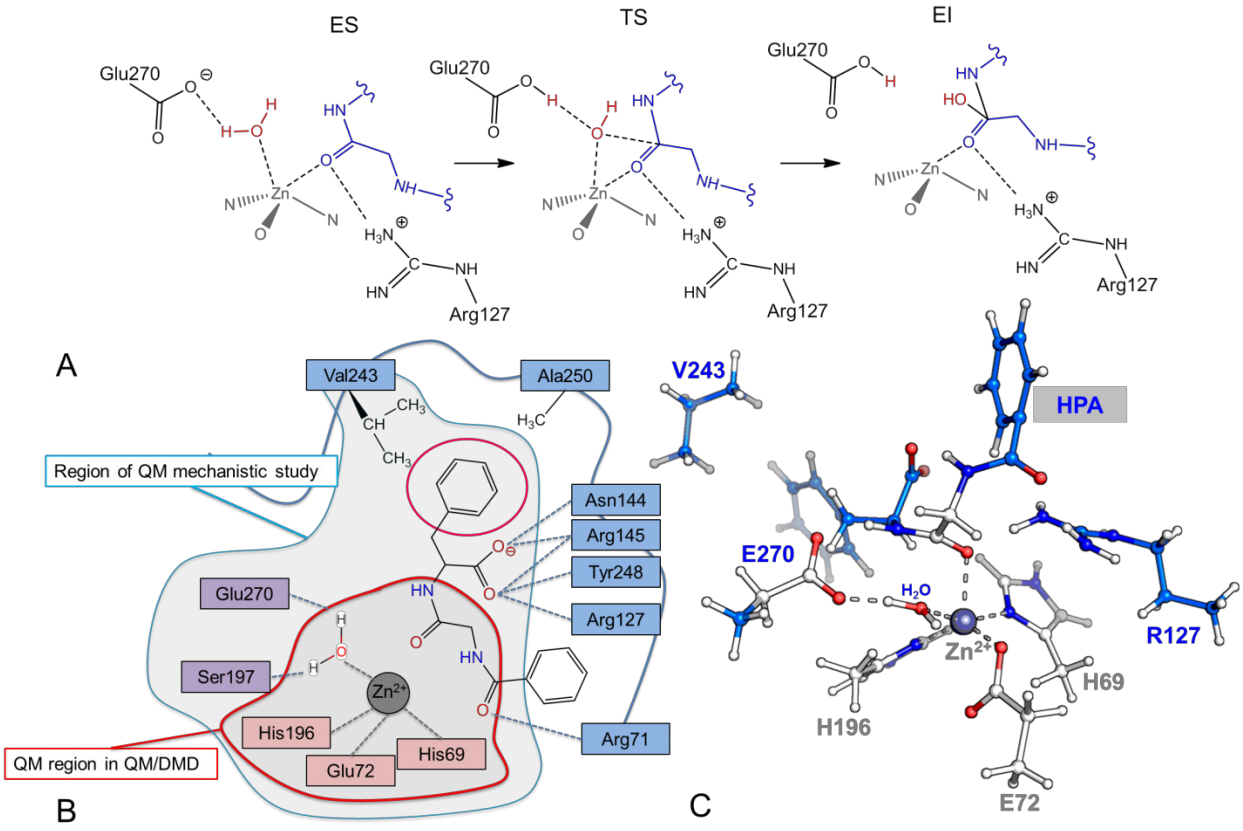


Figure 7.1. (A) Water-promoted mechanism of peptide hydrolysis in Carboxypeptidase A (CPA) with truncated peptide. Both Arg127 and Glu270 are thought to play vital roles in the reaction mechanism, Glu270 provides a hydrogen bond acceptor for the zinc-bound water and Arg127 stabilizing the negative charge build up on the tetrahedral intermediate. (B) Active site of CPA with region included in the QM mechanistic study and the truncated region used in QM/DMD simulation the hippuryl-L-phenylalanine (hippuryl-L-Phe). The pink circle indicates the bulky hydrophobic group that binds in the hydrophobic S' subsite. (C) Regions used in QM/DMD simulation and mechanistic study. The gray residues (H69, E72, H196) and part of peptide substrate represent the QM/DMD-region used during the simulation. The blue region includes the additional residues (V243, R127) and parts of the peptide added into the QM mechanistic study.

CPA is a well-studied enzyme from a structural and mechanistic perspective, yet the exact mechanism has still not been fully resolved. There are two potential pathways the hydrolysis could follow: via an anhydride mechanism or via a water-promoted mechanism. The anhydride mechanism is thought to compete with water-promoted mechanism for hydrolysis of

non-natural substrates, such as esters[32], however, for peptide hydrolysis the water-promoted one is believed to be the energetically favored one.[33] (Figure 7.1A) Several computational works have shed light on the mechanism of peptide hydrolysis in CPA. For the QM mechanistic study of reaction profile of CPA, most works utilize small cluster models or semi-empirical methods.[34,35,36] With the advancement of more efficient QM/MM methods, these robust methods are well-suited to describe the metal-ligand bonds and electrostatic contribution from the extended hydrogen bond network . The hybrid dynamics method that we will use, QM/DMD,[37] has tremendous success in modeling metalloenzymes, allowing sufficient sampling within the nanosecond regime and without the need for the parameterization of the metal.

In this study, we used QM/DMD to evaluate several peptide/protein mutations. Based on the predicted structures, one combination, V243R,hippuryl-L-glutamate, (V243R_FpepD), was predicted to replicate the positioning of the substrate with respect to the catalytic center that is conducive to successful hydrolysis. To gain further insight into the differences between the rate-determining steps in the native and V243R_FpepD mutant CPA, we employ a quantum theory of atom in molecules (QTAIM) analysis, investigating the topology of the electronic charge density to gain a deeper understanding of the bonding interactions where the chemical reaction takes place. This analysis gives a clue about possible reactivity. The subsequent mechanistic study further showed this CPA variant to be catalytically active. Finally, we tested its activity experimentally, and confirmed theoretical predictions.

7.2. METHODS

7.2.1. QM/DMD EQUILIBRATION OF THE NATIVE AND MUTANT PROTEINS

For the structural equilibration of the CPA variants with the bound substrates, this study employs our QM/DMD[37] method, where DMD is short for discrete molecular dynamics.[38,39,40,41] QM/DMD is a hybrid method, a variant of QM/MM, that efficiently captures metalloprotein

dynamics on the order of tens of nanoseconds, in conjunction with the quantum mechanical description of the active site. It has been shown to perform exceptionally well for recapitulating and recovery of native protein structures down to the subtle structural details at the active sites,[37] electronic properties of the bound metals, [42] protein conformational responses to substrate binding and metal replacement,[37, 43] and mechanistic studies of enzymatic reactions.[37,44] The main attractive feature of QM/DMD is its affordability as compared to that of other capable QM/MM methods. The strength of the method is in part due to DMD, a remarkably successful classical force field based method for aggressive and fast sampling biological molecules and their complexes. DMD incorporates solvation through an implicit averaged way into the force field in use.

The system is split up into three different boundary regions referred to as: QM-only, the DMD-only, and the shared QM-DMD region. DMD samples the entire protein except for the metal and atoms immediately coordinated to it, which are held frozen during the sampling. Due to this choice, there is no need for the parameterization of the force field to describe the metal, which is then free to change the coordination environment in the course of the simulation. The QM part of the simulations consists of periodic relaxations following the *ab initio* gradients for nuclear motions of the larger active site that includes the metal(s), their ligand(s), the substrate(s), and may include other important amino acids near the reactive center (shared QM-DMD region) (Figure 7.1B-C). The QM/DMD boundary is thus “breathing”, and goes around the metal and its immediate coordinating residues, or around the larger active site, depending on the stage of the simulation. The QM/DMD partition is described in Figure 7.1B. The chosen active site is leaving the dangling edges of the peptide and other residues to be managed by DMD, where the extended hydrogen-bonding network can be described. The QM region is expanded for the mechanistic study. The presence of the shared QM-DMD region permits for the simple and efficient communication between the QM and DMD regions. The geometric

information can be passed between the regions, and also the DMD potential gets slightly re-parameterized on-the-fly, placing the centers of the potential wells as dictated by the QM prediction. The DMD and QM stages alternate to convergence, which is tracked by QM and DMD energies, and backbone root mean squared deviation (RMSD) (Figure 7.2).

The sampling is important for the mutagenesis part, where we mutated the C-terminal hydrophobic group (Figure 7.1B, pink circle) on the peptide and the S' subsite binding pocket. The coupling of rigorous QM treatment of the metal-containing center, including all mechanistically important substrate-metal distances, with the sampling of protein backbone while maintaining important intermolecular interactions (hydrogen bonds, hydrophobic interactions, etc.), is the strength of our QM/DMD method.

The initial structure of CPA were obtained from the Protein Data Bank (PDB) (CPA PDB code: 6CPA[45]). The PDB contains a bound inhibitor, o-(((1r)-((n-phenylmethoxycarbonyl-l-alanyl)amino)ethyl)hydroxyphosphono)-l-benzylacetic acid (ZAAP(O)F), which served as a scaffold to build the substrate, as done in other computational works.[46,31,33]

7.2.2. CHOICE OF MUTATIONS

The mutations in the protein/peptide systems were chosen with the aim of achieving sequence-specific interactions, where the protein would selectively bind and cleave the modified substrate. The specificity of CPA is defined by the interactions of the C-terminus side chain (Phe, in the case of the native CPA) with the protein pocket, and also by the resultant proper positioning of the substrate with respect to Zn and other catalytic residues. For the binding pocket, the following mutations were tried: (1) V243R,hippuryl-L-glutamate (2) V243K,hippuryl-L-glutamate, (3) V243R,hippuryl-L-aspartate, and (4) V243K,hippuryl-L-aspartate. In a purely experimental approach, Edge *et al*. utilized site-directed mutagenesis to mutate away residues in the binding pocket to reverse the polarity of substrate/protein interaction.[16] In a similar spirit,

we are mutating from bulky hydrophobic groups to charged ones is, so as to completely alter the environment of the pocket, and hypothesize these interactions to be favorable. We will then assess the catalytic performance of these variants computationally, and test predictions experimentally, thus testing the rigor of our computational methodology for the application in enzyme design. Interestingly, designing human enzymes with altered substrate specificity has the potential to be used in antibody-directed enzyme prodrug therapy (ADEPT).[47,48]

7.2.3. MECHANISTIC STUDIES OF NATIVE AND MUTANT CPA

The larger active site (Figure 7.1B) was used for the mechanistic investigation on the cluster model, after protein equilibration. All QM calculations, both in the mechanistic study and within QM/DMD, are done using *Turbomole*,[49] utilizing the Resolution of Identity[50] and multipole accelerated resolution of identity[51] to speed up the calculations. The unrestricted Density Functional Theory (DFT)[52] was used for QM, due to the size of mechanistically important region. For most calculations, the exchange and correlation is treated with the TPSS functional[53,54] with a double ζ quality basis set, def2-SVP,[55] for H, C, N, O and S and a triple ζ quality basis set, def2-TZVPP,[56] for the metal, Zn. To account for solvation and screening effects introduced by the protein, the Conductor-like Screening Model (COSMO) continuum solvent model[57] with a dielectric of 20.0. The nature of each stationary point was confirmed with a full Hessian calculation, where selected residues are frozen to mimic the constraint imposed by the protein. The frozen residues are the alpha carbons and added hydrogens, to saturate the dangling valencies, and the two nitrogens on the peptide fragment. Following a frequency calculation, transition states were identified as one negative frequency with the normal mode going along the reaction coordinate; in this case of hydrolysis the nucleophilic attack of the zinc-bound water on the carbon of the peptide bond carbonyl. Single points with BP86,[58,59,60] B3LYP[61,62] and TPSSh[63] functional and def2-TZVPP basis set were performed on all stationary points

(Appendix 9.4, Table 9.4.1). Natural population analysis (NPA)[64] calculations were done to get the partial charges on all the stationary points.

7.2.4. CHARGE DENSITY ANALYSIS OF NATIVE AND MUTANT TRANSITION STATES

The electronic charge densities of the native enzyme and V243R_FpepD mutant were calculated using the Amsterdam Density Functional Package (ADF) version 2014.01[65,66,67] using similar computational parameters as in the QM mechanistic study. A TPSS functional[54,63] and COSMO solvent model[57] with a dielectric of 20.0 were used. A double ζ quality basis set, DZP, was employed for all atoms except the metal, which was calculated using a triple ζ quality basis set, TZP.[68] The Bondalyzer add-on package in Tecplot[69] was then used to analyze the calculated charge densities.

The quantum theory of atoms in molecules (QTAIM) can be used to understand bonding interactions based on the topology of the charge density (ρ(r)).[70] There are four types of stable critical points (CPs) in a 3D scalar field. Maxima generally occur at atomic nuclei and are thus called nuclear CPs. Cage CPs occur inside cages of atoms and are local minima in ρ(r). There are also two types of stable saddle points in a 3D scalar field. The first has positive curvature in 1 direction and is deemed a bond CP. A ridge of charge density can be drawn by following gradient paths starting at bond CPs that terminate at nuclear CPs. These ridges are called bond paths since they possess the topological properties imagined for a chemical bond. Saddle points with positive curvature in 2 directions are topologically required within rings of bond paths and are therefore called ring CPs. The amount of ρ(r) at bond CPs has been found to correlate to bond strength, especially for hydrogen bonds.[71]

QTAIM can also be used to calculate atomic charges.[70] Atomic basins can be realized by following all gradient paths that terminate at specific atomic nuclei. The volume that these gradient paths reside in defines the volume of each atom. Atomic basins are bounded by zero-

flux surfaces (ZFSs) in the gradient of the charge density with edges that connect bond, ring, and cage CPs. Bader charges are calculated by integrating the number of electrons in each atomic basin and subtracting this from the number of electrons that the isolated atom contains.

Recent developments to QTAIM allow for the definition of volumes of space that bound critical points other than nuclear CPs. Bond bundles are volumes in the charge density that contain a single bond CP and are found by first determining the pair-wise connections of ring, cage, and nuclear CPs with ridges of charge density.[72,73] ZFSs can be constructed that connect these 1D ridges and form a volume of space that defines a bonding region. The shape, size, and valence electron count of bond bundles has been used to understand chemical reactivity and stability.[74,75]

7.3. EXPERIMENTAL METHODS

CPA is a protein normally expressed and isolated in bovine (boCPA). To the best of our knowledge, boCPA is easily isolated through a bacterial system; the gene that codes for boCPA does not exist in plasmid vector form. However, relatively recently a variation of CPA was isolated in a fungal entomopathogen *Metarhizium anisopliae*.[76,77] The bacterial CPA has 4 residues that differ from the bovine CPA (boCPA) in the S' binding pocket: (written the following format - boCPA/meCPA) Ala250/Val257, Ile243/Val250, Leu203/Met209 and Ser194/Asp200. The I243V is the mutation made in boCPA. Due to the experimental setup, QM/DMD simulations and mechanistic studies were repeated with the native Ile243 replaced with Val243 to have equal comparison with theory and experiment. Proteins were grown and purified by the UCLA-DOE Institute at UCLA. For purification protocol, refer to reference [78].

7.4. RESULTS AND DISCUSSION

7.4.1. STRUCTURAL COMPARISON OF CPA NATIVE AND MUTANTS

From the four mutants tested with QM/DMD, two yielded stable RMSD during the simulation, proper hydrogen-bonding network in the binding pocket, and a reactive substrate orientation. For the V243R_FpepE mutant, the reactive carbonyl of the peptide substrate loses coordination with the zinc multiple times throughout the duration of the simulations (Appendix Figure 9.4.1A). Without this key interaction, the enzyme is no longer active. V243K_FpepD mutant, the substrate in the binding pocket rotated into a non-reactive orientation and maintained that position for the duration of the simulation (Appendix Figure 9.4.2B). These structural peculiarities justified the decision to remove them from our mutant search and focus on the two remaining mutations, V243R_FpepD and V243K_FpepE.

QM/DMD simulations of both V243R_FpepD and V243K_FpepE mutants both converged with hydrogen bonds in the substrate binding pockets that agree well with experiment and previous computational studies on the native CPA (Table 7.1). In particular, the distance between the carbonyl oxygen and zinc, Zn-O, (Figure 7.1B) remained in a reactive range: 2.10 ± 0.15 Å, 3.08 ± 0.12 Å, and 3.03 ± 0.39 Å, for the native, V243R_FpepD, and V243K_FpepE variants, respectively. This can be compared to previous computational studies that uses a parameterized force field to model the Zn^{2+}, producing the Zn-O distance of 4.19 Å.[33] The hydrogen bond between the R127 and carbonyl oxygen bound to zinc, R127-O1, is a vital interaction for the catalytic activity of the enzyme.[31,33] The native CPA shows the R127-O1 bond distance to be 2.52 ± 0.46 Å (standard deviation resulted from QM/DMD sampling), comparable to the previously reported 2.64 ± 0.52 Å. The two mutant peptidases display larger deviations in R(R127-O1), and longer hydrogen bonds. The V243R_FpepD variant is closer to the native CPA, with R(R127-O1) = 3.46 ± 0.73 Å, and V243K_FpepE has R(R127-O1) of 4.58 ± 0.69 Å, which is quite high. We thus predicted V243R_FpepD to be the best candidate to show catalytic activity, and moved forward with the mechanistic study on both mutants and the native CPA.

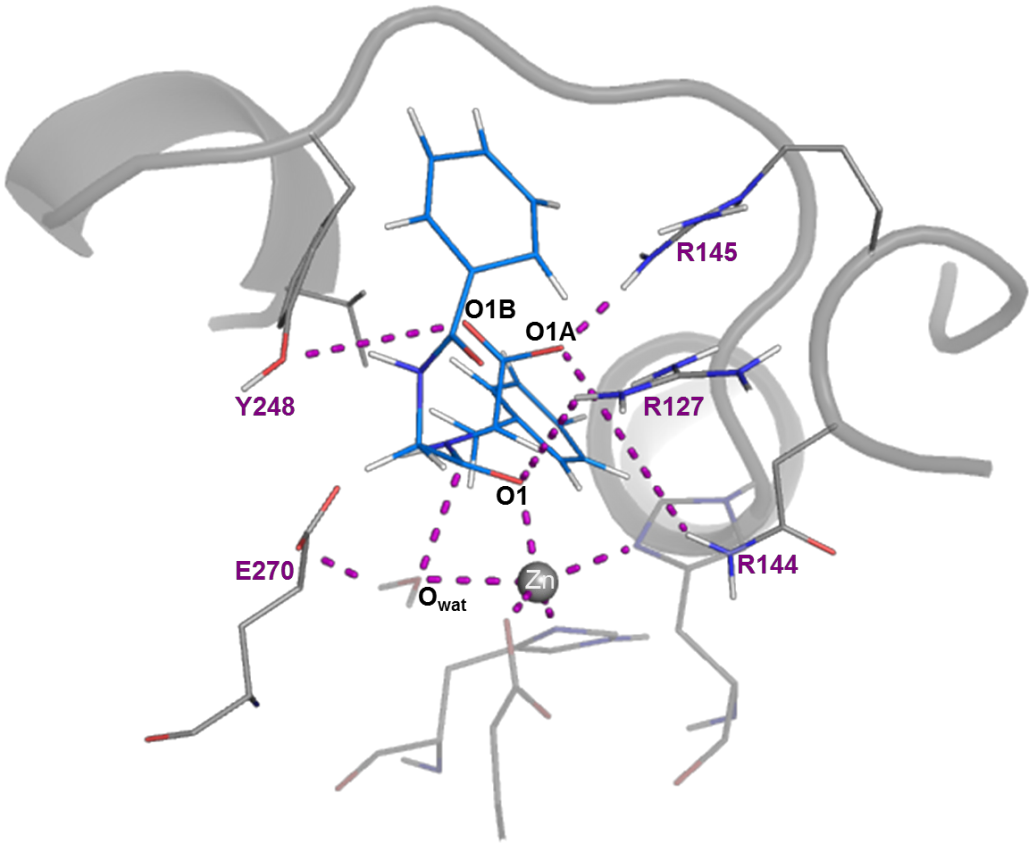


Figure 7.2. Representative binding pocket of native CPA and native substrate hippuryl-L-phenylalanine (hippuryl-L-Phe), showcasing important bond distances in binding and positioning of the substrate and evaluating potential mutants. Distances shown in Table 7.1.

Bond between atoms	FEF (Å)	V243R_FpepD (Å)	V243K_FpepE (Å)	Ref (QM/MM ES)
R127 (1HH1) --- O1	2.52 ± 0.46	3.46 ± 0.73	4.58 ± 0.69	2.64 ± 0.52
N144 (HND) --- O1A	6.57 ± 1.29	5.45 ± 0.98	5.84 ± 0.46	1.85 ± 0.16
R145(1HH1) --- O1A	2.09 ± 0.34	3.03 ± 1.40	2.80 ± 0.50	1.77 ± 0.18
R145(1HH1) --- O1B	3.56 ± 0.69	4.35 ± 0.93	4.13 ± 0.69	1.77 ± 0.18
Y248 (HO) --- O1B	4.62 ± 0.97	2.59 ± 0.76	2.05 ± 0.20	1.67 ± 0.11
Zn --- O1	**2.10 ± 0.15**	**3.08 ± 0.12**	**3.03 ± 0.39**	**4.19 ± 0.41**
Zn --- O_{WAT}	2.35 ± 0.33	2.20 ± 0.00	2.20 ± 0.00	2.13 ± 0.08
O_{WAT} --- C1	2.45 ± 0.35	3.00 ± 0.03	2.69 ± 0.19	2.83 ± 0.22
E270 (OE1) --- O_{WAT} (H)	3.16 ± 0.75	2.13 ± 0.38	2.55 ± 0.60	1.77 ± 0.14
R127 (NH1) --- O1	3.38 ± 0.36	4.01 ± 0.64	4.68 ± 0.51	n/a

Table 7.1. Hydrogen-bonding interactions in the substrate binding pocket of native and mutant CPA, compared to previous QM/MM simulation (Ref [33]).

From the structural data obtained with QM/DMD, the specific interactions in the binding pocket can be monitored. CPA naturally has a slight preference for bulky hydrophobic groups. For the native enzyme, we track the bond distances of the C-terminal phenylalanine (CP3-6) and the pocket valine, V243 (CG1, CG2, C1). (Figure 7.3, Table 7.2) Ranging from 4.97 ± 0.94 Å to 7.39 ± 1.22 Å, QM/DMD structures show the hydrophobic interactions are maintained throughout the simulation, with large standard deviations that could be a result of the non-specific nature of the binding pocket, accommodating a wide variety of hydrophobic residues (Table 7.2).

For the V243K_FpepE, the distance between the N of the NH_3 group of the pocket lysine, K243, and the O of the C-terminal glutamate side chain is 5.29 ± 0.64 Å. This distance being so large (not within a meaningful range for a strong H-bond), combined with the large R127-O1 also suggest that this variant is not viable.

For the V243R_FpepD mutant, the shortest hydrogen bond between the peptide and R243 (H of arginine and OD2 of peptide) is 3.61 ± 0.99 Å. All the hydrogen bonds between the peptide and pocket residue of the protein, specifically the two donors from the OD1 and OD2 of the peptide and the acceptors from the NH_2 groups from R243, fluctuate between ~3.6 to 5.8 Å. (Figure 7.3, Table 7.2) This range of distances further validates this mutant as a potentially active enzyme, subject to experimental testing.

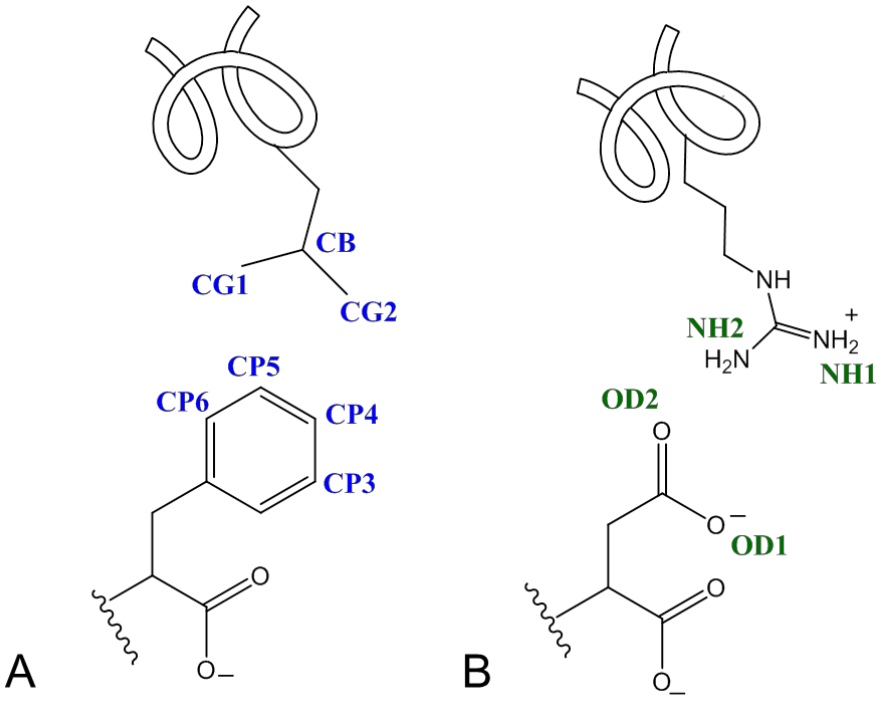


Figure 7.3. Cartoon representations of truncated bonding pockets of the native CPA (A) and V243R_FpepD mutant (B).

Native binding pocket	Bond distances (Å)	Mutant binding pocket	Bond distances (Å)
CG1 V243 -- CP3 PEP	7.09 ± 1.15	NH1 R243 -- OD1 PEP	5.17 ± 0.98
CG1 V243 -- CP4 PEP	6.18 ± 1.06	NH1 R243 -- OD2 PEP	5.07 ± 0.73
CG1 V243 -- CP5 PEP	4.97 ± 0.94	NH2 R243 -- OD1 PEP	4.34 ± 0.66
CG2 V243 -- CP3 PEP	7.39 ± 1.22	NH2 R243 -- OD2 PEP	3.73 ± 0.79
CG2 V243 -- CP4 PEP	6.38 ± 1.09	1HH1_ARG -- OD1_PEP	4.93 ± 1.10
CG2 V243 -- CP5 PEP	5.15 ± 0.94	2HH1_ARG -- OD1_PEP	5.85 ± 1.06
CG1 V243 -- CP3 PEP	7.09 ± 1.15	1HH2_ARG -- OD1_PEP	4.16 ± 0.75
CG1 V243 -- CP4 PEP	6.18 ± 1.06	2HH2_ARG -- OD1_PEP	4.52 ± 0.75
		1HH1_ARG -- OD2_PEP	4.99 ± 0.80
		2HH1_ARG -- OD2_PEP	5.82 ± 0.76
		1HH2_ARG -- OD2_PEP	3.72 ± 0.93
		2HH2_ARG -- OD2_PEP	3.61 ± 0.99

Table 7.2: Important bond distances of residues in the binding pocket of native CPA and V243R_FpepD mutant. See Figure 7.3 for atom label names.

7.4.2. MECHANISTIC STUDY OF CPA NATIVE, V243R_FPEPD AND V243K_FPEPE MUTANTS

Following structural comparison of the native and mutant enzymes, we proceeded with a full QM mechanistic study of an active site model. Previous works emphasize that the exact mechanism of CPA is not fully understood, with a current debate on whether the peptide hydrolysis is accomplished through a promoted water or anhydride mechanism. However, computations[34] have shown the water-promoted to yield the lowest barrier for the rate-determining step, nucleophilic attack of peptide backbone carbonyl carbon by water (Figure 7.1A). For mechanistic studies, the QM active site was extended to include the entire peptide, including the R127 residue in the binding pocket responsible for sequence specificity, and other residues known to stabilize the transition state (Figure 7.1B,C). Also, only the rate-determining step is considered here due to the complication of the hydrogen source (zinc bound water or solvent) for the second step of the reaction, protonation of the backbone amide. For the native and two mutant systems, a single structure from the QM/DMD trajectory was chosen based off of the lowest QM and DMD energies (Figure 7.4).

For the native enzyme, the Zn-O1 distance begins at 2.109 Å and continues to shorten to 2.035 Å and 2.012 Å, as the mechanism proceeds through the nucleophilic attack on the carbonyl C1 by the Zn-bound water, from the reactants to TS1 and to the tetrahedral intermediate (EI) (Table 7.1). The O_{wat}-C1 distance begins at 2.636 Å, shortens to 1.853 Å at TS1, and ends at 1.500 Å at EI, indicating a slightly late transition state. The barrier for the transition state is calculated to be 13.8 kcal/mol, in agreement with the experimental value, calculated using transition-state theory, of 15.7 kcal/mol (Figure 7.4).[79] The zinc-bound water detaches from the zinc as the tetrahedral intermediate forms and the C1-O_{wat} bond decreases to 1.500 Å, forming a new bond. The O1- H_{R127} bond shortens from 1.661 Å to 1.590 Å as the hydrogen bond increases in strength due to the buildup of charge on that carbonyl oxygen,

going from -0.83, -0.91 then -0.99. (Full table of charges in Appendix 9.4, Table 9.4.2) Our geometric parameters agree well with a previous study on a similar system, which was using a parameterized QM/MM.[34]

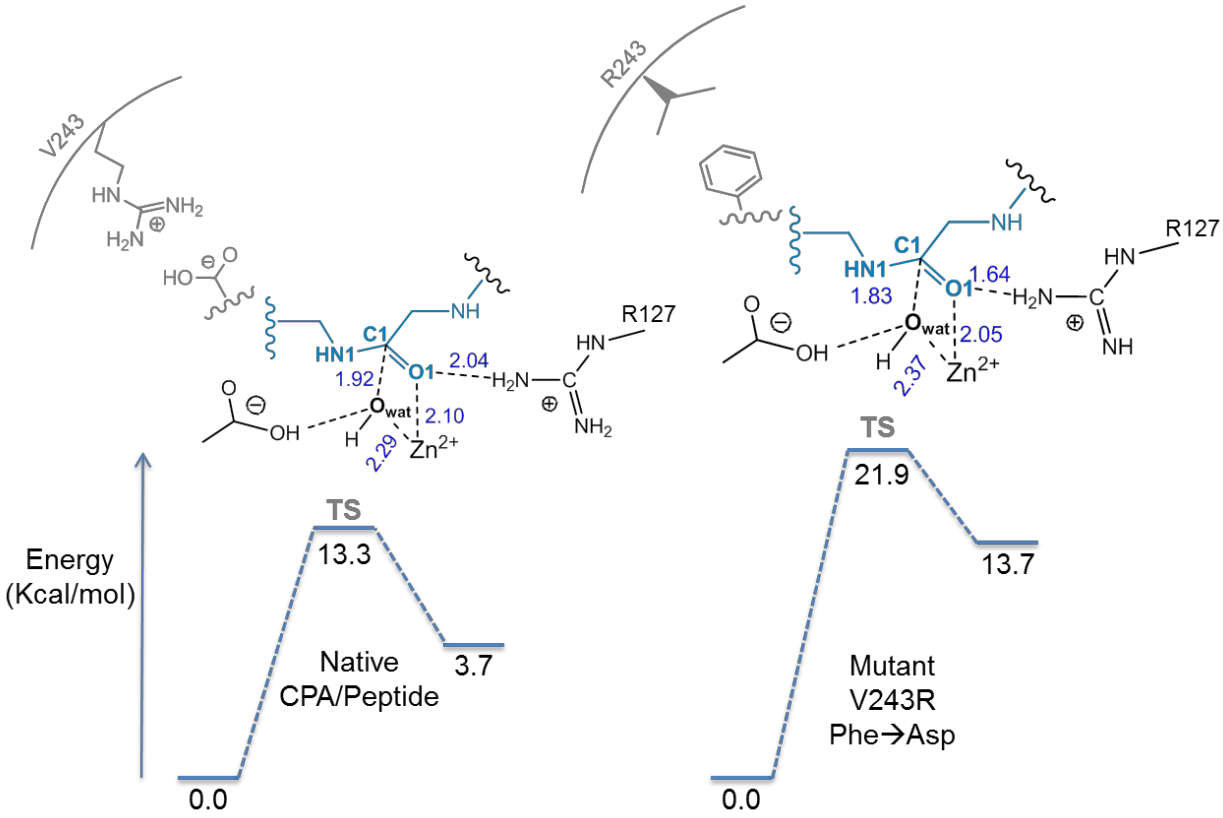


Figure 7.4. Mechanism of native peptidase/peptide and mutant peptidase/peptide systems. Single point energies reported are calculated with TPSSh/def2-TZVPP for all atoms from TPSS/def2-SVP(C,O,N,H) and def2-TZVPP (Zn^{2+}) optimized structures.

For the I243R_FpepD mutant, similar trends in varying bond distances are present throughout the reaction mechanism. The barrier for this transition is 21.6 kcal/mol, higher than the native, but still in good known barriers for systems that catalyze a similar hydrolysis reactions. The O1- H_{R127} bond is slightly elongated in comparison to the native CPA, starting at 2.247 Å and shortening to 1.904 Å in the intermediate. This could be due to the restructuring of

the binding pocket in response to the V243R mutation, which has a slight perturbation on the hydrogen-bonding network. The native and I243R_FpepD mutant were carried through to experimental validation, using the native enzyme as a control.

For the V243K_FpepE mutant, the mechanistic study yielded a reaction barrier of 28.7 kcal/mol, for the rate-determining step. This barrier is too high, and therefore, we eliminated V243K_FpepE from being tested experimentally.

	Native (Mutant - V243R_FpepD)			Ref		
	ES	**TS1**	**EI**	**ES**	**TS1**	**EI**
Zn---O1	2.11 (2.64)	2.04 (2.10)	2.01 (2.02)	5.51	2.06	1.94
Zn---O2	2.13 (2.09)	2.40 (2.29)	3.18 (2.95)	1.93	2.19	3.30
C1---O1	1.29 (1.27)	1.31 (1.31)	1.36 (1.37)	1.24	1.30	1.37
O_{wat}---C1	2.64 (2.68)	1.85 (1.92)	1.50 (1.52)	5.38	1.88	1.37
N1---C1	1.66 (1.33)	1.59 (1.39)	1.75 (1.44)	1.34	1.37	1.44
O1---H_{R127}	1.66 (2.25)	1.59 (2.04)	1.75 (1.90)	1.83	1.89	1.67
O1---N_{R127}	2.70 (2.85)	2.65 (3.03)	2.74 (2.03)	n/a	n/a	n/a

Table 7.3. Geometric parameters of stationary points of the water-mediated hydrolysis of the native and V243R_FpepD mutant (in parenthesis) obtained at the TPSS/def2-SVP (H,C,N,O) and def2-TZVPP (Zn^{2+}).

7.4.3. COMPARISON OF NATIVE ENZYME AND V243R_FPEPD REACTIVITY THROUGH QTAIM

Two key components to the catalytic activity of CPA are the correct positioning of the water molecule for nucleophilic attack and the stabilization of the partially charged substrate carbonyl. In both the native CPA and V243R_FpepD the water molecule is positioned near the substrate carbonyl due to its interactions with the Zn ion and hydrogen bonding to Glu270. The

native enzyme has additional hydrogen bonding from the second water hydrogen to Glu72. The donation of the second hydrogen atom in the native enzyme results in a slightly lower Bader charge on the water oxygen (see Table 7.4). This should make the native water molecule a better nucleophile as is indicated by the lower activation energy in the native enzyme as compared to V243R_FpepD.

	ES		**TS1**		**TI**	
	Native	Mutant	Native	Mutant	Native	Mutant
O_{wat}	-1.181	-1.135	-1.014	-1.038	-1.003	-1.003
C1	1.12	1.162	1.1	1.08	1.117	1.073
O1	-1.087	-1.103	-1.102	-1.102	-1.089	-1.085
Zn	1.314	1.308	1.297	1.304	1.25	1.262

Table 7.4. Bader charges of atoms involved in the chemical reaction of hydrolysis. See Figure 7.4 for labeling.

The partially negative carbonyl oxygen on the substrate is stabilized through two interactions: hydrogen bonding to R127 and coordination to the Zn center. The stabilizing interaction from the hydrogen bond between the R127 and O1 in the reactants is decreased in V243R_FpepD as compared to the native enzyme, but still present. This is indicated by the charge density at the H_R127-O1 bond CP, which has decreased from 0.049 in the native enzyme to 0.013 e/bohr^3 in the mutant (Table 7.5). Additionally, the O1-H_R127-N_R127 angle has decreased from being almost linear, 175°, to 147.2° in V243R_FpepD.

	ES		TS1		TI	
	ρ(r) at bond CPs (Bond length)		ρ(r) at bond CPs (Bond length)		ρ(r) at bond CPs (Bond length)	
	Native	**Mutant**	**Native**	**Mutant**	**Native**	**Mutant**
C1---O_{wat}	0.0166 (2.636)	0.0155 (2.675)	0.0953 (1.853)	0.0817 (1.919)	0.217 (1.500)	0.204 (1.523)
C1---O1	0.352 (1.285)	0.368 (1.268)	0.338 (1.312)	0.342 (1.306)	0.301 (1.364)	0.298 (1.369)
Zn---O1	0.0575 (2.109)	0.0183 (2.640)	0.0699 (2.035)	0.0599 (2.104)	0.0762 (2.011)	0.0746 (2.021)
Zn---O2	0.0558 (2.134)	0.0596 (2.088)	0.0327 (2.400)	0.0395 (2.294)	n/a (3.175)	n/a (2.946)
H_{R127}---O1	0.0487 (1.661)	0.0128 (2.247)	0.0596 (1.590)	0.0205 (2.038)	0.0412 (1.749)	0.0287 (1.905)
Ring CP	0.0166	0.0129	0.0262	0.0269	n/a	n/a
H2-O1	n/a	n/a	n/a	n/a	0.0199 (2.090)	n/a

Table 7.5. Charge density (ρ(r)) at the bond critical points (bond CPs) and bond lengths (in parenthesis below) of the bonding interactions of interest for the chemical reaction.

Similarly, the native enzyme has greater stabilization of the carbonyl oxygen from the Zn ion as compared to V243R_FpepD. This can be seen in the amount of charge density at the Zn-O1 bond CP and also in the size of the Zn-O1 bond bundle. The charge density drops from 0.0575 in the native enzyme to 0.0183 e/bohr^3 at the bond CP in V243R_FpepD. A cut plane of the bond bundle in the plane of the Zn-O1-ring CP is shown in Figure 7.5. These points define three vertices of the bond bundle that represents the volume of space the Zn-O1 bond occupies The pictured portion of the bond bundle in the native enzyme is larger than in V243R_FPepD which generally indicates a higher number of electrons in this region and therefore a more stabilizing bonding interaction.

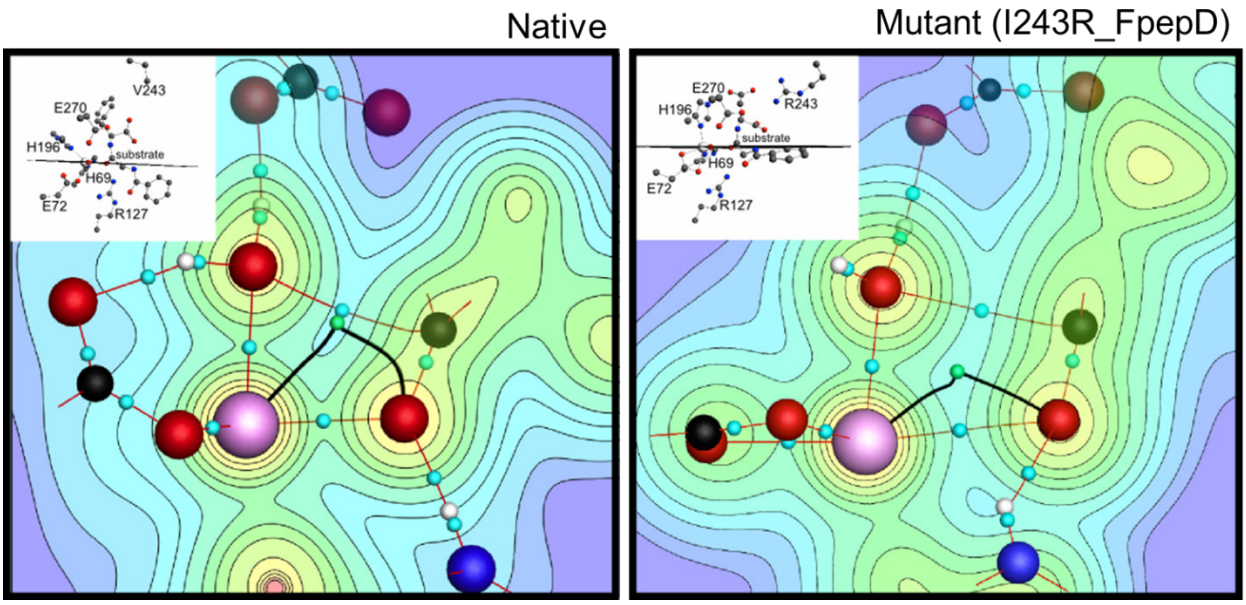


Figure 7.5 The upper left inset of each image shows the orientation of the cutplane that intersects O1, the ring CP, and the Zn ion in the native (left) and mutant (right) CPA. Contours in ρ(r) are drawn on the cut planes. Red lines indicate bond paths, thick black lines show approximate edges of the Zn-O1 bond bundle. The following coloring scheme is used: Zn-purple, O-red, C-black, H-white, N-blue, bond CP-cyan, ring CP-green.

In both the native CPA and V243_FpepD, the Zn ion and hydrogen bonding network are able to stabilize the partial charges on the substrate carbonyl. Throughout this step of the reaction the charge on the oxygen varies by only 0.178 electrons and 0.132 electrons for the native and mutated enzyme, respectively. Overall, the hydrogen bonding networks in both the native and mutated enzyme allow for the reaction to take place. The higher activation energy in V243_FpepD may be explained by slightly less stabilizing interactions from both hydrogen bonding and the coordination of the substrate carbonyl to the Zn ion.

In the rate-determining step of the reaction the Zn-O2 bonding interaction is broken, opening the tetrahedral ring that is pictured in Figure 7.5. Topologically, this occurs by the ring CP in the tetrahedral intermediate physically moving towards the Zn-O2 bond CP until it ultimately annihilates the bond CP. While the ring CP in V243R_FpepD is geometrically closer to the Zn-O2 bond CP, the charge density region between the ring and bond CP in the native

enzyme is “flatter”. The change in ρ(r) between the ring-bond CPs is only 0.0392 e/bohr^3 in the native enzyme while it is 0.0467 e/bohr^3 in V243R_FpepD. It is therefore less energy intensive for the native enzyme ring CP to annihilate the bond CP than in the mutant enzyme.

7.5. EXPERIMENTAL VALIDATION

Please see Appendix 9.4, specifically Chart 9.4.1. and 9.4.2.

7.6. CONCLUSION

Multi-scale computational tools are vital to the success of designing bio-catalytic materials such as artificial metalloenzymes. Our efforts begin with parsing experimentally determined X-ray structures of well-folded proteins into our fast and efficient hybrid dynamics software, QM/DMD, that employs a quantum mechanical description of the metal-containing active site coupled to sufficient backbone sampling reaching the nanosecond timescale. QM/DMD accurately models small-scale electronic properties of the metal and large-scale protein motions induced by substrate binding and/or mutations to the protein sequence. Here we employ QM/DMD to redesign carboxypeptidase A (CPA) to selectively hydrolyze a specific peptide sequence. QM/DMD has previously been used to capture fine structural and electronic properties of metal-containing proteins used in cluster model QM mechanistic studies. However, this is the first case where QM/DMD is used to completely redesign an enzyme. CPA is a Zn^{2+}-dependent exopeptidase that hydrolyzes C-terminal peptides with an aromatic or aliphatic side chain. The substrate docks into CPA via an extended hydrogen bond network, that is residue specific, and a hydrophobic pocket, where the interaction between the peptide substrate and protein is not residue specific. Through QM/DMD, we performed a series of *in silico* mutations to the hydrophobic pocket of CPA and the C-terminal residue of the peptide, to design a mutant protein/peptide system that would be catalytic and not affect the overall stability of the protein to

the point of denaturing. Following extensive QM/DMD sampling and cluster model Density Functional Theory (DFT) mechanistic studies, investigating specifically the water-promoted pathway for hydrolysis, of the native and various CPA mutants, one mutation, V243R_FpepD was selected and chosen for further analysis. For the rate-determining step of the hydrolysis, the native and V243R_FpepD mutant yield a barrier of 13.3 kcal/mol and 21.6 kcal/mol, respectively, in good agreement with the experimental value of 15.7 kcal/mol. Additional analysis through an extension of the quantum theory of atoms in molecules (QTAIM) sheds light on the differences between the catalytic barriers in native and mutant CPA, particularly by investigating the electronic charge densities of the two enzyme/substrate systems.

7.7. REFERENCES

1 Thayer, A. M., Biocatalysts. *Chemical & Engineering News* 2012, pp 13-18.

2 Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440* (7086), 940-943.

3 Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010,** *329* (5989), 309-313.

4 Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008,** *319* (5868), 1387-1391.

5 Lerner, R.; Benkovic, S.; Schultz, P., At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **1991,** *252* (5006), 659-667.

6 Hilvert, D., CRITICAL ANALYSIS OF ANTIBODY CATALYSIS. *Annual Review of Biochemistry* **2000,** *69* (1), 751-793.

7 Vasileiou, C.; Vaezeslami, S.; Crist, R. M.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Protein Design: Reengineering Cellular Retinoic Acid Binding Protein II into a Rhodopsin Protein Mimic. *Journal of the American Chemical Society* **2007,** *129* (19), 6140-6148.

8 Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D.

S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-195.

9 Bolon, D. N.; Mayo, S. L., Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences* **2001,** *98* (25), 14274-14279.

10 Dahiyat, B. I.; Mayo, S. L., Protein design automation. *Protein Science* **1996,** *5* (5), 895-903.

11 Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D., New algorithms and an in silico benchmark for computational enzyme design. *Protein Science* **2006,** *15* (12), 2785-2794.

12 Alexandrova, A. N.; Jorgensen, W. L., Origin of the Activity Drop with the E50D Variant of Catalytic Antibody 34E4 for Kemp Elimination. *The Journal of Physical Chemistry B* **2008,** *113* (2), 497-504.

13 Gordon, S. R.; Stanley, E. J.; Wolf, S.; Toland, A.; Wu, S. J.; Hadidi, D.; Mills, J. H.; Baker, D.; Pultz, I. S.; Siegel, J. B., Computational Design of an α-Gliadin Peptidase. *Journal of the American Chemical Society* **2012,** *134* (50), 20513-20520.

14 Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D., Protein Structure Prediction Using Rosetta. In *Methods in Enzymology*, Ludwig, B.; Michael, L. J., Eds. Academic Press: 2004; Vol. Volume 383, pp 66-93.

15 Das, R.; Baker, D., Macromolecular Modeling with Rosetta. *Annual Review of Biochemistry* **2008,** *77* (1), 363-382.

16 Edge, M.; Forder, C.; Hennam, J.; Lee, I.; Tonge, D.; Hardern, I.; Fitton, J.; Eckersley, K.; East, S.; Shufflebotham, A.; Blakey, D.; Slater, A., Engineered human carboxypeptidase B enzymes that hydrolyse hippuryl-L-glutamic acid: reversed-polarity mutants. *Protein Engineering* **1998,** *11* (12), 1229-1234.

17 Kaplan, J.; DeGrado, W. F., De novo design of catalytic proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (32), 11566-11570.

18 Maglio, O.; Nastri, F.; Martin de Rosales, R. T.; Faiella, M.; Pavone, V.; DeGrado, W. F.; Lombardi, A., Diiron-containing metalloproteins: Developing functional models. *Comptes Rendus Chimie* **2007,** *10* (8), 703-720.

19 Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A., An artificial di-iron oxo-protein with phenol oxidase activity. *Nature Chemical Biology* **2009,** *5* (12), 882-884.

20 Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T. G.; Solomon, E. I.; DeGrado, W. F., Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat Chem* **2012,** *4* (11), 900-906.

21 Pétra, P. H., Bovine procarboxypeptidase and carboxypeptidase A. In *Methods in Enzymology*, Gertrude E. Perlmann, L. L., Ed. Academic Press: 1970; Vol. Volume 19, pp 460-503.

22 (a) Reeck, G. R.; Walsh, K. A.; Neurath, H., Isolation and characterization of carboxypeptidases A and B from activated pancreatic juice. *Biochemistry* **1971,** *10* (25), 4690-4698; (b) Everitt, M. T.; Neurath, H., Rat peritoneal mast cell carboxypeptidase: localization, purification, and enzymatic properties. *FEBS Letters* **1980,** *110* (2), 292-296.

23 Bodwell, J. E.; Meyer, W. L., Purification and characterization of carboxypeptidase A from rat skeletal muscle. *Biochemistry* **1981,** *20* (10), 2767-2777.

24 Vallee, B. L., Galdes, A., Auld, D. S., Riordan, J. F., In *In Zinc Enzymes*, Spiro, T. G., Ed: Wiley: New York, 1983.

25 Lipscomb, W. N., Carboxypeptidase A mechanisms. *Proceedings of the National Academy of Sciences* **1980,** *77* (7), 3875-3878.

26 Christianson, D. W.; Lipscomb, W. N., Carboxypeptidase A. *Accounts of Chemical Research* **1989,** *22* (2), 62-69.

27 Natesh, R.; Schwager, S. L. U.; Sturrock, E. D.; Acharya, K. R., Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* **2003,** *421* (6922), 551-554.

28 Vallee, B. L.; Auld, D. S., Functional zinc-binding motifs in enzymes and DNA-binding proteins. *Faraday Discussions* **1992,** *93* (0), 47-65.

29 Vallee, B. L.; Auld, D. S., Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **1990,** *29* (24), 5647-5659.

30 Kilshtain-Vardi, A.; Glick, M.; Greenblatt, H. M.; Goldblum, A.; Shoham, G., Refined structure of bovine carboxypeptidase A at 1.25 A resolution. *Acta Crystallographica Section D* **2003,** *59* (2), 323-333.

31 Szeto, M. W. Y.; Mujika, J. I.; Zurek, J.; Mulholland, A. J.; Harvey, J. N., QM/MM study on the mechanism of peptide hydrolysis by carboxypeptidase A. *Journal of Molecular Structure: THEOCHEM* **2009,** *898* (1–3), 106-114.

32 Suh, J.; Park, T. H.; Hwang, B. K., Comparable rates for cleavage of amide and ester bonds through nucleophilic attack by carboxylate anion and general acid catalysis by metal-bound water in a carboxypeptidase A model. *Journal of the American Chemical Society* **1992,** *114* (13), 5141-5146.

33 Xu, D.; Guo, H., Quantum Mechanical/Molecular Mechanical and Density Functional Theory Studies of a Prototypical Zinc Peptidase (Carboxypeptidase A) Suggest a General Acid−General Base Mechanism. *Journal of the American Chemical Society* **2009,** *131* (28), 9780-9788.

34 Alex, A.; Clark, T., MO-Studies of enzyme reaction mechanisms. I. Model molecular orbital study of the cleavage of peptides by carboxypeptidase A. *Journal of Computational Chemistry* **1992,** *13* (6), 704-717.

35 Alvarez-Santos, S.; González-Lafont, A.; Lluch, J. M.; Oliva, B.; Avilés, F. X., On the water-promoted mechanism of peptide cleavage by carboxypeptidase A. A theoretical study. *Canadian Journal of Chemistry* **1994,** *72* (10), 2077-2083.

36 (a) Alvarez-Santos, S.; Gonzalez-Lafont, A.; M. Lluch, J.; Oliva, B.; X. Aviles, F., Theoretical study of the role of arginine 127 in the water-promoted mechanism of peptide cleavage by carboxypeptidase A. *New Journal of Chemistry* **1998,** *22* (4), 319-326; (b) Vardi-Kilshtain, A.; Shoham, G. I. L.; Goldblum, A., Anhydride formation is not a valid mechanism for peptide cleavage by carboxypeptidase-A: a semiempirical reaction pathway study. *Molecular Physics* **2003,** *101* (17), 2715-2724.

37 Sparta, M.; Shirvanyants, D.; Ding, F.; Dokholyan, Nikolay V.; Alexandrova, Anastassia N., Hybrid Dynamics Simulation Engine for Metalloproteins. *Biophysical Journal* **2012,** *103* (4), 767-776.

38 Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I., Discrete molecular dynamics studies of the folding of a protein-like model. *Folding and Design* **1998,** *3* (6), 577-587.

39 Dokholyan, N. V., Studies of folding and misfolding using simplified models. *Current Opinion in Structural Biology* **2006,** *16* (1), 79-85.

40 Ding, F.; Guo, W.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J.-E., Reconstruction of the src-SH3 Protein Domain Transition State Ensemble using Multiscale Molecular Dynamics Simulations. *Journal of Molecular Biology* **2005,** *350* (5), 1035-1050.

41 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V., Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* **2008,** *16* (7), 1010-1018.

42 Sparta, M.; Alexandrova, A. N., How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. *PLoS ONE* **2012,** *7* (10), e47172.

43 Valdez, C. E.; Alexandrova, A. N., Why Urease Is a Di-Nickel Enzyme whereas the CcrA β-Lactamase Is a Di-Zinc Enzyme. *The Journal of Physical Chemistry B* **2012,** *116* (35), 10649-10656.

44 Sparta, M.; Valdez, C. E.; Alexandrova, A. N., Metal-Dependent Activity of Fe and Ni Acireductone Dioxygenases: How Two Electrons Reroute the Catalytic Pathway. *Journal of Molecular Biology* **2013,** *425* (16), 3007-3018.

45 Kim, H.; Lipscomb, W. N., Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. *Biochemistry* **1990,** *29* (23), 5546-5555.

46 Wu, S.; Zhang, C.; Xu, D.; Guo, H., Catalysis of Carboxypeptidase A: Promoted-Water versus Nucleophilic Pathways. *The Journal of Physical Chemistry B* **2010,** *114* (28), 9259-9267.

47 Laethem, R. M. B., T. A.; Cory, M.; Elwell, L.; Moxham, C. P.; Ray, P. H.; Walton, L. M.; Smith, G. K., *Arch. Biochem. Biophys.* **1996,** *332*, 8-12.

48 Smith, G. K. e. a., *Trends Pharmacol. Sci.* **1997,** *9*, 299-304.

49 2011, T. V., a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH,. 1989–2007, Turbomole GmbH, since 2007**,** *available from http://www.turbomole.com*.

50 Arnim, M. V. A., R., Performance of parallel TURBOMOLE for density functional calculations. *Journal of Computational Chemistry* **1998,** *19*, 1746-1757.

51 Sierka, M.; Hogekamp, A.; Ahlrichs, R., Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *The Journal of Chemical Physics* **2003,** *118* (20), 9136-9148.

52 Dirac, P. A. M., *Quantum Mechanics of Many-Electron Systems*. 1929; Vol. 123, p 714-733.

53 Perdew, J. P.; Wang, Y., Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **1992,** *45* (23), 13244-13249.

54 Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E., Climbing the Density Functional Ladder: Nonempirical Meta\char21{}Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters* **2003,** *91* (14), 146401.

55 Schäfer, A.; Horn, H.; Ahlrichs, R., Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics* **1992,** *97* (4), 2571-2577.

56 Weigend, F. A., R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005,** *7*, 3297-3305.

57 Klamt, A.; Schuurmann, G., COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2* **1993,** (5), 799-805.

58 (a) Slater, J. C., A Simplification of the Hartree-Fock Method. *Physical Review* **1951,** *81* (3), 385-390; (b) Vosko, S. H.; Wilk, L.; Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **1980,** *58* (8), 1200-1211.

59 Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988,** *38* (6), 3098-3100.

60 Perdew, J. P., Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B* **1986,** *33* (12), 8822-8824.

61 Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.

62 Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010,** *132* (15), 154104.

63 Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P., Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *The Journal of Chemical Physics* **2003,** *119* (23), 12129-12137.

64 Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *The Journal of Chemical Physics* **1985,** *83* (2), 735-746.

65 te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T., Chemistry with ADF. *Journal of Computational Chemistry* **2001,** *22* (9), 931-967.

66 Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J., Towards an order-N DFT method. *Theor Chem Acc* **1998,** *99* (6), 391-403.

67 ADF2014. *SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com*

68 Van Lenthe, E.; Baerends, E. J., Optimized Slater-type basis sets for the elements 1–118. *Journal of Computational Chemistry* **2003,** *24* (9), 1142-1156.

69 Tecplot. *http://www.tecplot.com/ (2014)*.

70 Bader, R. F. W., Atoms in Molecules: A Quantum Theory. *Oxford University Press, Oxford* **1990**.

71 Boyd, R. J.; Choi, S. C., Hydrogen bonding between nitriles and hydrogen halides and the topological properties of molecular charge distributions. *Chemical physics letters* **1986,** *129* (1), 62-65.

72 Eberhart, M. E., A quantum description of the chemical bond. *Phil. Mag. B.* **2001,** *81*, 721-729.

73 Jones, T. E.; Eberhart, M. E., The bond bundle in open systems. *International Journal of Quantum Chemistry* **2010,** *110* (8), 1500-1505.

74 Jones, T. E.; Eberhart, M. E.; Imlay, S.; Mackey, C., Bond Bundles and the Origins of Functionality. *The Journal of Physical Chemistry A* **2011,** *115* (45), 12582-12585.

75 Jones, T. E., Nucleophilic Substitution: A Charge Density Perspective. *The Journal of Physical Chemistry A* **2012,** *116* (16), 4233-4237.

76 Joshi, L.; Leger, R. J. S., Cloning, Expression, and Substrate Specificity of MeCPA, a Zinc Carboxypeptidase That Is Secreted into Infected Tissues by the Fungal Entomopathogen Metarhizium anisopliae. *Journal of Biological Chemistry* **1999,** *274* (14), 9803-9811.

77 Stleger, R. J.; Bidochka, M. J.; Roberts, D. W., Characterization of a Novel Carboxypeptidase Produced by the Entomopathogenic Fungus Metarhizium anisopliae. *Archives of Biochemistry and Biophysics* **1994,** *314* (2), 392-398.

78 Austin, B. P.; Waugh, D. S., Isolation of Metarhizium anisopliae carboxypeptidase A with native disulfide bonds from the cytosol of Escherichia coli BL21(DE3). *Protein Expression and Purification* **2012,** *82* (1), 116-124.

79 Gardell, S. J.; Craik, C. S.; Hilvert, D.; Urdea, M. S.; Rutter, W. J., Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. *Nature* **1985,** *317* (6037), 551-555.

CHAPTER 8

Eris-QM/DMD: a Metalloprotein Design Tool

8.1. INTRODUCTION

The overall structure and stability of an enzyme is critical for its catalytic function. The protein must have the correct scaffold to accommodate the catalytic reaction, while maintaining the structural integrity in its environment, whether *in vivo* or *in vitro*. Thus, the task of enzyme design, done by nature or by humans, involves optimizing both activity and specificity, and stability, typically achieved through mutagenesis. The effect of even a single mutation at the active site on catalysis can be dramatic; for example, the substitution of a cysteine with serine in cytochrome P450 investigated by Wang *et al*. alters the protein enough to accommodate a wider variety of substrates for the cyclopropanation reaction *in vivo*.[1] The effects of isolated mutations on the protein stability varies, depending on the context in which the mutated amino acid is located.[2]

When approaching protein design computationally, the two tasks can be largely separated.[3,4] The desired reaction mechanism and rate can be pre-calculated on a small model, constrained, and hopefully mostly retained in the designed protein (the “theo-zyme” approach). The remaining task in the design has to do with optimizing protein stability, given the organization of the active site. Often times, metalloprotein function and stability are antagonistically related,[5] therefore one must carefully balance both the effects of catalysis and structure during the design phase. Additional refinements may include taking care of further pre-organization of the protein for stabilizing the transition state of the reaction through long-range electrostatics.[6] Assessing the effect of mutagenesis on enzyme stability in terms of its free energy is thus a key component, and it is very difficult. Intuitively evaluating whether one or two seemingly innocent mutations could cause the entire protein to denature is impossible. Directly calculating free energy, by summing up all Boltzmann weights of all the energy states in a system, is impractical due to the number of degrees of freedom and the required amount of sampling. Additionally, when it comes to metalloenzymes, the evaluation of the energy of each state requires quantum mechanical treatment of the metal-containing area, calling for hybrid

quantum-classical (QM/MM) methods, which are computationally costly.

However, the free energy difference between related systems is more amendable to evaluation, at least when done with a classical Hamiltonian. Classical molecular dynamics simulations do provide a tool for evaluating this change, ΔΔG.[7] Nivón *et al*. have developed a classical Monte Carlo sampling algorithm for automating the protein design process, but it does not allow for efficient backbone sampling.[8] The need for a flexible backbone description has been highlighted as necessary in the design process.[9,10] Dokholyan *et al.* have developed *Eris*, a protein stability prediction software that can move the protein backbone; it utilizes their slightly coarse-grained physical force field Medusa.[11] *Eris* evaluates the change in protein stability upon introduction of mutation(s) by making a thermodynamic cycle going through the unfolded states of the native protein and the mutant.[2,12] The computational speed without a loss of accuracy makes *Eris* a useful tool for gauging mutagenesis *in silico* prior to experimental efforts. *Rosetta*[13,14] is another popular design tool that, using a classical force field, samples through the rotameric and chemical spaces of the amino acids in the designed region of the protein. *Rosetta* has been successful in many cases of enzyme design.[15,16,17]

Although a great advancement in the field is thus obvious, when it comes to the design of metalloenzymes, there is a problem, especially when the metal has an incomplete population of the *d*-atomic orbitals (AOs) and exhibits such complications as mixed spin states, which intimately impact the active site geometry and the outcome of the catalytic process. In these cases, a quality QM description of the metal is required. Only some metals are parameterized in classical force fields. For example, the Medusa11 force field used in *Eris* can describe only copper (Cu^{2+}) and zinc (Zn^{2+}), which pose hardly any complications for computation, as compared to such metals as Fe^{2+} and Mn^{2+}. Thus, assessing ΔΔG of metalloenzymes upon mutagenesis presently cannot be done with sufficient accuracy. To summarize, the minimal "wish-list" for metalloenzyme design boils down to combining the evaluation of DD*G* upon

mutagenesis with the extensive backbone sampling and QM description of the metal center. Beyond this, some metal-containing sites are not amendable to quality QM treatment even when most state-of-the-art methodology is employed (for example the Fe_4S_4 or FeMoCo clusters). This constitutes a further issue, not addressed in the present work.

The quantum mechanical/molecular mechanical (QM/MM) approach, first introduced by Levitt and Warshel[18] to calculate the energies of intermediate states in enzymatic reactions, can provide an additional level of accuracy in modeling metalloproteins. Free energy calculations with QM/MM techniques have been developed, and will continue to be an important area of development. Jorgensen *et al*. have developed a QM/MM Monte Carlo free energy perturbation (FEP) method to calculate the free energy changes in biomolecular systems with adequate sampling. At the moment, it uses a semiempricial QM Hamiltonian, which is usually insufficient for metals.[19] Warshel *et al.* developed the Empirical Valence Bond (EVB) theory, and calculated free energy profiles of several different metalloenzymatic reactions. However, metal-ligand interactions are described in EVB through the harmonic oscillator approximation, and the metal thus cannot attach, detach, or exchange ligands, or change electronic state during the simulation.[20,21] Alhambra *et al*. also used QM/MM techniques to calculate the potential mean force, with umbrella sampling, along a fixed coordinate of the dehydrogenation by Zn^{2+}-dependent alcohol dehydrogenase, but again only with a semiempirical treatment of the metal and very limited sampling.[22] Yang *et al*. used QM/MM FEP calculations to determine the reaction path of a Mg^{2+}-dependent enolase.[23,24] The classical dynamics used in these QM/MM methods approach a picosecond limit to the timescale of their trajectories. Our own hybrid dynamics software, quantum mechanics/discrete molecular dynamics (QM/DMD),[25] performs record sampling for metalloproteins, and can routinely access such events as closure of protein loops,[26] and active site restructuring upon metal replacement or removal.[27,28] It is one of the most promising methods, and a good candidate for incorporation in a metalloenzyme design protocol.

The purpose of this work is to devise a metalloprotein design tool that takes advantage of the *Eris* algorithm for the evaluation of DD*G*, in conjunction with the sampling and structure prediction done with QM/DMD. The overall protein structure impacts the *Eris* DD*G* prediction, and this structure is coupled with the coordination and electronic structure of the metal. The coupling is captured by QM/DMD, and the metal is treated quantum mechanically within QM/DMD. The developed method is called *Eris-QM/DMD*. Due the involvement of QM and extensive backbone sampling, it is a unique method whose major use is envisioned in metalloprotein design. Testing of the method is achieved by assessing the DD*G* upon restoration of the native sequences in scrambled active sites of active natural metalloenzymes, representatives of different enzyme classes. This inverse design aims at the known evolutionary outcome (which is the native sequence), and thus tests the quality of the method. One prediction generated by *Eris-QM/DMD* turned out unexpected, however, and this result was further tested and confirmed experimentally.

8.2. METHODS

8.2.1. TEST SYSTEMS

The test systems to which *Eris-QM/DMD* will be applied are Ni^{2+}-dependent acireductone dioxygenase (ARD) and Zn^{2+}-dependent carboxypeptidase A (CPA), with the bound substrates. The active sites of ARD and CPA are shown in Figure 8.1. For both systems, we previously explicated reaction mechanisms, and realized the amino acids in the active sites that play critical roles in catalysis.[29,30,31] Therefore, they are our well-understood and trusted models. The chosen enzymes are also representatives of different classes: an oxidase, and a hydrolase. The initial structures of ARD and CPA were obtained from the Protein Data Bank (PDB) (ARD PDB code: 2HJI[32], CPA PDB code: 6CPA[33]). For ARD, since the PDB structure does not contain the substrate, 2-dihydroxy-3-keto-5-methylthiopentene, docked, we manually built it into the protein according to experimental and theoretical predictions.[34,35]**Error!**

Bookmark not defined. For CPA, the PDB contains a bound inhibitor, o-(((1r)-((n-phenylmethoxycarbonyl-l-alanyl)amino)ethyl)hydroxyphosphono)-l-benzylacetic acid (ZAAP(O)F), and therefore we built the substrate using the inhibitor as a scaffold as done by other computational works.[30,36,37,38] These initial native proteins with bound substrates were relaxed using QM/DMD, according to the scheme described below.

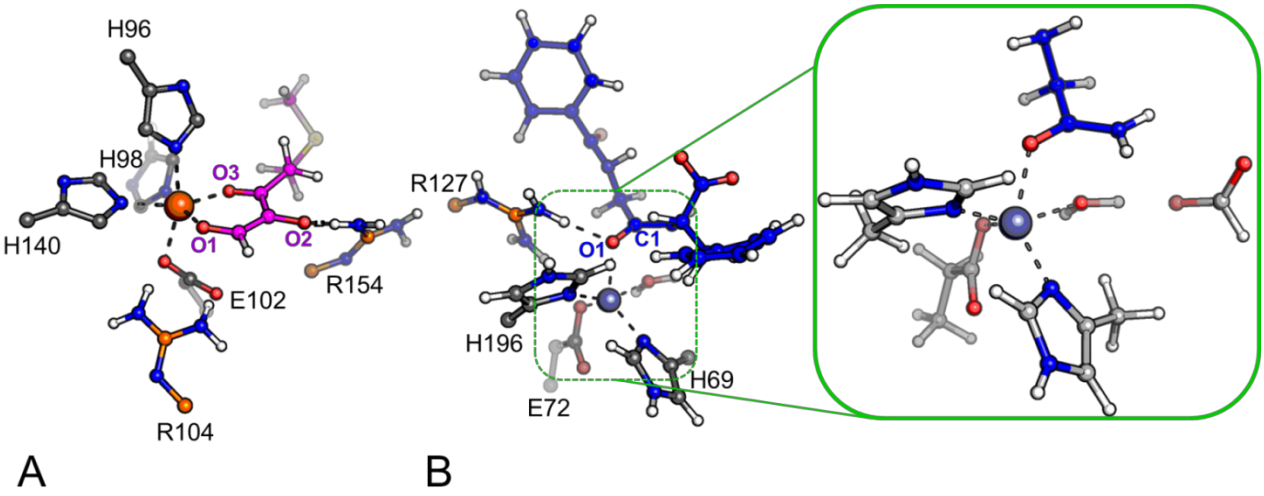


Figure 8.1. (A) Active site used in QM/DMD simulations of acireductone dioxygenase (ARD). (B) Full active site of carboxypeptidase A (CPA) with native substrate hippuryl-L-phenylalanine (hippuryl-L-Phe) (blue) along with the truncated QM model used in the QM/DMD simulations for computational speed up (in the green inset).

8.2.2. QM/DMD EQUILIBRATION OF THE NATIVE AND MUTANT PROTEINS

For sampling of protein variants, before and within the *Eris-QM/DMD* procedure, this study employs the QM/DMD[25] method, where DMD is short for discrete molecular dynamics.[39,40,41,42] QM/DMD is a hybrid method, a variant of QM/MM, that efficiently captures metalloprotein dynamics on the order of tens of nanoseconds, in conjunction with the QM description of the active site. It has been shown to perform exceptionally well for recapitulating and recovering native protein structures all the way down to the subtle structural details at the

active sites,[25] electronic properties of the bound metals,[25,27] protein conformational responses to substrate binding and metal replacement, [25,43] and mechanistic studies of enzymatic reactions.[25,43] The main attractive feature of QM/DMD is its affordability as compared to that of other capable QM/MM methods.[44,45,46,47,48,49,50,51] The strength of the method is in part due to DMD,[16-19] a remarkably successful classical force field-based method for sampling biological molecules and their complexes. DMD incorporates solvation through an implicit averaged way into the force field in use.

The protein system is divided into three different boundary regions handled by different levels of theory: DMD-only, QM-only, and QM-DMD-jointly. DMD samples the entire protein minus the QM-only region consisting of frozen metal(s) and the atoms immediately coordinating the metal(s). This approach eliminates the need for parameterization of the force field for the metal-ligand interactions. The QM-DMD boundary once in a while expands to include a larger active site (for ARD and CPA, those are depicted in Figure 8.1A&B: metal(s), ligands(s), substrate(s) and potentially other residues close to the chemically active area. This is the QM-DMD-region, which is optimized at a QM level. After that, the boundary shrinks back, and the QM-structural information is propagated to the rest of the protein through DMD, while DMD potentials get slightly reparameterized on-the-fly to match the QM predictions. The QM/DMD simulation is thus iterative, and proceeds to convergence in terms of QM and DMD energies, and backbone and active site RMSDs. Further details of the QM/DMD method can be found in previous publications.

8.2.3. QM METHODOLOGY USED IN QM/DMD SIMULATIONS

QM calculations are done using *Turbomole*.[52] All QM is calculated at the unrestricted Density Functional Theory (DFT) level, as a good compromise between accuracy and speed. The exchange and correlation are treated with the pure meta-GGA TPSS functional[53,54] with the empirical dispersion correction developed my Grimme *et al*.[55] To speed up calculations, we

employed the implementation of Resolution of Identity[56] and multipole accelerated resolution of identity.[57] For H, C, N, O and S, a double ζ quality basis set, def2-SVP,[58] was used and a triple ζ quality one, def2-TZVPP,[59] was used for the metals, Ni and Zn for ARD and CPA, respectively. Solvation and screening effects introduced by the protein were treated and approximated by the Conductor-like Screening Model (COSMO)[60] continuum solvent model with a dielectric of 20.0. In our numerous works on assessing structures and mechanisms of metalloenzymes, this level of theory was found to be sufficient for predicting the structures, and for the accurate reaction energetics and electronic properties of the metals a hybrid functional was then used.

8.2.4. *ERIS-QM/DMD* SIMULATIONS

Eris, in its original formulation for non-metal proteins, evaluates the effect of mutagenesis of one or several amino acids (collectively labeled A→B) on the overall protein stability, $\Delta\Delta G_{A\rightarrow B}$. First, *Eris* calculates the free energy change in the native protein (A) and the mutant (B) associated with their folding:

$G_{unfolded,A} - G_{folded,A} = \Delta G_A$,

$G_{unfolded,B} - G_{folded,B} = \Delta G_B$,

and also the ($\Delta G_{unfolded,A\rightarrow B} = G_{unfolded,A} - G_{unfolded,B}$) term.

$G_{unfolded}$ and G_{folded} are calculated as linear sums over the reference energies of all the amino acids in their structural context, found via an empirical parameterized expression, as reported previously.[61,62] The empirical approach is an alternative to more involved techniques such as FEP, but it is much cheaper, and the relative accuracy can be argued, depending on the extent of the structural change, and how amendable it is to FEP. More radical mutations cannot be treated very reliably in FEP, where gradual changes are preferred in accord with Zwanzig formalism. From the thermodynamic cycle, the desired $\Delta\Delta G_{A\rightarrow B}$ (or simply ΔΔG) is then obtained as

$\Delta\Delta G_{A\rightarrow B} = \Delta G_{unfolded,A\rightarrow B} + \Delta G_B - \Delta G_A$.

In this procedure, *Eris* relies on sampling and relaxation to obtain the equilibrated structures of the folded proteins. Following the mutation, the rotameric states of all amino acids in the protein are randomized. Sampling then occurs through the discrete rotameric space of amino acids, using Monte Carlo simulated annealing (each stochastic simulation is run twenty times, to accumulate enough statistics). *Eris* also goes through an additional subrotamer sampling at the final annealing temperature, and conjugate-gradient minimization. The backbone sampling in the original *Eris* can be either not done at all, or done via a modest backbone dihedral angles relaxation when backbone strains are detected upon mutations. Neither of the two approaches is satisfactory for metalloenzymes (results reported in SI), apparently reflecting the effect of the metal coordination on the overall protein structure. Additionally, *Eris* cannot deal with the changes in the coordination of the metal resulting from the effect of the protein. These coupled effects can be treated only when the metal coordination is dealt with at a QM level, and the backbone is extensively sampled within the same simulation. Hence, we coupled *Eris* with QM/DMD for sampling.

For each mutation, we run a QM/DMD simulation, allowing the protein to equilibrate and fully respond to the mutation, including the response of the metal coordination and protein backbone. Each such simulation is done over ~40 ns, as translated from the DMD time-units. The convergence of the simulations is monitors via the QM energies of the lager active site, the DMD energies of the entire protein minus the QM-only region, the backbone RMSD, and the all-atom RMSD of the active site. Trajectories are reported in the SI. At the end of every iteration of the QM/DMD trajectory (after the QM-phase), a sample structure is taken. This structure is mutated back to the native sequence, and an additional DMD-only simulation is run for 0.1 ns or until the convergence. The additional DMD step allows the protein to equilibrate with complete flexible backbone sampling (no angle or dihedral constraints). The parent mutant structure from the QM/DMD simulation and the native structure resultant from the pure DMD run are fed into

Eris to access the ΔΔG of mutagenesis. The obtained DD*G* values are then averaged over the considered set of iterations. Performing the *Eris* calculation at every iteration of QM/DMD allows us to track the convergence in terms of ΔΔG of mutagenesis. As will be shown below, the averaged values of DD*G* vary significantly, depending on the convergence of QM/DMD. When the average is taken over converged part of the trajectory, the values of DD*G* behave less erratically and make physical sense, and the standard deviation becomes smaller. The results indicate that QM/MM sampling is a required ingredient when it comes to predicting DD*G* of mutagenesis for metalloproteins.

We test the transferability of *Eris*-QM/DMD to metalloenzymes of a different functionalities and a different protein scaffolds. The first protein is ARD, a dioxygenase that plays a role in the methionine salvage pathway found in all organisms.[63] Depending on whether Fe^{2+} or Ni^{2+} is bound to the active, ARD oxidizes the substrate, 2-dihydroxy-3-keto-5-(methylthio)pentene (acireductone), into two products or three products, respectively, and this mechanistic bifurcation was recently explained.[31] In this mechanistic study it was realized that the two residues, R104 and R154, in the binding site stabilize, hold, and polarize the doubly deprotonated substrate and also transition state in the binding pocket, by donating hydrogen bonds to the substrate and one of the metal-coordinating residues, E102. Without these residues the substrate would coordinate in a different orientation, promoting a different mechanism. Here, R104 and R154 were mutated to aromatic and aliphatic, charged and uncharged amino acids (A, E, I, F), to test the performance of *Eris*-QM/DMD toward this diverse set. The method is expected to show that the return to the native sequence is favorable (has a negative ΔΔG value). If this would be not a test but a design problem, this would mean predicting the catalytically important mutations in the active site.

The second test enzyme is a hydrolase, CPA, an exopeptidase produced in the pancreas that preferentially cleaves C-terminal alphatic and aromatic amino acids from dietary proteins.[64] CPA employs a single Zn^{2+} that plays the role of a Lewis acid, polarizing the bound

carbonyl oxygen on the peptide backbone substrate, and the nucleophilic water molecule (Figure 8.3B). In CPA, an arginine, R127, is known both experimentally and computationally, to stabilize the transition state. As Zn^{2+} polarizes the carbonyl bond of the peptide, R127 stabilizes the negative charge that builds up on the oxygen at the reactants state, and more so - during the formation of the tetrahedral intermediate. The very similar set of mutations (A, E, I, K) was applied to R127.

We note that the amino acids that we will scramble in inverse design interact both with the substrates and with the rate-determining transition states (TSs) of the catalyzed reactions, and the interactions become stronger in the TS. Even though in our mechanistic studies we found the needed TSs, remaining on the saddle point during QM/DMD equilibration in *Eris*-QM/DMD was a tedious task. Therefore, the work was done on the complexes of the proteins with the bound substrates instead. Note, however, that this trick would not work if the scrambled amino acids would interact only with the TS, and not with the reactants. In such cases, keeping the system at or near the TS would be required, and it is certainly possible.

8.3. RESULTS AND DISCUSSION

8.3.1. ARD: NATIVE ARD FULLY RECOVERED, ALL EXCEPT FOR ONE MUTATION

The R104A, R154A mutant of ARD involves replacing the arginines with a small alanines, removing the stabilizing interactions with the substrate and the transition state. A significant amount of sampling is required to reach the convergence (Appendix 9.4, Figure 9.4.1). Note that the larger fluctuations of DMD energies are normal and have to do with the abrupt form of the potentials. The importance of sampling manifests itself in the ΔΔG values corresponding to mutations back to arginines. In Figure 8.2A, two representative sets of results are shown: the dashed line represents the ΔΔG values averaged over structures resulting from iterations 1-40 of QM/DMD, and the solid line shows those for averaging over iterations 41-80 (roughly the last 20-40 ns of dynamics). Notice that in both cases mutating both Ala back to Arg

is predicted to be favorable, but the value of ΔΔG is significantly more negative for the second half of the trajectory. The obtained negative ΔΔG of -9.3 kcal/mol indicates that the restoration of the two Arg residues if favorable for the overall protein stability, as expected. The poorer packing in the active site in the presence of two Glu can be seen with a naked eye (Figure 8.2C). Table 9.4.1 in Appendix 9.4 contains the results with the corresponding standard deviations.

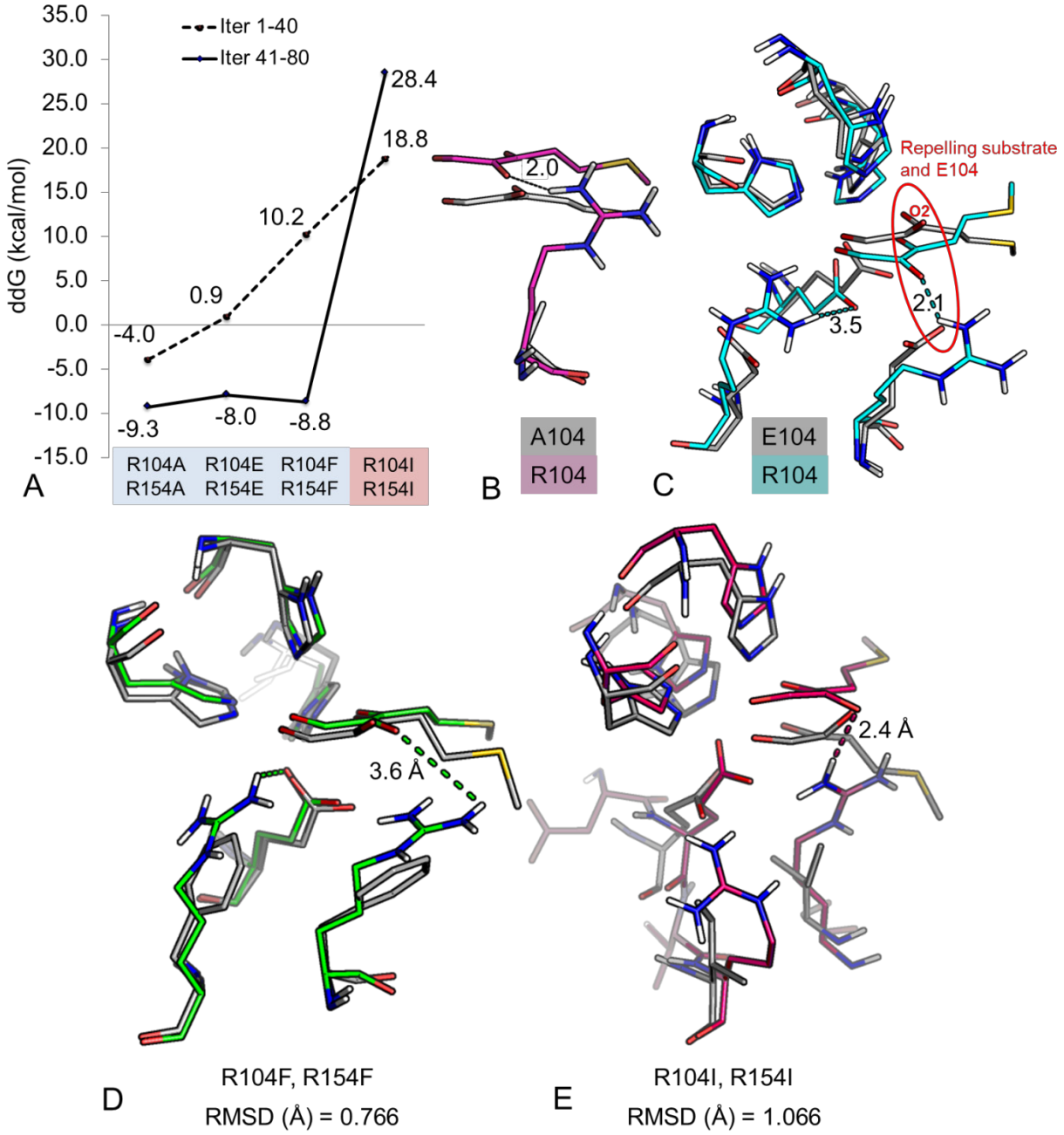


Figure 8.2. (A) The ΔΔG of mutation back to the native sequence for ARD, averaged over iterations 1-40 (dashed line) and iterations 41-80 (solid line) of QM/DMD. Negative values indicate that the native sequence is preferred over the mutant (the expected result). (B-E) Overlays of the active sites of the QM/DMD equilibrated mutants and the ARD active site recovered by *Eris*-QM/DMD: (B) R104A, R154A; (C) R104E, R154E; (D) R104F, R154F; (E) R104I, R154I. The original sequence is shown in cyan, and the mutant is shown in grey. The recovered contain the essential hydrogen bonds that R104 and R154 form in the active site.

The R104E, R154E mutant. QM/DMD simulations reveal a distorted substrate, enforced by the negative charges of glutamate (E104) residues clashing with the double deprotonated substrate bound to the metal. The double negative charge on the substrate repeals the glutamate, inducing a tilted distortion of the oxygen (O2) on the substrate relative to the metal. (Figure 8.2C) This distortion has an overall destabilizing effect on the active site of ARD. When the arginines are reintroduced into the protein, the hydrogen bonds of R104 to O2 and R154 to D102 are recovered with bond distances of 2.1 and 3.5 Å, respectively. The importance of sampling manifests itself in the ΔΔG values in Figure 8.2A, where a dramatic decrease in energy occurs after ~20ns of simulation time. The R104E, R154E mutation has an overall destabilizing effect with ΔΔG of -8.0 kcal/mol with the reintroduction of the arginines. (Figure 8.2C)

The R104F, R154F mutant undergoes a small amount of structural rearrangement when compared to the initial x-ray structure (RMSD = 0.766 Å) even though the phenylalanine residues have the potential to π-stack with each other or with other residues in the proteins and cause structural changes. (Figure 8.2D) Compared to R104E, R154E simulation, a similar pattern in ΔΔG values are seen in the R104F, R154F mutation only with a larger spread between the first half and second half of the simulation. During the first half of the simulations (twenty iterations), the average ΔΔG associated with reintroducing the arginines into the protein is 10.2 kcal/mol however, as the simulation progresses, a stabilizing ΔΔG of -8.8 kcal/mol is seen. The protein undergoes some significant rearrangement to recover the hydrogen bonds donated by the arginines, as shown by the overlay of the R104 and F104. The hydrogen bond distance between R104 and the O2 of the substrate averages to 3.6 Å, farther than what is seen in other mutants though.

The R104I, R154I mutant. Surprisingly, the only mutation Eris/QMDMD predicts having a larger stabilizing effect on protein than the native arginines is the R104I, R154I mutation. (Figure 8.2E) During the first twenty iterations of the simulations (~20 ns if simulation time), the average

ΔΔG of reintroducing the arginines is 18.8 kcal/mol, costly and destabilizing. As the simulation progresses, the ΔΔG continues to grow to 28.4 kcal/mol, becoming more and more unfavorable. Large structural rearrangements occur when isoleucines are introduced into the protein, more pronounced than any of the other mutations, with an RMSD of 1.066 Å when compared x-ray structure. The hydrophobic isoleucine residues could induce some internal protein structure reorganization/stabilization. The large protein stabilization could also be a weakness in our method.

8.3.2. CPA: *ERIS*-QMDMD RECOVERS NATIVE SEQUENCE IN ALL MUTANTS

For CPA, the residue we chose to focus for our *in silico* mutagenesis study is R127. (Figure 8.3) This arginine residue not only provides a hydrogen bond donor for the negative charge that builds up on the carbonyl oxygen (O1) on the peptide backbone during the reaction mechanism but also aids in the docking of the substrate to the Zn^{2+} (Figure 8.3B). In the crystal structure of the protein, with bound inhibitor, the N-O bond length between the arginine nitrogen and inhibitor zinc-bound oxygen is 2.85 Å. In computational works, this N-O bond distance is 2.61 Å[37] and 2.68 Å.[38] The N-O bond distance is a variable we track as we reintroduce the arginine back into the four chosen mutants: alanine (R127A), glutamate (R127E), lysine (R127K) and isoleucine (R127I).

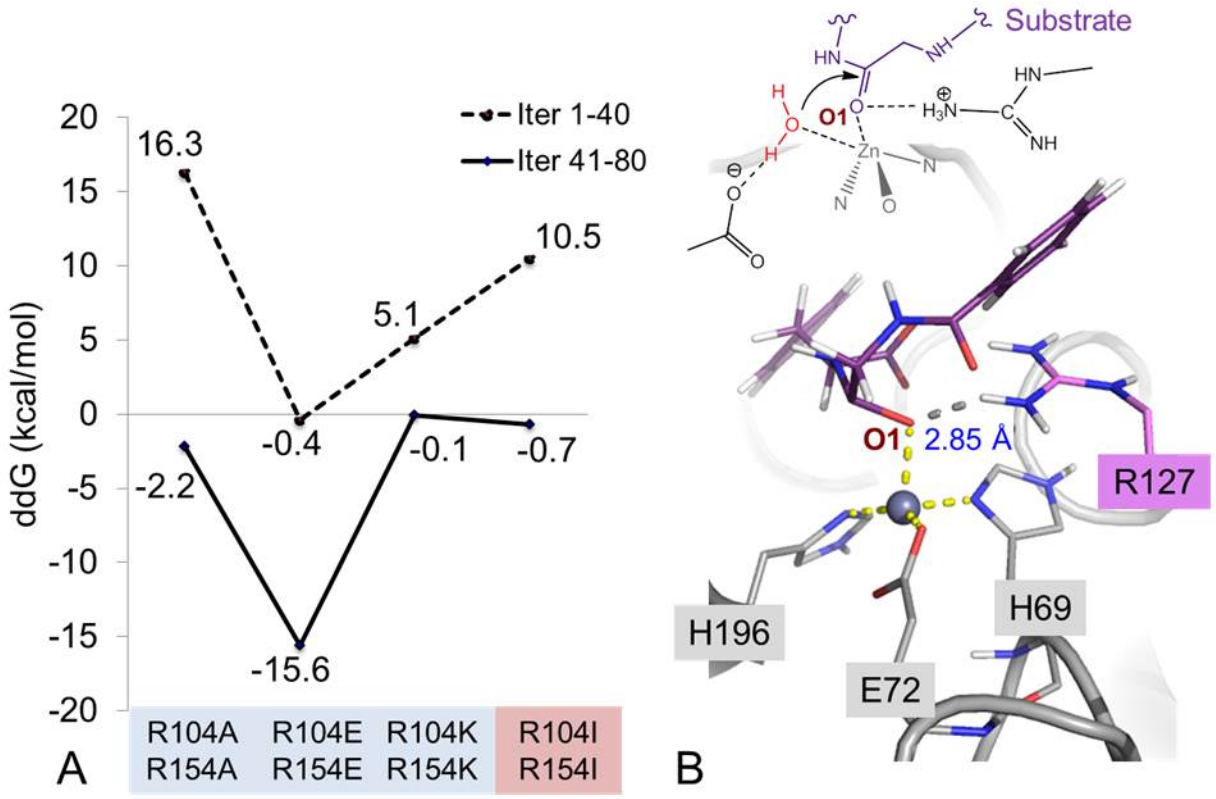


Figure 8.3 (A) The ΔΔG of mutation is calculated for each QM/DMD mutation, where $\Delta\Delta G=\Delta G_{mutant}-\Delta G_{native}$. In this simple equation, the arginine is considered the "mutant" since the study began with a series of mutations (A, E, K and I) where are aim to use *Eris*-QM/DMD to recover the native R residues in those positions (B) CPA active site with docked substrate hippuryl-L-phenylalanine. The carbonyl oxygen (O1) on the peptide forms an important hydrogen bond with R127 that assists in the binding of the substrate and stabilizes the negative charge on the oxygen during the reaction mechanism when a nucleophile, the Zn-bound water, acts the carbonyl carbon on the peptide backbone.

For the *R127A mutant*, the ΔΔG associated with substituting the alanine mutant back to arginine begins at 16.3 kcal/mol and steadily decreases to -2.2 kca/mol over the course of the simulation. This mutation involves the largest size difference between the two residues therefore, the backbone was allowed more freedom during the initial simulation. Hence, this is why the spread between the two ΔΔG values is so large as equilibration of the protein backbone takes time as it samples with the native arginine residue.

For the *R127E mutant*, replacing the positivity charged hydrogen bond donating amine groups on arginine with the negatively charge carboxy groups of glutamate is a large

destabilizing effect speaking from chemical intuition. The effect is manifested through tracking the ΔΔG of reinstalling the arginine as within the first part of the simulation, an average of -0.4 kcal/mol stabilization is seen. Although this change is within the margin of error in our method (~ 1 kcal/mol), the ΔΔG continues to decrease steadily to -15.6 kcal/mol as the protein equilibrates. The protein does not undergo large fluctuations, with an RMSD, between the native and mutant proteins, of roughly 0.7 Å as also seen by the recovering N-O between the amine nitrogen of R127 and carbonyl oxygen of the peptide bound to Zn, with a distance of 2.9 Å that agrees with experiment and other computations.

For the *R127I mutant*, the averaged ΔΔG upon the reintroduction of R127 is 10.5 kcal/mol and continues to decrease over the course of the simulation to -0.7 kcal/mol. (Figure 8.4A) There are a few structures that show the arginine-peptide hydrogen bond recovered with a comparable distance to the x-ray structure, 3.4 Å and 2.85 Å, respectively. (Figure 8.4C) The ΔΔG of -0.7 kcal/mol, although negative, shows only a slight preference for the native arginine over isoleucine, which could be another manifestation of a weakness in our method, for modeling isoleucines specifically, as was seen in the ARD mutant as well. However, the experimental design for testing this mutant is beyond the scope of this paper and was not pursued.

For our last mutant, *R127K mutant*, we matched residues relatively similar in size and charge, arginine and lysine, and predict that replacing arginine with lysine would not have a dramatic effect on protein stability of CPA. The *Eris*-QM/DMD simulation shows lysine is able to maintain the hydrogen bond to the peptide. However, this hydrogen bond is elongated, and subsequently when arginine is reintroduced into the protein, very small fluctuations restore it back to its native position. With an averaged ΔΔG of -0.1 kcal/mol, that is exactly what we see. The arginines overlay almost exactly with the lysine (Figure 8.4D). *Eris*-QM/DMD does have a slight preference for arginine, an indication that evolution chose the best positively charged hydrogen bond donor for this protein.

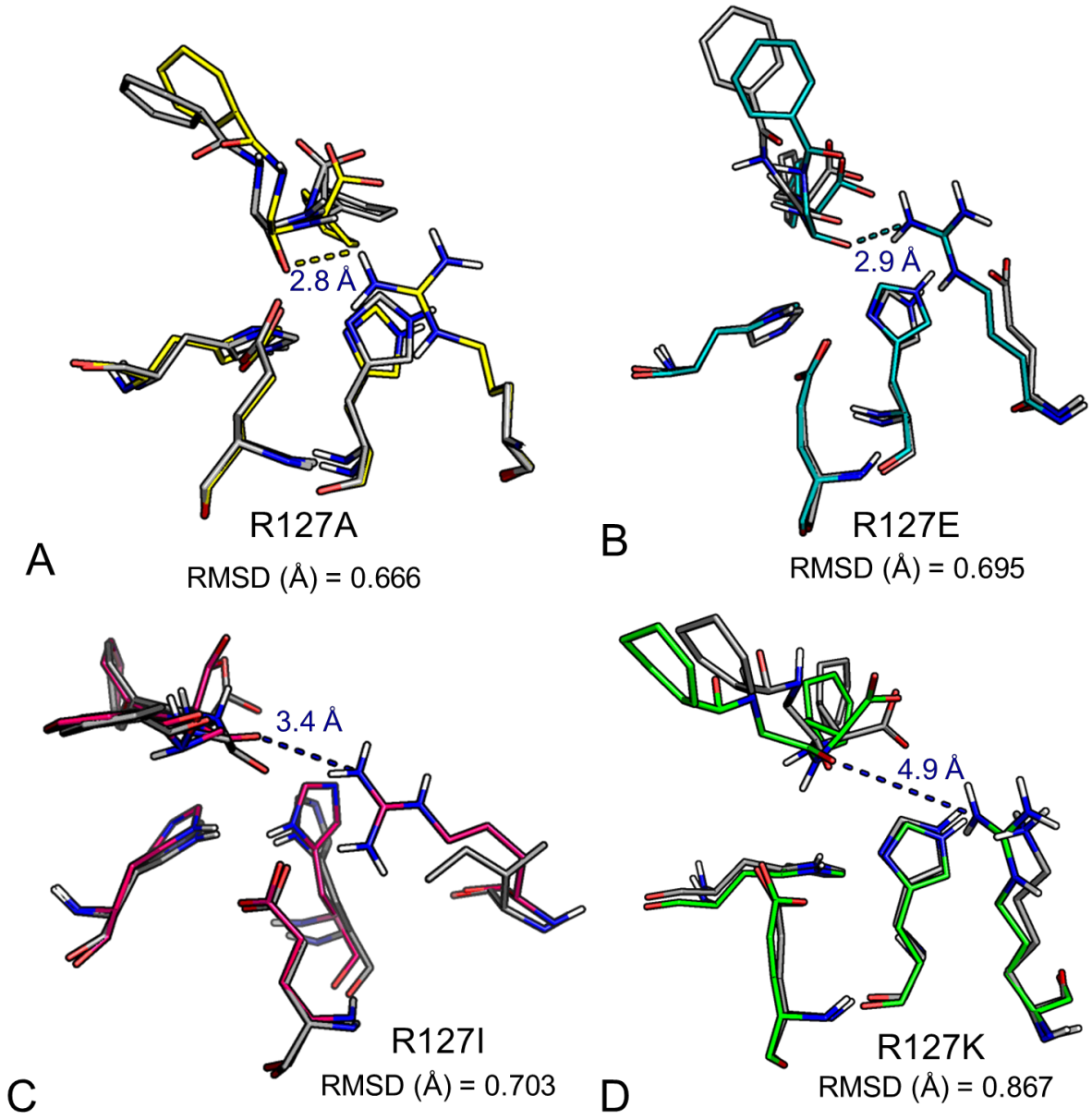


Figure 8.4. The four mutant CPA enzymes shown overlaid with native representative structure: (A) R127A, (B) R127E, (C) R127I and (D) R127K.

8.4. CONCLUSIONS

Here we debut our efficient and accurate metalloprotein design method, *Eris*-QM/DMD. *Eris*-QM/DMD provides a full QM description of the metal-containing active site and flexible backbone treatment of the protein during the *in silico* mutagenesis process. The method couples our fast QM/MM method, QM/DMD, which provides a multi-scale description of the

system with *ab initio* description of the metal and chemically active components and efficient backbone sampling with discrete molecular dynamics (DMD), with *Eris,* which measures protein stability when a mutation is introduced into a protein. We introduce mutations, known to hinder the activity and therefore stability of the structure, to two well-studied metalloenzymes. The first one is acireductone dioxygenase (ARD) that requires two arginines (R104, R154) to stabilize the binding of the doubly deprotonated substrate to the metal center. The other is carboxypeptidase A (CPA), a peptidase that possesses an arginine (R127) positioned by the carbonyl oxygen of the Zn-bound peptide, providing a hydrogen bond, and subsequently, a stabilizing residue for the buildup of negative charge during the reaction mechanism. After each QM/DMD simulation of each mutant, the native sequence(s) are reintroduced and the ΔΔG of that change is evaluated. A negative ΔΔG corresponds to *Eris*-QM/DMD having the ability recapitulate the known most active and stable protein structure.

For ARD, three of the pair mutations, R104A-R154A, R104E-R154E and R104F-R154F, yield negative ΔΔG values when the arginine residues are reintroduced back into the protein, showing the ability of our method to predict the most stable and active protein structure for a specific reaction. However, one mutation, R104I-R154I, was predicted by *Eris*-QM/DMD to be the more preferred residue. This stabilization could be due to some reordering in the protein or a weakness in our method. For CPA, for all four mutations (R127A, R127E, R127I and R127K) introduced and replaced with the native residue, *Eris*-QM/DMD reported the natural residue is preferred for the R127A, R127E and R127K mutants however, a slight preference for isoleucine is seen again. The preference is much smaller (3.4 kcal/mol) compared to the ARD mutation and could also be a weakness in our method. The lack of an electrostatic embedding scheme, a point charge representation of the whole protein and therefore the vital binding pocket for the substrate, within our QM/DMD method is a very natural next step in the further development and improvement of our methods.

8.5. REFERENCES

1 Z. J. Wang, N. E. Peck, H. Renata, F. H. Arnold, *Chem. Sci.* 2014, **5**, 598–601.

2 Yin, S.; Ding, F.; Dokholyan, N. V. 2007, Structure, 15, 1567-1576.

3 Han, K. E.; Bystroff, C.; Baker, D. Three-Dimensional Structures and Contexts Associated with Recurrent Amino Acid Sequence Patterns Protein Sci. 1997, 6, 1587–1590

4 Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De Novo Enzyme Design Using Rosetta3. PLoS ONE 6(5): e19230. doi:10.1371/journal.pone.0019230

5 Maglio O, Nastri F, Pavone V, Lombardi A, DeGrado WF. Proc Natl Acad Sci USA. 2003;100:3772–3777.

6 Warshel, A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. J. Biol. Chem. 1998, 273, 27035-27038

7 Kollman, P. *Chem. Rev.* **93**, 2395–2417 (1993)

8 Nivón, L. G.; Bjelic, S.; King, C,; Baker, D. Proteins: Structure, Function and Bioinformatics, 2014, 82, 858-866.

9 Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B (2011) A Generic Program for Multistate Protein Design. PLoS ONE 6(7): e20937.

10 Sterner, R. & Hocker, B. Catalytic versatility, stability, and evolution of the (βα)8-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055 (2005

11 Ding, F.; Dokholyan, N.V. PloS Comput. Biol. 2, e85 (2006)

12 Yin, S., Ding, F., and Dokholyan, N. V., "Eris: An automated estimator of protein stability", *Nature Methods*, 4:466-467 (2007)

13 Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D. Accurate Computer Based Design of a New Backbone Conformation in the Second Turn of Protein L J. Mol. Biol. 2002, 315, 471– 477

14 Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D.Design of a Novel Globular Protein Fold with Atomic-Level Accuracy Science 2003, 302, 1364– 1368

15 Kuhlman, B.; Baker, D.Native Protein Sequences Are Close to Optimal for Their Structures Proc. Natl. Acad. Sci. U.S.A. 2000, 97, 10383– 10388

16 DiMaio, F.; Terwilliger, T. C.; Read, R. J.; Wlodawer, A.; Oberdorfer, G.; Wagner, U.; Valkov, E.; Alon, A.; Fass, D.; Axelrod, H. L.; Das, D.; Vorobiev, S. M.; Iwai, H.; Pokkuluri, P. R.; Baker, D.Improved Molecular Replacement by Density- and Energy Guided Protein Structure Optimization Nature 2011, 473, 540– 543

17 Huang, P.-S.; Oberdorfer, G.; Xu, C.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D.High Thermodynamic Stability of Parametrically Designed Helical Bundles Science 2014, 346, 481– 485

18 Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227

19 (1) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005** 26, 1689 1700 (2) Acevedo, O.; Jorgensen, W. L. *Acc. Chem. Res.*, **2010**, *43* (1), pp 142–151 (3) Jorgensen, W. L.; Thomas, L. L. *J. Chem. Theory Comput.* **2008**, 4, 869– 876

20 Warshel, A. Computer Modeling of Chemical Reactions in Enzymes and Solutions; Wiley: New York, 1997.

21 S. C. L. Kamerlin & A. Warshel (2011) The Empirical Valence Bond Model: Theory and Applications. *Wiley Interdisciplinary Reviews-Computational Molecular Science* 1(1):30

22 Alhambra, C.; Corchado, J.; Sanchez, M.L.; Gao, J.; Truhar, D.G.; “Quantum dynamics of hydride transfer in enzyme catalysis” J. Am. Chem. Soc. 2000, 122, 8197-8203.

23 Liu, H, Zhang, Y, Yang, W. “How is the active site of enolase organized to catalyze two different reaction steps?” J. Am. Chem. Soc. 2000, 122, 6560-6570.

24 (1) Zhang, Y.; Lee, T.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46−54. (2) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483−3492.

25 Sparta, M., Ding, F., Shirvanyants, D., Dokholyan, N. V. & Alexandrova, A. N. (2012) Hybrid dynamics simulation engine for metalloproteins. Biophys. J. 103, 767-776.

26 Valdez, C. E.; Sparta, M.; Alexandrova, A. N. The role of the flexible L43-S54 protein loop of the CcrA metallo-beta-lactamase in binding structurally dissimilar beta-lactam antibiotics. 2013 J. Chem. Theor. Comput. 9, 730-737

27 Sparta, M.; Alexandrova, A. N. (2012) How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. PLoS ONE 7, e47172.

28 Nedd, S.; Redler, R. L.; Proctor, E. A.; Dokholyan, N. V.;* Alexandrova, A. N.* Cu,Zn Superoxide Dismutase without Zn is Folded but Catalytically Inactive. 2014, J. Mol. Biol., 426, 4112-4124.

29 Szeto, M. W. Y.; Mujika, J. I.; Zurek, J.; Mulholland, A. J.; Harvey, J. N., QM/MM study on the mechanism of peptide hydrolysis by carboxypeptidase A. *Journal of Molecular Structure: THEOCHEM* **2009,** *898* (1–3), 106-114.

30 Xu, D.; Guo, H.; J. Am. Chem. Soc. 2009, 131, 9780-9788.

31 Sparta, M.; Valdez, C. E.; Alexandrova, A. N. Metal-dependent activity of Fe and Ni acireductone dioxygenases: how two electrons reroute the catalytic pathway. 2013 J. Mol. Biol., 245, 3007-3018.

32 T. Ju, R. Goldsmith, S. Chai, M.J. Maroney, S.S. Pochapsky, T.C. Pochapsky. One protein, two enzymes revisited: a structural entropy switch interconverts the two isoforms of acireductone dioxygenase. J. Mol. Biol., 393 (2006), pp. 823–834

33 Kim, H.; Lipscomb, W. N. Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. 1990, Biochemistry, 29, 5546-5555

34 T.C. Pochapsky, T. Ju, M. Dang, R. Beaulieu, G.M. Pagani, B. OuYang ,in: A. Sigel, H. Sigel, R.K.O. Sigel (Eds.), Metal Ions in Life Sciences2, Wiley-VCH, Weinheim, Germany (2007), pp. 473–500

35 W.J. Wray, H.R. Abeles. The methionine salvage pathway in *Klebsiella pneumoniae* and rat liver. Identification and characterization of two novel dioxygenases. J. Biol. Chem., 270 (1995), pp. 3147–3150

36 Wu, S.; Zhang, C.; Xu, D.; Guo, H. J. Phy. Chem. B. 2010, 114, 9259-9267

37 Szeto, M. W. Y.; Mujika, J. I.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. Journal of Molecular Structure: THEOCHEM, 898, (2009) 106-114.

38 Valdez, C. E.; Nguyen, H.; Morgenstern, A.; Eberhart, M. E.; Loo, J. A.; Alexandrova, A. N. Computational redesign of the Carboxypeptidase A (CPA) and Experimental Validation. *Manuscript in preparation.*

39 Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. (1998) Molecular dynamics studies of folding of a protein-like model. Fold. Des. 3, 577–587.

40 Dokholyan, N. V. (2006) Studies of folding and misfolding using simplified models. Curr. Opin. Struct. Biol. 16, 79–85.

41 Ding, F., Guo, W. H., Dokholyan, N. V., Shakhnovich, E. I. & Shea, J. E. (2005) Reconstruction of the src-SH3 Protein Domain Transition State Ensemble using Multiscale Molecular Dynamics Simulations. J. Mol. Biol. 350, 1035–1050.

42 Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. (2008) Ab initio folding of proteins with all atom discrete molecular dynamics. Structure 16, 1010–1018.

43 Valdez, C. E.; Alexandrova, A. N. (2012) Why Urease Is a Di-Nickel Enzyme whereas the CcrA β-Lactamase Is a Di-Zinc Enzyme. J. Phys. Chem. B 116, 10649-10656.

44 Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718−730

45 Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

46 Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. In Review in Computational Chemistry; VCH: New York, 1995; Vol. 7.

47 Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.

48 Liu, H.; Muller-Plathe, F.; van Gunsteren, W. F. J. Mol. Biol. 1996, 261, 454−469

49 Bash, P. A.; Ho, L. L.; Mackerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci.* **1996**, *93*, 3698−3703.

50 Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.

51 Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462−3471

52 Turbomole V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, Turbomole GmbH, since 2007, available from http://www.turbomole.com

53 Perdew, J. P.; Wang, Y., Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **1992,** *45* (23), 13244-13249.

54 Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Phys. Rev. Lett. 2003, 91, 146401– 146401

55 Grimme, S. J. Comput. Chem. 2006, 27, 1787– 1799.

56 M.V. Arnim, R. Ahlrichs. Performance of parallel TURBOMOLE for density functional calculations. J. Comput. Chem., 19 (1998), pp. 1746–1757

57 M. Sierka, A. Hogekamp, R. Ahlrichs; Fast evaluation of the coulomb potential for electron densities using multipole accelerated resolution of identity approximation. J. Chem. Phys., 118 (2003), pp. 9136–9148

58 A. Schafer, H. Horn, R. Ahlrichs Fully optimized contracted gaussian basis sets for atoms Li to Kr J. Chem. Phys., 97 (1992), pp. 2571–2577

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

60 A. Klamt, G.J. Schüürmann COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient J. Chem. Soc. Perkin Trans. 2, 5 (1993), pp. 799–805

61 Yin, S.; Ding, F.; Dokholyan, N. V. Modeling backbone flexibility improves protein stability estimation. (2007) Structure, 15, 1567-1576.

62 Ding, F.; Dokholyan, N.V. (2006). Emergence of protein fold families through rational design. PLoS. Comput. Biol. 2, e85. Published online July 7, 2006.

63 (a) Myers, W. R., Wray, J. W., Fish, S. & Abeles, R. H. (1993) Purification and characterization of an enzyme involved in oxidative carbon-carbon bond cleavage reactions in the methionine salvage pathway of *Klebsiella pneumoniae*. J. Biol. Chem. 268, 24785-24791. (b) Oram S. W., Ai J., Pagani G. M., Hitchens M. R., Stern J. A., Eggener S., Pins M., Xiao W., Cai X., Haleem R., Jiang F., Pochapsky T. C., Hedstrom L., Wang Z. (2007) Expression and Function of the Human Androgen-Responsive Gene *ADI1* in Prostate Cancer. Neoplasia 9, 643–651.

64 Vallee, B. L., Galdes, A., Auld, D. S., Riordan, J. F., In *In Zinc Enzymes*, Spiro, T. G., Ed: Wiley: New York, 1983.

9. APPENDIX

9.1. SUPPORTING INFORMATION FOR CHAPTER 3

We attempted to calculate the relative free energies of the studied reactions, as opposed to just energies. The problem in doing so is that our studied systems are constrained during geometry optimization in a way that preserves the positions of the atoms directly attached to the protein backbone. In this context, the predicted lowest frequency vibrations corresponding to large "breathing" motion of the complexes are unphysical, because they do not involve the motion of the entire protein, and thus their frequencies are likely radically overestimated. The contributions of those modes to the entropy are very large, and so the entropy is estimated incorrectly. It is also noteworthy that Turbomole, used for these calculations, does not provide entropies for constrained complexes, which is thus justifiable. Nevertheless, we calculated vibrational and rotational entropies using our own stand-alone code, using the vibrational frequencies are rotational constants calculated by Turbomole. Free energies of reactions were thus calculated. The results do not appear in the main text, for the reason that will become apparent shortly. Here is how entropic contributions were calculated:

The individual contributions to the entropy of the constrained systems from rotational motions, vibrational motions and translational motions are as follows (Equation 1): (Source McQuarrie)[1]

$$S_{total} = S_{rot} + S_{trans} + S_{vib} + S_{elec} \quad (1)$$

For the purposes of this article, we can safely assume we the systems exist at the ground state therefore $S_{elec} = 0$. For the rotational contribution, we used the partition function for a non-linear polyatomic molecule (Equation 2). Since all systems have C_s symmetry, σ is set to one.

$$S_{rot} = R\left[\ln\left(\frac{\sqrt{\pi}}{\sigma_r}\left(\frac{T^{3/2}}{\sqrt{\Theta_x\Theta_y\Theta_z}}\right)\right) + \frac{3}{2}\right] \quad (2)$$

The translational contribution was calculated within the ideal gas approximation (Equation 3).

$$S_{trans} = R\left[\ln\left(\left(\frac{2\pi m k_B T}{h^2}\right)^{3/2}\frac{k_B T}{P}\right) + \frac{5}{2}\right] \quad (3)$$

The vibrational contribution, where there are K real normal modes (all imaginary frequencies were omitted), was calculated within the harmonic approximation. The vibrational partition function was calculated by choosing the bottom of the well as the zero of energy (Equation 4).

$$S_{vib} = R\sum_K\left(\frac{\Theta_{i,K}/T}{e^{\Theta_{i,K}/T}-1} - \ln\left(1-e^{-\Theta_{i,K}/T}\right)\right) \quad (4)$$

ΔG for the reaction was computed. Tables S1 and S2 are the expanded versions of Tables 1, 3, and 4 in the main text. In it, we report both energies and free energies of reaction, for comparison. One may see the drastic disagreement and lack of correlation between the values. For example Urease/lactam, the relative energies of the TS and intermediate changed from 7.5 kcal/mol and 6.6 kcal/mol to 4.3 kcal/mol and 6.0 kcal/mol, respectively. As was mentioned, this is due to the low lying frequencies that do not correspond to any normal modes of the system, which introduce errors to the vibrational partition function. We therefore omit all free energy results in the main text of the article, and report only energies.

	Urease/urea				**Urease/β-lactam**			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
TPSS/def2-SVP (kcal/mol)	7.1	0.0	14.3	14.2	2.8	0.0	7.5	6.6
B3LYP/def2-SV(P) (kcal/mol)	5.9	0.0	17.8	17.7	1.8	0.0	12.2	9.6
ΔG (kcal/mol)[a]	5.9	0.0	15.0	15.1	1.4	0.0	7.1	4.6

[a] Values calculated using the TPSS/def2-SVP results.

TABLE 9.1.1. Relative free energy (kcal/mol) for the Urease systems (see Figure 5.3., main text), where the thermal correction was added to the TPSS/def2-SVP data.

	β -Lactamase/urea				β -Lactamase/ β -lactam			
	Monodentate	Bidentate	TS	Interm.	Monodentate	Bidentate	TS	Interm.
TPSS/def2-SVP (kcal/mol)	0.0	8.6	19.6	n/a	0.0	0.6	15.7	3.7
B3LYP/def2-SV(P) (kcal/mol)	0.0	9.3	29.2		0.0	0.1	18.4	2.2
ΔG (kcal/mol)[a]	0.0	10.1	21.0		0.0	0.3	3.8	2.5

[a] Values calculated using the TPSS/def2-SVP results.

TABLE 9.1.2. Relative free energy (kcal/mol) for the Lactamase systems (see Figure 5. 4., main text), where the thermal correction was added to the TPSS/def2-SVP data.

	Ni-sub β -Lactamase/ β -lactam			
	Monodentate	Bidentate	TS	Interm.
TPSS/def2-SVP (kcal/mol)	0.0	-1.5	8.7	-8.3
B3LYP/def2-SV(P) (kcal/mol)	0.0	-0.9	12.4	-4.8
ΔG (kcal/mol)[a]	0.0	-0.3	10.7	-14.5

[a] Values calculated using the TPSS/def2-SVP results.

TABLE 9.1.3. Relative free energy (kcal/mol) for the Ni-sub β-Lactamase systems (see Figure 5.4., main text), where the thermal correction was added to the TPSS/def2-SVP data.

REFERENCES

1 Simon, J.D. and McQuarrie, D. A.; *Molecular Thermodynamics*. University Science Books, 1999

9.2. SUPPORTING INFORMATION FOR CHAPTER 5

	Singlet	Triplet	Quintet
Initial + dioxygen	11.5 / 42.9 / 43.5	0.0 / 0.0 / 0.0	36.6 / 37.7 / 38.8
Reactant	3.5 / 38.0 / 41.7	-8.0 / -3.3 / 0.1	32.9 / 42.8 / 51.6
TS	4.8 / 44.6 / 48.5	-6.6 / 0.4 / 4.2	35.2 / 47.8 / 52.1
Product	-84.6 / -75.2 / -82.8	-77.3 / -83.2 / -91.1	-19.8 / -11.7 / -13.0

Table 9.2.1. The reaction profile for ARD was computed using BP86/def2-SVP (def2-TZVPP for Ni^{2+}) followed with single point calculations with TPSSh/def2-TZVPP and b3lyp/def2-TZVPP. Values are listed at BP86/TPSSh/b3lyp

	Singlet	Triplet	Quintet	Septet
Initial + dioxygen	0.7 / 10.6 / 19.2	3.5 / 11.0 / 18.3	0.0 / 0.0 / 0.0	33.3 / 36.8 / 39.8
Reactant	-7.6 / 6.7 / 27.5	-5.1 / 6.8 / 16.3	-8.8 / -5.0 / -1.7	12.9 / 19.3 / 26.5
TS1	-6.6 / 11.1 / 25.0	-4.6 / 8.5 / 19.5	-8.2 / -2.7 / 1.1	14.7 / 22.9 / 30.7
Spilt	*	-34.2 / -21.0 / -11.9	-33.9 / -16.6 / -9.3	-32.3 / -24.8 / -21.1
TS2	*	-31.6 / -18.4 / -8.1	-30.7 / -11.9 / -2.4	-29.3 / -21.9 / -16.9
Product	-92.0 / -89.6 / -89.5	-108.3 / -98.5 / -98.4	-114.3 / -113.0 / -119.0	-84.8 / -82.4 / -85.2

Table 9.2.2. The reaction profile for ARD' was computed using BP86/def2-SVP (def2-TZVPP for Ni^{2+}) followed by single point calculations with TPSSh/def2-TZVPP and b3lyp/def2-TZVPP. Values are listed at BP86/TPSSh/b3lyp

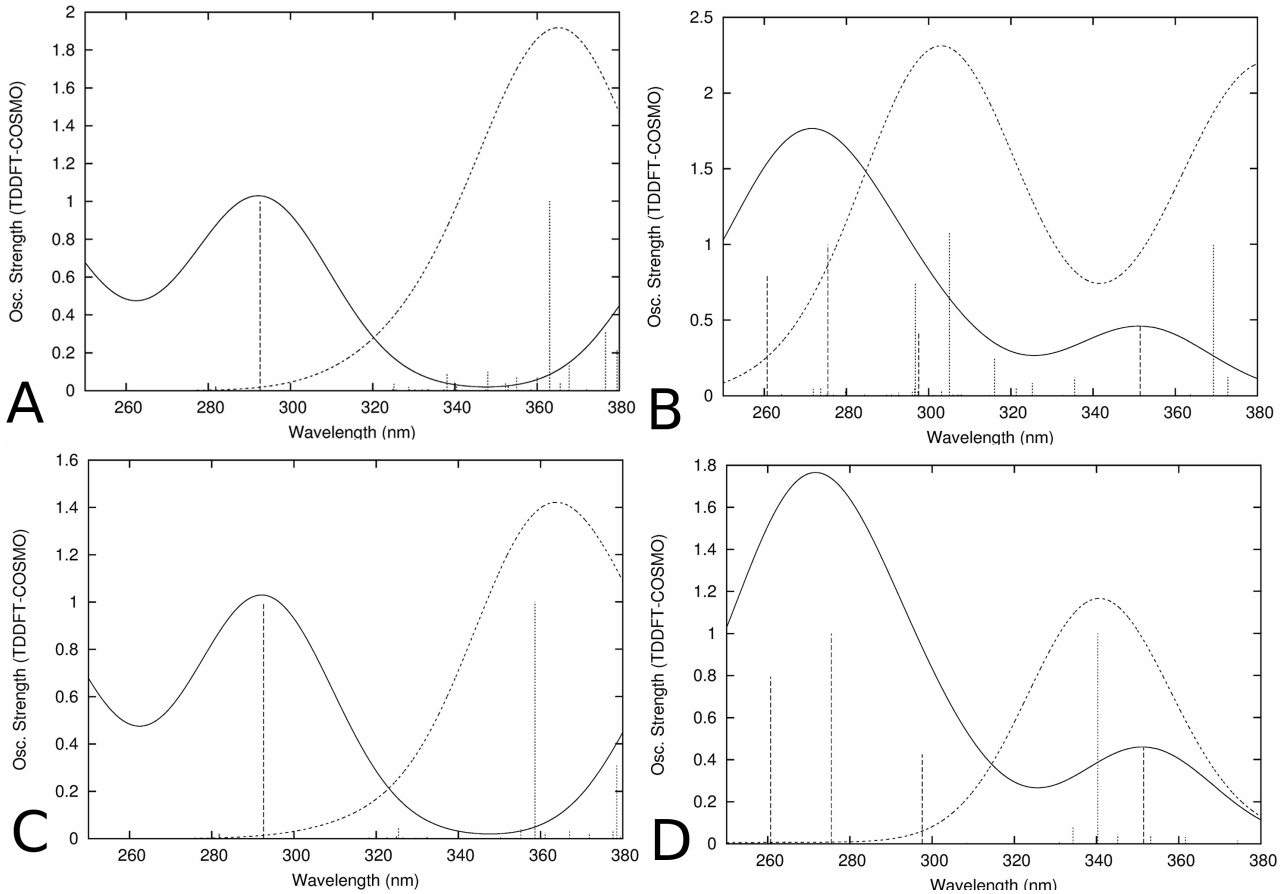


Figure 9.2.1. Calculated UV-vis spectra shows that when the substrate binds to the iron atom as a dianion (dotted lines), the λ_{max} becomes red shifted when compared to substrate in solution as a monoanion (solid lines). (A, B) TDDFT calculated spectra of the 6-membered ring adduct (dotted line) compared to free substrate in solution (solid line) calculated at with bp86/def2-SVP (def2-TZVPP for Zn), and b3lyp//def2-SVP (def2-TZVPP for Zn), respectively. (C, D) TD-DFT calculated spectra of the 5-membered ring adduct (dotted line) compared to free substrate in solution (solid line) calculated at with bp86/def2-SVP (def2-TZVPP for Zn) and b3lyp//def2-SVP (def2-TZVPP for Zn), respectively.

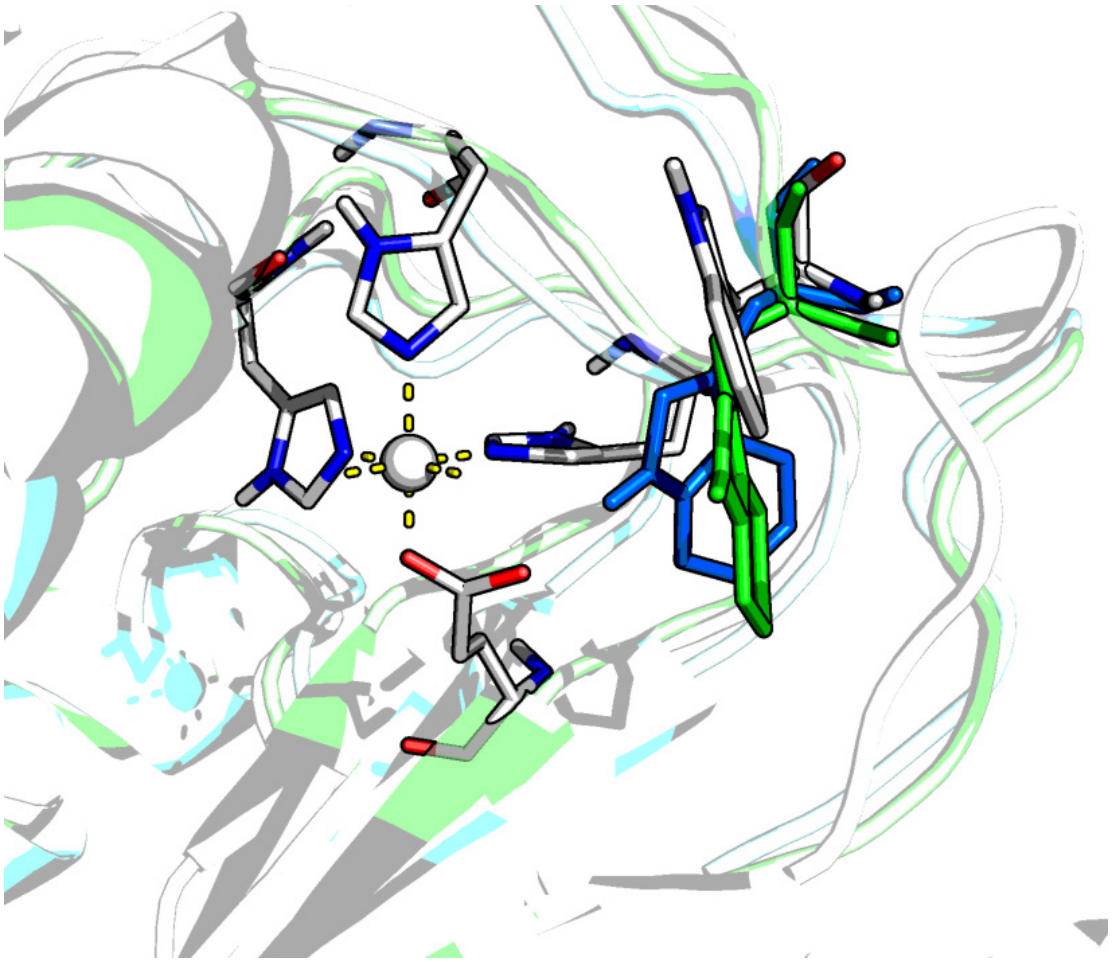


Figure 9.2.2. Representative structures from QM/DMD simulations with nickel (green) and iron (blue) are overlaid with the original x-ray structure (PDB: 1ZRR) (white). Paying particular attention the orientation of Trp162 relative to the metal-containing active site (shown in the stick representation), the metal species does not strongly effect the position of the Trp162 residue.

9.3. SUPPORTING INFORMATION FOR CHAPTER 6

	Doublet			Quartet		
	BP86	TPSSh	B3LYP	BP86	TPSSh	B3LYP
Initial	-3794.155424	-3796.18441	-3794.803152	-3794.1385	-3796.157812	-3794.775826
dioxygen	-150.2259316	-150.3907904	-150.3299122	-150.2259316	-150.3907904	-150.3299122
Initial + dioxygen	-3944.381356	-3946.5752	-3945.133065	-3944.364432	-3946.548602	-3945.105738
Reactant	-3944.386189	-3946.575836	-3945.126412	-3944.379544	-3946.591355	-3945.149491
TS1	-3944.384017	-3946.569897	-3945.119103	-3944.36391	-3946.574553	-3945.131801
Spilt	-3944.446622	-3946.623126	-3945.167274	-3944.364108	-3946.576206	-3945.133777
TS2	-3944.43912	-3946.615772	-3945.156057	-3944.365158	-3946.573443	-3945.132988
Product (Fe-ARD)	-3944.540803	-3946.73073	-3945.297059	-3944.424938	-3946.60336	-3945.150755
Product (Ni-ARD)	*	*	*	-3944.51853	-3946.744234	-3945.31981

Table 9.3.1. The reaction profile for Co-ARD in double, quartet and sextet (only relevant structures) was computed using BP86/def2-SVP (def2-TZVPP for Co^{2+}) followed by single point calculations with TPSSh/def2-TZVPP and B3LYP/def2-TZVPP.

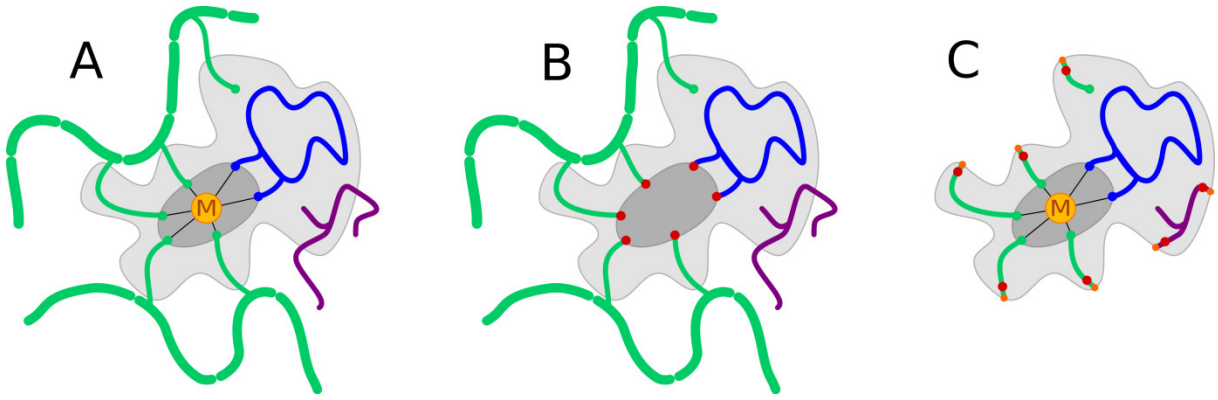


Figure 9.3.1. The protein is partitioned into DMD-only, QM-only and QM-DMD "breathing" region. A) The QM/DMD boundary contains the metal, coordinating residues and any substrates, cofactors or other regions of the protein important for catalysis. B) DMD samples the green protein backbone and the light gray region that includes amino acid residues, substrates and any other regions that need DMD sampling.

The only region DMD does not work on is the dark gray region where the metal and the atoms coordinating the metal are held frozen (red dots specify the atoms coordinating the metal. C) QM works exclusively on the light gray and dark gray regions, where the residues are truncated and frozen at the C_β positions (red dots). The “breathing” region is the light gray region that spends time in both DMD and QM machinery, hence providing the communication between the different methods.

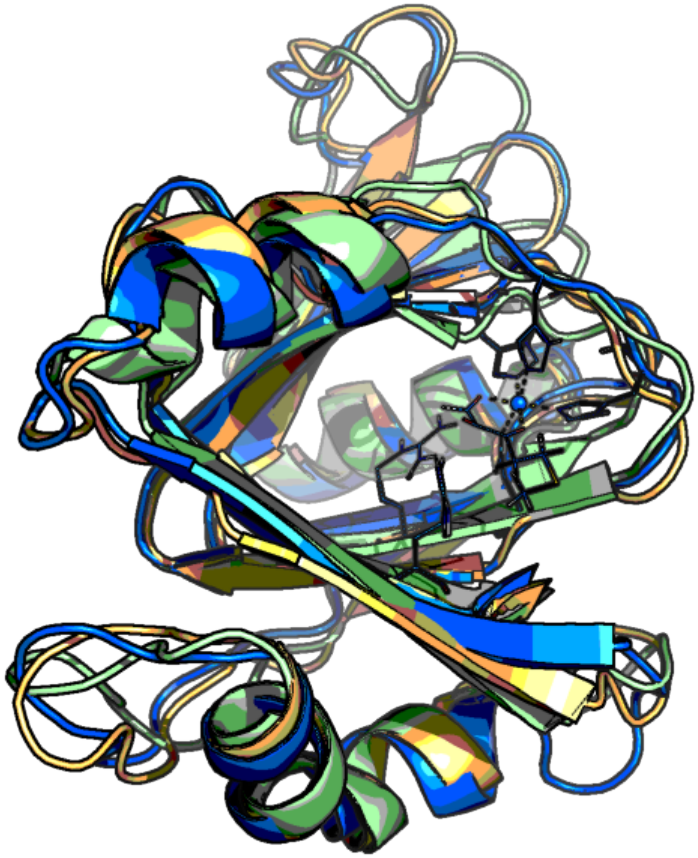


Figure 9.3.2. Representative structures from QM/DMD simulations with Ni-ARD (green protein), Fe-ARD (yellow protein) and Co-ARD (blue protein) show how the metal species does not strongly effect the backbone position. Structures are aligned by RMSD of the Cα.

9.4. SUPPLEMENTAL INFORMATION FOR CHAPTER 7

Native peptidase/peptide			
	ES	TS1	TI
TPSS	0.0	13.8	5.3
BP86	0.0	13.9	7.4
B3LYP	0.0	15.8	6.2
TPSSh	0.0	13.3	3.7
V243R_FpepD mutant peptidase/peptide			
	ES	TS1	TI
TPSS	0.0	17.9	10.1
BP86	0.0	13.0	9.8
B3LYP	0.0	21.9	13.6
TPSSh	0.0	21.6	13.7
V243K_FpepE mutant peptidase/peptide			
	ES	TS1	TI
TPSS	0.0	28.6	18.9
BP86	0.0	17.2	19.0
B3LYP	0.0	29.7	19.2
TPSSh	0.0	27.4	15.0

Table 9.4.1. For native and two mutants (V243R_FpepD, V243K_FpepE) Single points of stationary points with the TPSS, BP86, B3LYP, and TPSSh functional and def2-TZVPP basis set for all atoms from structures geometrically optimized with TPSS/def2-SVP (H,C,N,O) and def2-TZVPP (Zn^{2+}). Energies are in kcal/mol.

Native peptidase/peptide			
	ES	TS1	TI
H69	0.05011	0.0461	0.05171
E72	-0.84435	-0.86923	-0.88227
R127	0.93467	0.9008	0.94108
H196	0.06622	0.07172	0.07382
V243	0.00026	0.00017	6E-05
E270	-0.84731	-0.58525	-0.55422
Peptide	-0.97994	-1.24874	-1.52329
Wat	-0.06968	0.00425	0.23132
Zn	1.69001	1.68024	1.66181
V243R_FpepD mutant peptidase/peptide			
	ES	TS1	TI
H69	0.06817	*0.0461*	0.04592
E72	-0.83326	*-0.86923*	-0.84521
R127	0.93307	*0.9008*	0.92049
H196	-0.03165	*0.07172*	-0.0183
R243	0.9049	*0.00017*	0.8666
E270	-0.84356	*-0.58525*	-0.57727
Peptide	-1.84552	*-1.24874*	-2.26343
Wat	-0.01874	*0.00425*	0.20675
Zn	1.66653	*1.68024*	1.66447

Table 9.4.2: Natural population (NPA) charges of residues in native and V243R_FpepD mutant peptidase/peptide system calculated with TPSSh functional and def2-TZVPP basis set for all atoms.

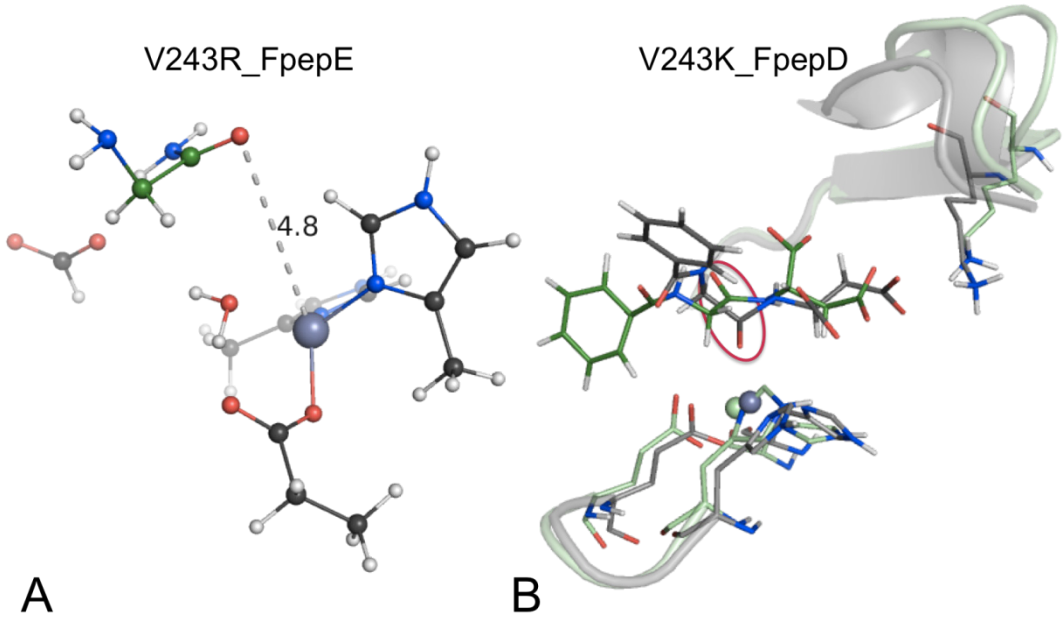


Figure 9.4.1. (A) Loose coordination of the V243R_FpepE mutant eliminates this mutant as a potential redesign candidate. The average carbonyl and zinc distance is 4.8 Å. (B) The mutant peptide in the V243K_FpepD completely flips around in the binding pocket and looses the key carbonyl-zinc interaction. A comparison of the first structure (gray protein) and final structure (green protein) with the pink circle highlighting the carbonyl that looses coordination with the zinc upon flipping of the substrate.

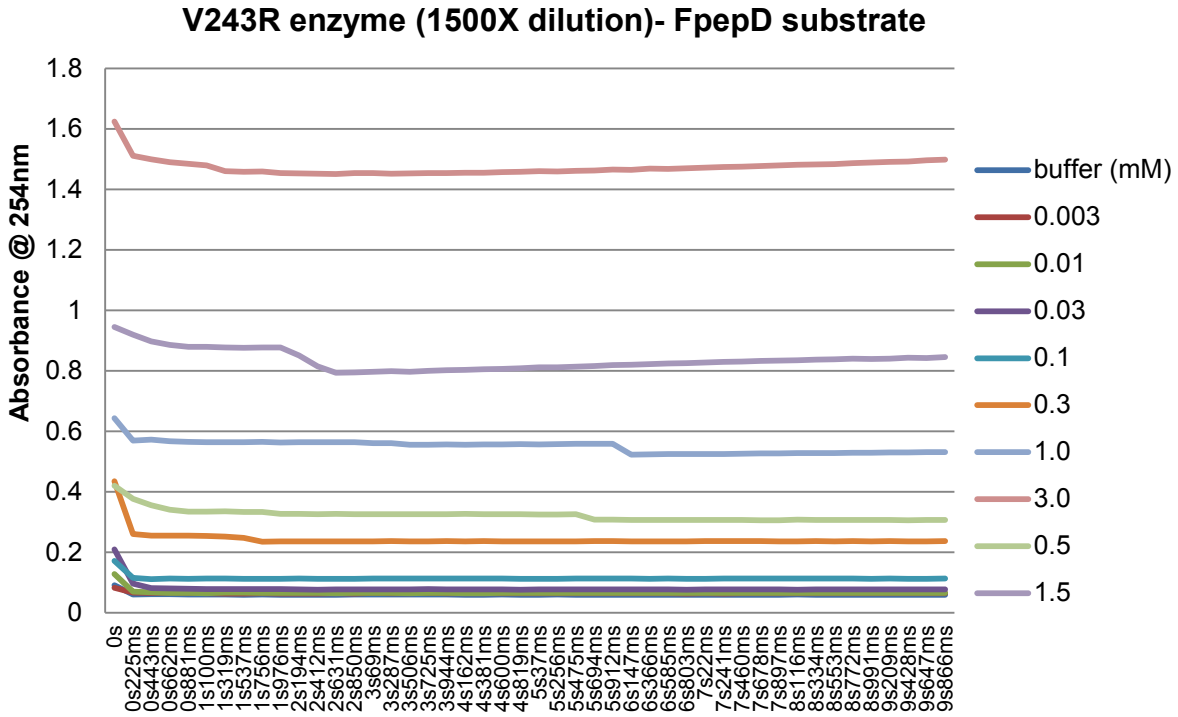


Chart 9.4.1. UV-vis spectra of the V243R_FpepD mutant peptidase/peptide system. The decrease in absorbance @ 254nm corresponds to the cleavage of the amide bond of the substrate under Michaelis-Menten kinetics conditions ([S] >>> [E]). The legend shows the varying substrate concentrations the assay is run at, from 0.003mM – 3.0mM.

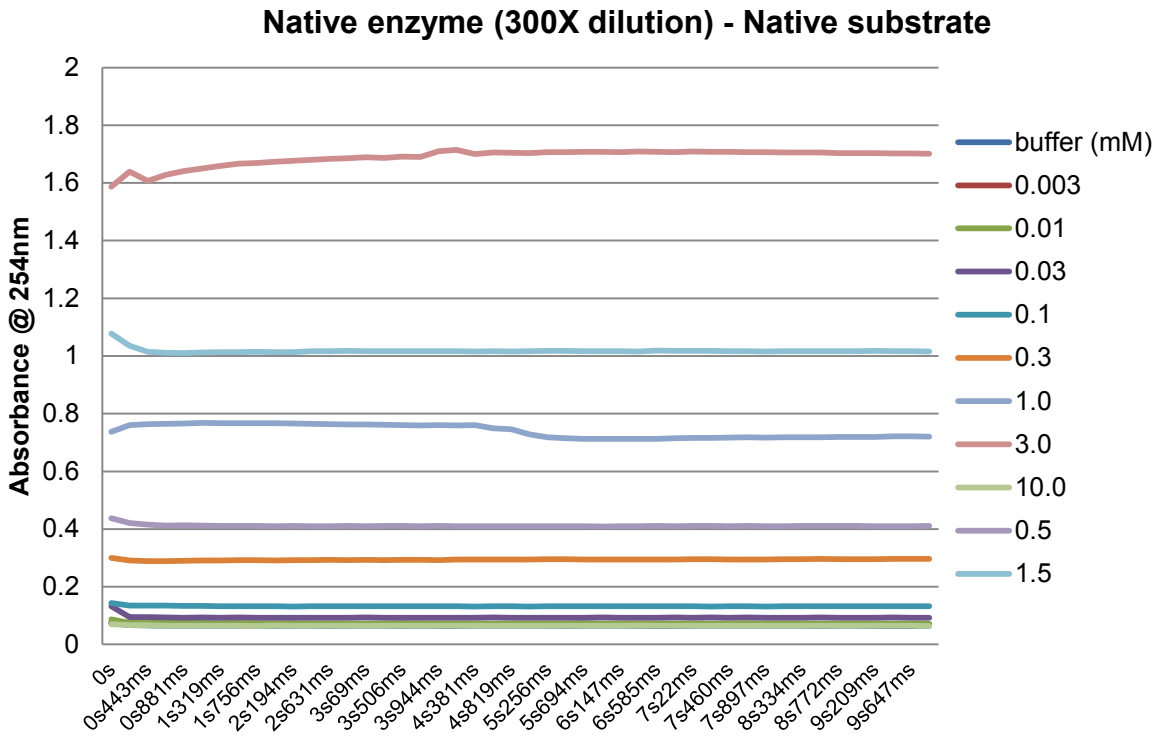


Chart 9.4.2. UV-vis spectra of the native CPA peptidase/peptide system. The decrease in absorbance @ 254nm corresponds to the cleavage of the amide bond of the substrate under Michaelis-Menten kinetics conditions ([S] >>> [E]). The legend shows the varying substrate concentrations the assay is run at, from 0.003mM – 3.0mM. Experimental trouble shooting is underway to further address the curves where an increase of absorbance is seen. This could be due to a combination of different reasons: the enzymes is cleaving itself, materials have been contaminated or grown old or something altogether more complicated

9.5. SUPPLEMENTARY INFORMATION FOR CHAPTER 8

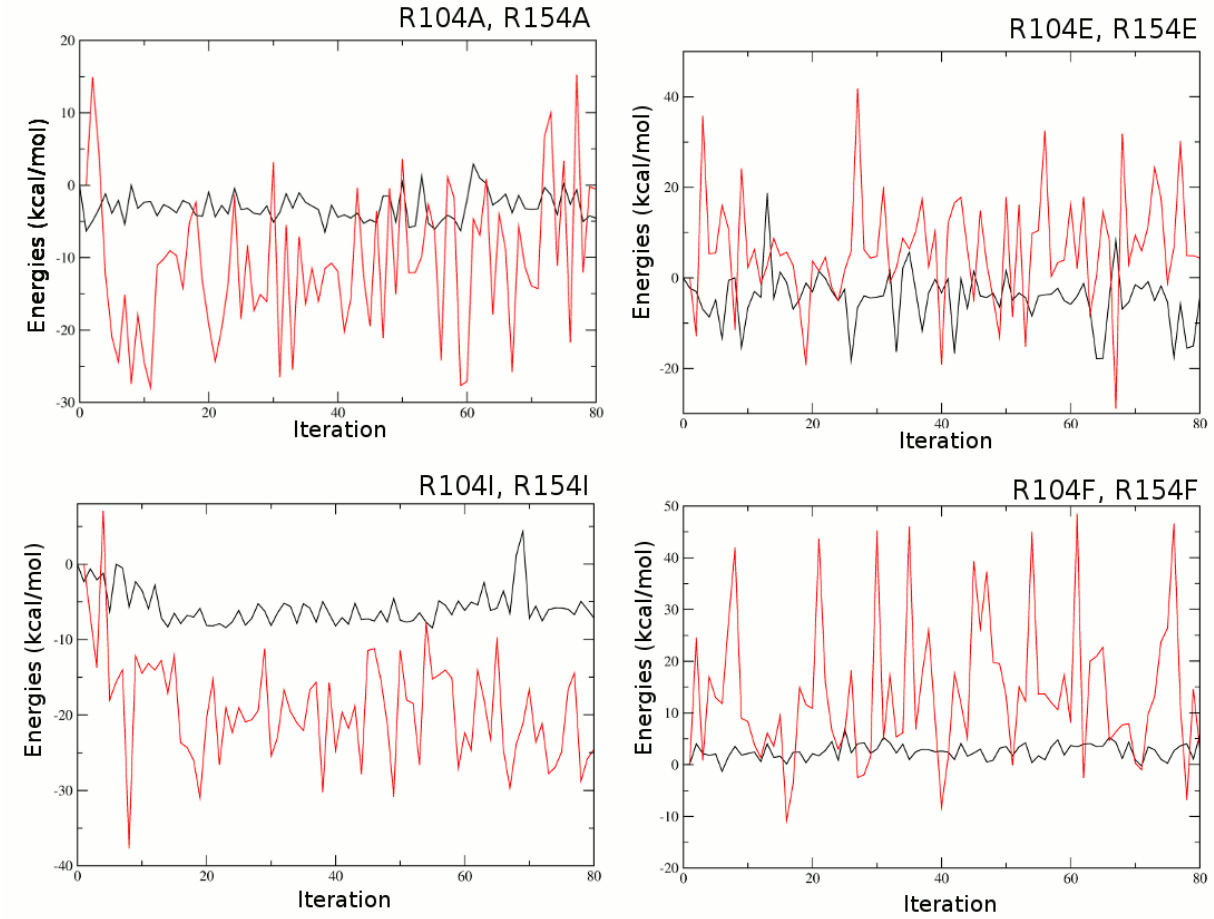


Figure 9.5.1. QM/DMD convergence data for each corresponding mutation, R104A, R154A (top left), R104E, R154E (top right), R104I, R154I (bottom left) and R104F, R154F (bottom right). The black line shows the relative QM energies for each iteration, with Iteration_0 structure as the reference, and the red lines shows the DMD energies for each iteration, with Iteration_1 as the reference structure.

	Mutations			
ARD (R104, R154)	R → A	R → E	R → F	R → I
ΔΔG (kcal/mol) ± std dev (Iterations 1-40)	-4.0 ± 23.6	0.9 ± 24.2	10.2 ± 21.6	18.8 ± 22.0
ΔΔG (kcal/mol) ± std dev (Iterations 41-80)	-9.3 ± 15.4	-8.0 ± 23.2	-8.8 ± 18.9	28.4 ± 14.2
CPA (R127)	R → A	R → E	R → K	R → I
ΔΔG (kcal/mol) ± Std Dev (Iterations 1-40)	16.3 ± 55.5	-0.4 ± 72.3	5.1 ± 54.2	10.5 ± 29.3
ΔΔG (kcal/mol) ± std dev (Iterations 41-80)	-2.2 ± 51.4	-15.6 ± 58.7	-0.1 ± 31.2	-0.7 ± 29.5

Table 9.5.1. Corresponding ΔΔG (kcal/mol) and standard deviations from the acireductone dioxygenase (ARD) and carboxypeptidase A (CPA) test cases.