UCLA
UCLA Electronic Theses and Dissertations

Title
Multiscale Modeling of Metalloenzymes: Design and Evolution

Permalink
https://escholarship.org/uc/item/6c0848pm

Author
Nechay, Michael

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Multiscale Modeling of Metalloenzymes: Design and Evolution

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Michael Raymond Nechay

2017

© Copyright by

Michael Raymond Nechay

2017

ABSTRACT OF THE DISSERTATION

Multiscale Modeling of Metalloenzymes: Design and Evolution

by

Michael Raymond Nechay

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2017

Professor Anastassia N. Alexandrova, Chair

With just a simple alphabet of natural amino acids and common metals, enzymes perform a spectacular number of reactions necessary for life on earth with enviable ease – with neither the extreme conditions nor chemical waste seen from many industrial reactors in human-developed catalysis. Principally, protein structure and related function is not well understood beyond being decisive for its unique selectivity and efficiency. A reliable treatment of larger protein movements/sampling coupled with precise quantum mechanical treatment of metals and bonds breaking/forming (multiscale) is still an open problem in Computational Chemistry. Metalloenzymes can be particularly challenging to model so we present some of the latest in multiscale modeling techniques. We have developed methods sensitive enough to study "selection" of similar metals in enzymes such as HDAC8, where previous literature failed to conclude which metal is active during in vivo catalysis, while fast enough to study larger protein movements including fold stability. Eminently, multiscale modeling opens the discussion of engineering enzymes to cater to modern day needs in catalysis. Society has not only developed needs in catalysis outside the scope of natural evolution (e.g., drug synthesis, energy

conversion) but also has access to more of the periodic table than nature has had a chance to explore. Thus, combining the efficacy of natural enzymes with modern catalytic processes has enormous potential. We have studied iridium, a rare metal with thermodynamic advantages over other metals in promoting catalysis of hydroamination and transfer hydrogenation. In the context of an enzyme we predict catalytic rates near and even exceeding existing organometallic catalysts with further design for specificity available. We are in collaboration with an experimental group in designing a protein fold which both directly accommodates and uses iridium in catalysis. We hope these foundations will support metalloenzyme design efforts towards novel chemical transformations performed as efficiently and environmentally soundly as nature has shown us is possible.

The dissertation of Michael Raymond Nechay is approved.

Paula Loredana Diaconescu

Caius Gabriel Radu

Anastassia N. Alexandrova, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

CHAPTER 1 COMPUTATIONAL TREATMENT OF METALLOPROTEINS 1
1.1. INTRODUCTION 2
1.2. METHODS FOR METALLOPROTEIN TREATMENT 7
1.2.1. QUANTUM MECHANICS/DISCRETE MOLECULAR DYNAMICS (QM/DMD) 7
1.2.2. APPROACH TO METALLOENZYME DESIGN: GENERAL OUTLINE 11
1.2.3. EREBUS 12
1.2.4. MULTI-SCALE DESIGN TOOL 13
1.2.5. ENTROPY EVALUATION 15
1.3. EXAMPLES OF APPLICATIONS 18
1.3.1. LARGE-SCALE MOTION IMPORTANT FOR ENZYME CATALYSIS 18
1.3.2. CHANGES IN THE COORDINATION GEOMETRY OF THE METAL 24
1.3.3. METALLOENZYMES DESIGN 26
1.4. FUTURE ASPIRATIONS 27
CHAPTER 2 HISTONE DEACETYLASE 8: CHARACTERIZATION OF PHYSIOLOGICAL DIVALENT METAL CATALYSIS 30
2.1. INTRODUCTION 31
2.2. COMPUTATIONAL METHODS 37
2.2.1. QM/DMD SIMULATIONS 37
2.2.2. LARGE QM SUBSYSTEM REQUIREMENT 40
2.2.3. MECHANISTIC STUDY 43
2.2.4. ELECTRON CHARGE DENSITY ANALYSIS 44
2.3. RESULTS AND DISCUSSION 45
2.3.1. CATALYTIC MECHANISM 45
2.3.2. WHY SUCH A DIFFERENCE FOR DIFFERENT METALS? 49
2.3.3. ANALYSIS OF CONTRIBUTIONS OF PROTEIN PARTS ON REACTION PROFILES 50
2.3.3.1. ROLE OF POTASSIUM ION 51
2.3.3.2. ROLE OF THE CATALYTIC HIS-ASP DYAD CONTACTS 52
2.3.3.3. CONTRIBUTIONS OUTSIDE QM SUBSYSTEM 53
2.3.4. METAL BINDING 55
2.3.5. METAL BINDING AFFINITY AND REACTIVITY 58
2.4. CONCLUSIONS 61

CHAPTER 3 COMPUTATIONAL APPROACH ON METAL SUBSTITUTION: SERUM TRANSFERRIN 63
3.1. INTRODUCTION 64
3.2. COMPUTATIONAL METHODS 65
3.2.1. FREE ENERGY OF METAL SUBSTITUTION 66
3.2.2. QM/DMD SAMPLING OF THE PROTEIN 68
3.3. RESULTS AND DISCUSSION 69
3.3.1 METAL GEOMETRY 69
3.3.2. LARGER STRUCTURAL DIFFERENCES OF SERUM TRANSFERRIN 73
3.3.3. DIFFERENCES IN METAL BINDING ENERGIES 76
3.3.3.1 METAL BINDING TO “ACID” FORM 76
3.3.3.2 METAL BINDING TO “DOUBLE” FORM 78
3.3.3.3 METAL BINDING TO “PRTR” FORM 79
3.4. CONCLUSIONS 80
CHAPTER 4 IRIDIUM ENZYME DESIGN – HYDROAMINATION 82
4.1. INTRODUCTION 83
4.2. METHODS 85
4.2.1. TEST SYSTEM 85
4.2.2. OVERVIEW OF DESIGN 87
4.2.3. COMPUTATIONAL DETAILS 89
4.3. RESULTS AND DISCUSSION 89
4.3.1. AB INITIO DESIGN OF METAL-BINDING COMPLEXES 89
4.3.2. COMBINATORIAL SEARCH FOR SUITABLE PROTEIN SCAFFOLDS 90
4.3.3. TEST FOR METAL-BINDING AND RETAINMENT AND POCKET REDESIGN 92
4.3.4. IN SILICO TESTING FOR CATALYTIC PROFICIENCY 94
4.4. CONCLUSIONS 95
CHAPTER 5 GUIDED EVOLUTION OF PROTEIN STRUCTURE 96
5.1. INTRODUCTION 97
5.2. METHODS 98
5.2.1. TEST SYSTEM 98
5.2.2. DMD EQUILIBRATION OF THE ORIGINAL FOLD 99
5.2.3. FACTORIAL DESIGN OF EXPERIMENT 101
5.3. RESULTS AND DISCUSSION 103

5.3.1. METAL BINDING RESIDUES CAN BE REMOVED WITHOUT SIGNIFICANT EFFECT..... 103
5.3.2. MUTATIONS OF SURFACE AND LINKER HAVE SIGNIFICANT EFFECTS ON STRUCTURE..... 106
5.3.3. PEPTIDE STRUCTURE CAN BE OPTIMIZED WITH PAIRS OF H-BONDING MUTATIONS..... 108
5.3.3.1. KEEPING ARGININE RESIDUES “IN PLACE” 109
5.3.3.2. MODIFYING OTHER SALT BRIDGING RESIDUES 110
5.3.3.3. SUMMARY OF HYDROGEN BOND NETWORK MUTATIONS 110
5.4. TESTING FOR METAL-BINDING AND RETAINMENT..... 117
CHAPTER 6 APPENDIX 120
6.1. SUPPORTING INFORMATION FOR CHAPTER 2..... 120
CHAPTER 7 REFERENCES 168

LIST OF FIGURES

Table 2.1. Computed reaction profiles with different methods and choices for the active sites. 47

Table 2.2. ΔΔG of binding between metal ions and HDAC8, relative to Co^{2+}.....57

Table 2.3. Predicted relative total catalytic activities. Binding affinity energies were taken relative to Co^{2+}. Normalized values appear on the right column, relative to the predicted most active metal ion-protein complex.....60

Table 3.1. Summary of substrate and protonation differences between otherwise identical structures of sTf studied.....66

Table 4.1. First column contains protein data bank entries the publicly available rigid substructure search tool, Erebus, output from a small cluster of amino acid atoms taken from geometry optimization of the transition state of an amino acid-iridium(III) complex catalyzing hydroamination. The second column indicates any metal present in the protein (though there does not have to be for Erebus to find amino acids in a certain arrangement), the third column indicates the molecular weight of the protein, and the last column indicates the available resolution of the .pdb file ("Solution" indicates NMR structure). 91

Table 5.1. β0 represents an average RMSD of all simulations and the other β values indicate the average magnitude and direction of RMSD change when the residue is mutated to a non-metal binding residue and the opposite if it isn't. For example, if residues marked C and E are mutated and the rest stay the same, the average RMSD would be the sum of the average of all simulations with βC and βE added and βC and βA, βB, and βD subtracted (4.42 – 0.480 – 0.128 + 0.228 – 0.002 + 0.108). Overall, only A was found to have a statistically

significant impact on RMSD ($p < 0.05$). No significant interaction effects were observed. 105

Table 5.2. As in Table 5.1., β_0 represents an average RMSD of all simulations and the other β values indicate the average magnitude and direction of RMSD change when the residue is mutated as described above. All main effects were statistically significant ($p < 0.05$) and with a large positive magnitude indicating that every mutation attempted resulted in a large increase of RMSD, thus structure deviated considerably from the target structure. No significant interactions were observed. 108

Table 5.3. Summary of mutation sets toward goal of optimizing the hydrogen bond network of the MID1sc2 peptide. 111

Table 5.4. Summary of multi-way ANOVA results for hydrogen bond network optimization. See Table 5.3. and the corresponding discussion for descriptions of all categories – multiplied terms are higher order interaction terms where the coefficient is added when all labels in the term are present. 112

Table 6.1. List of literature binding energies with regards to a series of chelating agents for all of the relevant metal ions studied in HDAC8 (in kcal/mol). 162

Table 6.2. Calculated ΔG (kcal/mol) - DTPA metal swapped with EGTA 163

Table 6.3. Experimental ΔG (kcal/mol) - DTPA metal swapped with EGTA 164

Table 6.4. Calculated ΔG – Experimental ΔG (kcal/mol) - DTPA metal swapped with EGTA 165

Table 6.5. Comparison of energies at various dialectric constant values from interior of protein (dielectric = 4) to water (dielectric= 80) 166

Figure 1.1. 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. .. 9

Figure 1.2. 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) Wi represents the accuracy of a match to the input scaffold, more specifically Wi 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.....13

Figure 1.3. Analysis of the experimental binding thermodynamics for approximately 100 protein-ligand complexes. Reprinted with permission from Reference [68].....16

Figure 1.4. 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.....20

Figure 1.5. 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 with permission from Reference 42.....23

Figure 1.6. 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.....24

Figure 1.7. The redox reaction in Ni-SOD is modulated through the changing coordination environment of the histidine (H1) residue.....25

Figure 2.1. The active site of HDAC8 extracted from the crystal structure (PDB code 2V5W), highlighting the most critical residues relevant to catalysis within the active site. In this work, F306 was reverted to tyrosine to facilitate catalysis (yellow, Y306F). Blue residues participate directly in catalysis or coordination of the divalent metal ion, the orange residue is the crystal structure substrate, represented in purple is a nearby K ion surrounded by a small net of residues holding it in place.....33

Figure 2.2. Schematic reaction diagrams for all previously considered HDAC8 mechanisms: (a) relies on the unusual initial protonation state of the coordinating His; reaction. Further, concerted addition to the carbonyl with transfer of proton to His142, as in (a), is found less favorable due to electrostatic effects of the nearby protein (as discussed in "Contributions Outside QM Subsystem" later in this paper). (b) was originally considered by the Zhang group,[41, 42] and reaction (c) was studied by Chen et al[60] as an alternative to the Zhang mechanism. Reactions (b) and (c) are the mechanisms most studied in this paper. The relevant 35

Figure 2.3. Typical convergence plots of QM/DMD runs. The top left graph displays the energy of the QM active site and the top right displays the DMD energy of the protein as a function of QM/DMD iteration number, respectively. Note that sporadic high-energy structures would be discarded in the iterative procedure. The bottom left displays the convergence of backbone RMSD, while the bottom right exhibits the convergence of the active site RMSD. A levelled off RMSD graph indicates the lowest energy structures can be weighed based on Boltzmann statistics. This simulation with Zn^{2+} in the active site, and it is representative of all studied systems. 39

Figure 2.4. Diagram of residues included in the QM profile in previous, and the present studies. The bound substrate is shown in orange. Original work by the Zhang group[41, 42, 59] included residues colored in blue. The investigation by Chen et al.[60] included the blue residues, as well as the additional residues D176 and D183 in green. The current study expands on both of these by including additional residues colored grey, and the K^+ ion colored purple, as justified in the text. 42

Figure 2.5. Reaction pathway for HDAC8 and a variety of divalent metal ions (TPSS/Def2-TZVPP+Gibbs free energy corrections//TPSS/Def2-DZVPP, COSMO).....46

Figure 2.6. Critical points and bond paths of interest in the active site of HDAC8 with Zn (left) and Mn (right). When Zn is present in the active site a bond path forms between the water oxygen and the carbon atom from the substrate carbonyl. This topologically necessitates a ring critical point to exist in the active site as well. When Mn is present, no ring CP is found as there is not a bond path between the water and substrate carbonyl. Ni, Fe, and Co give the same topology as Zn, while Mg has the same topology as Mn. Sphere coloring is as follows: C-black, N-blue, O-red, Metal-grey, bond CP-cyan, ring CP-green. 50

Figure 2.7. Highlights the difference in Gibbs Free Energies for the most active Co^{2+}-HDAC8, relative to the reactant state, upon electrostatic embedding. An increase in energy indicates embedding has a destabilizing effect, while decrease in energy indicated stabilizing effect of embedding. States ts1c and His142 Intermediate are only considered in mechanism (c), while ts1b, His143 Intermediate, and ts2b are considered in the current mechanism. 54

Figure 2.8. Schematic of all considered thermodynamic cycles exploited for relative binding affinities. 56

Figure 2.9. Trends between QTAIM $\mathcal{E}(r)$ at bond CP and $\Delta\,\Delta G$ of metal swapping.....58

Figure 3.1. As a state function, differences in Gibbs free energy of a protein binding two metals from water can use a combination of just a couple DFT calculations with experimental stability constant data. This avoids various otherwise problematic computational challenges of modeling a metal in water. 67

Figure 3.2. Different protonation states of the active site – note Tyr is deprotonated when attached directly to the metal 68

Figure 3.3. The distribution and outliers of octahedral angle variance for each metal in each protonation state of the protein. Several metals show significant deviation from

the octahedral geometry. Acid forms of Zn(II) and Cr(III) omitted for now due to calculation error. 70

Figure 3.4. Three snapshots of the metal binding site of protonation state “phys”. Zn^{2+} (left) and Fe^{2+} (right) lose a tyrosine during the simulation (middle is Zn^{2+} with octahedral geometry shown for reference). 71

Figure 3.5. Finer details of Figure 3.3. 72

Figure 3.6. Ti(IV) overcoordinates in the “double” form (left) and in the other forms distorts from octahedral as shown (right). 73

Figure 3.7. Tracking the distance between the subdomains throughout a simulation is just one way to track a conformational change from “open” to “closed”. The minimum difference between backbone atoms of residues 41-44, 11, 12, 288, 289 on one subdomain with 177-180 on the other (pointed to by the arrows here) is calculated for every structure and plotted across simulation time in Figure 3.8. .74

Figure 3.8. Distance between subdomains plotted over the course of all simulations for each metal 75

Figure 3.9. Stereoscopic view of the Co(III) variant (blue color) and Fe(III) variant (tan color) of the lowest energy structures QM/DMD discovered for the “acid” variant. 77

Figure 3.10. View of Co(II) bound to the “acid” form of serum transferrin with lysine – aspartic acid proton transfers highlighted in purple 78

Figure 3.11. Lowest energy conformations of the Fe(III) and Co(III) metal binding sites of the “double” form”. 79

Figure 3.12. Fe(III) rejects water for aspartate in the active site effectively helping repack the active site relative to Co(III) which retains the originally bound water molecule. .80

Figure 4.1. Typical catalytic cycle of hydroamination via an iridium catalyst to be mimicked by an enzyme 85

Figure 4.2 The rate determining step of hydroamination modeled in silico in this work and then our design scheme was used to find an enzyme that could preferably stabilize the transition state.....86

Figure 4.3. General enzyme design scheme to find and adapt a protein for a desired chemical transformation.....87

Figure 4.4. Scheme of protein (pdb accession code 1V5R) with metal bound vs metal and substrate with glutamate residue showing strong affinity to the positive charge on the bound molecule.....93

Figure 4.5. Calculations at TPSS with def2-TZVPP for iridium and def2-SVP for other atoms. Cosmo dielectric constant ε = 80 because of the solvent exposed nature of the site.....94

Figure 5.1. Protein design scheme of Chapter 4 modified to include the protein evolution scheme described in this chapter.....98

Figure 5.2. The original published MID1 structure1 is shown on the left and is composed of two symmetric alpha helix bundles arranged perpendicular and bound by two metal atoms. The modified MIDsc2 structure is shown on the right to consist of a both parts connected by a linker, thus yielding a single peptide where point mutations can be applied assymetrically (one set of metal binding residues was removed for MIDsc2). Note that MIDsc2 was unable to be crystallized but is the model structure future catalysis was meant to be based on and that is the motivation for this work and so will be often referred to as the “target structure” throughout this text.....99

Figure 5.3. Sixteen replicates of the original peptide (MID1sc2) underwent 55 nanoseconds of Discrete Molecular Dynamics – a short annealing step at 310 K followed by a longer run at 275K. RMSD of the structure vs the original structure is plotted over

this simulation for all replicates and significant deviations observed for a fraction of the runs.....100

Figure 5.4. Significant deviations from the original (target) structure were observed to include undesired salt bridge formation, secondary structure forming in the linker along with undesired hydrogen bonds, and a parallel peptide bundle. 101

Figure 5.5. The five potentially metal-binding residues are mutated to non-metal binding residues (namely, to the amide residue asparagine) in a full factorial design. From the full factorial design, multi-way ANOVA can help determine main and interaction effects of residue mutation on structure. 104

Figure 5.6. Two sets of mutations, labeled LF (link fix) and APB (avoid peptide bundle), are identified as potentially mutable towards reducing secondary structure linker formation and peptide bundle formation, respectively, and are tested here via a full factorial design and multi-way ANOVA on the results. 107

Figure 5.7. Five sets of mutations, labelled A-E, were identified as potential contributors to protein fold stability via hydrogen bond networks. 109

Figure 5.8. The distribution of pairwise distances between residues 16 and 88 across all simulations with and without the Trp16Asp mutation. A sharp peak around 2 Angstroms, typical of a hydrogen bond, corroborates with the improved stability as a small success in the design process. 113

Figure 5.9. Pairwise distances of residue pairs 23 + 21 and 23 + 26 between mutation set C (red) and no mutation set C (blue) confirming that a hydrogen bond forms as designed. 115

Figure 5.10. A sampling of structures at the two minima outside of the global minima occupied primarily by the Ala21Glu mutant as in Figure 5.10. 116

Figure 5.11. A sampling of structures at the two minima outside of the global minima occupied primarily by the peptides without the Ala21Glu mutation as in Figure 5.10. 116

Figure 5.12 The peptide structure on the left was one of many DMD structures which give an accurate depiction of how the peptide may behave in solution. QM/DMD quickly came to equilibrium by binding two residues to iridium(III). 117

Figure 5.13. Equilibration of the peptide was rapid and reasonable structures were discovered within only 8 iterations of QM/DMD. 118

Figure 5.14. Several independent QM/DMD runs showing convergence in the binding mode of iridium(III) to the structure. 119

Figure 5.15. (Left) A theorized model of the tripeptide complex used by the Hilvert lab for a transfer hydrogenation reaction and (Right) a transition state complex for the same reaction, determined experimentally using only ligands found in proteins. 119

Figure 6.1. All divalent metals in this study coordinated by EGTA and DTPA, for use as a reference in determining metal binding affinities to HDAC8. Shaded areas surrounding each complex represent the solvent accessible surface area and highlight the solvent exclusion of these chelating complexes. 166

Figure 6.2. Catalytic model utilizing a two-state ensemble for $K_{binding}$ over a range of possible binding affinities for Co. In this model, $Kbinding(M) = e - (\Delta GCo + \Delta\Delta GM)/RT1 + e - (\Delta GCo + \Delta\Delta GM)/RT$, where ΔGCo is the binding affinity of Co and $\Delta\Delta GM$ is the binding affinity of metal M relative to Co. 167

VITA

2007-2008	Associate of Science Cuyahoga Community College Cleveland, OH
2008-2012	Bachelor of Science, Chemical and Biomolecular Engineering, With Honors in Engineering The Ohio State University, Columbus, OH
2012-2017	Theoretical and Computational Chemistry graduate student at UCLA
2016	NSF Supplement: ”Non-physiological metals and their clusters as active sites in artificial metalloenzymes”. 2016

CHAPTER 1

Computational Treatment of Metalloproteins

1.1. INTRODUCTION

Metalloproteins present a considerable challenge for modeling, especially when the starting point is far from thermodynamic equilibrium. Examples include formidable problems such as metalloprotein folding and structure prediction upon metal addition, removal, or even just replacement; metalloenzyme design, where stabilization of a transition state of the catalyzed reaction in the specific binding pocket around the metal needs to be achieved; docking to metal-containing sites and design of metalloenzyme inhibitors. Even more conservative computations, such as elucidations of the mechanisms and energetics of the reaction catalyzed by natural metalloenzymes, are often nontrivial. The reason is the vast span of time- and length-scales over which these proteins operate, and thus the resultant difficulties in estimating their energies and free energies. It is required to perform extensive sampling, properly treat the electronic structure of the bound metal or metals, and seamlessly merge the required techniques to assess energies and entropies, or their changes, for the entire system. Additionally, the machinery needs to be computationally affordable. Although a great advancement has been made over the years, including some of the seminal works resulting in the 2013 Nobel Prize in chemistry, many aforementioned exciting applications remain far from reach. We review the methodology on the forefront of the field, including several promising methods developed in our lab that bring us closer to the desired modern goals. We further highlight their performance by a few examples of applications.

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.

1.2. METHODS FOR METALLOPROTEIN TREATMENT

1.2.1. 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 1). DMD is a flavor of MD that approximates the continuous interaction potentials in classical MD with square-well potentials, (Figure 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 1A). Complete details can be found in earlier works.[43] (Review of DMD by Liz – add one more sentence about DMD) 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 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 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 1.1B). 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.

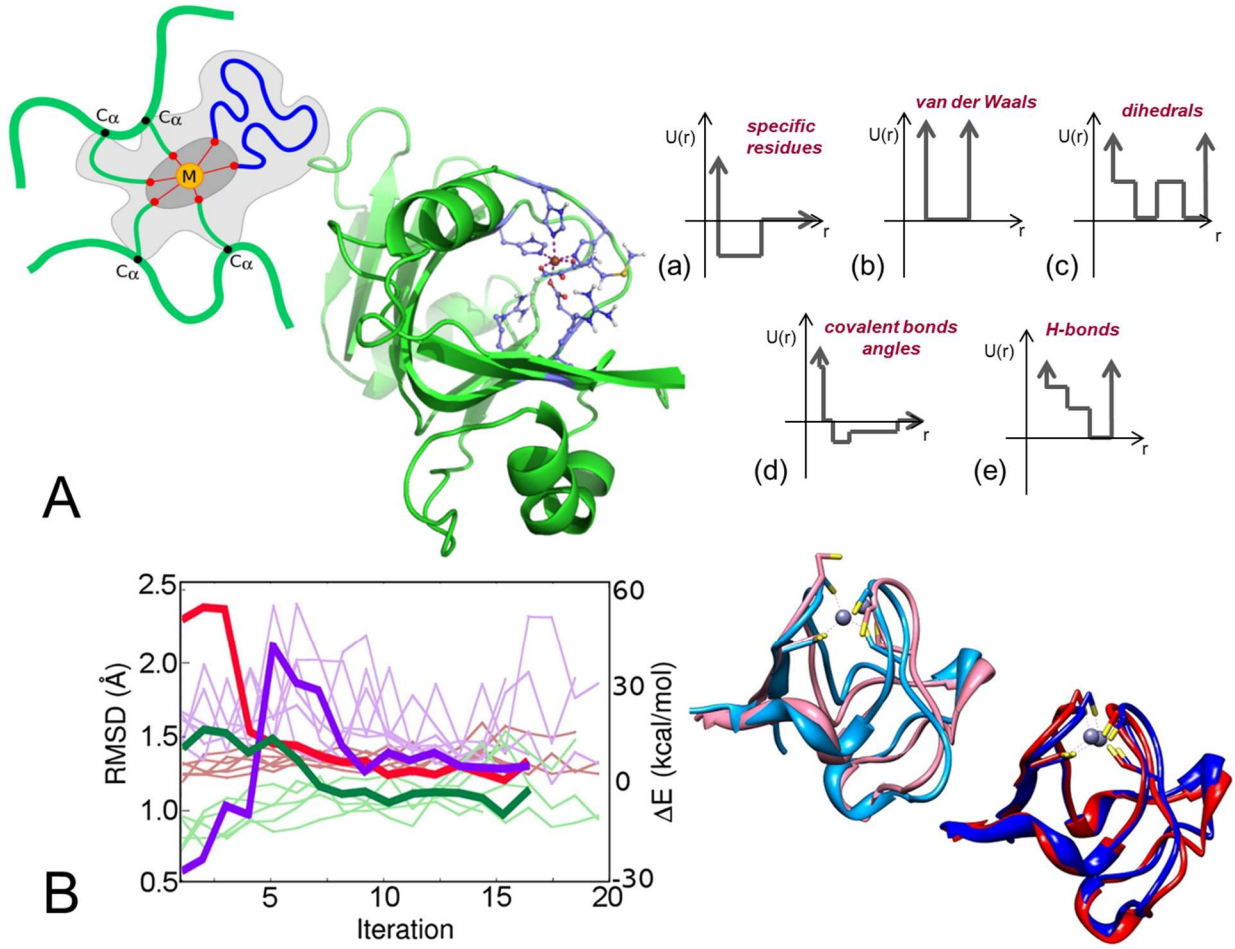


Figure 1.1. QM/DMD. (A) The unique feature of QM/DMD is the "breathing" QM-DMD boundary: the light grey area on the scheme is managed in both the QM and DMD regions. The alpha carbons of the "breathing" residues are held frozen during the pure QM phase, and the atoms directly coordinated to the metal (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.

1.2.2. 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.

1.2.3. 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 2A was found in the natural protein in Figure 2B. The *Erebus* search is based on sub-graph isomorphism algorithm with a user-defined allowed structural uncertainty, as illustrated in Fig. 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.

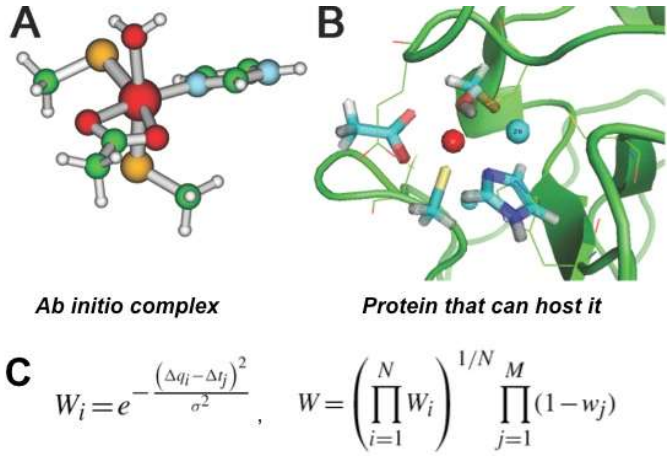


Figure 1.2. Illustration of Erebus. (A) A metal-ligand complex calculated using DFT. Hypothetically, this complex is desired for catalysis inside a protein pocket. (B) The active site of di-Zn β-lactamase found by Erebus to contain the Asp, two Cys, His residues, and a water molecule in the geometry close to the one desired from (A). Notice that it is already a Zn-binding site. (C) Wi represents the accuracy of a match to the input scaffold, more specifically Wi 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.

1.2.4. 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)

1.2.5. 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 3).[68]

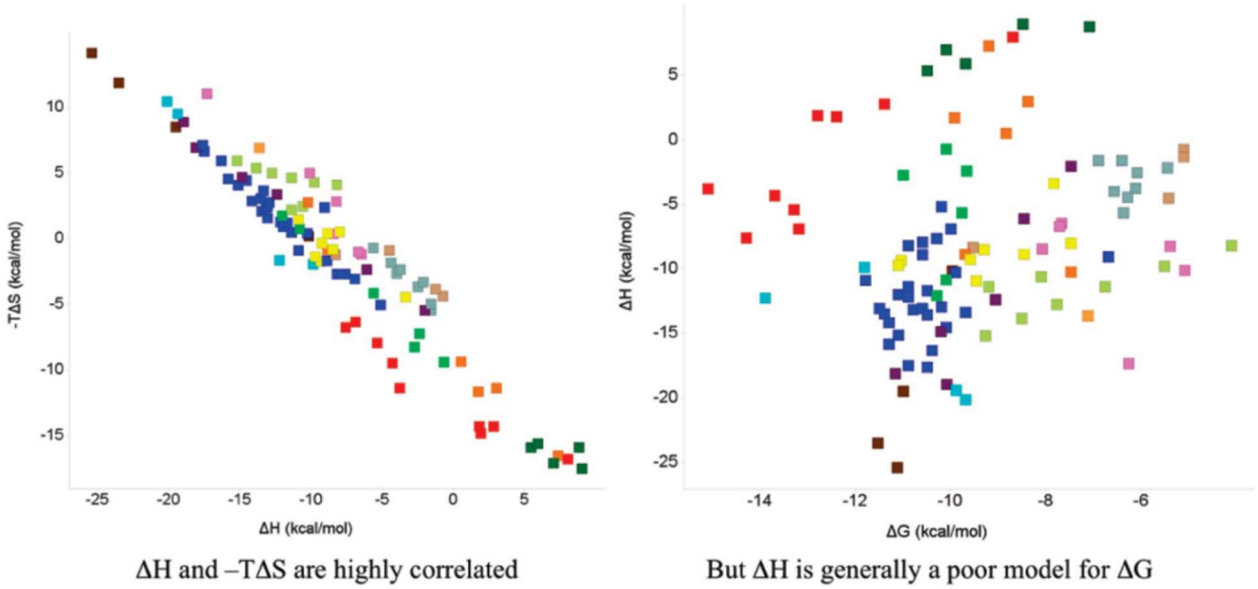


Figure 1.3. Analysis of the experimental binding thermodynamics for approximately 100 protein-ligand complexes. Reprinted with permission from Reference [68].

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.

1.3. EXAMPLES OF APPLICATIONS

1.3.1. LARGE-SCALE MOTION IMPORTANT FOR ENZYME 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 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

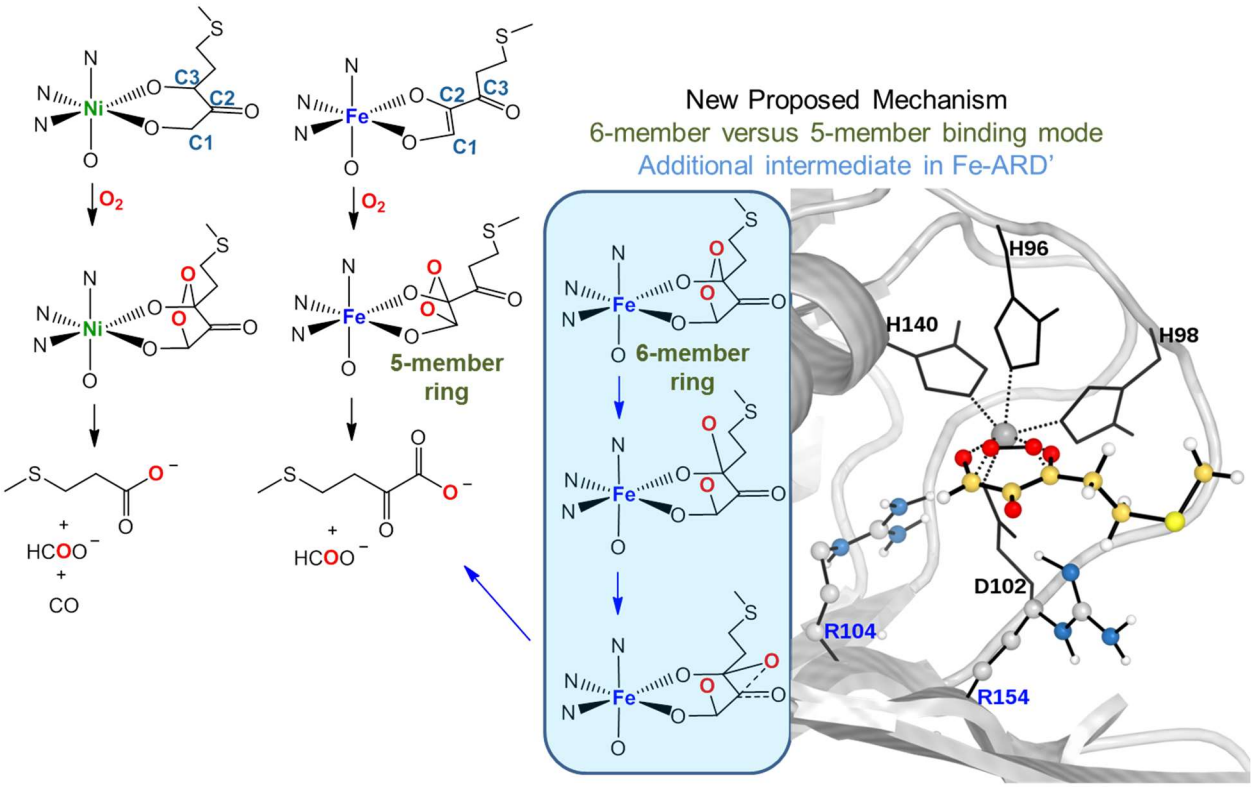


Figure 1.4. The old hypothesis on product differentiation of Fe-ARD' and Ni-ARD was disproved through our QM/DMD simulations coupled with DFT mechanistic studies and shown to stem purely from electronic structure of the metal, not the binding mode of the substrate to the metal. However, the residues 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 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 □* 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 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 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 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.

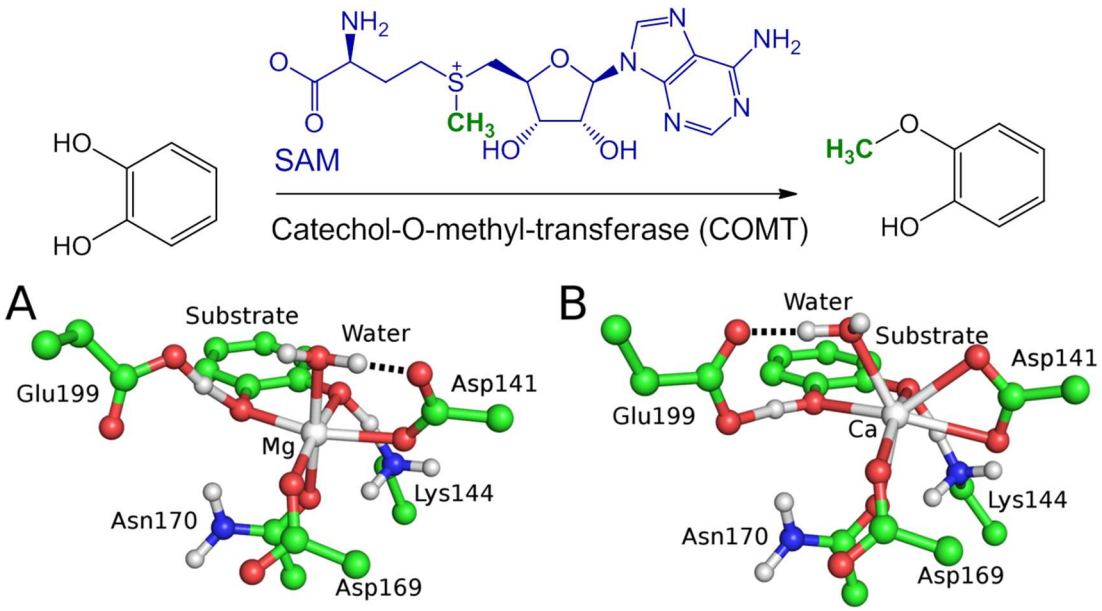
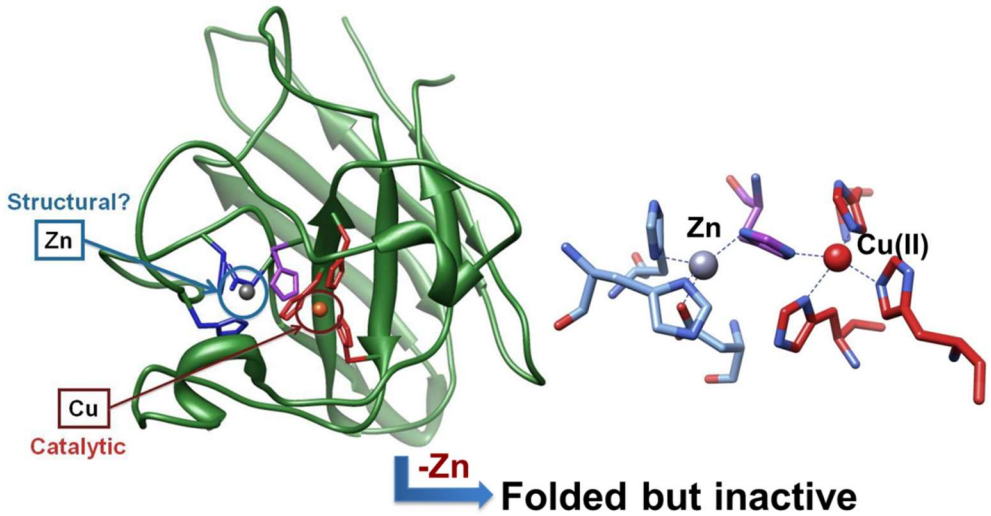


Figure 1.5. Catechol-O-methyl-transferase (COMT) catalyzes the methyl transfer from the cofactor SAM to a catechol motif found in neurotransmitters. (A) Native COMT binds Mg^{2+} , which positions the substrate in the proper orientation toward SAM, for methyl transfer. However, when Ca^{2+} (B) is bound, the active site distorts, putting reacting parts out of alignment, and thus leaving the enzyme inactive. Reprinted with permission from Reference 42.

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.



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

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

Figure 1.6. 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

1.3.2. 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 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. 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.

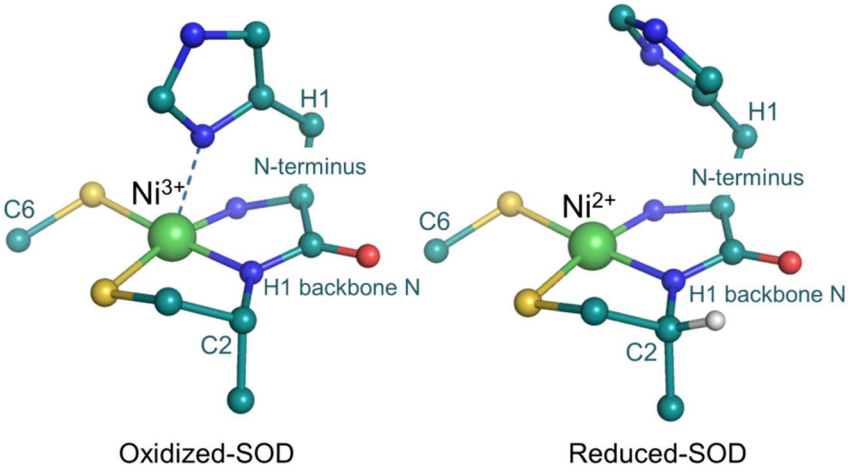


Figure 1.7. The redox reaction in Ni-SOD is modulated through the changing coordination environment of the histidine (H1) residue.

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 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.

1.3.3. 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.

1.4. 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.

CHAPTER 2

Histone Deacetylase 8:

Characterization of Physiological Divalent Metal Catalysis

2.1. INTRODUCTION

Histone Deacetylases (HDACs) are responsible for the removal of acetyl groups from histones resulting in gene silencing. Overexpression of HDACs is associated with cancer, and their inhibitors are of particular interest as chemotherapeutics. However, HDACs remain a target of mechanistic debate. HDAC class 8 is the most studied HDAC, and of particular importance due to its human oncological relevance. HDAC8 has traditionally been considered to be a Zn-dependent enzyme. However, recent experimental assays have challenged this assumption and shown that HDAC8 is catalytically active with a variety of different metals, and that it may be a Fe-dependent enzyme in vivo. We studied two opposing mechanisms utilizing a series of divalent metal ions in physiological abundance (Zn^{2+}, Fe^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+}, and Mg^{2+}). Extensive sampling of the entire protein with different bound metals was done with mixed quantum-classical QM/DMD method. Density Functional Theory (DFT) on an unusually large cluster model was used to describe the active site and reaction mechanism. We have found that the reaction profile of HDAC8 is similar among all metals tested, and follows one of the previously published mechanisms, but the rate-determining step is different from the one previously claimed. We further provide a scheme for estimating the metal binding affinities to the protein. We use the Quantum Theory of Atoms in Molecules (QTAIM) to understand the different binding affinities for each metal in HDAC8 as well as the ability of each metal to bind and properly orient the substrate for deacetylation. The combination of this data with the catalytic rate constants is required to reproduce the experimentally observed trend in metal-depending performance. We predict Co^{2+} and Zn^{2+} to be the most active metals in HDAC8, followed by Fe^{2+}, and Mn^{2+} and Mg^{2+} to be the least active.

The acetylation of lysine residues is an important reversible post-translational modification that modulates protein function, affecting a variety of cellular processes.[113-117] Proteomic surveys[118-120] have identified acetyl-lysine residues in diverse groups of proteins, including transcription factors,[121, 122] cell signaling proteins,[123] metabolic enzymes (most prominently acetyl-CoA synthase[124-126]), structural proteins in the cytoskeleton,[127, 128] and HIV viral proteins.[129, 130]

One of the first discovered examples of lysine acetylation was that occurring in histones,[131, 132] the predominant protein components of chromatin. Acetylation of histones has been linked to gene regulation: the addition of an acetyl moiety to histone lysine residues gives rise to an open chromatin structure that facilitates DNA transcription, while the removal of acetyl from histone acetyl-lysine residues is associated with closed chromatin structure, transcriptional repression, and gene silencing.[133] The enzymes responsible for the addition and removal of acetyl groups are known as histone acetyltransferases (HATs) and histone deacetylases (HDACs)[134-138] for historical reasons, although it is now recognized that some of them may act on non-histone proteins as well.[113, 115, 116]

Because overexpression of HDACs is associated with pathological states, including cancer, HDACs are of interest as drug targets.[136, 139-142] Inhibition of HDACs results in histone hyperacetylation and transcriptional activation of genes that are linked to growth arrest and apoptosis in tumor cells.[143] Suberoylanilide hydroxamic acid (SAHA), a HDAC inhibitor that coordinates directly to the catalytic metal ion of zinc-dependent HDACs, was approved by the FDA as an anti-cancer drug in 2006.[136]

HDACs are classified into four subtypes based on phylogenetic similarity.[144] Classes I, II, and IV are zinc-dependent hydrolases, while members of Class III are NAD^+-dependent and do not contain a catalytic metal ion. Of the zinc-dependent HDACs, histone deacetylase 8 (HDAC8)

is the best-characterized by experiment.[145-151] X-ray crystal structures of several HDAC8 enzyme-inhibitor and mutant-substrate complexes have been obtained.[145-148]

The crystal structure of the HDAC8 Y306F mutant, in complex with an acetylated lysine substrate (PDB accession code 2V5W),[147] illustrates the key residues involved in deacetylation (Figure 2.1.). The active site of HDAC8 contains a metal-binding center that is coordinated to one histidine (H180) and two aspartate (D178 and D267) residues. A water molecule, which is present in the crystal structure coordinates with the zinc and is associated with two additional histidine residues, H142 and H143. These histidines form dyads with the aspartates

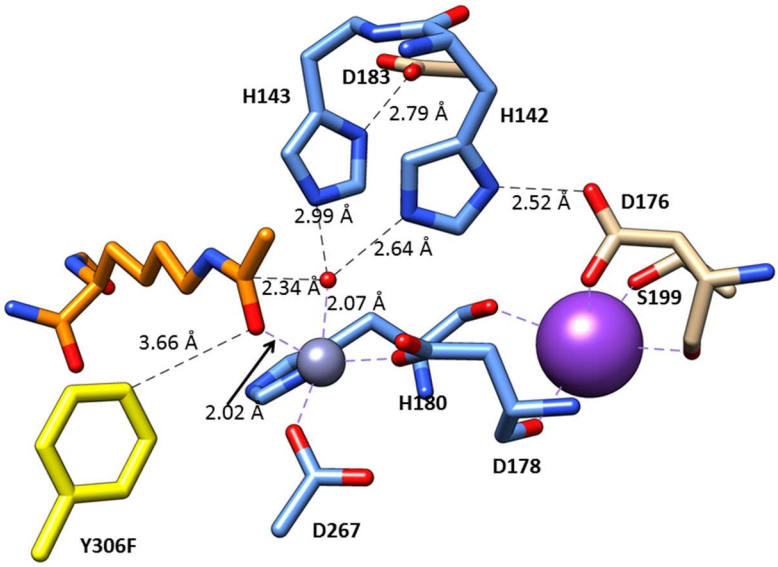


Figure 2.1. The active site of HDAC8 extracted from the crystal structure (PDB code 2V5W), highlighting the most critical residues relevant to catalysis within the active site. In this work, F306 was reverted to tyrosine to facilitate catalysis (yellow, Y306F). Blue residues participate directly in catalysis or coordination of the divalent metal ion, the orange residue is the crystal structure substrate, represented in purple is a nearby K ion surrounded by a small net of residues holding it in place.

D176 and D183, respectively. Given that the mutation of residue 306 from tyrosine to phenylalanine is sufficient to render the enzyme inactive, this residue must also play a role in the deacetylation process. In the crystal structure, the acetyl-lysine substrate appears coordinated to the zinc, with a carbonyl oxygen (R(Zn-O) = 2.02 Å).

Previously suggested reaction mechanisms for the deacetylation process include a general acid-base catalytic pair mechanism,[137, 152] and a proton shuttle catalytic mechanism.[153] These mechanisms were hypothesized based on previous studies of histone deacetylase-like protein (HDLP),[134, 154] a homologue of HDAC found in hyperthermophilic bacteria. In the general acid-base mechanism (Figure 2.2., mechanism (a)), a singly-protonated H142 acts as a base, abstracting a proton from the water molecule, while a doubly-protonated H143 acts as an acid, donating a proton to the substrate nitrogen.[134, 135] The acetyl group leaves in conjunction with nitrogen protonation. This mechanism relies on an unusual (at neutral pH) protonation state for H143 that is not supported by theoretical calculations.[153, 155] In the proton shuttle mechanism (Figure 2.2., mechanism (b)), H142 and H143 are initially singly protonated. H143 alternates between acting as a base and acting as an acid, first abstracting a proton from water and subsequently transferring it to the substrate nitrogen.[153]

While HDAC8 is of interest as a drug target inhibitor for pathological treatments, there has been a small amount of theoretical analysis surrounding the mechanism by which HDAC8 catalysis functions. Parallels have been drawn between the active site of serine proteases and HDAC8.[134] The nature of the active site has led to a variety of mechanistic speculations and hypotheses that have been examined theoretically.[153, 154] Most of these prior investigations utilized mixed quantum mechanical/molecular mechanical (QM/MM) methods to elucidate HDAC8's mode of catalysis. Expensive ab initio Born-Oppenheimer molecular dynamics has been used,[157] but the computation expense was so high that the active site had to be significantly truncated, and the effect of the nearby potassium ion and potassium-binding pocket were included only as

an electrostatic correction post-sampling. The role K^+ plays in HDAC8 catalysis is a contentious matter. Theoretical investigations have concluded that this K^+ has a stabilizing electrostatic effect;[153] however, experimental studies have claimed potentially inhibitory effects.[150] These mechanistic studies have also sought to examine the impact of the

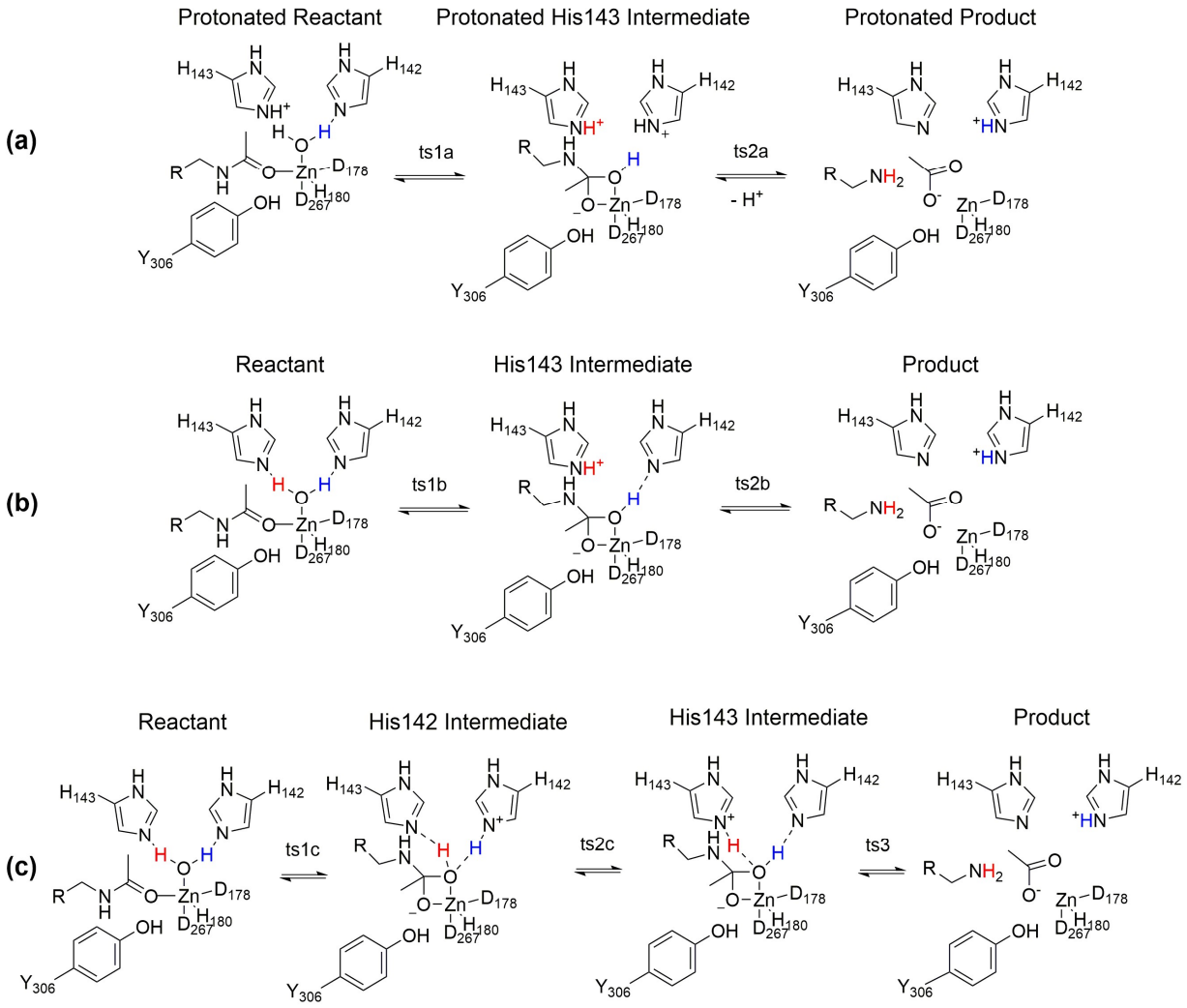


Figure 2.2. Schematic reaction diagrams for all previously considered HDAC8 mechanisms: (a) relies on the unusual initial protonation state of the coordinating His; reaction. Further, concerted addition to the carbonyl with transfer of proton to His142, as in (a), is found less favorable due to electrostatic effects of the nearby protein (as discussed in "Contributions Outside QM Subsystem" later in this paper). (b) was originally considered by the Zhang group,[153, 154] and reaction (c) was studied by Chen et al[172] as an

alternative to the Zhang mechanism. Reactions (b) and (c) are the mechanisms most studied in this paper. The relevant protons responsible for facilitating these mechanisms are highlighted in red and blue. Mechanism (a) was not heavily considered due to the requirement that His residues take on an unlikely protonation state at neutral pH.

His-Asp dyads that coordinate directly to an active site water likely involved in catalysis. Generally, it has been found that these dyad contacts are necessary to facilitate adequate basicity of the participating His residues. Computational studies thus far have centered exclusively on the hypothesis that HDAC8 functions as a native Zn^{2+} enzyme. However, Gantt et al[149] demonstrated that HDAC8 functions with a variety of physiological divalent metals, and suggested that HDAC8 may be an Fe^{2+} dependent enzyme, contrary to all published theoretical models, currently.

In this study, we examine how HDAC8 may function when utilizing several different divalent metal ions, Co^{2+}, Fe^{2+}, Zn^{2+}, Ni^{2+}, Mg^{2+}, and Mn^{2+}. Analyzing a variety of metals is important because it has previously been shown that different metals are capable of facilitating alternative enzymatic functions.[156] Our model relies on mixed quantum-classical description, and utilizes the largest published QM active site, which reveals significant differences in the catalytic profile from previously published theoretical results. We also examine the role that the nearby K^+ and its binding pocket, and the active site His-Asp dyads may play in catalysis. We will also discuss the capabilities of each metal to bind to HDAC8, and how that may influence apparent catalysis in vivo.

2.2. COMPUTATIONAL METHODS

The initial structure of HDAC8 was obtained from the crystal structure of the Y306F mutant (PDB accession code 2V5W)[147]. Residue 306 was transformed from phenylalanine to tyrosine using UCSF Chimera. The protonation states of all residues were chosen in accord with their pKa values at neutral pH, including singly protonated His180 on the δ site, and singly protonated H142 and H143 on the ε site.

2.2.1. QM/DMD SIMULATIONS

Different metal variants of HDAC8 might not necessarily have identical equilibrium structures, and therefore it was important to perform full statistical mechanical equilibration, with the quantum mechanical (QM) treatment of the metals, and the sampling of the protein backbone. Mixed quantum-classical simulations were carried out using QM/DMD,[157] a hybrid quantum mechanics (QM)/discrete molecular dynamics (DMD) method. This method provides the advantages of fast sampling of protein conformations without the need to rely on parameterization of the classical force field for metal ions. It has been shown to perform very well in the structural and mechanistic studies of natural and modified metalloenzymes.[156, 158-162] In QM/DMD, QM calculations are performed using density functional theory (DFT) as implemented in Turbomole v6.5.[163] Sampling in the classical regime is carried out using DMD,[155] a variant of a classical MD based on solving energy and momentum conservation equations rather than Newtonian equations of motion. The protein is partitioned into three regions: (1) a QM region, in which atoms can only be moved via a QM optimization (these are the metal and atoms immediately bound to it); (2) a mixed QM-DMD region, in which atoms can be moved by QM or by DMD, depending on the stage of the simulation (this region is discussed separately in the section below; it effectively constitutes a large cluster model of the active site); and (3) a DMD-only region, containing the

bulk of the protein. The DMD parameters for the QM-DMD region are adjusted on-the-fly, based on the QM calculations. Solvent effects are included implicitly in the force field of DMD. In the QM region, solvent has been modeled using COSMO[164] using a dielectric constant (ε) of 4 (in consideration of the buried active site), and, separately, non-polarizable point charges (AMBER force field) have been included. All zinc states had their electronic energy and COSMO corrections recomputed using ε of 8 and 20 - the resulting energy range was found to be within the expected error of DFT calculations (see Table 6.1.5. of Appendix).

Beyond QM/DMD iteration 0, which is a DFT optimization performed on the QM region of the initial structure, each iteration of QM/DMD consists of a DMD step followed by a QM step. To improve sampling, the DMD simulation begins with annealing at the start of each iteration. QM-only atoms are held fixed and a few additional constraints (such as the maximum allowed distance variation between histidine-aspartate dyads) are applied, in accord with previously reported recommendations.[157, 158, 160, 161, 165] After the annealing, the temperature is kept low for the final 10,000 DMD time units (0.5 ns), when the data – an ensemble of structures – is collected.

The Kabsch[166] RMSD is computed for all pairwise structures to quantify their geometric similarity. Based on the RMSDs, a hierarchical clustering algorithm is used to group the structures into distinct clusters. For this study, the number of clusters was set as 3. The lowest DMD energy structure within each cluster proceeds to the QM step. The QM step begins by expanding the QM-DMD boundary and cutting the QM-DMD region from the DMD-generated structures. Atoms at the QM-DMD boundary are capped with hydrogen atoms. To maintain the positions dictated by the protein backbone, the boundary atoms and the capping hydrogens are frozen for the duration of the QM step. DFT single point energy calculations are carried out using the TPSS functional[167, 168]. The def2-TZVPP basis set[169] is used for the metal atoms, while all other atoms are described by def2-SVP[169]. The Grimme dispersion correction[170] is included for all QM calculations. Structures are assigned a scoring index based on their deviation from the lowest energy QM

region and the lowest energy DMD structure found during the iteration. An ideal structure would have both the lowest QM energy and the lowest DMD energy, but the best structure found during each iteration may represent a compromise between the two. As in previous studies, the QM and DMD portions were weighted equally when assigning a scoring value. The QM region of the best-scoring structure is optimized under the same conditions described for the single point energy calculations, followed by removing the capping hydrogens, reinstallation of the QM region into the protein, and shrinking the QM-DMD boundary to the QM-only region. The resulting DMD- and QM-optimized structure is used as a starting structure for the next QM/DMD iteration. The simulations were propagated to convergence in terms of the active site all-atoms RMSD, backbone RMSD, QM and DMD energies, which typically was reached within 20-25 iterations, roughly corresponding to 10 ns of dynamics.

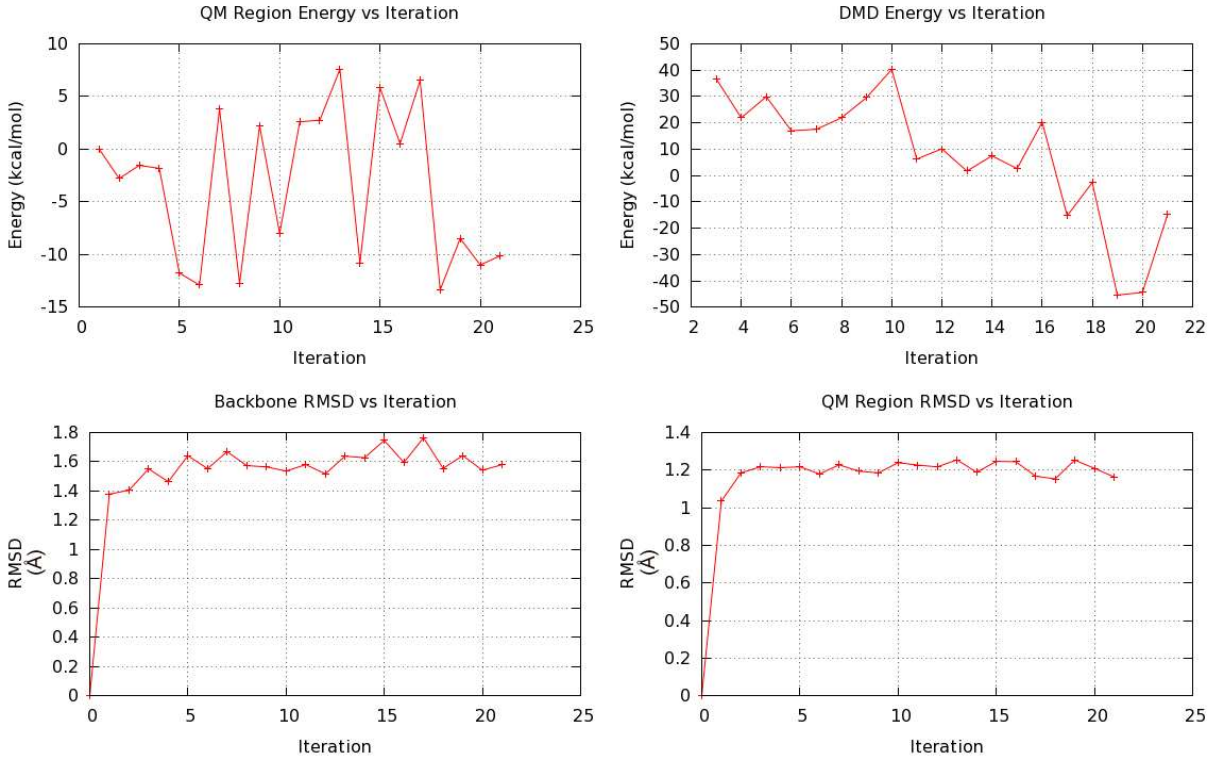


Figure 2.3. Typical convergence plots of QM/DMD runs. The top left graph displays the energy of the QM active site and the top right displays the DMD energy of the protein as a function of QM/DMD iteration

number, respectively. Note that sporadic high-energy structures would be discarded in the iterative procedure. The bottom left displays the convergence of backbone RMSD, while the bottom right exhibits the convergence of the active site RMSD. A levelled off RMSD graph indicates the lowest energy structures can be weighed based on Boltzmann statistics. This simulation with Zn^{2+} in the active site, and it is representative of all studied systems.

Figure 2.3. shows typical results of a QM/DMD run, specifically for HDAC8 with Zn^{2+} bound to the divalent site. Across all of the metals in this study, metal coordination largely remains similar as in the Zn^{2+} case. One notable exception is Ni^{2+} which assumes a larger coordination number; Ni^{2+} becomes over-coordinated by acquiring the second oxygen ligand of Asp178.

2.2.2. LARGE QM SUBSYSTEM REQUIREMENT

The choice of QM-DMD subsystem sizes was found to be critical, with the minimal models producing clearly wrong results. The smallest 112-atom QM-DMD subsystem shown in blue in Figure 2.4. was originally studied. This seemingly reasonable QM-DMD region appeared to be problematic, because upon optimization at the QM level (with any choice for the basis set and DFT method) the substrate acetyl oxygen dissociated from Zn^{2+} (R(O-Zn) = 3.09 Å), whereas in the crystal structure of the Y306F mutant this coordination is present. Over the course of 40 QM/DMD iterations, the substrate remained uncoordinated to the metal, and coordinated instead to Y306. Of course, it is possible that the Y306F mutation itself is responsible for the difference in the substrate binding, and the substrate detachment in the presence of Y306 is therefore mechanistically meaningful (note that the F306Y mutant is catalytically inactive). To rule out this possibility, we also simulated the Y306F mutant in the presence of the substrate. It was found that the substrate still dissociates from the zinc into an unreactive conformation (contrary to the crystal structure), thereby discrediting the choice for the smaller QM-DMD system. Interestingly, in previous studies, this effect was never reported. The implications for the reaction mechanism

must be profound, as the detached substrate is not a position conducive to catalysis. We find that a larger QM region is absolutely necessary.

The QM-DMD subsystem was then expanded to include a total of 165 atoms, as seen in Figure 2.4. Of particular importance is a conserved K^+ ion located within 7 Å of the active site that was investigated in a previous QM/MM study.[153] Single point calculations that included K^+ revealed stabilization of transition states. Furthermore, complexes lacking K^+ were found to have a longer zinc-substrate distance of 2.23 Å, compared to 2.13 Å in the wild-type structure. Thus, K^+ with its minimal coordination sphere was included in our QM system. We argue that without the coordination sphere, the electric field of K^+ would not be shielded, and thus would be exaggerated. All-together, the QM-DMD region consisted of the sidechains of H142, H143, D267, and Y306, the backbone chain spanning L177, D178 (with its sidechain), L179, H180 (with its sidechain), H181, G182, and D183 (with its sidechain), sidechain of S199, backbone of L200, sidechains of the dyad residues D183 and D176 connecting to metal-coordinating H142 and H143, respectively, and the additional G304 and G151. The backbone of the latter two amino acids provided hydrogen bonds to D178, and allow for its single-dentate coordination to the metal, which in turn left a place for substrate binding. Simulations done with this choice for the QM-DMD region showed a bonding distance of ca. 2.15 Å between Zn^{2+} and the substrate, which is very reasonable. We find this choice for the QM region justified and sufficient.

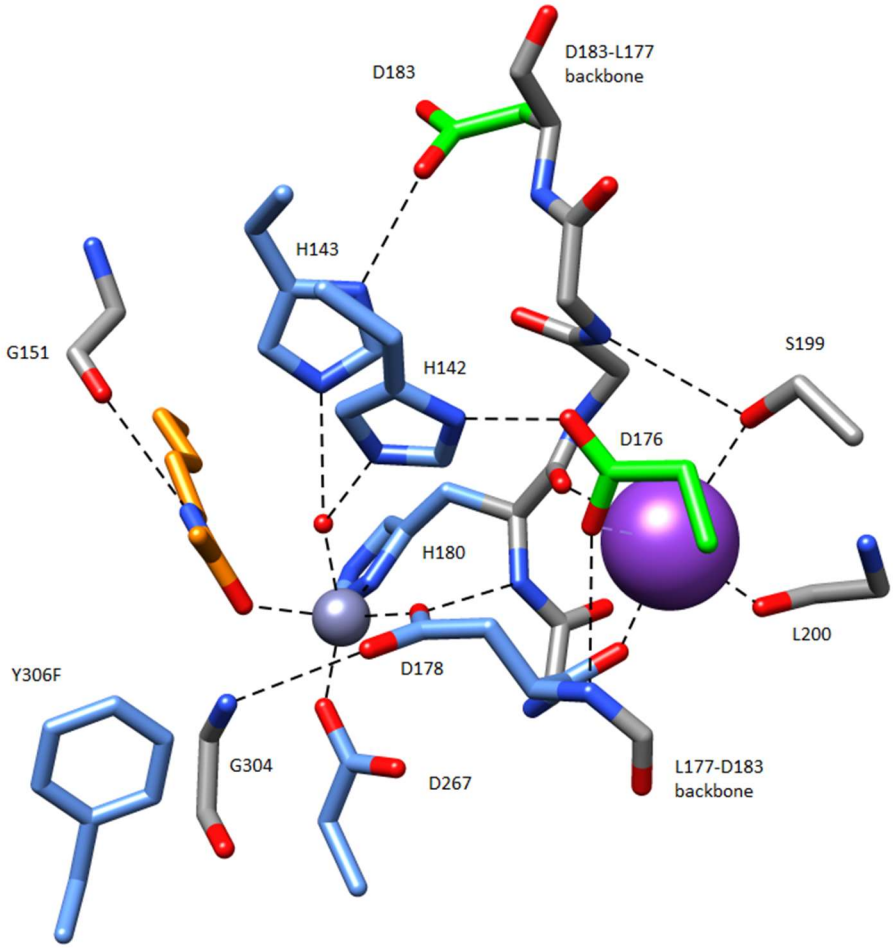


Figure 2.4. Diagram of residues included in the QM profile in previous, and the present studies. The bound substrate is shown in orange. Original work by the Zhang group[153, 154, 171] included residues colored in blue. The investigation by Chen et al.[60] included the blue residues, as well as the additional residues D176 and D183 in green. The current study expands on both of these by including additional residues colored grey, and the K^+ ion colored purple, as justified in the text.

We note that the choice for the QM-DMD region expands upon prior investigations significantly, and represents the largest studied QM active site. Previous investigations by the Zhang group[153, 154, 171] (Figure 2.4., blue) included H142, H143, D178, H180, D267, and Y306, but this provided an incomplete description of the dyad network attached to H142 and H143, whose

basicity is the key to the first step of the reaction. The complete QM description of this network was provided by Chen et. al.,[172] who included D176 and D183. However, in their work, the influence of the potassium ion was included only via single point calculations. Additional parts not previously considered to be of importance but now included are: G182, S199, G303, the carbonyl groups of G304, L200, G151, and the backbones spanning residues L177-D183 (Figure 2.4., grey).

2.2.3. MECHANISTIC STUDY

The QM-DMD region described in the previous section was also used for the subsequent QM-only mechanistic study. All calculations of the stationary points on the reaction profile have been optimized with Turbomole v6.5 using the TPSS functional and Def2-TZVPP basis set for all metals, and Def2-SVP basis sets for non-metal atoms. Spin-unrestricted calculations were performed on all active sites. Re-equilibration of all metals, except Zn and Mg, in low- and high-spin states was attempted. High-spin states were consistently preferred by >10 kcal/mol for all points along the reaction profiles, thus reaction pathways involving low-spin states were considered uncompetitive and are not included. Vibrational frequency analysis was done using the same level of theory, and the nature of all transition states was determined by the presence of a single imaginary frequency aligned with the reaction coordinate. Single point energy calculations were carried out using the larger Def2-TZVPP basis set for all atoms. We note that the size of the QM region necessary in this work is very large, and more expensive DFT calculations, for example hybrid functionals, are beyond reach in this context. Nevertheless, we choose to compromise in this way, in view of the identified chemical significance of the large cluster model.

2.2.4. ELECTRON CHARGE DENSITY ANALYSIS

The Quantum Theory of Atoms in Molecules (QTAIM) can be used to understand bonding interactions based on the topology and geometry of the charge density (ρ(r)).[173] Topological classifications can be made based on critical points (CPs), i.e. points where the gradient of the charge density vanishes. There are four types of stable CPs in a 3D scalar field such as ρ(r)). Maxima generally occur at atomic nuclei and are thus called nuclear CPs. Cage CPs occur inside atomic cages and are local minima in ρ(r). There are also two types of saddle points in ρ(r). The first, a ring CP, has positive curvature in two directions, and is topologically required inside rings of atoms. The other saddle point has one positive curvature and is called a bond CP. This name derives from the fact that the presence of a bond CP is indicative of a ridge of charge density originating at bond CPs and terminating at nuclear CPs. These ridges are called bond paths since they possess the topological properties imagined for a chemical bond. The magnitude of the charge density at bond CPs often correlates to the strength of interactions between atoms.[173-175] In this work, we used QTAIM to investigate ρ(r) for different metal variants, to elucidate the electronic origin of the calculated mechanistic differences.

The charge density for the reactant geometries of HDAC8 with all 6 metal ions was calculated with the Amsterdam Density Functional Package (ADF) version 2014.01[176-178] using similar computational parameters as in the QM mechanistic study. A TPSS functional[168, 179] and COSMO solvent model with a dielectric constant of 4.0 were utilized. 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.[180] Spin-unrestricted calculations were performed on the Zn and Mg active sites, while all other metals were calculated in their high-spin states. The Bondalyzer add-on package in Tecplot[181] was then used to analyze the calculated charge densities.

Born-Oppenheimer molecular dynamics (BOMD) simulations were done for the full DFT sampling of the complexes of the studied metals with organic chelators. This was needed for the assessment of the relative metal binding affinities to the HDAC8 protein, as described below.

2.3. RESULTS AND DISCUSSION

2.3.1. CATALYTIC MECHANISM

A step toward understanding the chemistry of HDAC8 is elucidating the nature of the catalytic metal ion. Most often it is assumed to be Zn^{2+}. Since the enzyme is functional with Zn^{2+}, this metal becomes a natural suspect, due to its abundance, low toxicity, flexibility of coordination,[182] and fast ligand exchange.[183] However, recent experimental evidence indicates that HDAC with other divalent metals exhibits similar efficacy, sometimes surpassing that of Zn^{2+}. It is particularly noteworthy that Co^{2+} and Fe^{2+} were found to be more active than Zn^{2+}.[37] In this work, we study the Zn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, and Mg^{2+} variants of HDAC8.

In QM/DMD simulations, all metals generally exhibited a distorted square pyramidal geometry. Representative structures were taken from QM/DMD trajectories for the mechanistic study (the corresponding geometries of the active sites are given in the Appendix 6.1.). We examined mechanism (b) and (c) (Figure 2.2.). While part of the mechanism (c) (ts1c) was found to be viable, tsc2 corresponding to the subsequent concerted proton shuttling was not found, thus further characterization of mechanism (c) was not pursued.

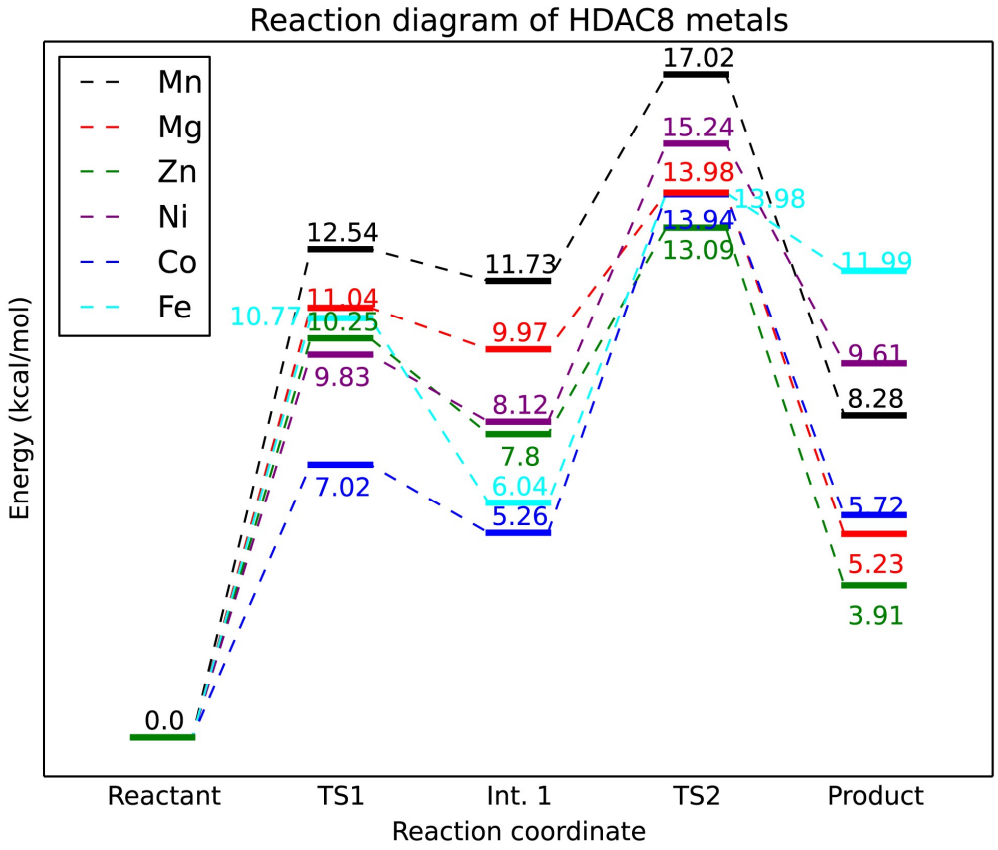


Figure 2.5. Reaction pathway for HDAC8 and a variety of divalent metal ions (TPSS/Def2-TZVPP+Gibbs free energy corrections//TPSS/Def2-DZVPP, COSMO).

Table 2.1. Computed reaction profiles with different methods and choices for the active sites.

	TS1		Int		TS2	
Full active site, TPSS/Def2-SVP + Gibbs free energy corrections, COSMO	**Zn^{2+}:**	**10.25**	**Zn^{2+}:**	**7.80**	**Zn^{2+}:**	**13.09**
	Fe^{2+}:	**10.77**	**Fe^{2+}:**	**6.04**	**Fe^{2+}:**	**13.98**
	Co^{2+}:	**7.02**	**Co^{2+}:**	**5.26**	**Co^{2+}:**	**13.94**
	Ni^{2+}:	**9.83**	**Ni^{2+}:**	**8.12**	**Ni^{2+}:**	**15.24**
	Mn^{2+}:	**12.54**	**Mn^{2+}:**	**11.73**	**Mn^{2+}:**	**17.02**
	Mg^{2+}:	**11.04**	**Mg^{2+}:**	**9.97**	**Mg^{2+}:**	**13.98**
Full active site, TPSS/Def2-TZVPP + Gibbs free energy corrections[a], COSMO	**Zn^{2+}:**	**11.42**	**Zn^{2+}:**	**10.45**	**Zn^{2+}:**	**16.01**
	Fe^{2+}:	**12.77**	**Fe^{2+}:**	**8.50**	**Fe^{2+}:**	**16.68**
	Co^{2+}:	**10.52**	**Co^{2+}:**	**7.43**	**Co^{2+}:**	**16.82**
	Ni^{2+}:	**11.50**	**Ni^{2+}:**	**10.71**	**Ni^{2+}:**	**17.75**
	Mn^{2+}:	**14.55**	**Mn^{2+}:**	**14.54**	**Mn^{2+}:**	**20.15**
	Mg^{2+}:	**13.35**	**Mg^{2+}:**	**13.22**	**Mg^{2+}:**	**17.47**
No K^{+}[b]	**Zn^{2+}:**	**11.38**	**Zn^{2+}:**	**8.52**	**Zn^{2+}:**	**17.00**
	Fe^{2+}:	**12.23**	**Fe^{2+}:**	**8.85**	**Fe^{2+}:**	**17.25**
	Co^{2+}:	**11.53**	**Co^{2+}:**	**8.07**	**Co^{2+}:**	**18.24**
	Ni^{2+}:	**11.58**	**Ni^{2+}:**	**9.56**	**Ni^{2+}:**	**18.26**
	Mn^{2+}:	**14.19**	**Mn^{2+}:**	**12.64**	**Mn^{2+}:**	**21.17**
	Mg^{2+}:	**12.85**	**Mg^{2+}:**	**11.01**	**Mg^{2+}:**	**18.34**
No K^{+} nor ligands coordinating it[b]	**Zn^{2+}:**	**9.98**	**Zn^{2+}:**	**5.25**	**Zn^{2+}:**	**13.74**
	Fe^{2+}:	**11.74**	**Fe^{2+}:**	**5.661**	**Fe^{2+}:**	**15.40**
	Co^{2+}:	**9.60**	**Co^{2+}:**	**4.60**	**Co^{2+}:**	**13.05**
	Ni^{2+}:	**11.37**	**Ni^{2+}:**	**7.77**	**Ni^{2+}:**	**16.12**

	Mn^{2+}: 12.86	**Mn^{2+}: 8.55**	**Mn^{2+}: 17.72**
	Mg^{2+}: 11.44	**Mg^{2+}: 7.45**	**Mg^{2+}: 14.83**
No K^+ site, and no D176 and D267[b]	**Zn^{2+}: 22.92**	**Zn^{2+}: 20.95**	**Zn^{2+}: 22.95**
	Fe^{2+}: 23.13	**Fe^{2+}: 20.45**	**Fe^{2+}: 19.75**
	Co^{2+}: 23.32	**Co^{2+}: 20.95**	**Co^{2+}: 22.95**
	Ni^{2+}: 24.10	**Ni^{2+}: 22.24**	**Ni^{2+}: 23.18**
	Mn^{2+}: 25.88	**Mn^{2+}: 23.19**	**Mn^{2+}: 25.40**
	Mg^{2+}: 25.11	**Mg^{2+}: 23.07**	**Mg^{2+}: 22.88**

[*] For all calculations, Def2-TZVPP basis set was used for all metals

[a] Gibbs free energy corrections (ZPE + entropic and thermal corrections) were calculated with the Def2-SVP basis set.

[b] Single point energies based on the frozen geometries of complete active sites, computed at the TPSS/Def2-SVP level of theory without any free energy corrections.

Mechanism (b) was found to be a viable catalytic pathway for all studied metals (Figure 2.5., Table 2.1.). The second transition state (ts2b) appears to be rate limiting. This result marks a pronounced deviation from previously published models, which claimed the first transition states, ts1b, to be rate-limiting. This may be an artifact of the QM region being too small in the earlier study. We further find that Zn^{2+} facilitates the fastest catalytic path, followed by Co^{2+}, Ni^{2+}, Mg^{2+}, Fe^{2+}, and Mn^{2+}. This is a notable contrast to experimental assays published by Gantt and coworkers[149] who found Co^{2+} and Fe^{2+} to be the most active, followed by Zn^{2+}, and little reactivity for Ni^{2+} and Mn^{2+}. Because the proton transfer step is rate limiting, it would be reasonable to expect some tunneling of the proton to contribute to the kinetics. A quantum tunneling correction factor to the rate constant, developed by R. P. Bell,[184] was calculated to be approximately 1.4,

thus it is expected tunneling has some noticeable effect on the kinetics, however, the identity of the metal does not significantly alter this correction.

2.3.2. WHY SUCH A DIFFERENCE FOR DIFFERENT METALS?

The main role of the metals in the reaction comes in the first step, in ts1b, where the carbonyl of the substrate is activated for nucleophilic attack. This determines both the activation barrier and the energy of the resultant intermediate, thereby shaping the entire reaction profile. We examine ts1b using QTAIM, to shed light on the electronic effects leading to the metal-dependent performance. The lower energy barriers for Zn^{2+}, Ni^{2+}, and Co^{2+} may be in part due to the higher number of occupied d orbitals than for Mn^{2+} (Mg^{2+} obviously lacking any). However, Fe^{2+} is an interesting outlier. The greater number of d electrons implies that orbitals higher up in the enzyme's molecular orbital manifold will be filled. In general, orbitals higher in the manifold are characterized by a greater number of antibonding orbitals and hence more interatomic nodal planes. It is orbitals of this type that stabilize topological rings and cages by contributing density along bond paths though not along nodal planes. This combination leads to deeper rings and greater curvature at bond CPs. In the plane of a ring of nuclei as shown in Figure 2.6., the ring point is a minimum in the charge density. In order for a ring point to exist in this region, a bond path must form between the water molecule and substrate carbonyl.

A bond path is present in the reactant state between the substrate carbonyl and water molecule when Zn^{2+}, Ni^{2+}, Fe^{2+}, or Co^{2+} is placed in the active site of HDAC8. This bond path is not present for Mg^{2+} or Mn^{2+}. While this bond is not indicative of a strong interaction (based on low values of charge density at the bond CP as well as the curved nature of the bond path), it still indicates a stabilizing interaction between the water molecule and substrate. It can be seen in mechanism (b) that when a proton is transferred from the water molecule to H143, a bond forms

between the carbonyl carbon and water oxygen. This bond has already begun to form in the reactant state for Zn^{2+}, Ni^{2+}, Fe^{2+}, and Co^{2+}, which facilitates the removal of a water hydrogen and ultimately lowers the activation energy of this reaction step. However, the correlation of this topological effect with the barrier height is not bulletproof, keeping in mind Fe^{2+}.

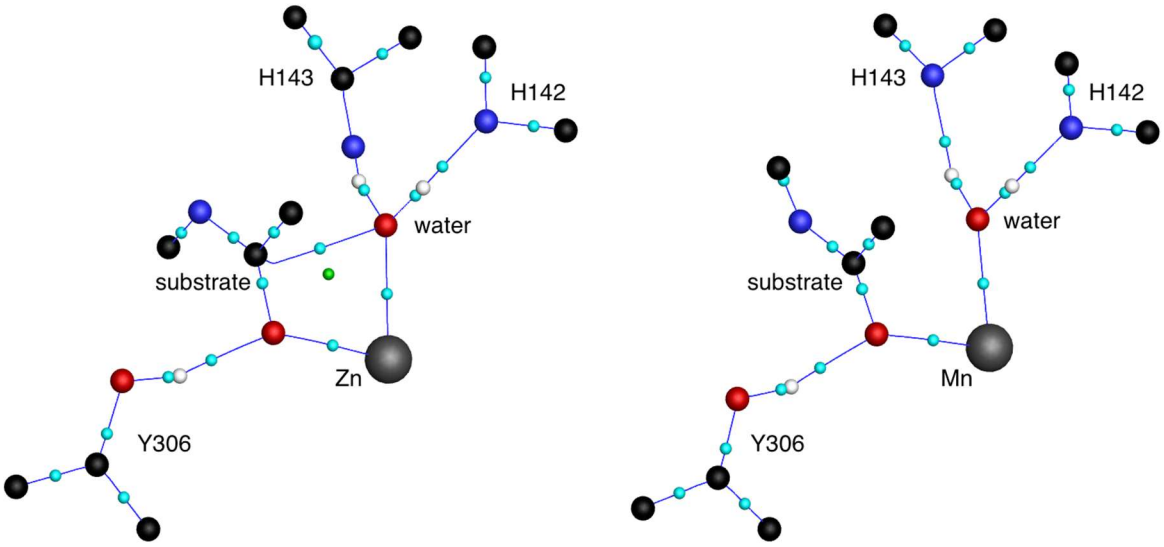


Figure 2.6. Critical points and bond paths of interest in the active site of HDAC8 with Zn (left) and Mn (right). When Zn is present in the active site a bond path forms between the water oxygen and the carbon atom from the substrate carbonyl. This topologically necessitates a ring critical point to exist in the active site as well. When Mn is present, no ring CP is found as there is not a bond path between the water and substrate carbonyl. Ni, Fe, and Co give the same topology as Zn, while Mg has the same topology as Mn. Sphere coloring is as follows: C-black, N-blue, O-red, Metal-grey, bond CP-cyan, ring CP-green.

2.3.3. ANALYSIS OF CONTRIBUTIONS OF PROTEIN PARTS ON REACTION PROFILES

The individual effects of various amino acids in the active site and their role in producing the computed energetics was analyzed. Specifically, we omit certain portions of the active site in the QM site and calculate the energies of the resultant structures without re-optimization. It is recognized that we are making an approximation that the nature of the stationary points would

not change significantly upon such changes. This decomposition analysis is highly qualitative, and the goal here is to put all the effects on similar footing with respect to the computational methodology, in order to be able to discuss their respective contribution. Results are summarized in Table 2.1.

2.3.3.1. ROLE OF POTASSIUM ION

Previous studies have suggested that K^+ may play an important electrostatic role in HDAC8 catalysis, but observations have been conflicting.[150, 153] It has been generally found that K^+ decreased the rate-limiting transition state barriers, and facilitated a more efficient catalytic path. A reaction pathway using QM active site in this work without K^+ calculated with a Def2-TZVPP basis set can be found in Table 2.1.

Within our model, the K^+ ion does influence the catalytic pathway, albeit to a lesser degree than previously thought. It has a mild stabilizing effect on the rate-limiting transition state: for Mn^{2+}, Co^{2+}, and Ni^{2+}, the stabilization is less than 0.5 kcal/mol, while for Zn^{2+} there is a destabilization by ~0.1 kcal/mol. Mg^{2+} remained relatively unaffected by the presence or absence of K^+ in its rate-limiting transition state. It was also found that K^+ had mild destabilizing influence on ts1b and Int1 with all metals, while having little impact on the product state. Prior investigations have reported more significant stabilizing effects on the rate-limiting step of approximately 4 kcal/mol.[153] Our findings suggest that the impact of potassium as a stabilizing electrostatic component may have been exaggerated in the literature. However, K^+, being so proximal to the active site, might play a critical role in facilitating general active site rearrangement during the reaction.

To additionally assess the role of the K^+ binding site as a whole, a significantly contracted active site was built; it contained the sidechains of the residues H142, H143, D178, H180, D267, Y306, D176, and D183 (numbering as in the pdb structure 2V5W), resembling the previously

published model by Chen et. al.[172] (Figure 2.4., blue and green residues). In this case, we found that K^+ brings significant stability to the system when the QM region is reduced. In general, the rate-limiting transition state was stabilized by an additional 3–4 kcal/mol. The ts1b was found to be stabilized by approximately 1–2 kcal/mol. Most apparent, the intermediate state was stabilized by ~5 kcal/mol for all metals, while the product state was destabilized by ~1 kcal/mol. These findings suggest that K^+ is an important component of the active site, but the effect of this ion needs to be considered in the presence of the shielding ligand environment.

2.3.3.2. ROLE OF THE CATALYTIC HIS-ASP DYAD CONTACTS

The D176 and D267 residues might play a role in positioning and polarizing H142 and H143, both of which were suspected to be catalytic bases in the reaction. The contributions the catalytic dyad contacts lend towards HDAC8 catalysis have been a matter of discussion.[172] An even more contracted QM site was constructed to characterize the effect of the H142-D176 and H143-D267 dyad contacts. This site excluded the D176 and D267 dyad residues. A visual representation of this region can be seen in Figure 2.4. (blue residues), and resembles some of the first published QM regions.[154] This system in our calculations concurs with current literature, demonstrating the critical importance of these dyad contacts. The His-Asp dyad residues provide exceptional stability to the catalytic pathway. In the absence of these residues, the positioning and protonation sites of H142 and H143 would be adversely affected. Additionally, their removal results in destabilization of the ts2b rate-limiting transition state on the order of ~5 kcal/mol, and would render HDAC8 extremely slow. However, more generous destabilization is afforded to many other states of the system, which alter the general catalytic landscape. Specifically, the intermediate state increases in energy by 10—12 kcal/mol. Most notably, the ts1b state increases in energy by ~12 kcal/mol for all metals, rendering this the new rate-limiting transition state for all metals, consistent with the results of prior theoretical investigations considering the same

mechanism. Though these effects are exaggerated without geometry optimizations, they still suggest that these dyad contacts are critical to the catalytic function of HDAC8. They most apparently serve to increase the basicity of H142 and H143 and provide stabilization of intermediate states via electrostatics and polarizing hydrogen bonds. These results also demonstrate that these dyad contacts are likely crucial for communicating with the nearby K^+ binding region. Although the K^+ binding pocket has an overall destabilizing electrostatic effect, it likely allows for configurational rearrangement of the Asp dyad contacts to better coordinate to their respective His partners. This in turn alters the entire landscape of the predicted catalytic path, shifting the rate-limiting transition state from ts1b to ts2b by increasing the basicity of these dyad His residues. The increased basicity however, means deprotonation of the His residue to the substrate will require higher energy. When these contacts are removed, the energetic landscape reverts back to a shape consistent with, albeit higher in energy, the earliest published theoretical work. Thus, HDAC8 balances electrostatic destabilization with favorable configurational rearrangement to produce the catalytic profile published here.

2.3.3.3. CONTRIBUTIONS OUTSIDE QM SUBSYSTEM

Early in the mechanistic studies published on HDAC8, the Zhang group showed significant electrostatic contributions of several nearby residues.[154] Most of the residues cited in the Zhang group's early paper have been explicitly included in the 165-atom QM subsystem presented here, sans a His-Glu salt bridge whose effects cancelled in the early paper. In order to test for further electrostatic contributions, COSMO solvation of the larger active site was replaced with a system of point charges derived from the AMBER force field,[185-188] as implemented in Nwchem.[189] The resulting point charges were transferred back to Turbomole for recalculation of all intermediates and transition states for the mechanism (b). Current implementations prohibit the combination of implicit solvation and electrostatic embedding, and thus we are forced to consider these effects

separately. This, however, does not alter our ultimate conclusions, which will be presented below. The mechanism was found to not change, but there are some important energetic differences summarized by the example in Figure 2.7.

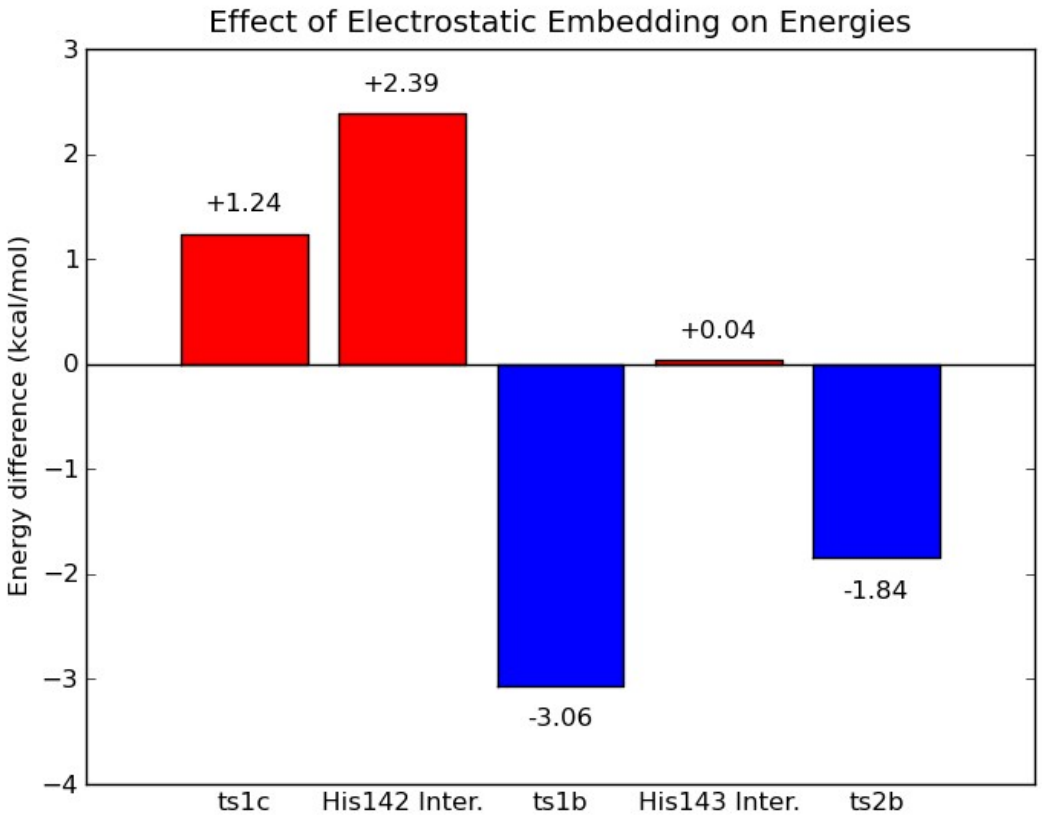


Figure 2.7. Highlights the difference in Gibbs Free Energies for the most active Co^{2+}-HDAC8, relative to the reactant state, upon electrostatic embedding. An increase in energy indicates embedding has a destabilizing effect, while decrease in energy indicated stabilizing effect of embedding. States ts1c and His142 Intermediate are only considered in mechanism (c), while ts1b, His143 Intermediate, and ts2b are considered in the current mechanism.

As Figure 2.7. shows, the electrostatic environment, treated as explicit non-polarizable point charges, favors mechanism (b), which starts with the proton shuttled directly to H143, while disfavors mechanism (c) where the proton first moves to the H142. Thus, the electrostatic

contributions from outside the already-large QM subsystem do not significantly alter the mechanism of action, however they still help direct the relative pathway.

2.3.4. METAL BINDING

We have also explored how metal binding affinity may affect the overall reactivity of HDAC8's deacetylating action. Gantt and coworkers have shown that several of the metal ions in this investigation were inactive in vitro,[149] however, many of the computed rate-limiting transition state energies found in this paper are well under the ceiling of catalytic inactivity. It is possible that the binding affinity of each metal towards HDAC8 could have an effect on in vitro catalytic activity, and may play an important role in in vivo metal selection. Here, we attempt to elucidate how significant this effect might be.

To this end, first, we have compared the accuracy of DFT to the experiment, in assessing metal binding, using the experimentally known stability constants of the complexes of the studied metals with EGTA and DTPA (Table 6.1.1. of Appendix).[190] From these stability constants, the free energies of reactions (1-4) in Figure 2.8. are known (Table 6.1.1. of Appendix). Further, from ΔΔG of reactions (1) and (2), one can close the thermodynamic cycle for DPTA on the upper left in Figure 2.8. Analogously, one can close the cycle for GEDTA on the upper right in Figure 2.8., using ΔΔG of reactions (3) and (4). Through combination of these two cycles, one can then calculate the ΔG of metal swapping between the two chelators,

$$\text{DTPA-}M_a^{2-} + \text{GEDTA-}M_b^{2-} \rightarrow \text{DTPA-}M_b^{2-} + \text{GEDTA-}M_a^{2-}$$

thus bypassing the complicated calculation of the solvated metal ions, $M_a^{2+}(aq)$ and $M_b^{2+}(aq)$. The ΔG of metal swapping were calculated both from the experimental data, and with DFT. The structures of the metal–EGTA and metal–DTPA complexes that we used were geometry-optimized and also subjected to 5 ps of BOMD, to verify the ground state geometries (they are

shown in Figure 6.1.1. of Appendix). To avoid the additional desolvation complications, polydentate ligation of the chelating agent was selected such that the bound metal had no solvent access, preventing explicit water coordination, and allowing for the exclusive use of COSMO. The theoretical and experimental ΔG of metal swapping can be found in Table 6.1.2. of Appendix. The computed ∆G values were found to be in good agreement with experiment, having an average of 0.5 ± 1.5 kcal/mol error (Table 6.1.3. of Appendex). Despite small absolute errors, there is a notable bias across substitutions involving Mn^{2+} - the authors concede this source of error may be due to some deficiency in the representation of the metal ion in silico or in describing its coordination mode.

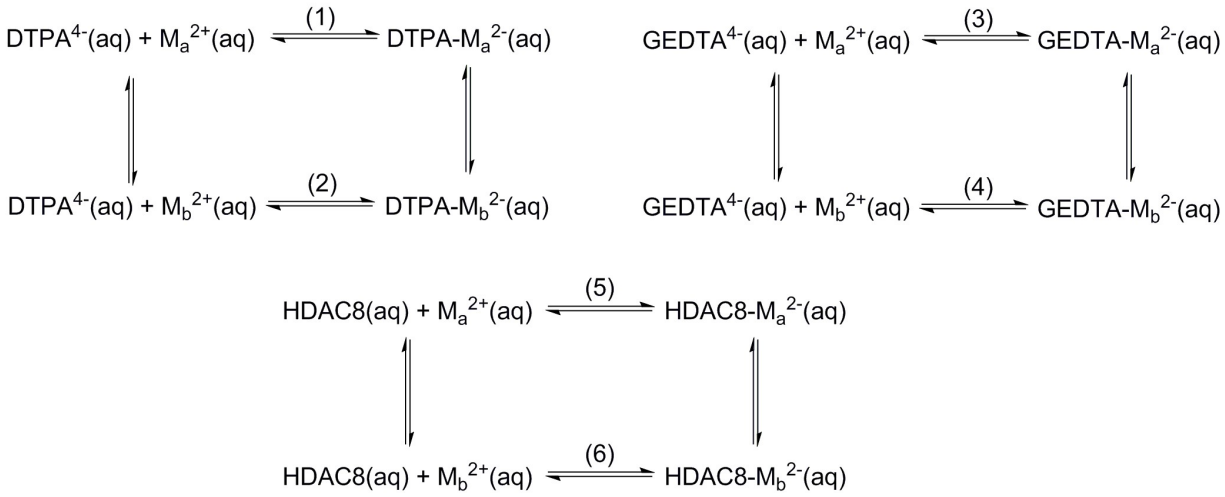


Figure 2.8. Schematic of all considered thermodynamic cycles exploited for relative binding affinities.

This methodology was then applied to a theoretical metal swap between DTPA and HDAC8:

$DTPA\text{-}M_b^{2-} + HDAC8\text{-}Co^{2-} \rightarrow DTPA\text{-}Co^{2-} + HDAC8\text{-}M_b^{2-}$

and using ΔG of reactions (1) and (2), and (5) and (6) in Figure 2.8., and with M_a set to Co^{2+} as a reference. The above reaction can be paired with the experimental ΔG of reactions shown in Table 6.1.1. of Appendix, to yield ΔG of the following process:

$M_b^{2+}(aq) + HDAC8\text{-}Co^{2-} \rightarrow Co^{2+}(aq) + HDAC8\text{-}M_b^{2-}$

This gives the desired relative affinities of the different metals to HDAC8, collected in Table 2.2.

Table 2.2. ΔΔG of binding between metal ions and HDAC8, relative to Co^{2+}.

Metal	□□G binding, kcal/mol
Co^{2+}	0.00
Fe^{2+}	3.52
Mg^{2+}	16.75
Mn^{2+}	6.28
Ni^{2+}	-4.33
Zn^{2+}	1.99

As one may see, Co^{2+} is predicted to bind to HDAC8 very strongly. It is followed by Zn^{2+} and Fe^{2+}. Mn^{2+} and Mg^{2+} have considerably smaller affinities for HDAC8. However, Ni^{2+} is calculated to have the highest binding affinity towards HDAC8 among all the studied metals. This ordering can be related to the active site geometries and corresponding electronic effects. A particularly direct interaction exists between the metal and H180, which is therefore highly important for metal binding. For this residue, the relationship between binding affinity and the amount of charge density at the Nε-metal bond CP is linear (Figure 2.9.). The charge density at the bond CP is a consequence of σ-bonding and π-back-donation with the available d-AOs on

the given metal. Mg, lacking any occupied d-AOs, has a significantly lower charge density at the bond CP and lower binding affinity than any of the transition metals, despite having a relatively short bond length (which is simply due to its smaller size). Thus, the geometric and electronic parameters of this Nε-metal bond can serve as predictors of the metal binding affinity in this case.

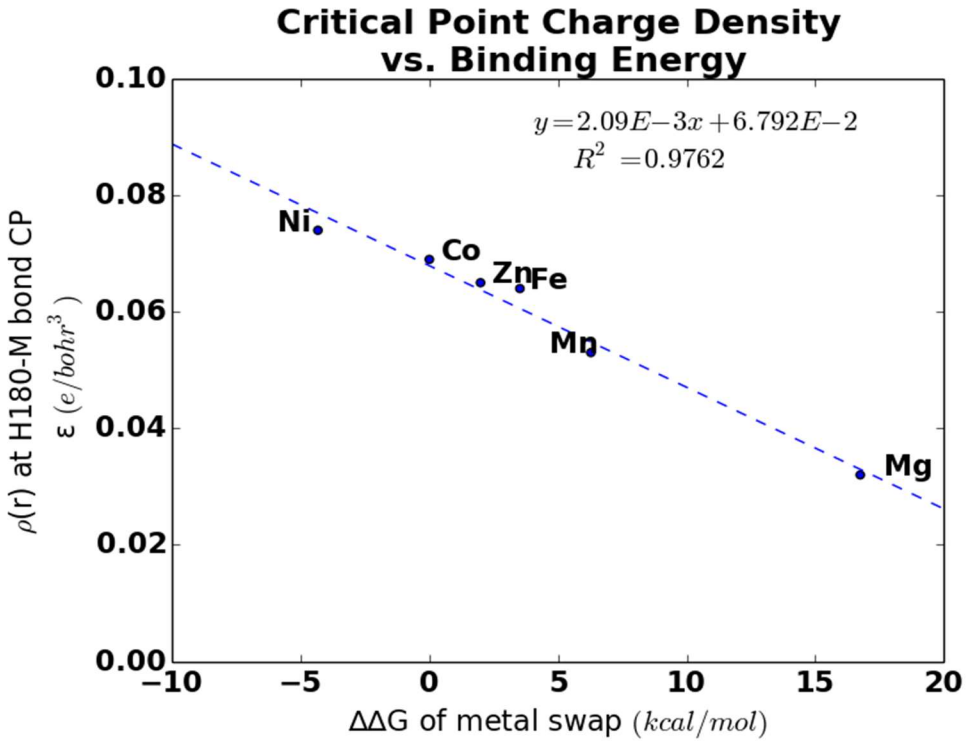


Figure 2.9. Trends between QTAIM Ɛ(r) at bond CP and ΔΔG of metal swapping.

2.3.5. METAL BINDING AFFINITY AND REACTIVITY

In protein assay experiments, the purified protein is often equilibrated in a high concentration solution of the desired metal ions. These ions are then purged from the mixture, leaving the protein-metal complex behind. In these conditions, the metal ions exist in equilibrium between protein and solution. If metal binding affinity is too small, this could render the system less active than demonstrated by computational studies. This difference in binding affinity could also play a key role in the overall catalytic activity of enzyme capable of utilizing different metals,

and elucidate its preference for any one particular ion. We now aim to evaluate the ability for the protein to retain a metal and subsequently use it for catalysis. Our simple probabilistic model captures the combined effects of metal-binding equilibration and subsequent catalysis relative to any particular ion. In the equation,

$$K_{rel} = k_{cat} \cdot K_{binding},$$

K_{rel} stands for the relative total catalytic activity of a particular protein-metal complex, k_{cat} represents the catalytic activity as determined by QM mechanistic analysis, and $K_{binding}$ embodies the relative equilibrium binding constant between a metal and an enzyme. These values can be represented as Boltzmann distributions utilizing the above computed quantities in catalysis and binding affinity to yield

$$K_{rel} = \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \cdot \exp\left(-\frac{\Delta\Delta G_{binding}}{RT}\right)$$

whose values are summarized in Table 2.3. An alternative two-state ensemble model for binding was also used to assess catalytic efficacy (Figure 6.1.2.), which agreed qualitatively with the current simple model for a reasonable range of binding affinities.

Table 2.3. Predicted relative total catalytic activities. Binding affinity energies were taken relative to Co^{2+}. Normalized values appear on the right column, relative to the predicted most active metal ion-protein complex.

	K_{rel}	Normalized	k_{cat} (s^{-1}) from experiment[37]
Ni^{2+}	1.89E-08	1.00	N/A
Co^{2+}	7.64E-11	4.03E-03	1.2
Zn^{2+}	1.27E-11	6.70E-04	0.90
Fe^{2+}	1.75E-13	9.24E-06	0.48
Mn^{2+}	1.37E-17	7.11E-10	N/A
Mg^{2+}	1.46E-23	7.71E-16	N/A

This model qualitatively predicts the trend seen in catalytic activity for these metals in experiment, though the trend is slightly perturbed,[149] either due to experimental errors, or inaccuracies of our model. Co^{2+} is predicted as the most active, followed by Zn^{2+} and Fe^{2+}. Additionally, non-catalytic metal Mn^{2+} is predicted as being 5 orders of magnitude less reactive than Fe^{2+}. This is due to both manganese's poor binding affinity to the active site of HDAC8 and higher ΔG_{rxn}. Zn^{2+} is predicted to be less catalytically active than Co^{2+}, despite its low reaction barrier, due to its poor affinity, while Co^{2+} exhibits high catalytic activity in spite of a higher reaction barrier. Fe^{2+} competes with Co^{2+} and Zn^{2+}, even though it favors a different rate-determining step and has a higher reaction barrier. However, Ni^{2+}, experimentally shown to be inactive, is a clear outlier, and is predicted to be exceptionally catalytic, due to high binding affinity. This may be an artifact of the simple model, since coordination of Ni^{2+} is different from those of other metals. Mg^{2+} examined in this work was not tested experimentally. We show that, even though it has

reasonable reaction energetics, its affinity to HDAC8 is very low, and Mg^{2+} is ultimately predicted by our calculations to have a very low catalytic activity. Thus, it is obvious that the d-AO structure is required in HDAC8, despite the simple Lewis acid catalysis performed by the metal in this enzyme.

2.4. CONCLUSIONS

We computationally studied the metal-dependent mechanism and performance of the Fe^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+}, and Mg^{2+} variants of the HDAC8 enzyme, in corroboration with and also beyond the existing experimental data. It was found that an unusually large QM region was required in both mixed quantum-classical dynamics simulations and subsequent mechanistic study (the QM region is nearly doubled as compared to all previous published studies), to properly describe this enzyme. This expansion has significant impacts on the overall catalytic ability of HDAC8, and predicts the rate-determining step of the reaction to be different from previous findings, in all cases but Fe^{2+} HDAC8. Instead of the nucleophilic attack, the second proton transfer step appears to be rate-determining. We have shown that the His-Asp dyads coordinating to the metal bound water are critical in facilitating catalysis, and defining the new rate-determining step. We also have shown that the nearby K^{+} ion, together with its coordination sphere has only a mild stabilizing electrostatic effect, in contrast to previously published results. We additionally devise a scheme for estimating the relative binding affinities of different metals to the protein. In combination with the reaction free energies, this quantity allows us to reproduce the experimentally observed trend in metal-dependent performance of HDAC8, with the exception of Ni^{2+}. Co^{2+} is predicted to be the most active catalytic metals in HDAC8, followed by Zn^{2+} and Fe^{2+}, whereas Mn^{2+} and Mg^{2+} are several orders of magnitude less active. We note that our model does not include any information about relative metal availability, and other reactivity in the cellular environment, which might further impact the apparent preference for different metal variants.

Supporting Information can be found in Appendix 6.1. including XYZ coordinates of all reactant states, literature binding energies and calculated “metal swap” energies, single point energies at different values of the dialectric value in COSMO, and a catalytic model utilizing a two-state ensemble for $K_{binding}$ over a range of possible binding affinities for Co.

CHAPTER 3

Computational Approach on Metal Substitution: Serum Transferrin

3.1. INTRODUCTION

Serum transferrin is one of a family of homologous iron-binding proteins (transferrins) that are commonly used throughout biological systems to control the level of free iron in biological fluids. However, serum transferrin is typically only about 39% saturated with iron[191] which allows it to bind to other metals such as the more strongly binding nonessential metal, Ti(IV).[192] Titanum is a particularly interesting metal that will be studied here that has also seen some 50 years of research as an anticancer therapeutic.

Like other transferrins, serum transferrin is a glycoprotein that consists of two major lobes which each contain a single high-affinity metal-binding site. The protein's metal site is adapted to bind ions with a large charge-to-size ratio, most notably Fe(III), with a synergistic anion, usually (bi)carbonate, and this binding is highly dependent on pH. The transport of a metal ion via serum transferrin starts with its binding to the transferrin receptor followed by receptor-mediated endocytosis and, notably for this work, releases the metal into the cell via acidification (by hydrogen ion pumps). The acidification opens the protein structure which triggers this release.

Titanium, primarily the dominant Ti(IV) oxidation state, is one metal that has received a lot of attention in literature due to its relative abundance in the body and anticancer properties. Although titanium is classically considered a nonbiological metal, its wide use in human applications (prosthetics, cosmetics, sunscreen, food, dye) has raised our exposure to the metal considerably. Also, our bodies retain much more of it than what would be expected based on its low solubility in water - blood serum concentration is 1-2 μM[193],[194] however, water solubility[195] is only 0.2 fM. Thus, the mechanisms of action in the body have been heavily disputed. Fe(III) and Ti(IV) have very similar ionic radii: 69 pm for low-spin Fe(III), 78.5 pm for high-spin Fe(III), 74.5 pm for Ti(IV) and due to Ti(IV) being a stronger Lewis Acid, so it is not surprising that it binds more strongly to iron transport proteins such as serum transferrin. Unfortunately, there are few tools which can easily probe Ti(IV), for several reasons.[196] For example, it is a d0 metal that is

invisible to standard spectroscopic and magnetic techniques (ultraviolet-visible spectrophotometry, electron paramagnetic resonance spectroscopy) and the nuclear magnetic resonance active isotopes have very low sensitivity and resolution.[197,198] Ti(IV) is thus an important target for computational work and metal transport via serum transferrin has previously been studied in silico with other metals.[199]

3.2. COMPUTATIONAL METHODS

We aim at modeling the relative binding energy and structural differences of the serum transferrin protein with a host of divalent to tetravalent metals, namely Zn(II), Fe(II), Ga(III), Co(III), Cr(III), Fe(III), Ti(IV). We used our recently developed metalloprotein sampling engine QM/DMD[200] (DMD standing for Discrete Molecular Dynamics[201]) methodology that includes both high level *ab initio* calculations of the protein site and efficient sampling of the three dimensional protein structure. Complete details on the QM/DMD machinery,[200] as well as another application to compute relative metal binding affinities by advantageously combining QM/DMD results with experimental data (for the case of HDAC8 catalysis),[202] can be found in previous work. QM/DMD has the ability to recapitulate native protein structure from native and distorted structures[200], capture large-scale motions of proteins,[203] and predict events such as ligand attachment[204] and detachment[205] so we believe it will give reliable results across the different metals and active site protonations studied in this work.

The initial structure of sTf consists of one of the two homologous lobes of the entire protein with a bound Al^{3+} present in the metal binding site. Residues were protonated to represent four different states of the protein and tabulated in Table 3.1. below. The primary differences between these structures are that two of the four structures are meant to represent the “open” form of serum transferrin (double/prtr) and two represent the “closed” form (acid/phys). Note that the open configurations are the form in which the metal is released and one tyrosine residue is replaced

with a water molecule. The closed forms have the standard coordination[206] of Fe^{3+} in transferrin (2 Tyr, 1 Asp, 1 His, bidentate CO_3^{2-}). Solvent was handled implicitly in DMD sampling and the most relevant waters for the area directly surrounding the metal were explicitly included in QM calculations and the remaining water treated implicitly via COSMO (COSMO stands for "COnductor-like Screening MOdel").

Table 3.1. Summary of substrate and protonation differences between otherwise identical structures of sTf studied.

Label	H_2O bound to M	Carbonate exists as:	LYS 204	LYS 294	TYR 186	HIS 117	HIS 117	HIS 271/ 287
prtr	✓	HCO_3^-	+	neutral	Not Metal-bound	+	neutral (δ)	+
double	✓	HCO_3^-	neutral	+	Not Metal-bound	+	neutral (δ)	+
acid		HCO_3^-	+	+	Metal-bound	+	neutral (δ)	+
phys		CO_3^{2-}	neutral	+	Metal-bound	neutral (ε)	neutral (ε)	neutral (ε)

3.2.1. FREE ENERGY OF METAL SUBSTITUTION

The eminent goal of this work is to expand on a previous successful calculation of the differences in free energy of binding a solvated metal to a protein (ΔΔG), as well as show structural differences between the different metal forms of the protein. The method predicting the ΔΔG of metal binding has been previously published by our lab in the context of enzyme catalysis of HDAC8[207] (also Chapter 2). In the aforementioned chapter, the calculated ΔΔG of metal binding between metal chelators was found to be accurate within 0.5-2 kcal/mol and ΔΔG of metals binding to the HDAC8 enzyme provided the missing link between activation energies of an HDAC8's catalytic cycle and the known experimental trend in catalytic activity.

The method used here to calculate ΔΔG only requires calculation of the ground state energy and vibrational/thermal corrections of metals separately bound to a chelator and to a protein molecule. Literature

stability constants of a metal binding from aqueous solution to the chelator can be converted into free energies and complete the thermodynamic equation for ΔΔG of metal binding, see Figure 3.1. below. This method worked well, in part, because literature reference data is used in place of the otherwise problematic step of modeling a solvated metal in silico.

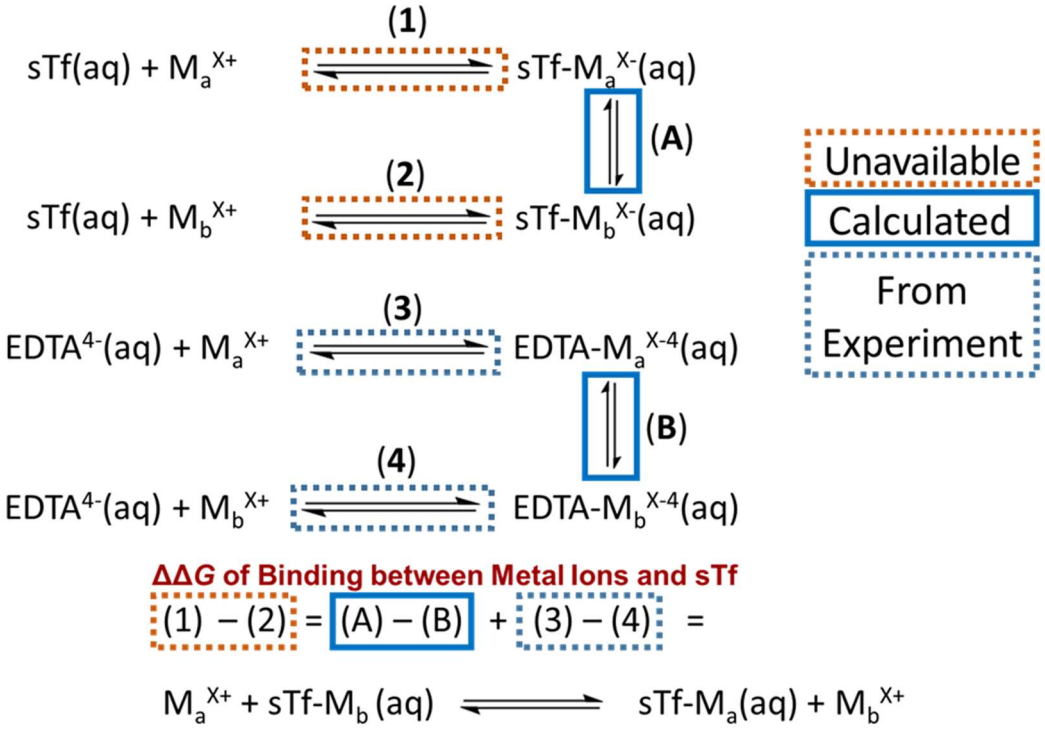


Figure 3.1. As a state function, differences in Gibbs free energy of a protein binding two metals from water can use a combination of just a couple DFT calculations with experimental stability constant data. This avoids various otherwise problematic computational challenges of modeling a metal in water.

Compared with the HDAC8 example, this work with serum transferrin will deal with a greater number of variables, namely more variety among the tested protein forms (several different protonation states) and in metal charge (divalent to tetravalent metals for sTrf, only divalent for HDAC8).

3.2.2. QM/DMD SAMPLING OF THE PROTEIN

5 replicate QM/DMD runs were run for a maximum of 40 QM/DMD iterations for every pairing of metal (Zn^{2+}, Fe^{2+}, Ga^{3+}, Co^{3+}, Cr^{3+}, Fe^{3+}, Ti^{4+}) and protonation state of protein (labeled: acid, prtr, double, phys – see Table 3.1. and Figure 3.2. below for a visual on the protonation states of the active site) A structure at the end of every iteration was collected and analyzed for structural data, and the lowest energy structures were passed onto full QM optimizations with vibrational corrections calculated and used in the subsequent metal binding energy analysis.

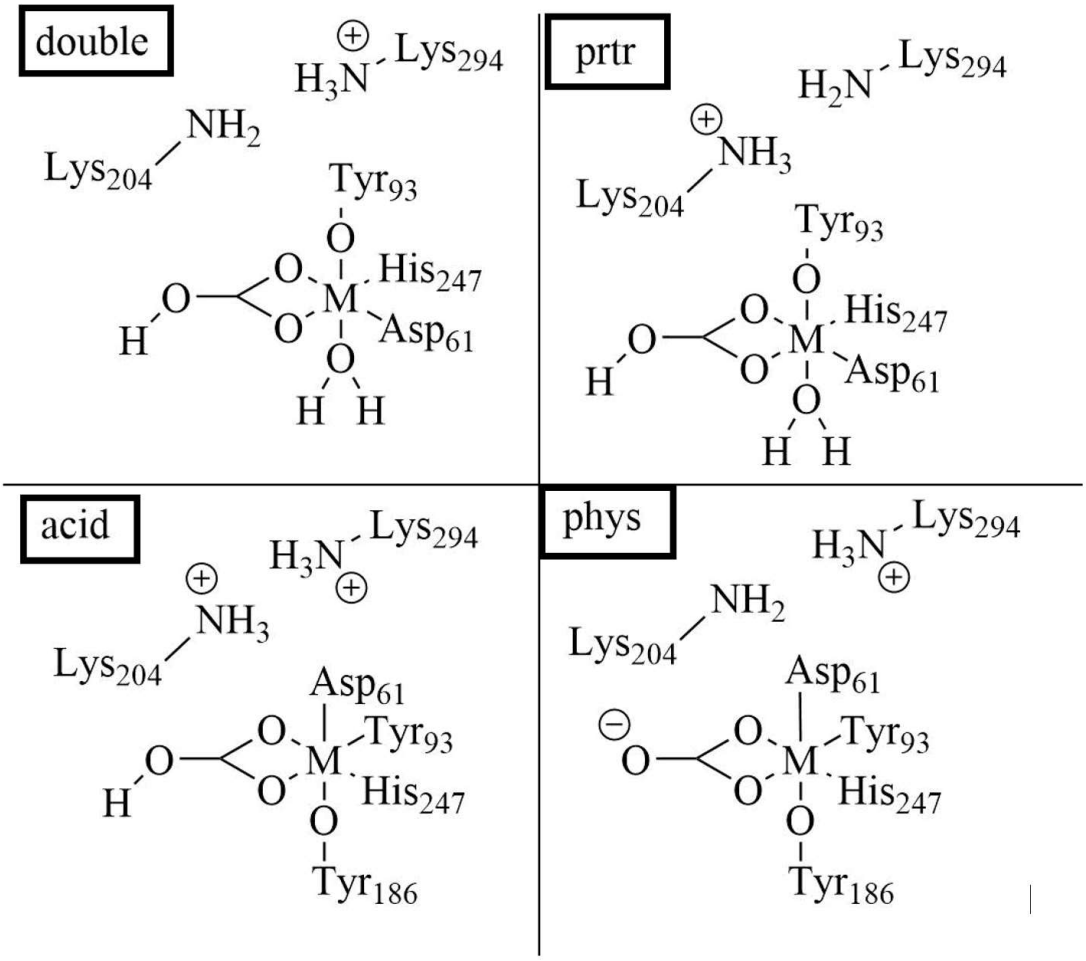


Figure 3.2. Different protonation states of the active site – note Tyr is deprotonated when attached directly to the metal

3.3. RESULTS AND DISCUSSION

3.3.1 METAL GEOMETRY

Metal replacement in a protein can have long ranging effects to protein structure and energy. Any changes would start at the core of the active site where different metals will take on different coordination geometries. Starting with the assumption that, like for Fe(III), serum transferrin and the synergistic anion provide an environment well suited for an octahedral geometry, we can calculate angle variance to determine the dispersion about 12 perfect 90 degree angles using the following formula:

$$\sigma_{\mathrm{oct}}^2 = \frac{1}{11}\sum_{i=1}^{12}\left(\theta_{\mathrm{i}} - 90^{\circ}\right)^2$$

The best 16 structures (based on an equal balance of DMD and QM energy) were selected from each metal/protein pairing and the distribution of angle variance for each group is plotted in Figure 3.3. below.

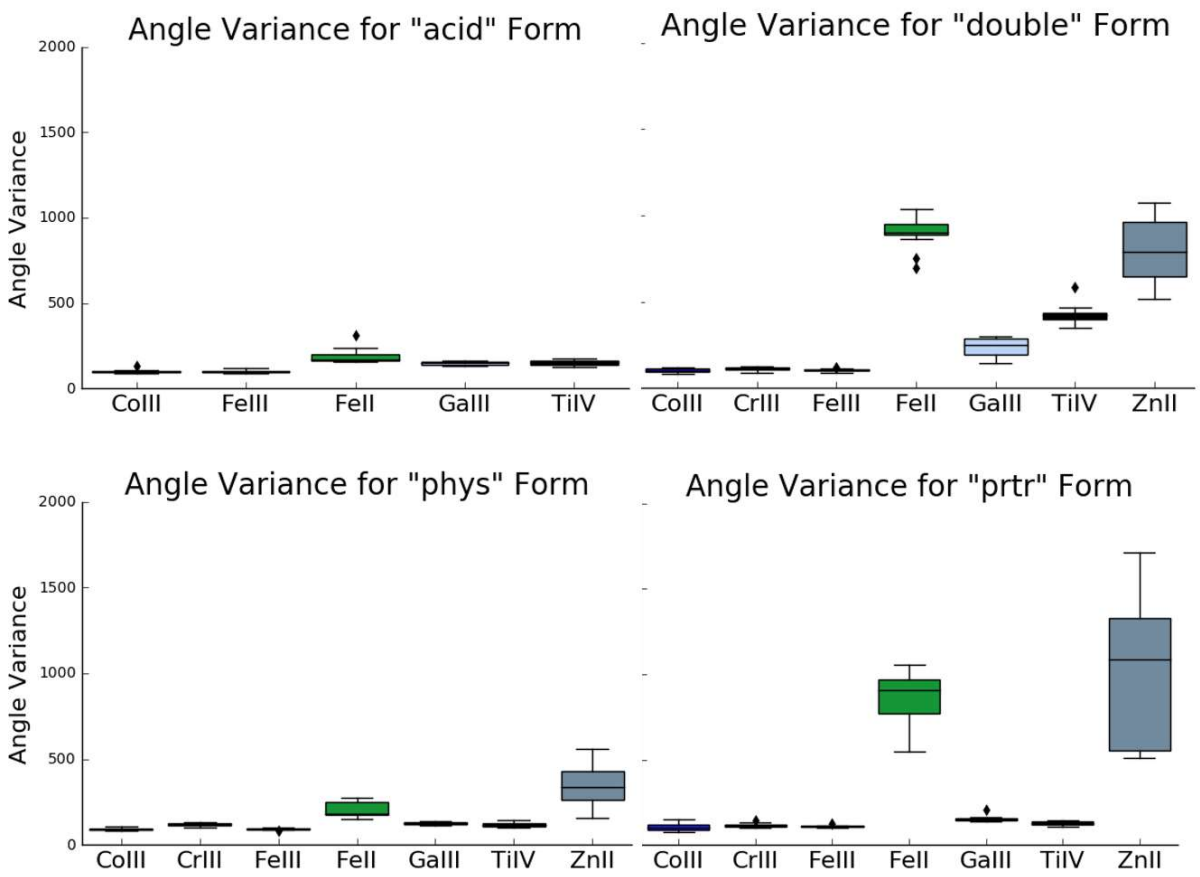


Figure 3.3. The distribution and outliers of octahedral angle variance for each metal in each protonation state of the protein. Several metals show significant deviation from the octahedral geometry. Acid forms of Zn(II) and Cr(III) omitted for now due to calculation error.

The outliers in this data are primarily Zn(II) and Fe(II), however, the behavior changes based on the protonation state of the molecule. For the “phys” and “acid” forms of the protein, these changes are minor with some structures close to octahedral but many lose one or more ligands (see Figure 3.4. below)

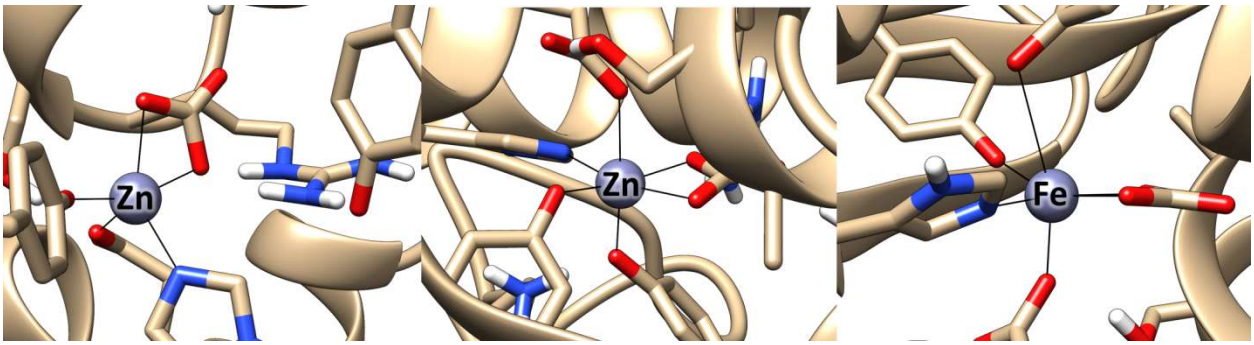


Figure 3.4. Three snapshots of the metal binding site of protonation state "phys". Zn^{2+} (left) and Fe^{2+} (right) lose a tyrosine during the simulation (middle is Zn^{2+} with octahedral geometry shown for reference).

The majority of angle variance in the protein's "double" form, particularly for Zn(II) and Fe(II), can similarly be described by the loss of the water molecule and monodentate bonding of the carbonate anion. In some snapshots these metals are found to have weakly bound octahedral forms, however the first coordination shell in the protein provides a charge of -3 when both tyrosine residues are bonded to a metal. Thus, the divalent metals Zn(II) and Fe(II) have the lowest charge out of the metals studied, easily shed ligands, and assume a tetrahedral geometry. Taking out the outliers, there are more subtle but decided changes in other geometries

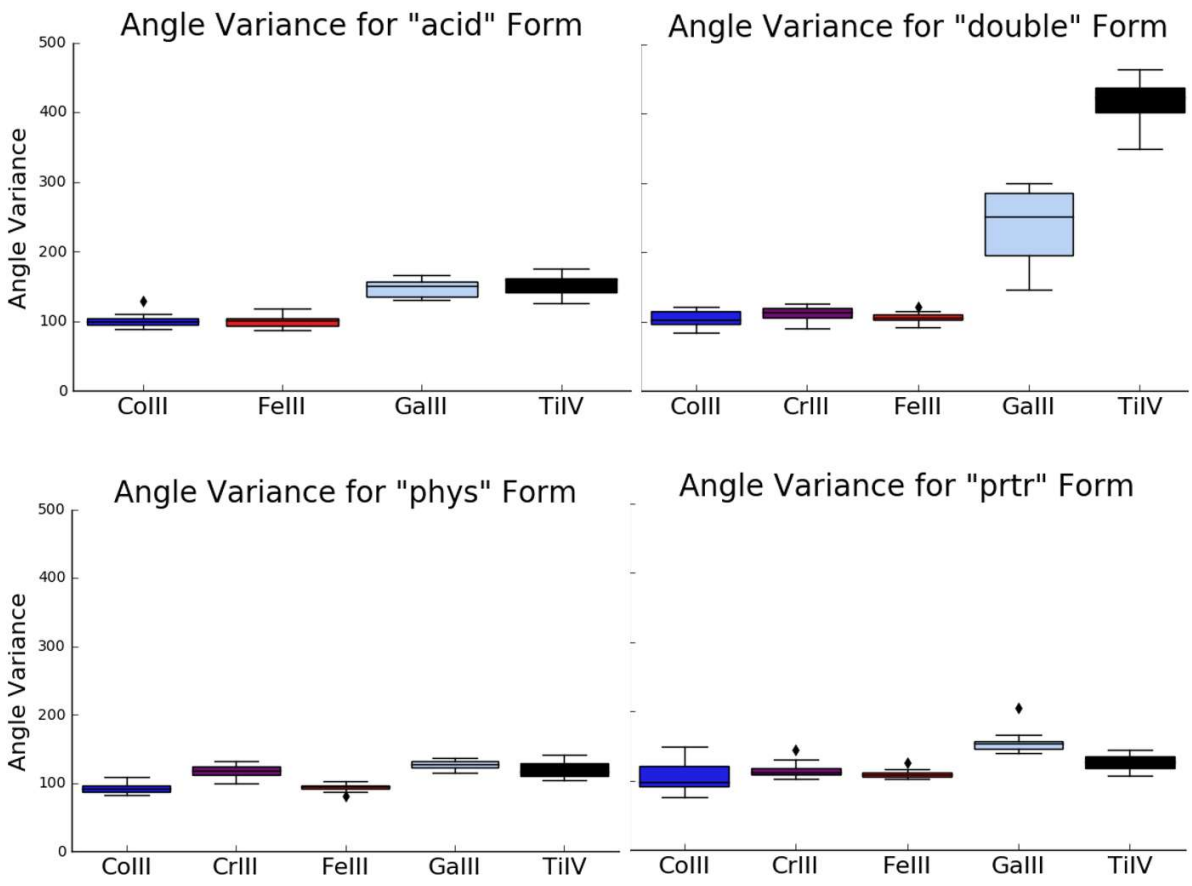


Figure 3.5. Finer details of Figure 3.3.

Ga(III), unlike to other trivalent metals, has a propensity to binds monodentate to the carbonate substrate especially in the “double” and “acid” forms. This is likely a result of the weaker overall binding character of Ga(III) which is known to be lower than Fe(III).

Ti(IV), with its stronger binding character than Fe(III), overcoordinates in the “double” form (shown in Figure 3.6. below), and, when it does not overcoordinate, typically presents octahedral or a distorted octahedral as shown in Figure 3.6.:

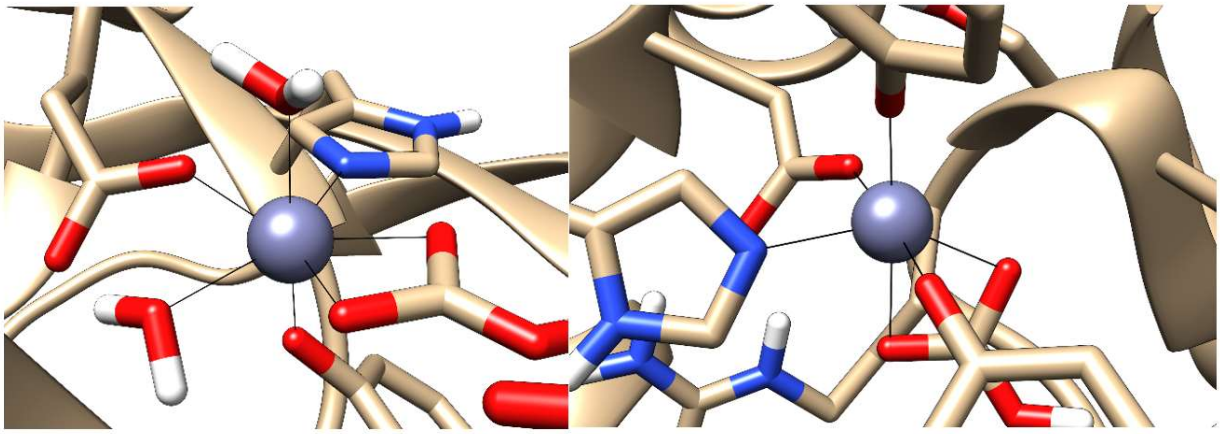


Figure 3.6. Ti(IV) overcoordinates in the “double” form (left) and in the other forms distorts from octahedral as shown (right).

3.3.2. LARGER STRUCTURAL DIFFERENCES OF SERUM TRANSFERRIN

In determining ΔΔG of metal binding (previously described and applied to HDAC8), large scale protein fluctuations in energy and entropy are assumed to be constant upon switching of the metal. When this assumption is reasonable, the issue of calculating free energy changes of a protein and balancing that energy appropriately with the free energy of a QM calculation is avoided. The flexible “open” state and the more tightly packed “closed” state of serum transferrin are two examples of conformations where this assumption would break down if a metal swap were to induce the change. One way we can quantify a change between the two states is to measure the distance between subdomains as illustrated in Figure 3.7. below.

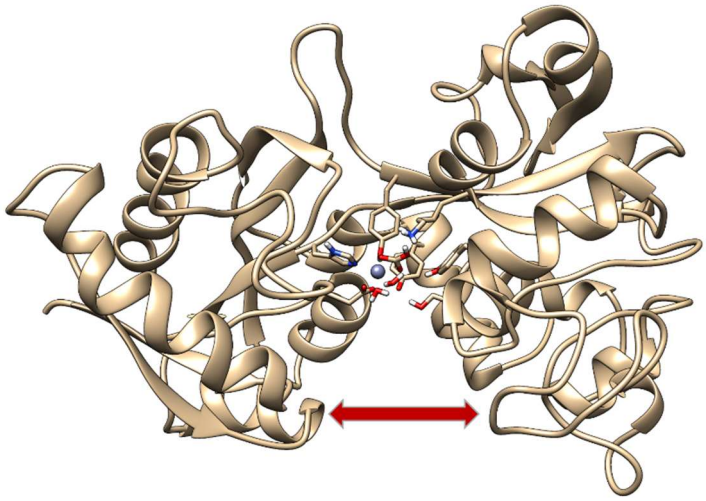


Figure 3.7. Tracking the distance between the subdomains throughout a simulation is just one way to track a conformational change from “open” to “closed”. The minimum difference between backbone atoms of residues 41-44, 11, 12, 288, 289 on one subdomain with 177-180 on the other (pointed to by the arrows here) is calculated for every structure and plotted across simulation time in Figure 3.8.

The two closed forms, “acid” and “phys”, show a tight packing of subdomain atoms which persists across all iterations and replicates for all metals. For the two open forms, all trivalent metals and Zn(II) similarly show some flexion in the subdomains but there was never a transition to the closed form across the simulation time. For both of the metals that showed coordination changes, the hairpin turn centered around residues 177-181 engages the turn centered around residues 286-190 by backbone hydrogen bonding of Lys 289 with the side chain of Thr 179. In the case of Fe(II), the loss of the water from the metal binding site leads to a slight repacking of the core of the protein which narrows the gap between subdomains but still leaves the metal binding site solvent accessible (see Figure 3.8. below). This slightly tighter closure of the structure was found to be preferred for Fe(II) on the basis of equal parts DMD and QM energies with 11 of the 16 lowest energy structures being in range of this interaction. For Ti(IV), the dispersion around octahedral geometry seems to allow more flexibility of the domains, however, in the lowest 16 energy structures of Ti(IV) bound to serum transferrim, only 1 of them displayed the aforementioned Lys 289/Thr179 hydrogen bond interaction.

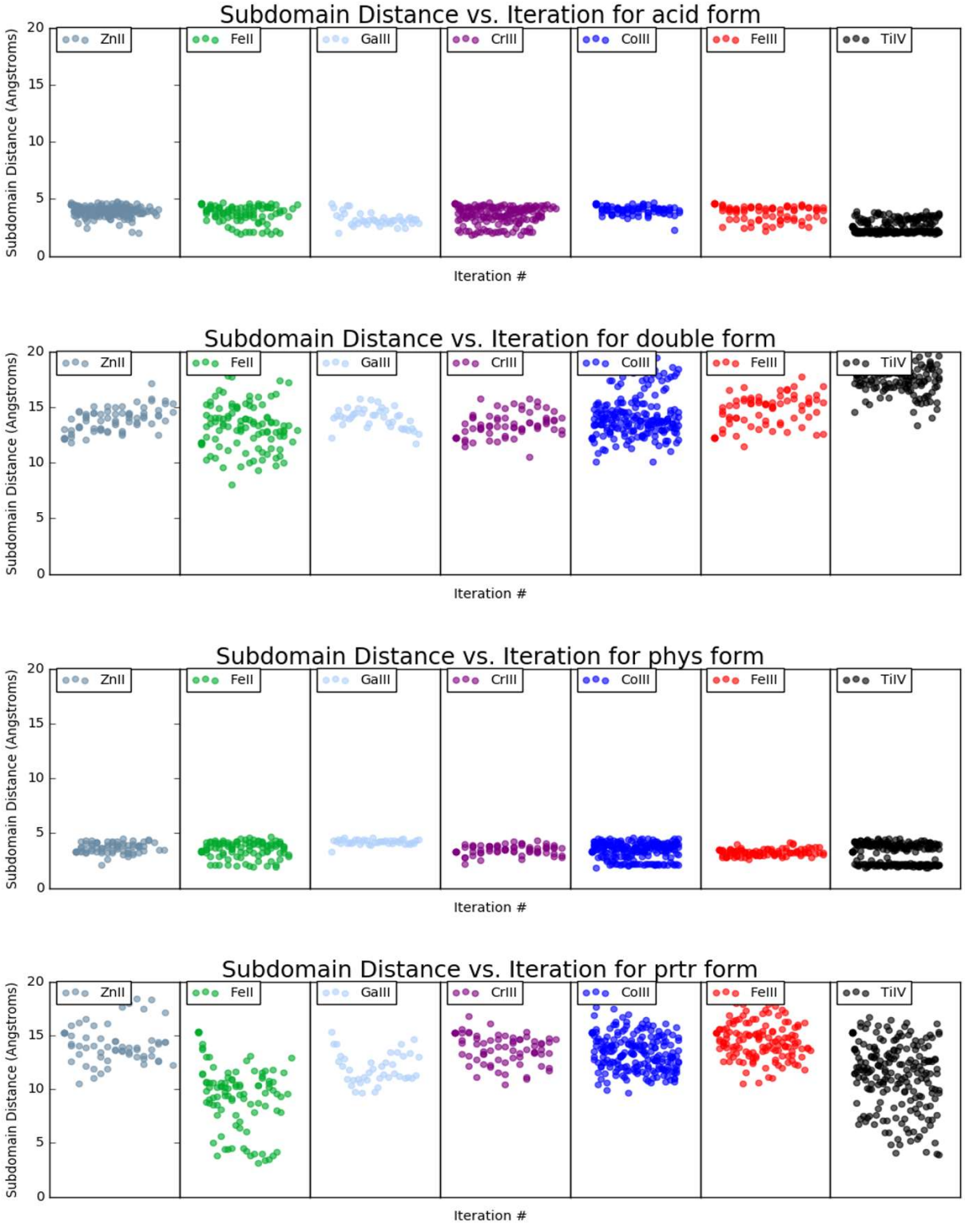


Figure 3.8. Distance between subdomains plotted over the course of all simulations for each metal

3.3.3. DIFFERENCES IN METAL BINDING ENERGIES

To compute binding energy, first the lowest energy conformer of QM/DMD's ensemble of structures is chosen and fully optimized at the QM level with vibrational frequencies also calculated for thermal free energies. A second set of calculations requires the global minimum of a chelator bound to each metal of interest. The chelator should be chosen such that its global minimum is consistent across the metals studied because, e.g., if a water molecule comes or goes between a chelator releasing one metal and binding another, the extra step would need to be accounted for somehow. A good start for the global minimum would be crystal structure data which is readily available for many metal chelator complexes. The crystal structure may be misleading as the chelator may behave differently as a solid than in water. The metal chelator chosen here was EDTA due to the readily available literature data for nearly all metals studied – except Ti(IV), but the data for Ti(III) is available and can be combined with ionization energies of Ti(III) to Ti(IV), so this is not an absolute limitation. The EDTA-metal studied here were first sampled with Born–Oppenheimer Molecular Dynamics and distorted along vibrational modes to help find the global minimum.

3.3.3.1 METAL BINDING TO "ACID" FORM

Fe(III) and Co(III) are both hard Lewis acids with similar ionizations energies and coordination geometries and have been predicted to bind to serum transferrin with nearly identical affinity (log K values: Fe(III) = 21.3, Co(III) = 21.5, Ga(III) = 19.6, Ti(IV) = 26.7).[208] Previously in the analysis of the QM/DMD simulations, Co(III) and Fe(III) variants behaved nearly identically however the computed $\Delta\Delta G$ of swapping Co(III) with Fe(III) in the "acid" form of the protein comes out to 33.09 kcal/mol. Figure 3.9. shows a stereoscopic view of these two metals bound to the protein and the structures aligned in order to show the great similarity of binding mode and residue packing of the active site. It might be surprising how different in energy these two structures are, but a closer look in Figure 3.10. gives one potential reason for this large energy discrepancy -

there are two proton transfers from lysine residues to aspartic acid residues. These proton exhanges were not given to the DMD procedure which does not change protonation of residues due to it placing discontinuous infinite energy wells on all covalent bonds. So, during a QM optimization step there must have been a near barrierless path for the proton to move (a transition uncommon in water). Due to the low solvent exposed nature of the “acid” form’s metal site, the dielectric constant for COSMO was indeed lowered (to 20) and thus the QM region was implicitly solvated with a solvent less polar than water. This could suggest that the closed active site of the “acid” form promotes more neutrally charged species and the Fe(III) variant is currently under-sampled as it did not find that minima yet (or there are limitation in COSMO, TPSS, or some other part of the QM calculation).

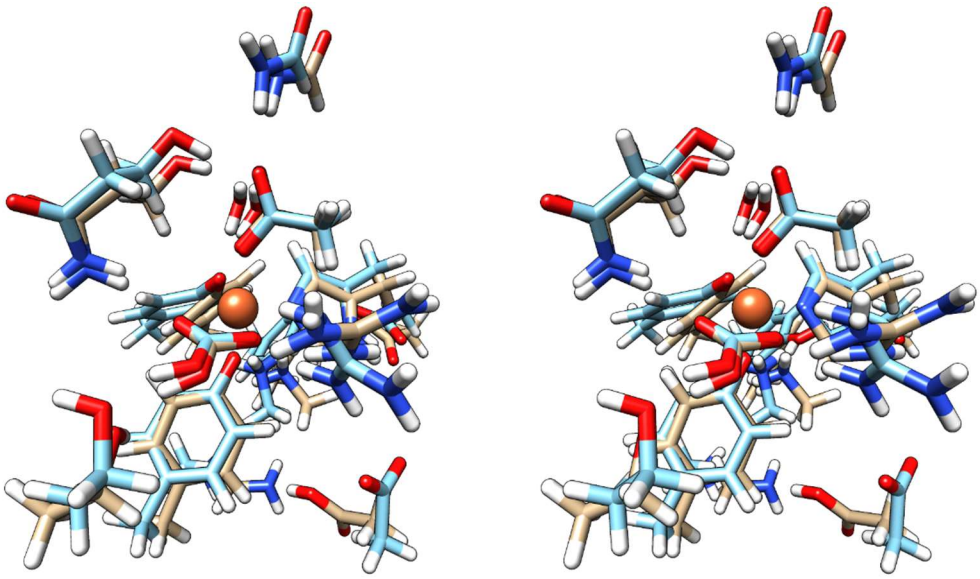


Figure 3.9. Stereoscopic view of the Co(III) variant (blue color) and Fe(III) variant (tan color) of the lowest energy structures QM/DMD discovered for the “acid” variant.

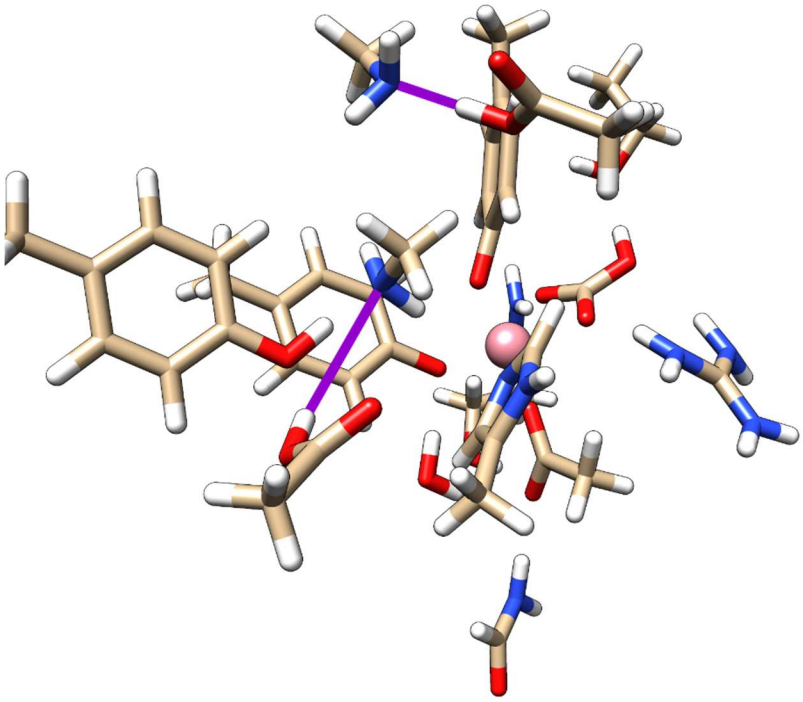


Figure 3.10. View of Co(II) bound to the "acid" form of serum transferrin with lysine – aspartic acid proton transfers highlighted in purple

3.3.3.2 METAL BINDING TO "DOUBLE" FORM

ΔΔG of swapping Co(III) with Fe(III) comes out to 12.54 kcal/mol for the "double" variant of the protein. Similar to the "acid" form, the coordination mode and immediate area surrounding each metal is similar between these metals. A closer look in Figure 3.11. shows that both metal binding sites have undergone a lysine – aspartic acid proton exchange and there is a slightly different network of hydrogen bonds between the two metal. At the time of this writing, the Fe(III) QM/DMD simulations did not reach the same sampling as the Co(III) so further sampling may result in structures of closer energy to Co(III).

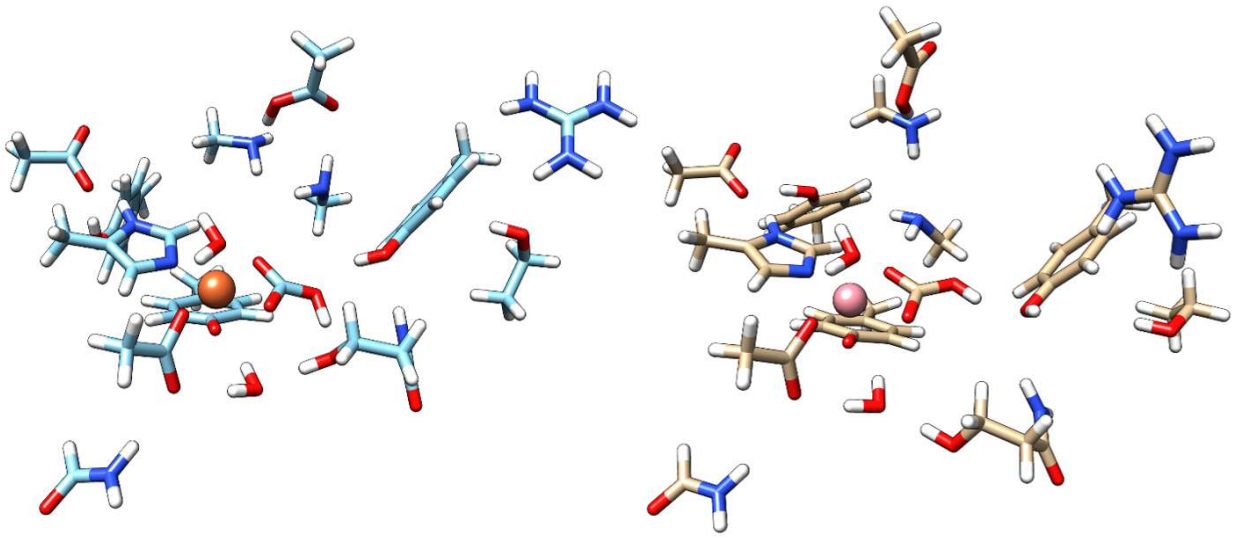


Figure 3.11. Lowest energy conformations of the Fe(III) and Co(III) metal binding sites of the “double” form”.

3.3.3.3 METAL BINDING TO “PRTR” FORM

ΔΔG of swapping Co(III) with Fe(III) comes out to 4.21 kcal/mol for the “prtr” variant of the protein. The “prtr” form is one of the more mobile “open” forms of the protein and we see evidence of changes Fe(III) and Co(III) underwent in Figure 3.12. Namely, the water bound to Fe(III) was replaced with an aspartate residue which resulted in some repacking of the metal site. In section 3.3.2 this was observed to noticeably close the metal site of the protein (but not as closed as “acid” or “double”).

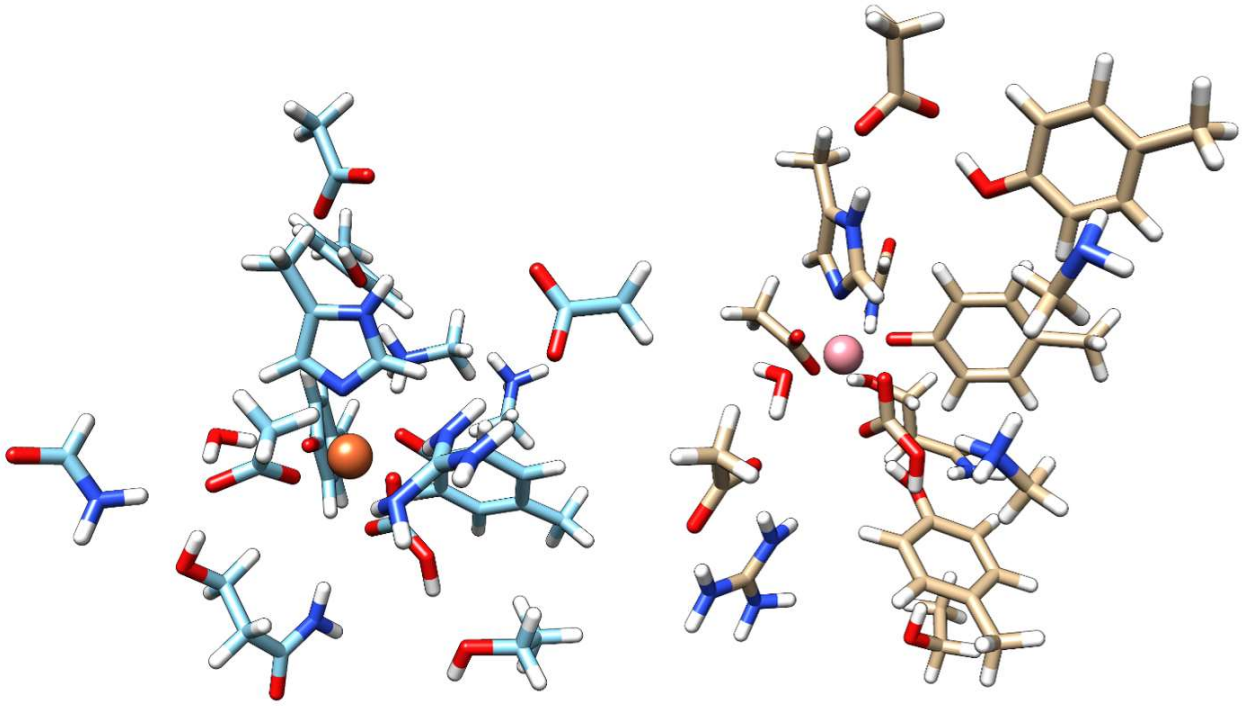


Figure 3.12. Fe(III) rejects water for aspartate in the active site effectively helping repack the active site relative to Co(III) which retains the originally bound water molecule.

3.4. CONCLUSIONS

The metal swap procedure described has shown potential for high accuracy (within 0.5 - 2 kcal/mol), however, comes with some important assumptions and experimental design considerations. The requirement of precise stability constants is typically not an issue with many references available on the subject (and usually to at least two significant figures - note that the error in the literature value needs to be propagated in the calculation). There are also the considerations in chelator selection that need to be made to ensure the chelator of choice completes the thermodynamic equations involved, e.g., if there is water displacement between metals that also needs to be modeled, or if an ionization potential is also needed to change between oxidation states such as with Ti(III)/Ti(IV).

Ideally for this method, energy differences of the larger protein structure may be assumed to be constant. This will be true in many cases and, if not, another method may need

to be employed to calculate that difference and balance with that of the QM region. Fortunately, QM/DMD is a hybrid engine designed for large biomolecules such as serum transferrin, so the full dynamics provided by DMD (no rigid backbone, long time scales) is perfect in order to confirm if there are or are not large conformational changes in the protein molecule.

Due to the vast number of degrees of freedom available to a protein, even QM/DMD's powerful sampling engine may need some time to properly sample the active site and converge on the true lowest energy structures for comparison. This is especially true due to the sensitivity of DFT to small changes in, e.g., hydrogen bond networks.

For large proteins, a sufficiently solvated model may require a large number of discrete and computationally expensive molecules to simulate alongside the system and QM/DMD is designed to treat water in a nearly completely implicit matter which speeds up calculation time. Paired with the this method's packaging of all metal solvation dynamics into an easily referenced literature value, this method truly has the potential to be one of the most efficient ways to calculate metal binding energy differences.

CHAPTER 4

Iridium Enzyme Design – Hydroamination

4.1. INTRODUCTION

Our lab's QM/DMD method is coupled with a protein design scheme to create a tool to design a binding pocket around a metal which yields an enzyme specific for a desired chemical reaction. Enzymes are some of the most efficient catalysts, with impressive stereoselectivity, operating at mild reaction conditions for a host of biotic reactions, and have even been used in industry in synthesizing drugs such as antibiotics[209]. However, the structural evolution of natural enzymes is guided by a narrow range of available materials to primarily meet biological needs. Modern chemical research has advantageously advanced catalysis using non-physiological metals in areas such as synthesis and energy conversion, and there is a wealth of potential in bridging these evolving efforts to yield catalysts of similar ideality that is typical of natural enzymes to biotic processes, but for a much wider range of reactions. The important precedent for this design scheme follows what is termed transition-state stabilization, originally described by Linus Pauling, which can be applied to development of catalytic antibodies and has already found some success in ab initio protein catalyst design[210].

Throughout the enzyme work in our lab, we use our own in-house mixed quantum mechanical-discrete molecular dynamics (QM/DMD) method to accurately and affordably capture organometallic interactions which are otherwise too expensive to model with traditional QM/MM schemes and too difficult to accurately represent in typical all-atom MD schemes. QM/DMD's primary bottleneck involves SCF calculations of the QM site which involves one or more metal atoms and the substrate/residue atoms most important for computational treatment - the rest of the protein is represented in the QM site by structural constraints and sometimes as a matrix of charges. In many cases, enzyme design involves considerable changes to a protein, from simple metal replacement at its core (this chapter), mutations near the active site, even extensive sets of mutations throughout the protein (Chapter 5), This materializes as one limitation of most QM/MM software but QM/DMD has the ability to reestablish a protein's dominant structure in an

efficient manner. This efficiency is in a large part because DMD focuses on very rapid sampling of protein structure. This is done by treating solvent implicitly, (thus scaling well to larger systems and time scales), using discretized potentials for interactions, and keeping the backbone flexible. A full description of enzyme reactivity must account for conformational effects as it is necessary for identifying specific conformations and their place in reaction cycles. This fact materializes as one limitation of most simulation QM/MM software but discrete molecular dynamics (DMD) component of our simulation achieves efficient sampling. For example, metal replacement in an enzyme can lead to large conformational changes that propagate past the active site so we avoid the rigid-backbone approximation, such as to pack the protein around the transition state of interest.. Thus, this work involves both advancing the computational methodology and applications using both physiological and nonphysiological metals. All of our computational methods efficiently solve the typical problem of managing the strong coupling of the metal and bulk protein otherwise problematic in silico, and allow us to test metalloenzyme efficacy in catalysis of both biotic and abiotic reactions.

To facilitate the discussion of the promising capabilities of in silico design using our methods, a preliminary model enzyme was developed that could accommodate a non-physiological metal to catalyze a reaction with an activation barrier lower than seen in current literature. Even though this model ultimately proved to be unsuccessful due to experimental limitations, it nevertheless showcased our methods and gave us a start for further improvements discussed in Chapter 5. The enzyme is inspired by current research in iridium catalysis, where iridium catalyzes hydroamination[211] (Figure 4.1.) and realizes advanced catalytic activity due to transition-state stabilization by the protein's tertiary structure by way of an electrostatic environment which stabilizes the transition state more than the related reactant and product. Specifically, in the course of rate-limiting step, 33 → 34 in Figure 4.1, the NH2 group develops a positive charge. This can be stabilized by providing an anionic residue that the group can interact

with in the corresponding area of the protein. Since the substrate aquires a positive charge and its electronic system is delocalized, additional mechanisms of stabilization can be conceived through π-stacking and other weak interactions of the aromatic system. The main catalytic effect would be coming from the metal bound in the desirable electronic state to facilitate catalysis. Further, through engineering of steric hindrance via the three dimensional protein structure, a desired chiral product can be preferred. These ideas led us to the first attempts to design an Ir enzyme catalyzing hydroamination.

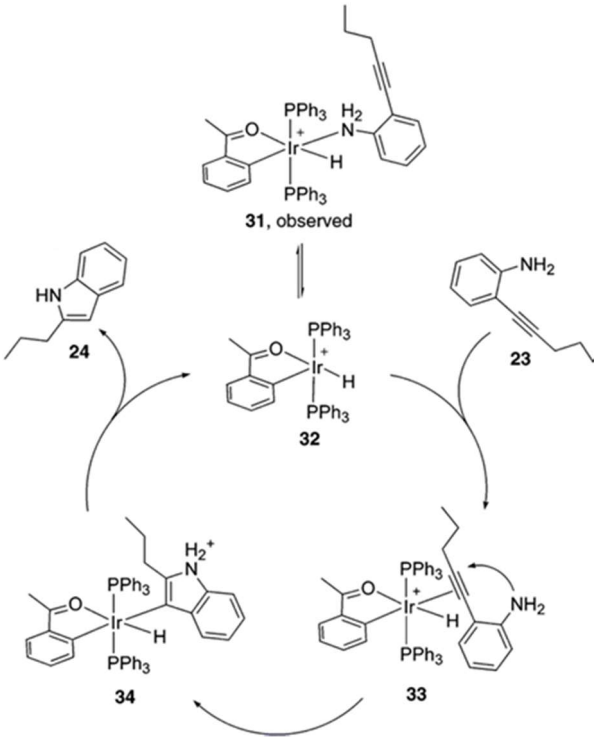


Figure 4.1. Typical catalytic cycle of hydroamination via an iridium catalyst to be mimicked by an enzyme

4.2. METHODS

4.2.1. TEST SYSTEM

The test system to which our primary enzyme design method will be applied is inspired by current research in iridium catalysis (where iridium catalyzes hydroamination[211] and the rate determining step is shown in Figure 4.3. The test case was chosen to be a small molecule with

only a single major step of a reaction besides the acid/base components which are known to not be the rate determining steps. Significantly, the charge redistribution in the rate limiting step of hydroamination provides a clear opportunity to design a binding pocket which favors the transition state. Thus, the proposed metalloenzyme realizes advanced catalytic activity due to transition-state stabilization by the protein's tertiary structure.

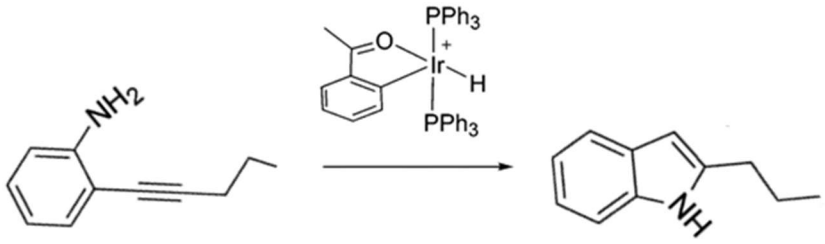


Figure 4.2 The rate determining step of hydroamination modeled in silico in this work and then our design scheme was used to find an enzyme that could preferably stabilize the transition state.

The design space available within a protein is what allows us to engineer transition state stabilization and in this case yielded a theoretical barrier lower than computed from the structures in the cited literature.

4.2.2. OVERVIEW OF DESIGN

An outline of the enzyme design protocol in use is briefly outlined in Figure 4.4. and in the proceeding paragraphs.

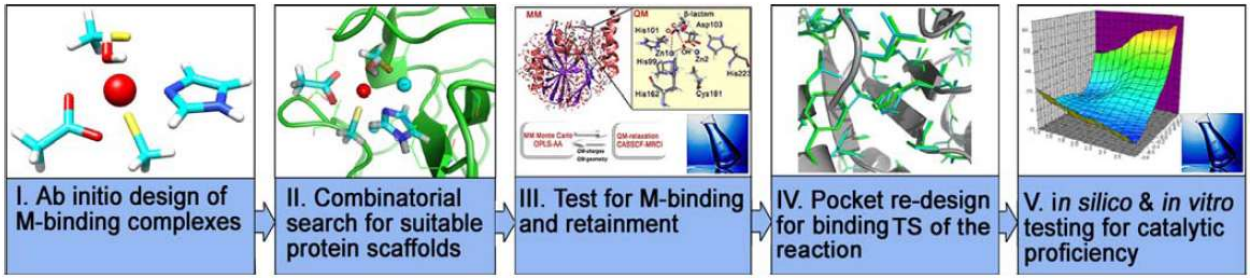


Figure 4.3. General enzyme design scheme to find and adapt a protein for a desired chemical transformation

The start of the enzyme design, ab initio design of metal-binding complexes is performed. In this scheme, however, the important first step of design centers around theoretical amino acid complexed to a metal in a specific geometry that would give some road to catalysis. Metallo-amino acid complexes have seen some use in literature as catalysts, but as part of the design process we are not limited by synthesis of potentially chiral candidates, and the reaction barrier can be further improved down the design pipeline. A thorough study of all permutations of metal binding amino acids are placed around the metal and the resulting electronic characteristics of the metal (e.g., simply the charge on the metal especially if it is to act as a Lewis Acid) is optimized within this set. The substrate must then be docked, and the reaction pathway observed computationally. Once the transition state of a semi- to very reasonable reaction pathway is discovered, the locations of the amino acid ligands are saved and brought to the next step in the process.

The combinatorial search for suitable protein scaffolds step advantageously uses the expansive and growing database of protein crystal structures available publicly via the protein data bank. We use a well-established rigid substructure search available for public use online,

Erebus,[212] which employs an iterated sorting and filtering scheme, which first scans the target structure and collects atom pairs that have an equivalent pair of atoms with matching distance as in the query structure. Given a set of atom coordinates from the ab initio design step, Erebus returns a list of proteins containing the best matching substructures. The reason we use the amino acid positions of the transition state is because our aim is to find a protein which preferentially binds the transition state, but the next step will confirm it can accommodate the reactants as well.

Once Erebus returns us a collection of proteins that have the right amino acids in the right locations such that the metal of interest can be placed in its preferred geometry, QM/DMD equilibration is used to ensure this metal and substrate of interest will bind to the protein in a desired way. For example, depending on the lability of the amino acid—metal bonds, there may be various observed stereochemistries and ideally there would be one lowest energy structure to proceed with in the design. The three dimensional structure also needs to be given time to respond to this perturbation before we have a good idea of what to do in the next step.

Once we have confidence that a protein can bind a metal and substrate in a certain way, we can re-design the binding pocket for optimal binding of the reaction's transition state (i.e. create a theo-zyme, or theoretical enzyme). Firstly, the amino acids in immediate proximity can be altered to include electrostatic, hydrogen bonding, π-stacking, and other non-covalent interactions. Moving further from the active site, more residues can be changed in order to best optimize the electrostatic environment of the substrate for the reaction.

Once equilibrated to the reactant form, the active site is studied using quantum mechanics in a package such as Turbomole in order to find the reaction pathway in full context of the enzyme. At this stage, larger protein movements can mostly be ignored as the rate of moving from the reactant to product or intermediate state is typically too fast for significant structural changes in the protein macromolecule to occur or even to promote the reaction[213] (but a site around the active

site can still be included in quantum mechanical optimizations). In vitro testing follows to confirm catalytic proficiency experimentally.

4.2.3. COMPUTATIONAL DETAILS

Density Functional Theory (DFT) of the Turbomole Quantum Chemistry program package was used for the QM calculations in this chapter. To optimize all starting geometries, intermediates, and saddle points, the nonempirical functional TPSS with DFT-D3[214] dispersion was used and is known to benchmark well for organometallic complexes.[215] For these starting calculations, the def2-TZVPP basis set was applied for all metal atoms (iridium in this case) and def2-SVP applied to all other atoms. Reported energies are single point calculations on the starting geometries but using the functional TPSSh (with exact Hartree Fock exchange) and def2-TZVPP basis set for all atoms instead. For all calculations the resolution identity approximation was used for acceleration and the COSMO implicit solvation scheme was applied with an appropriate dielectric constant (the same as water, in this case because of the solvent exposed nature of the active site.)

4.3. RESULTS AND DISCUSSION

4.3.1. AB INITIO DESIGN OF METAL-BINDING COMPLEXES

A mechanistic study of the literature iridium(III) complex of Figure 4.3. showed a free energy barrier of approximately 16 kcal/mol (computational details: TPSS with def2-TZVPP for iridium and def2-SVP for all other atoms) and followed a simple mechanistic step with the metal acting primarily as a Lewis acid. The first step was to find a combination of amino acid residues which would make the Ir protein candidate to follow the same mechanism and with a reasonable reaction barrier. Four functional heads of side chains present in natural amino acids were considered: carboxylate, imidazole, phenol, and deprotonated thiol. At least one site of iridium(III)'s octahedral binding geometry would need to be open to accepting a new ligand.

Several hundred unique permutations of ligands around the metal center were possible. The candidate pool can be narrowed by discarding candidates that would leave a large overall charge of the complex (the literature complex has a charge of +1). This small library was screened based on the natural charge on iridium(III) via a natural bond orbital[216] calculation that showed having two or more deprotonated thiol residues would most closely align the iridium(III)'s natural charge to that of the literature complex (~0.17 of literature complex, ~0.70 on amino acid complex). A mechanistic study of some of the top candidates showed an identical route to catalytic transformation of the substrate molecule to the product but with an energy barrier of approximately 2 kcal/mol higher. Thus, we found that several candidates involving 2-4 cysteine residues were promising candidates for ligation of Ir(III) in a protein, and the geometry of the corresponding transition state was used for the next step in the design process.

4.3.2. COMBINATORIAL SEARCH FOR SUITABLE PROTEIN SCAFFOLDS

With geometries of the desired amino acid residues at the transition state of the intended reaction, a combinatorial search was performed using the publicly accessible tool, Erebus.[212] Many structures are found, a portion of which are provided on the next page for discussion (Table 4.1.)

Table 4.1. First column contains protein data bank entries the publicly available rigid substructure search tool, Erebus, output from a small cluster of amino acid atoms taken from geometry optimization of the transition state of an amino acid-iridium(III) complex catalyzing hydroamination. The second column indicates any metal present in the protein (though there does not have to be for Erebus to find amino acids in a certain arrangement), the third column indicates the molecular weight of the protein, and the last column indicates the available resolution of the .pdb file ("Solution" indicates NMR structure).

Erebus search results			
pdb accession code	**Metals present**	**Weight (Da)**	**Resolution (Å)**
1H2B	Zn and subs (diff sites)	81,184	1.62
3E28	Zn and sulfate (diff sites)	159,556	2.50
3E24	Zn and sulfate (diff sites)	53,007	2.30
1R79	Zn	8,956	Solution
3HCT	Zn	31,412	2.10
1WFH	Zn	6,811	Solution
1QNI	multiple copper	394,559	2.40
1T3K	Zn	17,006	Solution
1X4W	Zn (2 of them)	7,555	Solution
1X4K	Zn	8,060	Solution
1WFF	Zn	9,280	Solution
1WIL	Zn (2 of them)	9,815	Solution
1Q4W	Zn and sub	43,167	1.93
1V0D	Zn,Mg,Pb	37,897	2.60
3MMK	Zn,Cl^-	37,963	2.16
3Q05	Zn	122,711	2.40
1Q1N	Zn	39,800	3.15
2DAR	Zn	9,898	Solution
2CSZ	Zn	83,510	Solution
2MIU	Zn	11,681	Solution
1NU1	Fe_2S_2, others	241,509	3.2
3HPV	Fe	140,714	2.3
1VQ2	Fe, sub	21,638	2.2
2EGP	Zn	8,379	Solution
1LI7	Zn	104,916	2.6
1V5R	Zn	10,718	Solution
1BRF	Fe	5,956	0.95

4.3.3. TEST FOR METAL-BINDING AND RETAINMENT AND POCKET REDESIGN

Multiple protein candidates were screened using QM/DMD. First, it is necessary to verify structural retainment of the protein upon metal binding. This is especially because many of these proteins originally had a tetrahedral zinc(II) present, thus unpredictable structural changes may occur in moving to an octahedral geometry of iridium(III). QM/DMD simulations across many of the candidates listed in Table 4.1. seemed to be able to accept iridium fairly well, particularly the ones with an aspartate or glutamate as a coordinating ligand. Because the structures returned were almost all tetrahedral structures, residues had to move from 109.4 deg angles to 90 deg angles, except in the case of an aspartate/glutmate's functional head which can bind bidentate to allow for less total distortion to propagate from the active site (leaving one substrate binding site remaining). And that what was consistently observed when replacing with the iridium(III). However, for the larger more packed proteins the change in coordination geometry still put noticeable strain on other parts of the protein (it could be imagined that these were most likely evolved to handle a specific type of metal). Similarly, discussed in Chapter 3 is one example of a protein that can undergo a minor but noticeable movement from open to closed structure on replacement of an octahedral metal with a tetrahedral one.

Upon structure equilibration with just the metal, the substrate is added to further verify that the structure is maintained and finally the product molecule is placed inside the active site. It is desired to equilibrate the protein structure in response to binding of both the substrate and reaction intermediate to see if structural reorganization needs to take place. One candidate, PDB accession code 1V5R, was identified as particularly promising because binding of the reaction intermediate invoked a key structural change in the proteins structure to bring a glutamate residue a bit closer to the substrate (Figure 4.6.)

One benefit of choosing the protein presented here was that, due to its small size, limited steric hindrance, and flexible shape, there wasn't a significant observed change in the overall tertiary structure upon these additions. In fact, the structural change that was discovered, movement of a GLU residue toward the active site, is what made this protein an attractive candidate for further study. However, it was not confirmed if it was more the change in coordination geometry that initiated that change or the positive charge that developed on the intermediate – iridium(III) moved quickly and smoothly to a non-distorted octahedral geometry as seen in the amino acid complexes.

It should be noted that Glu coming down to the substrate comes with some entropic penalty that would be harder to capture, but it was the key to catalysis later.

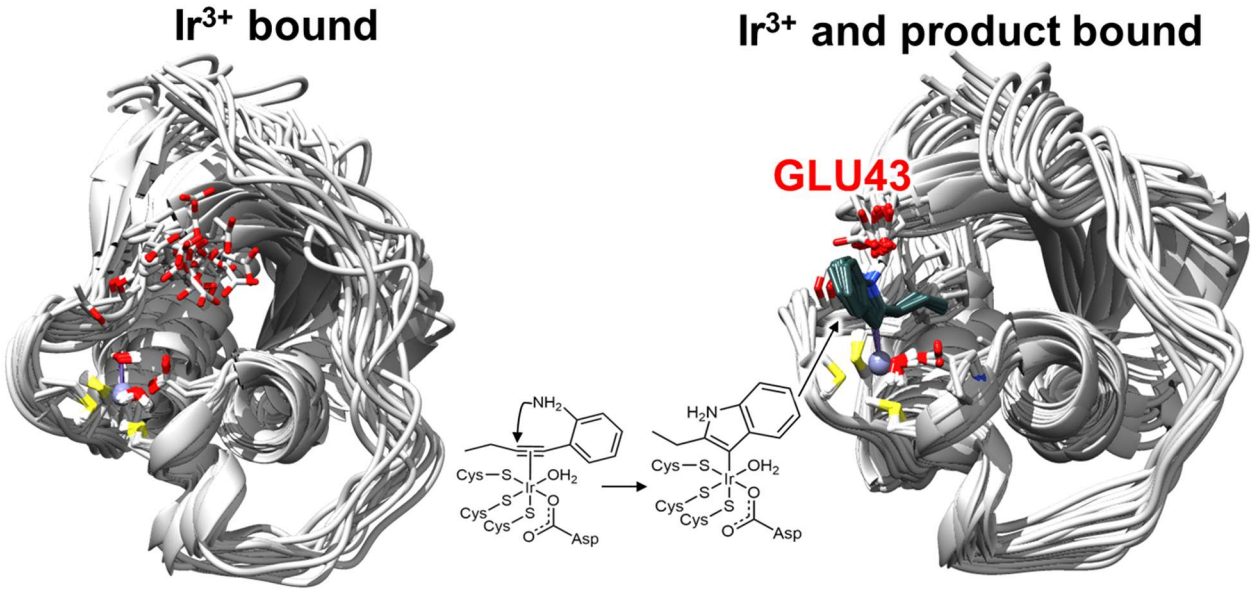


Figure 4.4. Scheme of protein (pdb accession code 1V5R) with metal bound vs metal and substrate with glutamate residue showing strong affinity to the positive charge on the bound molecule.

4.3.4. IN SILICO TESTING FOR CATALYTIC PROFICIENCY

Using the equilibrated structure for the intermediate and iridium(III) bound as a template, the active site was carved out of the equilibrated protein and prepared for mechanistic study. All residue bonds cut from the protein were capped with hydrogen atoms and frozen in place to simulate the protein's own constraints on the site during reaction. The reaction profile is shown in Figure 4.7.

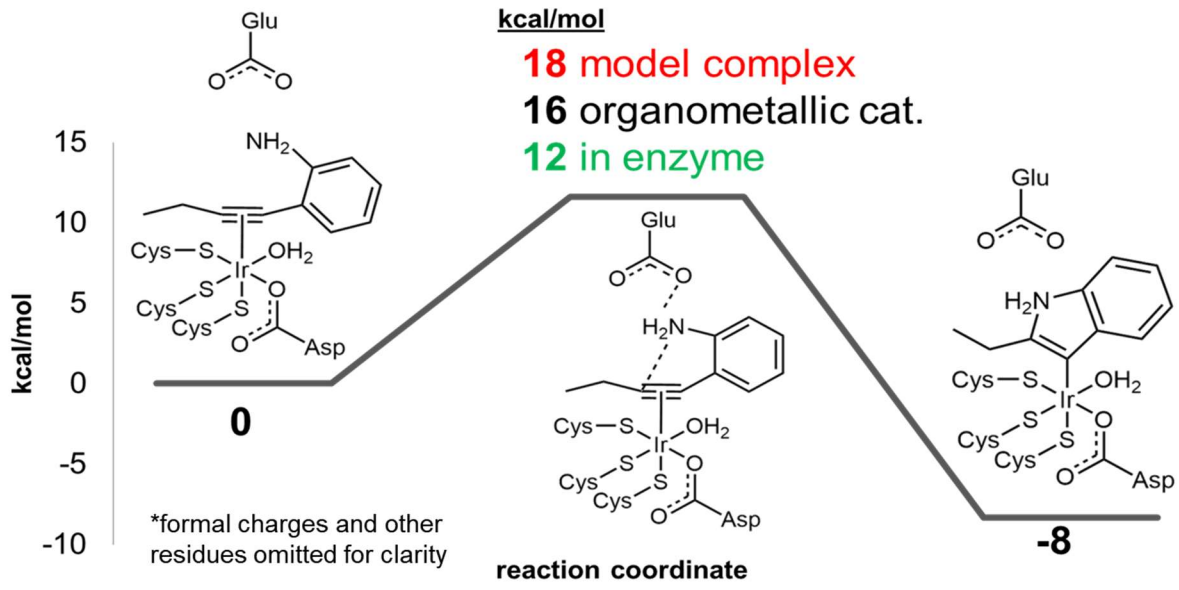


Figure 4.5. Calculations at TPSS with def2-TZVPP for iridium and def2-SVP for other atoms. Cosmo dielectric constant ε = 80 because of the solvent exposed nature of the site.

As noted before, the jump in catalytic activity does not fully account of any structural reorganizational energy required but clearly gives a boost to catalytic efficiency, possibly allowing faster catalysis than possible with the original organometallic complex in literature. Overall, this protein showed a great promise in inding Ir(III) and catalyzing hydroamination. Unfortunately, it was never experimentally realized despite the attempt being made, for reasons discussed below.

4.4. CONCLUSIONS

The full protein design scheme of this chapter showcases a set of tools largely unique to the Alexandrova lab toward designing a protein not only with impressive catalytic efficacy in silico, but with a particularly novel target of enzyme catalysis (a nonphysiological metal which has only rarely been reported to directly bind to a protein at all). Indeed, iridium(III) by itself is very difficult to bind to a protein and this led to experimental collaborators to have great difficulty in getting the iridum(III) bound to the protein and so the catalytic nature was never tested in vitro. Due in part to its high ionization energy and 18 electrons filling s, p, d orbitals, iridium(III) complexes have one of the most stable octahedral geometries of any other metal – hexa aqua iridium(III) waters have a residence time of 300 years making it the most inert out of all transition metal/water complexes.[217] There are ways in which iridium(III) can be incorporated into a protein – its Id mechanism for substation can be lowered by anionic ligands and ones which sterically push other ligands outside of the desired octahedral arrangement as in Pentamethylcyclopentadienyl iridium(III)[218]. This moiety in particular is receiving current attention in the lab of Professor Don Hilvert where further work will be tested, the beginnings of which are outlined in the following chapter.

Nonetheless, this scientific contribution is one of many ongoing research initiatives aimed at designing enzymes for abiotic chemical transformations. Among them are recent successes in producing Diels-Alder cycloaddition, proton transfer, multistep retroaldol transformations, and metal-dependent hydrolysis of phosphotriesters. While the efficiencies reported so far (k_{cat}/K_M = 0.1–100 $M^{-1}s^{-1}$) are typically low in the context of the best natural enzymes (10^6 – 10^8 $M^{-1}\,s^{-1}$), these works set an important precedence for this exciting new field of catalyst development and full reviews of these methods can be found in other recent literature.[219,220]

CHAPTER 5
Guided Evolution of Protein Structure

5.1. INTRODUCTION

The structure and stability of a particular three dimensional structure of a protein is critical for its catalytic function. Even single mutations can have a dramatic effects on protein structure/stability and these effects compound drastically with additional mutations. In the past several decades, tremendous progress has been seen in experimental determination of protein structure but it remains a challenge to tackle the problem of fully assessing the extensive number of permutations that may be involved in significant redesigns.

This work is an extension of the work by the Dokholyan lab in using the rapid simulation capabilities of Discrete Molecular Dynamics (DMD) to assess a protein structure's response to sets of mutations. Performed in silico, precise atomic positions can be monitored and evaluated via statistical tests to quantitatively determine deviations from a target variable. Here, we will use root-mean-square deviation (RMSD) of backbone atom positions from a target fold and a full factorial design of selected mutations to make numerical estimates of the impact of not only single mutations but the higher order effects multiple mutations may have on protein structure. This statistical model captures both the likelihoods that any set of mutations may bring the protein away from the target structure as well as the magnitude of these effects. Done in an iterative fashion, this in silico evolution of a protein can be managed holistically and guided in a quantitative way.

When approaching protein design computationally, this analysis can be part of a critical step toward producing particularly viable candidates for full computational and experimental testing, shown as step IIb. in Figure 5.1. below.

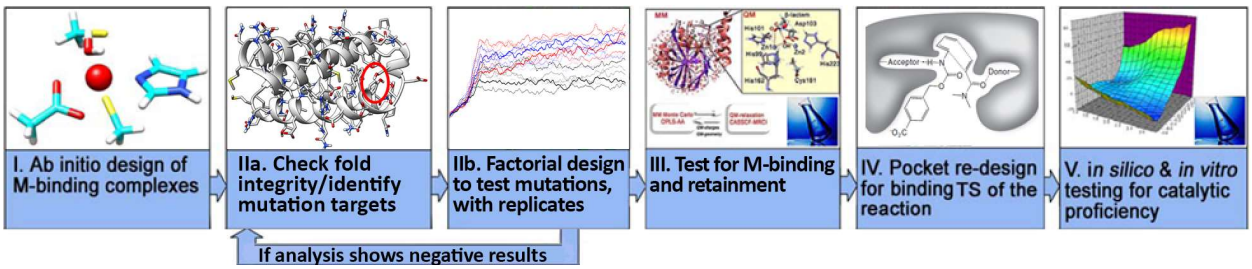


Figure 5.1. Protein design scheme of Chapter 4 modified to include the protein evolution scheme described in this chapter.

5.2. METHODS

5.2.1. TEST SYSTEM

The test system to which DMD and statistical testing will be applied is a small metalloprotein MID1sc2 which was redesigned by the Hilvert lab from a homodimer MID1[221] template by addition of a linker between the helices in order to make the structure a single protein molecule so that asymmetric mutations may be applied. Previously, a small number of mutations were experimentally found to impart appreciable enhancements to diels-alderase and esterase activity but the scaffold itself could not be crystallized, thus structural details were uncertain. The crystal structure of MID1 and a model of MID1sc2 are shown in figure 5.2. For MID1sc2, we identified significant deviations in the fold across many simulations which corroborates with experimental difficulties in crystallizing the protein. Note that the symmetrical metal binding residues are removed and an amino acid sequence was added between the bundles, bringing the structure from a dimer to one cohesive protein molecule. MID1sc2 was chosen for this study because of its small size, ease of production, ability to bind a metal and early successes in the original dimer showing catalytic activity[222] achieving catalytic activity via mutation yet to be published.

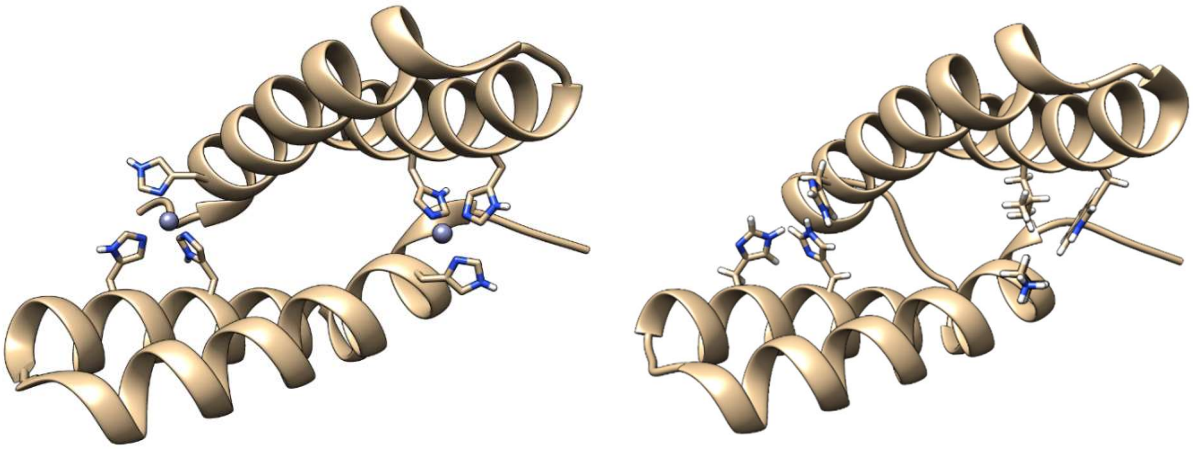


Figure 5.2. The original published MID1 structure221 is shown on the left and is composed of two symmetric alpha helix bundles arranged perpendicular and bound by two metal atoms. The modified MIDsc2 structure is shown on the right to consist of a both parts connected by a linker, thus yielding a single peptide where point mutations can be applied assymetrically (one set of metal binding residues was removed for MIDsc2). Note that MIDsc2 was unable to be crystallized but is the model structure future catalysis was meant to be based on and that is the motivation for this work and so will be often referred to as the "target structure" throughout this text.

5.2.2. DMD EQUILIBRATION OF THE ORIGINAL FOLD

The starting point for this experiment is the original MID1sc2 model structure provided by the Hilvert lab, shown on the right in Figure 5.2. containing two parallel alpha helices connected by a linker segment, and will be used as the target structure to maintain for the purposes of testing this procedure. Catalytic activity will be added at a later stage. Over the span of approximately 55 nanoseconds of simulation time, a short annealing procedure was conducted where temperature was gradually raised to 310 K, where major structural perturbations can quickly occur, and then is cooled and held at a modest temperature of 275 K. In this short simulation, significant fluctuations in protein structure were observed even at the lower temperature. Some simulations deviated considerably from the target which will help guide some of our goals in

producing a viable structure for later study. Figure 5.3. shows RMSD vs. simulation time for 16 replicates.

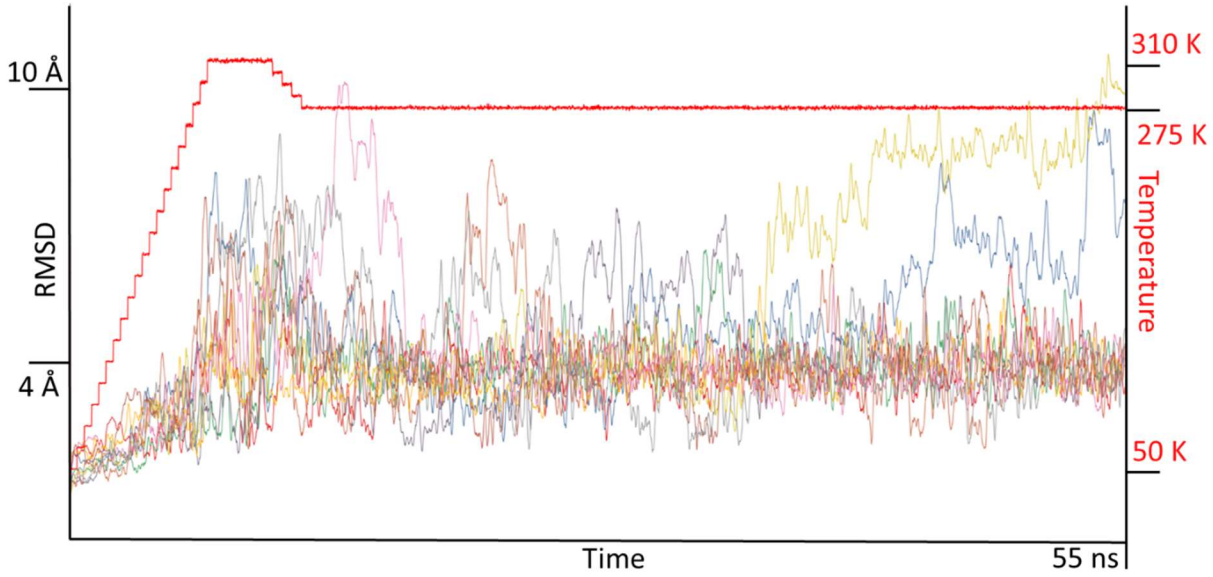


Figure 5.3. Sixteen replicates of the original peptide (MID1sc2) underwent 55 nanoseconds of Discrete Molecular Dynamics – a short annealing step at 310 K followed by a longer run at 275K. RMSD of the structure vs the original structure is plotted over this simulation for all replicates and significant deviations observed for a fraction of the runs.

A closer look at these structures shows several important changes in protein structure. Some large structural variations were found to originate from promiscuous hydrogen bonding, salt bridging, parallel peptide bundle formation, and issues with the added linker (Figure 5.4.). These issues will be addressed in the results and discussion section of this chapter.

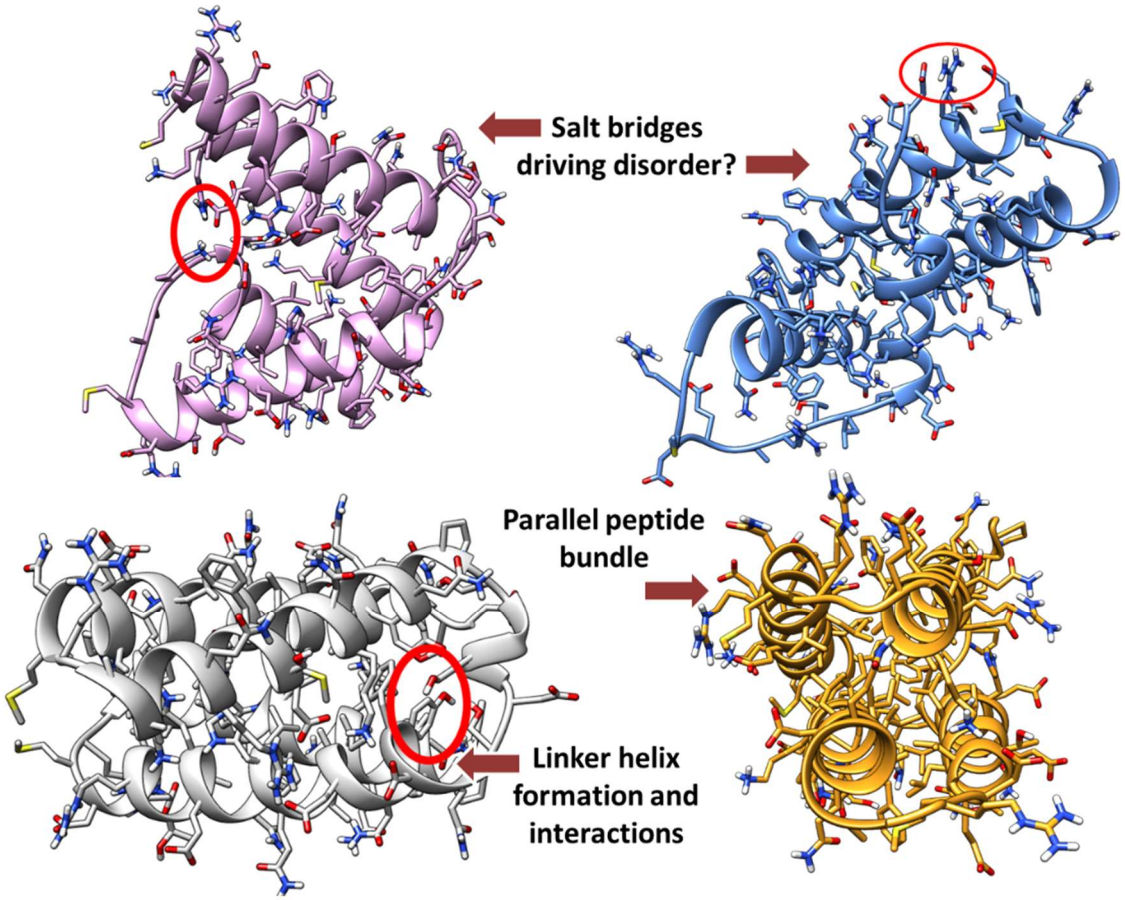


Figure 5.4. Significant deviations from the original (target) structure were observed to include undesired salt bridge formation, secondary structure forming in the linker along with undesired hydrogen bonds, and a parallel peptide bundle.

5.2.3. FACTORIAL DESIGN OF EXPERIMENT

An example of a full factorial design with two mutation sets, A and B, on a wild type protein would involve mutating the protein in four different ways and measuring a response variable: only set A, only set B, both sets A and B, and no mutation. We are interested if the observed response variable of the mutants is different than that of the wild type and by how much/in which direction. Mutants with either (but not both) set A or B may have an effect (a main effect) on this response variable. The response variable could be some measure of structural variation from the wild type, such as root-mean-square deviation (RMSD), and this effect can be measured as an average

difference of the response variable with and without mutation. When both sets of mutations are applied to the protein, the resulting effect on the response variable may be greater or less than the effect of both added together (an interaction effect). A linear model can describe main and higher order effects in this way as follows:

$$Response = \beta_0 + \beta_A A + \beta_B B + \beta_{AB} A * B$$

Where A and B are categorical variables with values of 0 (no mutation) or 1 (mutated).

Determining main effects and interaction effects between two independent variables on a continuous dependent variable can be done using two-way Analysis of Variance (ANOVA). In practice, characteristics of a response variable (such as RMSD) may not perfectly match all of the assumptions of two-way ANOVA, *e.g.,* the assumption of normality in the distribution of the response variable may not be satisfied if multiple minima are observed. However, ANOVA is known to be quite "robust" to violations and these assumptions can always be tested using separate techniques (e.g., the Shapiro-Wilk test for normality), can be forced by taking the responses as sample means (via the central limit theorem), or the results simply taken qualitatively if we're only looking to see if there is some statistically significant difference and in which direction.

Two-Way ANOVA can be expanded to include more variables such as the full linear model with mutations A, B, and C:

$$Response = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_{AB} A * B + \beta_{AC} A * C + \beta_{BC} B * C + \beta_{ABC} A * B * C$$

The main effect coefficients are the simple slopes of the mutation within each group and the interaction coefficients are the difference in slopes. Thus, with the higher order terms, the effect of mutations may vary depending on which other mutations are present.

ANOVA is essentially a statistical tool for splitting variability into component sources. By measuring variability within groups one has a baseline against which differences among group means can be compared. The resulting statistic is used to determine the probability this difference is due to pure chance (the p-value). In this study, the null hypothesis is that the structure of a mutant does not deviate from that of the protein without any mutations and we will reject the null hypothesis if the structure is significantly different (better or worse).

In this work, five mutation sets were identified for each trial which results in 32 possible combinations of mutation (or lack of mutation). 16 replicates of each are simulated using DMD to give 16 independent observations of each 32 sequences. The observations are taken as the RMSD against the original structure at the end of the simulation. DMD naturally samples the conformational space of the protein so there is a wealth of structural data across the simulation however, in order to preserve the independence of each observation, only a single representative structure at the end of the simulation is analyzed for the response variable (RMSD). The response variable could easily be changed to another item of interest, such as some measure of the geometry of the active site, distance between residues, etc.

5.3. RESULTS AND DISCUSSION

5.3.1. METAL BINDING RESIDUES CAN BE REMOVED WITHOUT SIGNIFICANT EFFECT

Glutamate, aspartate, and histidine residues all carry the potential to bind metals and extra residues of these types may interfere with the intended metal binding pocket in the protein. Thus, there is broad interest in switching metal binding residues to a residue such as asparagine. In the case of MID1sc, five of these residues are identified and labeled in figure 5.5. (ASP 25, GLU32, GLU 40, HIS 65, GLU81 labeled as A, B, C, D, and E, respectively.)

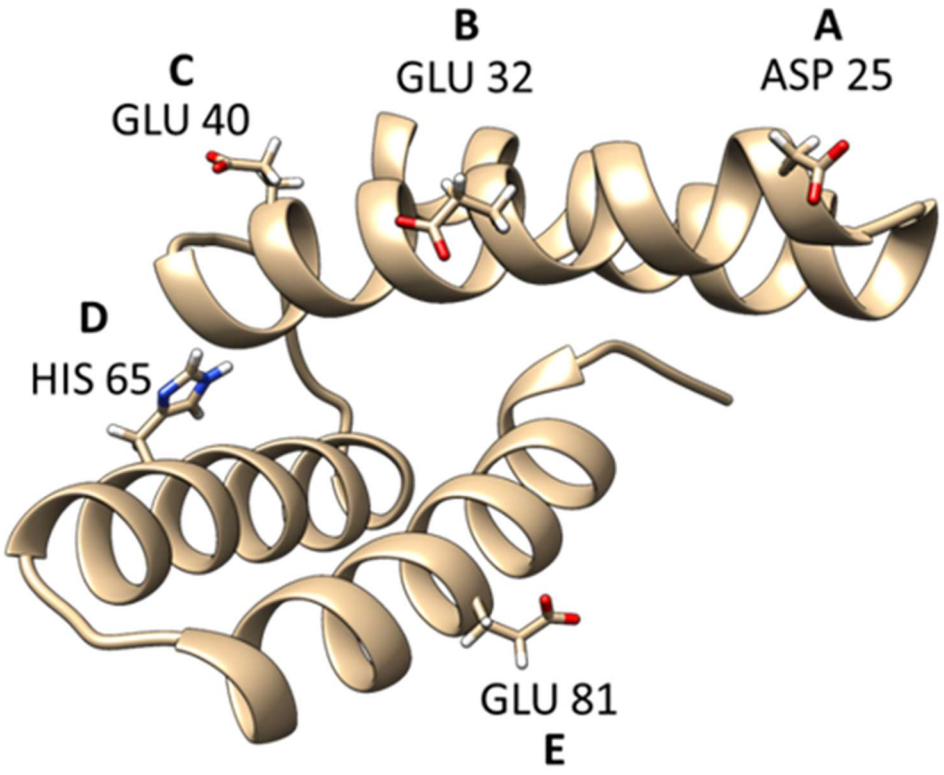


Figure 5.5. The five potentially metal-binding residues are mutated to non-metal binding residues (namely, to the amide residue asparagine) in a full factorial design. From the full factorial design, multi-way ANOVA can help determine main and interaction effects of residue mutation on structure.

The first illustration of this method will involve five single mutations. Every combination of these mutations (32 total) was simulated 16 times, with random initial velocities, and the response variable of RMSD vs the original structure was calculated using the structure at the end of each simulation. These responses were recorded and analyzed via multi-way ANOVA. A model with only the main effects is presented below and, as seen in Table 5.1., the results show no statistically significant interaction effects.

Table 5.1. β0 represents an average RMSD of all simulations and the other β values indicate the average magnitude and direction of RMSD change when the residue is mutated to a non-metal binding residue and the opposite if it isn't. For example, if residues marked C and E are mutated and the rest stay the same, the average RMSD would be the sum of the average of all simulations with βC and βE added and βC and βA, βB, and βD subtracted (4.42 – 0.480 – 0.128 + 0.228 – 0.002 + 0.108). Overall, only A was found to have a statistically significant impact on RMSD ($p < 0.05$). No significant interaction effects were observed.

$$RMSD\ (\text{Å}) = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_D D + \beta_E E$$

$$\beta_0 = 4.42, \quad \beta_A = 0.480, \quad \beta_B = 0.128, \quad \beta_C = 0.228, \quad \beta_D = 0.002, \quad \beta_E = 0.108$$

$$p_0 = 0.000, \quad p_A = 0.026, \quad p_B = 0.550, \quad p_C = 0.289, \quad p_D = 0.997, \quad p_E = 0.618$$

Based on the results above, we expect mutation A (Asp25Asn) to provide a significantly different response than the original protein with an RMSD difference of an estimated 0.480 Å (i.e., mutant deviates further from the target). Further, we quickly conclude that all of the other residues may also have a small positive effect on the RMSD response but these effects are not large enough to be statistically significant and may have come from pure chance due to the noise present in the data. Thus, we can decide to either discard the possibility of mutating ASP 25 to become non-metal binding or decide its detrimental structural effect is better than having that aspartate present. Note that there is no requirement to only have two categories for each mutation – if we wanted to test changing the residue to asparagine or a weak metal-binding residue such as serine a multi-way ANOVA could be performed but the full factorial design would require 3 instead of 2 categories multiplied 5 times for each mutation set times 16 replicates (from $16 \times 2^5 = 512$ simulations to $16 \times 3^5 = 3888$ simulations) thus, a balance must be struck between capturing

full effects (including interactions) of residues and size of the design space. One possibility is to first screen for main effects in a one-way ANOVA and expand the design space from there with a full or partial factorial design. In other words, start by checking if there is a significantly higher RMSD when all residues are mutated to any of various residues and then, one by one, mutating each set to find out which one was responsible, then not all combinations need to be tested (but some interaction effects could be missed, if there are any).

5.3.2. MUTATIONS OF SURFACE AND LINKER HAVE SIGNIFICANT EFFECTS ON STRUCTURE

A helix bundle is a small protein fold composed of alpha helices stacked next to each other and simulations suggested our bundle, which is intended to have perpendicular alpha helices, instead can switch to parallel alpha helices – a structure which has long been known as a sort of super-secondary structure,[223] however not our structural goal for eventual catalysis. One can imagine that changes to where the helices interface may disrupt this stacking and make it unfavorable. Also, the linker between the sets of alpha helices may interact with the rest of the peptide in an unknown way so mutations in that region also may provide additional stability of the target structure. The residues identified for the following analysis are shown explicitly and labeled LF (Link Fix) or APB (Avoid Peptide Bundle) in Figure 5.6.

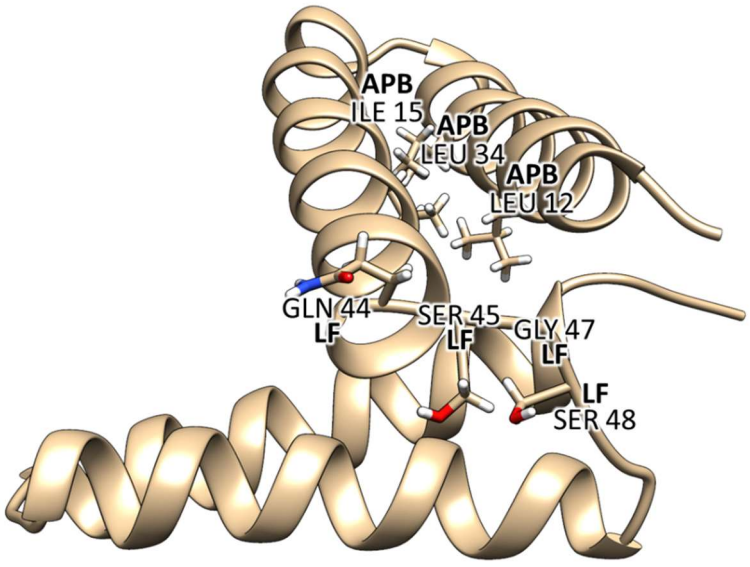


Figure 5.6. Two sets of mutations, labeled LF (link fix) and APB (avoid peptide bundle), are identified as potentially mutable towards reducing secondary structure linker formation and peptide bundle formation, respectively, and are tested here via a full factorial design and multi-way ANOVA on the results.

The reasoning for exploring these sites follows: it is well known that proline can help disrupt secondary structure formation[224] so GLN 44 and GLY 47 are considered for mutation to proline. By inspection of the supplied peptide's simulations, Ser 45 and Ser 48 were observed to partake in some unnecessary hydrogen bonding so those may be changed to glycine (another residue known to disrupt secondary structure alongside proline[225].) These three mutations are all part of mutation set LF in the equation below. For parallel peptide bundle disruption, we can try switching the larger hydrophobic residues to the smaller hydrophobic residue alanine (APB_1) or threonine (APB_2) and these two options compose the ternary categorical variable APB. Results can be summarized as below by the following equation and Table 5.2.:

Table 5.2. As in Table 5.1., β_0 represents an average RMSD of all simulations and the other β values indicate the average magnitude and direction of RMSD change when the residue is mutated as described above. All main effects were statistically significant ($p < 0.05$) and with a large positive magnitude indicating that every mutation attempted resulted in a large increase of RMSD, thus structure deviated considerably from the target structure. No significant interactions were observed.

$$RMSD\ (\text{Å}) = \beta_0 + \beta_{LF}LF + \beta_{APB_1}APB_1 + \beta_{APB_2}APB_2$$

$$\beta_0 = 7.636, \quad \beta_{LF} = 0.924, \quad \beta_{APB_1} = 3.503, \quad \beta_{APB_2} = 2.983$$

$$p_0 = 0.000, \quad p_{LF} = 0.007, \quad p_{APB_1} = 0.000, \quad p_{APB} = 0.000$$

Clearly, all of these residue combinations have deleterious effects on the protein's ability to retain the target structure. And again, we probably would have been satisfied with testing the mutations one by one to gauge just the main effects. A one-by-one approach would save computation time by not testing every pair of mutations against each other, however, we would not have been as confident, e.g., that we didn't miss some pair of residues that would have benefitted the structure in an ancillary way. With the above results of a full factorial design, we are more confident that none of the proposed mutation sets, in any arrangement tested, would help in maintaining the target structure.

5.3.3. PEPTIDE STRUCTURE CAN BE OPTIMIZED WITH PAIRS OF H-BONDING MUTATIONS

Hydrogen bonding, particularly salt bridge formation, is well known to contribute stability to the entropically unfavorable folded conformation of proteins.[226] In this section we will consider several residues, shown explicitly in Figure 5.7., and how we can tailor some of them to improve fold stability of the target structure.

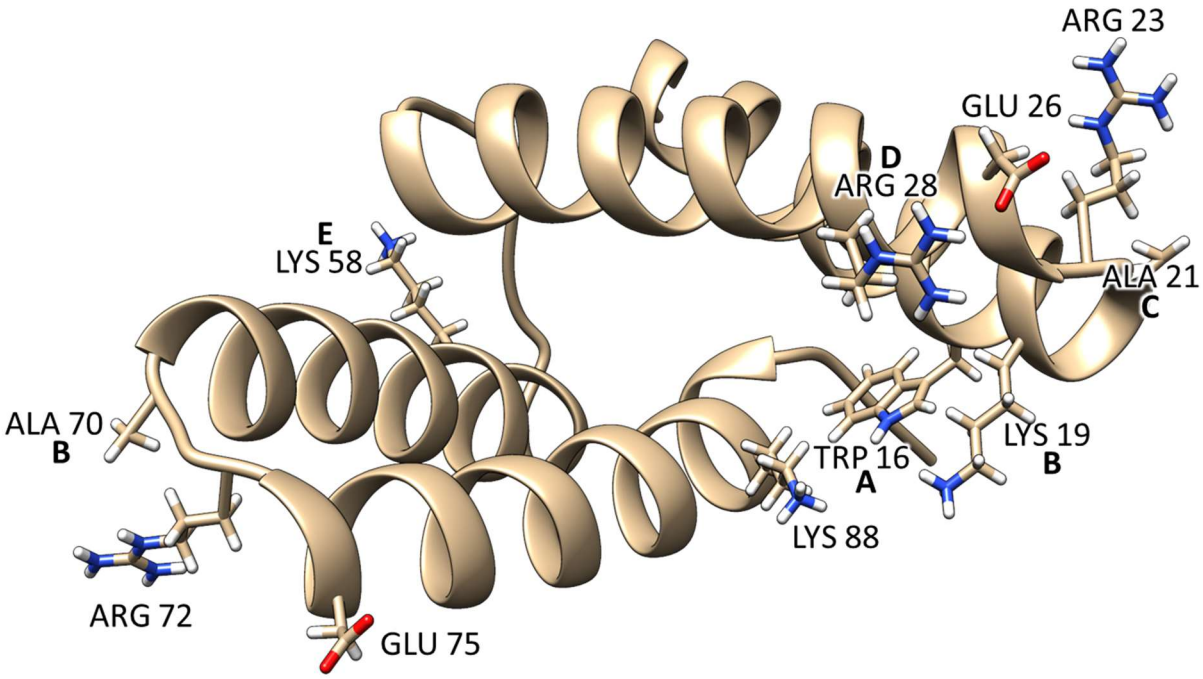


Figure 5.7. Five sets of mutations, labelled A-E, were identified as potential contributors to protein fold stability via hydrogen bond networks.

5.3.3.1. KEEPING ARGININE RESIDUES "IN PLACE"

In several of the simulations of the original peptide, Arg 72 and Arg 23 were observed to form salt bridges with other residues in the protein. This promiscuous behavior distorted the peptide considerably and so we can consider a couple of options to help avoid that: mutate both arginine residues to something less promiscuous (e.g., an uncharged residue), or try to discourage them from bridging in another way. Previously in this chapter, several negatively charged metal binding residues were removed in order to reduce metal-binding competition. Thus, removing more charged residues may create potential issues with the protein fold holding in water – ideally we'd like the keep changes to the overall surface charge minimal. Instead of removing the arginine residues, we will attempt to hold them in place by placing an additional hydrogen

bond acceptor nearby. In the case of Arg 23 and Arg 72, there are already aspartate residues nearby and by mutating the alanine on the side opposite to the aspartates to a charged residue such as glutamate, see Figure 5.7., would offer another stabilizing interaction to the arginine, help maintain the overall surface charge when paired with removing metal binding residues, and the new charged residues would hopefully be less promiscuous themselves due to the hydrogen bond network that would form. Mutations Ala21Glu and Ala70Glu will be considered mutation sets A and C in the subsequent analysis.

5.3.3.2. MODIFYING OTHER SALT BRIDGING RESIDUES

Similar to Arg 72 and Arg 23, residues Lys 19, Arg 28, and Lys 58 were all found to participate in hydrogen bonding that may be detrimental to the target structure. In these cases, neutralization seemed to be the clearest option so each was mutated to non-charged residues – glutamine for the case of the lysine residues and threonine for arginine. Lastly, visual inspection of the trajectories of the supplied structure showed considerable movement of the two alpha helix bundles so anything that could help anchor the sheets to stay in place may be helpful – Trp 16 and Lys 88 were identified as close and directional enough to form a salt bridge if only they were opposite charges. Thus, Trp 16 was chosen to be mutated to aspartate.

5.3.3.3. SUMMARY OF HYDROGEN BOND NETWORK MUTATIONS

As summarized in Table 5.2., all of these potential mutations were fit into 5 mutation sets and a full factorial design of experiment was created with each set having two categorical labels – mutated (as described previously) or not mutated. With 16 replicates this gives a total of 512 DMD simulations. All mutation sets were single mutations however multiple can be combined as in mutation set B. This combination is meant to reduce computational cost at low risk of missing an important interaction term - Ala70Glu and Ala21Glu are already the same mutation, just on different alpha helices in the linked peptide. If one coefficient were found to be significant we can reasonably expect the other to be similar and so the effect of Lys19Gln can still be inferred.

If Lys19Gln and Ala70Glu were studied separately we would have 6 sets of mutations and the number of simulations would rise from 512 to 1024. Another option would be to make that categorical of the ternary kind – no mutation, Lys19Gln, and Ala70Glu, for a total of 768 simulations.

Table 5.3. Summary of mutation sets toward goal of optimizing the hydrogen bond network of the MID1sc2 peptide.

Label	Mutation Set	Description	Categories
A	Trp16Asp	Designed salt bridge with Lys88	Not mutated/Mutated
B	Lys19Gln + Ala70Glu	Remove promiscuous H-bond + Keep Arg72 “in place”	Not mutated/Mutated
C	Ala21Glu	Keep Arg23 “in place”	Not mutated/Mutated
D	Arg28Thr	Remove promiscuous H-bond	Not mutated/Mutated
E	Lys58Gln	Remove promiscuous H-bond	Not mutated/Mutated

Using the same multi-way ANOVA procedure as before, we see a more complex set of results indicating some potentially important interactions between categorical variables. In Table 5.4., the ANOVA output is summarized with statistically significant ($p < 0.05$) values bolded. All main effects were included as well as potentially interesting interaction terms which will be discussed below.

Table 5.4. Summary of multi-way ANOVA results for hydrogen bond network optimization. See Table 5.3. and the corresponding discussion for descriptions of all categories – multiplied terms are higher order interaction terms where the coefficient is added when all labels in the term are present.

Label	Coefficient (β)	P-value
Constant	**4.2019**	**0.000**
A	**- 0.2762**	**0.000**
B	- 0.0435	0.549
C	- 0.0228	0.754
D	**0.1697**	**0.024**
E	0.0104	0.886
A*E	**0.1516**	**0.037**
B*C	-0.1319	0.070
B*E	0.0552	0.447
C*E	-0.1192	0.101
B*C*E	0.1128	0.121

A couple of the main effects are clear. Mutation set A, which introduces a potential salt bridge, lowers the RMSD vs the target structure by an average 0.28 Å, indicating a stabilization of the target structure. Along with Figure 5.8. below, we can confirm the salt bridge was formed and does help to stabilize the structure. Also of significance, Mutation set D seems detrimental to the fold by raising the RMSD by an average of 0.17 Å. Mutation set E doesn't appear to have much of an effect on its own, however when paired with mutation A, the pair was found to have a statistically significant effect on RMSD which cancels roughly half the main effect of mutation set A. This A*E interaction could suggest that the designed salt bridge of A not only stabilized the target structure by introducing a salt bridge as intended, but somehow also coerced the lysine on the other end of the peptide to stabilize. If this was via another salt bridge, that would explain why neutralizing lysine of mutation set E has a deleterious effect on A's main effect.

Confirmation that it was a hydrogen bond formed between residues 16 and 88 in mutation set C was observed by plotting residue distance frequency for all simulations with mutation set A with those without in Figure 5.8. below, which shows a sharp peak around 2 Angstroms (typical of a hydrogen bond) for those with mutation set A against a broader peak around 4 Angstroms for those simulations without the mutation.

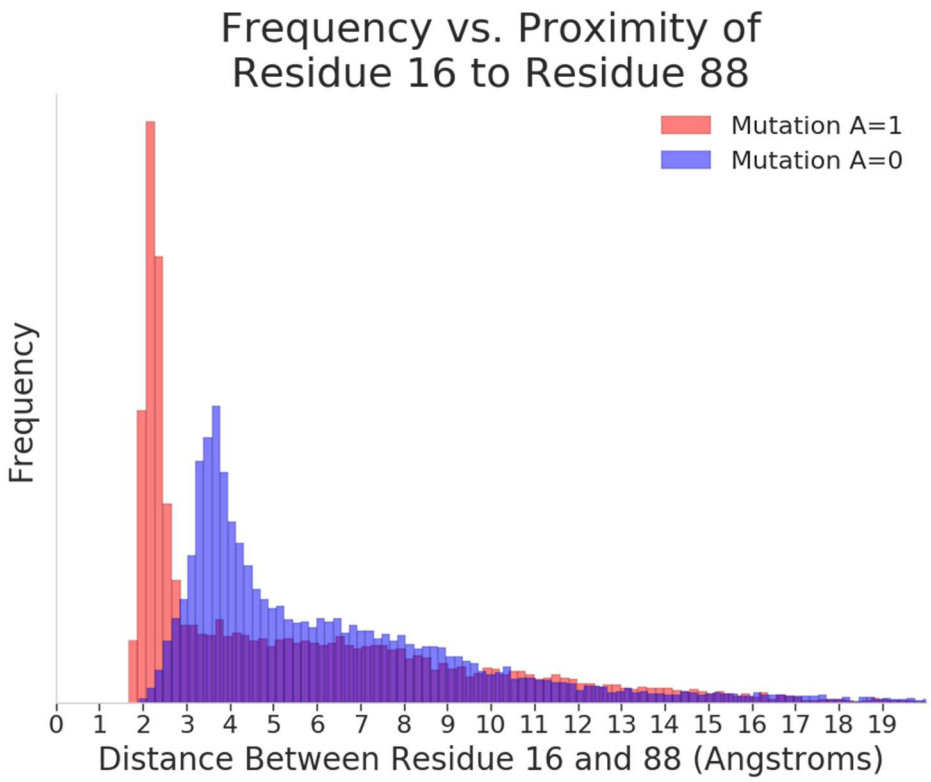


Figure 5.8. The distribution of pairwise distances between residues 16 and 88 across all simulations with and without the Trp16Asp mutation. A sharp peak around 2 Angstroms, typical of a hydrogen bond, corroborates with the improved stability as a small success in the design process.

The other coefficients did not reach a level of statistical significance in the ANOVA results, however, there are some sizable interaction effects to note. Mutation sets B and C were intended to help stabilize the fold primarily by preventing arginine residues from hydrogen bonding with other residues. According to the multiway ANOVA performed, the main effects of sets B and C on RMSD are statistically inconclusive, however, when paired together the main

and B*C interaction effects of the linear model sum to indicate a lowering of average RMSD toward the target structure by an average of (0.0435 + 0.0228 + 0.1319) = 0.1982 Å. This effect is larger than some of the individual statistical significant contributions but split between several variables which reduces the power of the test. Cases like this remind us that more statistical power may be necessary to determine higher order effects with statistical certainty, primarily by adding more replicate simulations.

The highest order effect included in the model, B*C*E, can be seen as a sort of “leveling off” of lower order effects. If one considers two scenarios – a protein mutated with sets B + C and another one with both C and E, we would expect a lowering of the average RMSD because the main and interaction (B*C and C*E) effects sum to negative. If all three mutations were present in the same molecule, B*C and C*E effects would stack and B*C*E is also added and approximately cancels out one of the two second order terms, thus, B*C*E isn’t necessarily a bad combination it just shows that the protein didn’t get any better when combining stabilizing mutations.

A look at pairwise distances confirms a hydrogen bond does form in the case of both mutation sets B and C (see Figure 5.9. for mutation set C) alongside a lower average RMSD to the target structure. The distribution also shows a multimodal nature of the distribution which suggests flexibility of the site. Figure 5.10. shows a sampling of structures at several of the observed minima and Ala21Glu (mutation C) visually showing a visually more ordered set of structures.

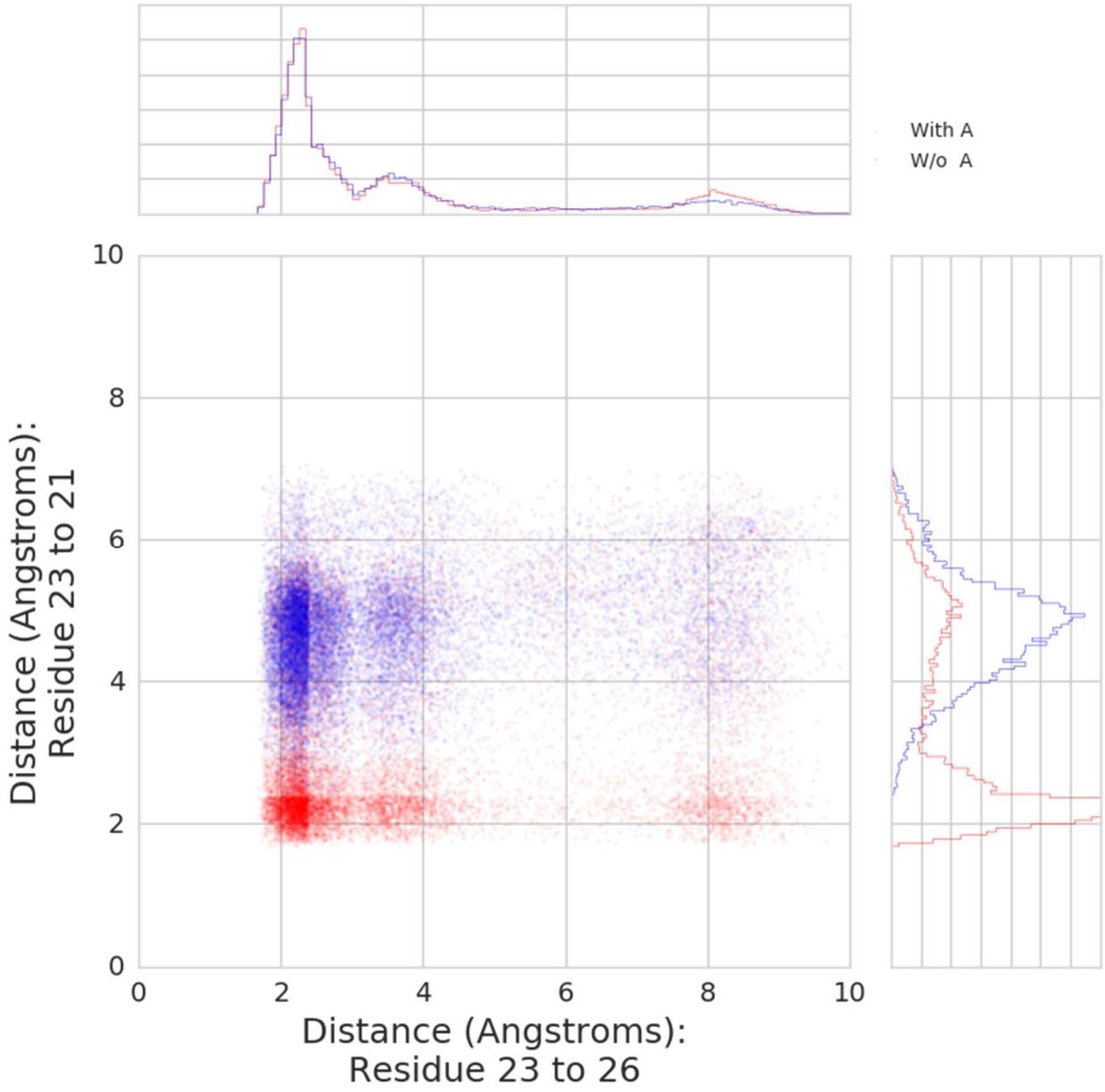


Figure 5.9. Pairwise distances of residue pairs 23 + 21 and 23 + 26 between mutation set C (red) and no mutation set C (blue) confirming that a hydrogen bond forms as designed.

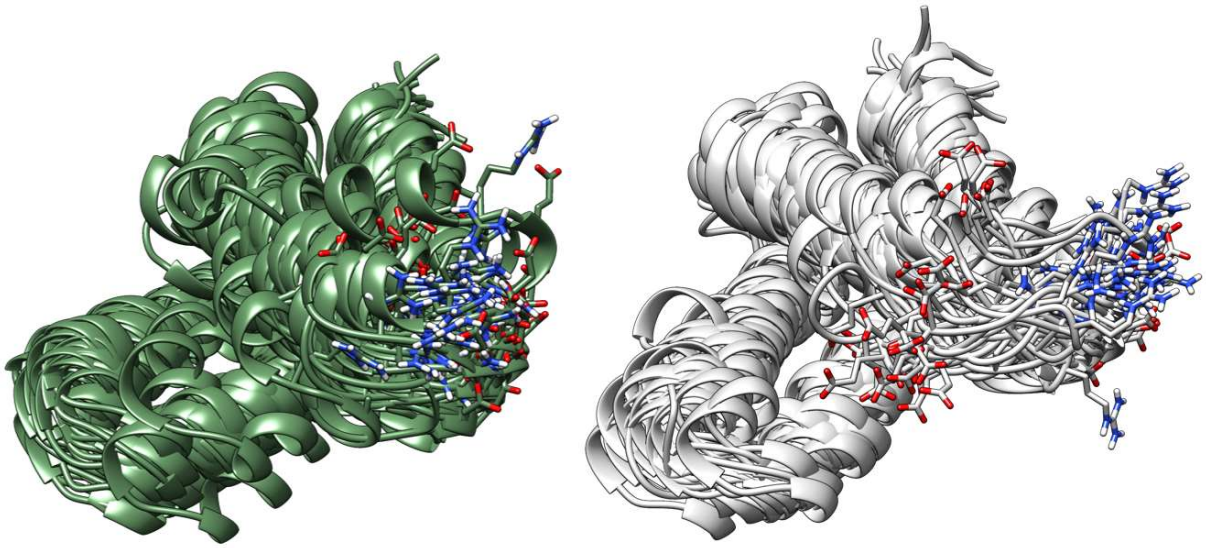


Figure 5.10. A sampling of structures at the two minima outside of the global minima occupied primarily by the Ala21Glu mutant as in Figure 5.10.

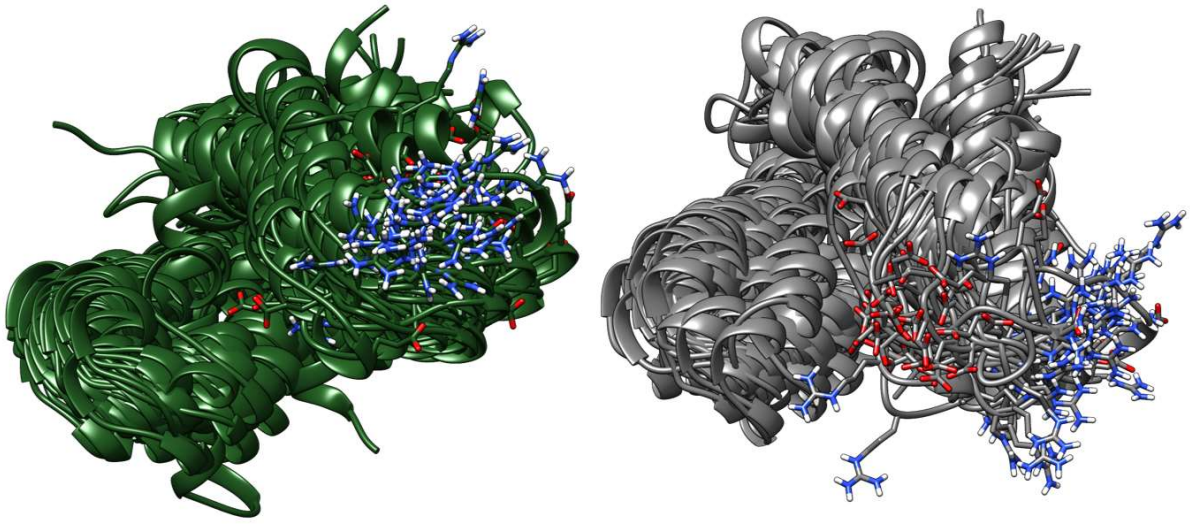


Figure 5.11. A sampling of structures at the two minima outside of the global minima occupied primarily by the peptides without the Ala21Glu mutation as in Figure 5.10.

5.4. TESTING FOR METAL-BINDING AND RETAINMENT

So far, we have increased confidence in the evolved peptide's tertiary structure, fold stability, and in its limited competing metal binding sites. A full QM/DMD procedure (discussed in depth previously) on the protein in the context of an iridium(III) – pentamethylcyclopentadiene (Cp*) complex gives the final link towards testing catalysis. Previously, iridium(III) showed great adaptability inside of an enzyme by yielding a lower activation energy for a hydroamination reaction than typical organic catalysts. This was due in part to the protein's imposed electrostatic environment (Chapter 4). Here, a form of iridium(III) which is more readily substituted will aim to be bound near the interior of this small peptide. To prepare this system for QM/DMD, an ensemble of DMD snapshots were taken from the promising peptide mutants and Ir(III)-Cp* was placed in the vicinity of the binding location of interest (see left snapshot of Figure 5.12.). Then, the QM/DMD procedure goes to work in rapidly equilibrating the structure via an iterative method.

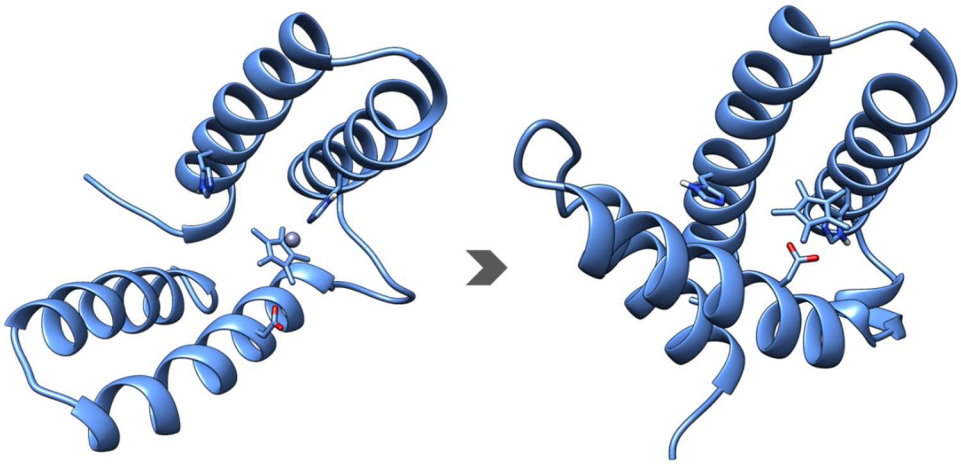


Figure 5.12 The peptide structure on the left was one of many DMD structures which give an accurate depiction of how the peptide may behave in solution. QM/DMD quickly came to equilibrium by binding two residues to iridium(III).

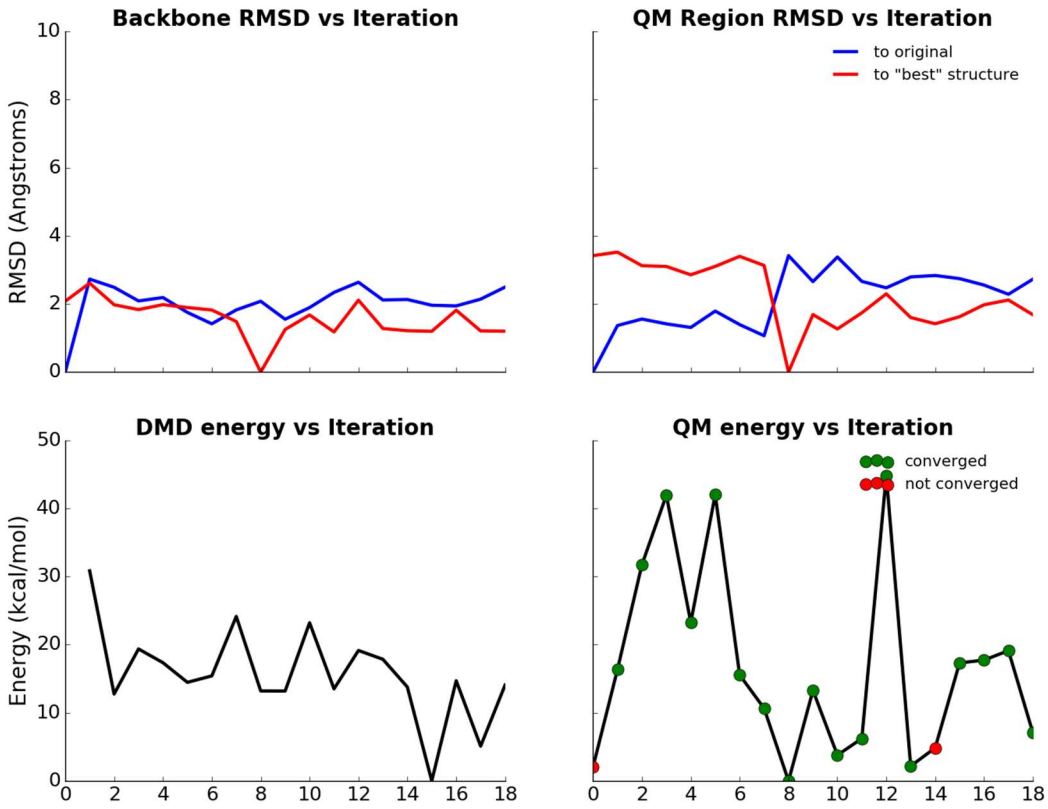


Figure 5.13. Equilibration of the peptide was rapid and reasonable structures were discovered within only 8 iterations of QM/DMD.

Out of a total of 16 structures, all starting from different poses and placement of the metal, the lowest energy structures were fairly unanimous in using His61 and Asp35 to bind to the metal (see Figure 5.14. below) This arrangement leaves room for one more site and likely with enough protein around the metal which can be manipulated for catalytic activity or substrate specificity.

The catalytic activity we hope to achieve stems from previous work by the Hilvert lab,[227] where a tripeptide binding Ir(III)-Cp* showed minor hydrogenase activity. The concept of a rare

nonphysiological metal such as iridium directly bound to a biological molecule as a cofactor for catalysis is still a very novel idea.

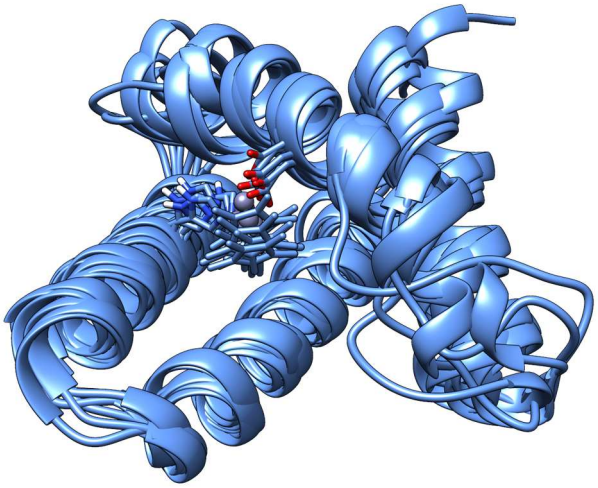


Figure 5.14. Several independent QM/DMD runs showing convergence in the binding mode of iridium(III) to the structure.

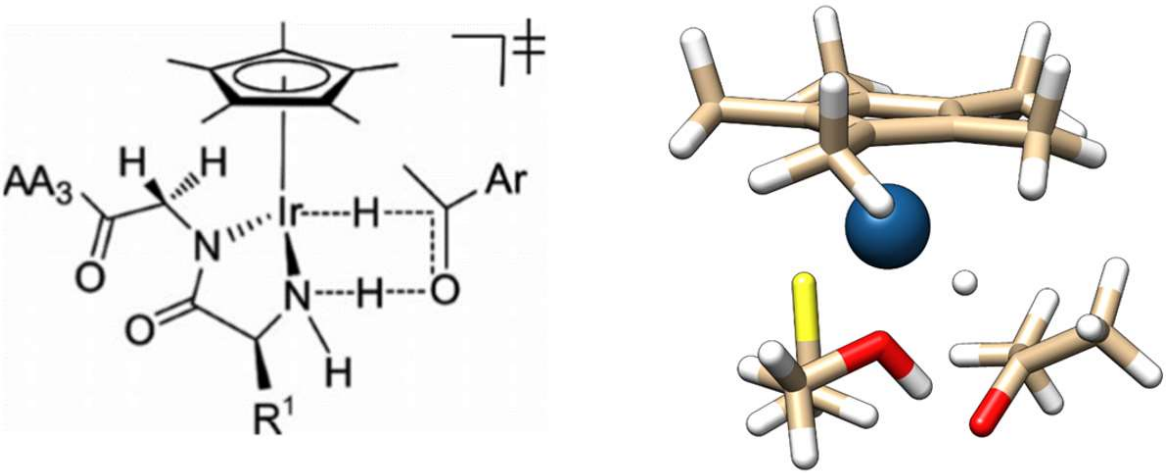


Figure 5.15. (Left) A theorized model of the tripeptide complex used by the Hilvert lab for a transfer hydrogenation reaction and (Right) a transition state complex for the same reaction, determined experimentally using only ligands found in proteins.

CHAPTER 6

APPENDIX

6.1. SUPPORTING INFORMATION FOR CHAPTER 2

XYZ coordinates of equilibrated reactant states (lowest QM-energy structures obtained from QM/DMD trajectories)

Co^{2+}

H	-0.7091596	-4.8591872	3.2845582
C	-0.5230898	-4.5844874	2.2532488
C	0.7945745	-3.9053774	2.0520402
N	1.3776761	-3.8204424	0.7997691
C	1.6612764	-3.2392696	2.9064872
C	2.5356204	-3.1334002	0.9262125
N	2.7348784	-2.7560500	2.1873146
H	0.9499303	-4.1016441	-0.1416903
H	-0.5727643	-5.5101475	1.6509742
H	-1.3627817	-3.9441726	1.9184819
H	1.5855286	-3.1009981	3.9860068
H	3.2222642	-2.9556115	0.0980455
H	2.3080388	-6.7834862	5.7034268
C	3.2847382	-6.4432764	6.0167766
C	4.0129441	-5.6464905	4.9789997
N	4.1121425	-6.0802458	3.6764314
C	4.7820615	-4.4913550	5.0406013
C	4.9096456	-5.2157983	3.0054639

N	5.3479505	-4.2338354	3.8015781
H	3.6448620	-6.9788532	3.3442653
H	3.1796968	-5.8567207	6.9454451
H	3.8541723	-7.3651506	6.2462303
H	4.9400029	-3.8383361	5.8997841
H	5.1105963	-5.3330057	1.9408551
H	-0.5415696	-3.3970680	-4.4447576
C	-0.1808598	-4.1972076	-3.8231578
C	0.5305633	-3.7297752	-2.5550814
O	1.3658347	-2.7726287	-2.6485193
O	0.2698709	-4.4027967	-1.5147827
H	0.5437115	-4.7693030	-4.4337465
H	-1.0119751	-4.8751872	-3.5725673
H	1.8977490	-0.1090000	-5.5861568
C	2.6112586	0.2730698	-4.8524872
O	3.4447946	1.1197679	-5.1907918
N	2.4974311	-0.1891340	-3.5764710
C	3.5004374	0.2135437	-2.5807378
C	4.8788034	-0.3893573	-2.9806205
O	4.9471635	-1.3649877	-3.7455573
C	2.9666251	-0.1857498	-1.1755791
C	3.9335948	-0.0515756	-0.0033090
O	3.7090033	0.7109920	0.9705822
O	4.9852283	-0.7929508	-0.0083162
H	1.8975082	-1.0032266	-3.3610853
H	3.6104940	1.3109298	-2.6278894
H	2.6390872	-1.2400603	-1.2215028

H	2.0774910	0.4294579	-0.9679962
H	8.0447554	0.1959598	-2.2970188
N	5.9897750	0.1940099	-2.4887833
O	8.3515790	-2.5133408	-3.1035625
C	7.5209081	-1.8374516	-2.4794493
H	5.9608254	0.9562405	-1.7745632
C	7.2898360	-0.3556098	-2.8349784
H	7.4876863	-0.2656621	-3.9170205
N	6.8550981	-2.2767107	-1.3855164
C	7.0865682	-3.5834286	-0.8188773
C	6.0267825	-4.5899866	-1.3241768
O	4.9461189	-4.2209592	-1.8032002
C	7.0932433	-3.4823466	0.7333446
C	8.0893554	-2.4721731	1.2327524
N	7.7519532	-1.1566601	1.5473461
C	9.4552115	-2.6189205	1.3992840
C	8.8784055	-0.5328346	1.8862116
N	9.9300542	-1.3869398	1.8115783
H	6.0975370	-1.6987887	-0.9752488
H	10.8991805	-1.1572599	2.0249494
H	8.0729272	-3.9235021	-1.1827692
H	7.3174840	-4.4725024	1.1638384
H	6.0937927	-3.1780094	1.0803466
H	10.1140604	-3.4738679	1.2562796
H	8.9444511	0.5143238	2.1792408
H	5.9194366	-7.8607456	-1.7443490
N	6.3393584	-5.9093764	-1.2496469

O	4.2034172	-6.9189760	0.3874693
C	4.1340423	-6.9552884	-0.8427208
H	7.2014962	-6.2080203	-0.7987956
C	5.3901570	-6.9192562	-1.7193990
H	5.1076457	-6.6792173	-2.7605016
N	2.9766983	-7.0334977	-1.5479761
C	1.7013021	-7.3000387	-0.8862991
C	1.6641628	-8.8022081	-0.5440940
O	1.2961524	-9.6536700	-1.3722133
H	3.0220609	-7.0032906	-2.5704433
H	0.8815326	-7.0283623	-1.5645196
H	1.6404394	-6.6687850	0.0153136
H	1.6063490	-10.6777240	1.8256590
N	2.1782282	-9.0675794	0.6716102
C	2.4445686	-10.3905942	1.2221494
C	3.7291212	-10.3794195	2.0633152
C	3.7180025	-9.3869325	3.2586587
O	4.4814795	-9.6039148	4.2197504
O	2.9309143	-8.3732425	3.1232629
H	2.4287540	-8.3193951	1.3407244
H	2.5111270	-11.1024046	0.3804751
H	4.5814027	-10.0872801	1.4184993
H	3.9439331	-11.3927444	2.4456046
H	1.7735990	-6.6652562	-6.1462066
C	2.0946388	-7.1276260	-5.2265470
O	3.1891161	-6.4140984	-4.6146198
H	4.0067446	-6.6788004	-5.0747467

H	2.3618081	-8.1833115	-5.4123312
H	1.2594966	-7.1056597	-4.5087185
H	4.6767974	-4.3759876	-7.9091156
C	5.2893070	-3.9249578	-7.1434560
O	4.8171485	-3.0619135	-6.3955753
H	7.0954860	-4.2313176	-6.2052966
N	6.5004164	-4.4741174	-7.0133760
H	6.8211599	-5.2089808	-7.6409324
H	7.4986058	4.4701774	-0.2574698
C	7.7040756	3.7413378	0.5462596
C	6.8586584	2.4624951	0.3520638
O	6.3122645	2.2795463	-0.7596784
O	6.8058312	1.6632952	1.3651719
H	8.7761305	3.4736152	0.5357913
H	7.4818988	4.1992997	1.5234162
H	0.7360996	2.7154584	2.9971984
C	1.6626890	3.1528582	3.3529482
O	1.6498298	4.0052759	4.2541664
H	4.2115176	4.2583576	2.7577484
N	2.7809913	2.7162755	2.7451105
C	4.0717578	3.2580982	3.1365382
H	2.7708207	1.9917231	2.0141178
H	4.8608334	2.6043533	2.7386542
H	4.1527528	3.2922787	4.2363838
H	8.8453850	7.5318458	4.9068372
C	9.5478446	6.7386462	5.0986072
C	8.8117203	5.4345447	5.2932816

C	8.5378925	4.5851420	4.2036962
C	8.3035501	5.0688763	6.5559349
C	7.7640546	3.4261996	4.3467065
C	7.5322994	3.9141421	6.7238502
C	7.2442616	3.0931070	5.6155084
O	6.4688775	1.9998976	5.8143086
H	6.3002941	1.5463119	4.9398150
H	10.2411394	6.6910747	4.2403996
H	10.1297758	7.0176578	5.9951012
H	8.9308306	4.8432962	3.2128133
H	8.5054429	5.7080056	7.4246242
H	7.5399214	2.7869813	3.4823613
H	7.1297920	3.6360326	7.7037094
H	8.8968350	-2.3677186	6.2765764
C	7.8532056	-2.1791488	6.4094064
O	5.7706477	0.4380241	3.7987952
C	5.1618012	-0.3305869	4.6181889
C	3.6631903	-0.2771412	4.7489046
N	5.8598925	-1.1101805	5.4435475
H	5.3807472	-1.6601134	6.1700495
C	7.3143136	-1.2611506	5.3167256
H	3.2304088	-0.2530273	3.7382651
H	3.2613187	-1.1314456	5.3148710
H	3.3933252	0.6649017	5.2591573
H	7.7790245	-0.2612116	5.3666064
H	7.5262930	-1.6706496	4.3121614
H	7.7020622	-1.7366399	7.4090854

H	7.3354413	-3.1533360	6.3808491
O	5.1300828	-1.8604183	2.5710437
H	4.1381604	-2.1285322	2.4485804
H	5.4820274	-2.6601844	3.0787317
Co	5.9252621	-0.1529061	1.7017240
K	3.6160630	-3.5281105	-3.9350245
C	5.4175302	-3.1939191	8.7233546
O	4.9405269	-2.5469627	7.8085429
N	5.3349970	-2.8549984	10.0039944
H	5.7378873	-3.4495743	10.7282207
H	5.9559966	-4.1099976	8.5299952
H	4.8371564	-2.0093200	10.2830672

Fe^{2+}

H	-5.4987569	-2.4957286	2.0776899
C	-5.3126870	-2.2210288	1.0463805
C	-3.9817093	-1.5684677	0.8543183
N	-3.3973797	-1.4599479	-0.3957346
C	-3.0992219	-0.9515022	1.7283794
C	-2.2220296	-0.8064257	-0.2482908
N	-2.0145475	-0.4738033	1.0240591
H	-3.8283951	-1.7176546	-1.3421310
H	-5.3822205	-3.1423844	0.4393702
H	-6.1390907	-1.5612317	0.7167756
H	-3.1718434	-0.8450081	2.8116908
H	-1.5311661	-0.6188362	-1.0707971
H	-2.4815585	-4.4200276	4.4965585
C	-1.5048591	-4.0798178	4.8099083
C	-0.7511629	-3.3090751	3.7715711
N	-0.6370546	-3.7608636	2.4768435
C	0.0423137	-2.1705483	3.8343198
C	0.1923834	-2.9230022	1.8108683
N	0.6377514	-1.9415011	2.6037822
H	-1.1185744	-4.6532656	2.1475392
H	-1.6147691	-3.4757816	5.7269351
H	-0.9458787	-5.0014764	5.0654798
H	0.1967359	-1.5087980	4.6874905
H	0.4085861	-3.0559501	0.7509878
H	-5.3311669	-1.0336095	-5.6516259
C	-4.9704571	-1.8337490	-5.0300261

C	-4.2548283	-1.3605252	-3.7656554
O	-3.4102585	-0.4128683	-3.8718860
O	-4.5217050	-2.0154548	-2.7157908
H	-4.2488883	-2.4086116	-5.6412722
H	-5.8028626	-2.5083879	-4.7745487
H	-2.8918483	2.2544586	-6.7930251
C	-2.1783387	2.6365284	-6.0593555
O	-1.3478883	3.4878109	-6.3936366
N	-2.2898685	2.1649321	-4.7866500
C	-1.2902902	2.5594201	-3.7831207
C	0.0913407	1.9680573	-4.1873461
O	0.1640800	0.9957288	-4.9565506
C	-1.8276505	2.1314933	-2.3878143
C	-0.8862468	2.2693676	-1.1952416
O	-1.1816385	2.9650715	-0.1914639
O	0.2128391	1.5963873	-1.2158395
H	-2.8905663	1.3495257	-4.5777511
H	-1.1865713	3.6581748	-3.8121425
H	-2.1334543	1.0711421	-2.4522406
H	-2.7339411	2.7222969	-2.1827764
H	3.2551582	2.5594184	-3.5038871
N	1.1997739	2.5551120	-3.6942408
O	3.5620873	-0.1474078	-4.3226539
C	2.7330301	0.5256972	-3.6930664
H	1.1690720	3.3048533	-2.9668357
C	2.5002387	2.0078488	-4.0418467
H	2.6966565	2.1011054	-5.1239301

N	2.0696349	0.0806623	-2.5999327
C	2.3084696	-1.2282938	-2.0429189
C	1.2422128	-2.2299930	-2.5365573
O	0.1584373	-1.8600022	-3.0079085
C	2.3432497	-1.1408491	-0.4902490
C	3.3358758	-0.1204790	-0.0089677
N	2.9875048	1.2033839	0.2512541
C	4.7008749	-0.2555266	0.1732278
C	4.1087397	1.8454917	0.5742651
N	5.1661535	0.9930903	0.5443456
H	1.3181619	0.6607568	-2.1821457
H	6.1317457	1.2344399	0.7604643
H	3.2887871	-1.5654404	-2.4251704
H	2.5868711	-2.1312651	-0.0714986
H	1.3465194	-0.8516714	-0.1221051
H	5.3653305	-1.1115704	0.0670575
H	4.1664332	2.9037213	0.8277047
H	1.1298394	-5.4972870	-2.9512173
N	1.5522130	-3.5490834	-2.4595480
O	-0.5822603	-4.5616830	-0.8175585
C	-0.6535675	-4.5918320	-2.0477791
H	2.4195080	-3.8488491	-2.0194223
C	0.6005597	-4.5557976	-2.9262673
H	0.3161651	-4.3165613	-3.9671308
N	-1.8123977	-4.6636278	-2.7512714
C	-3.0864054	-4.9313761	-2.0875099
C	-3.1218848	-6.4340339	-1.7471423

O	-3.4872517	-7.2842177	-2.5776861
H	-1.7683264	-4.6345763	-3.7737843
H	-3.9074361	-4.6604679	-2.7646066
H	-3.1465029	-4.3006315	-1.1855458
H	-3.1832482	-8.3142654	0.6187907
N	-2.6099852	-6.7019229	-0.5308313
C	-2.3450287	-8.0271356	0.0152810
C	-1.0579268	-8.0268395	0.8514658
C	-1.0654004	-7.0465325	2.0557514
O	-0.2991810	-7.2708771	3.0128402
O	-1.8535956	-6.0321546	1.9308706
H	-2.3584449	-5.9551833	0.1385721
H	-2.2838468	-8.7352972	-0.8299736
H	-0.2073175	-7.7295654	0.2067874
H	-0.8437701	-9.0445619	1.2223550
H	-3.0159983	-4.3017976	-7.3530749
C	-2.6949585	-4.7641674	-6.4334153
O	-1.5966740	-4.0546751	-5.8239518
H	-0.7807136	-4.3245405	-6.2839970
H	-2.4326241	-5.8211965	-6.6184357
H	-3.5290620	-4.7382811	-5.7144836
H	-0.1127999	-2.0125290	-9.1159839
C	0.4997098	-1.5614992	-8.3503243
O	0.0277744	-0.6975110	-7.6035686
H	2.3058887	-1.8678590	-7.4121649
N	1.7108191	-2.1106589	-8.2202443
H	2.0307539	-2.8464974	-8.8471358

H	2.7090085	6.8336360	-1.4643381
C	2.9144784	6.1047964	-0.6606087
C	2.0656698	4.8274887	-0.8273554
O	1.5195791	4.6199023	-1.9363252
O	2.0105401	4.0539196	0.2032743
H	3.9858023	5.8333349	-0.6820750
H	2.7088085	6.5714234	0.3160283
H	-4.0534977	5.0789170	1.7903300
C	-3.1269082	5.5163168	2.1460798
O	-3.1391597	6.3755022	3.0404493
H	-0.5780796	6.6218162	1.5508801
N	-2.0080619	5.0757568	1.5404874
C	-0.7178395	5.6215568	1.9296699
H	-2.0212765	4.3356844	0.8247482
H	0.0729925	4.9719754	1.5277954
H	-0.6328364	5.6551212	3.0293021
H	4.0557878	9.8953044	3.6999689
C	4.7582474	9.1021048	3.8917389
C	4.0225921	7.7973114	4.0827388
C	3.7503222	6.9505908	2.9907208
C	3.5132494	7.4283845	5.3438361
C	2.9754824	5.7918714	3.1292810
C	2.7410412	6.2735599	5.5075508
C	2.4538361	5.4567868	4.3962697
O	1.6761425	4.3629829	4.5908215
H	1.5105232	3.9132680	3.7160916
H	5.4530260	9.0561536	3.0345560

H	5.3388748	9.3797775	4.7894889
H	4.1452614	7.2105428	2.0011090
H	3.7150123	8.0648158	6.2144851
H	2.7523745	5.1552562	2.2623073
H	2.3375780	5.9926925	6.4861139
H	4.1072377	-0.0042600	5.0697081
C	3.0636083	0.1843098	5.2025381
O	0.9720216	2.7834962	2.5837898
C	0.3625311	2.0182340	3.4146814
C	-1.1344755	2.0710338	3.5378941
N	1.0676194	1.2573000	4.2468849
H	0.5918314	0.7079927	4.9771558
C	2.5221207	1.1035882	4.1117041
H	-1.5618823	2.0769413	2.5241235
H	-1.5359994	1.2252825	4.1163946
H	-1.4089107	3.0213340	4.0298944
H	2.9880069	2.1029297	4.1590123
H	2.7257621	0.6931759	3.1054297
H	2.9127251	0.6254501	6.2027870
H	2.5466621	-0.7901636	5.1722253
O	0.3682426	0.4359319	1.4122288
H	-0.6234963	0.1528227	1.3093235
H	0.7445611	-0.3709078	1.8904265
Fe	1.1142851	2.1944183	0.5542483
K	-1.1614471	-1.1735171	-5.1459663
C	0.6280834	-0.8304768	7.5166628
O	0.1509296	-0.1835041	6.6016746

N	0.5453997	-0.4915398	8.7971261
H	0.9482900	-1.0861157	9.5213524
H	1.1663994	-1.7465391	7.3231269
H	0.0482839	0.3546028	9.0759754

Mg^{2+}

H	-0.7091598	-4.8591886	3.2845591
C	-0.5230899	-4.5844887	2.2532494
C	0.7996937	-3.9194359	2.0562258
N	1.3793377	-3.8165491	0.8039423
C	1.6767061	-3.2842015	2.9223606
C	2.5473309	-3.1486806	0.9411060
N	2.7539705	-2.8018177	2.2099793
H	0.9486331	-4.0854157	-0.1399161
H	-0.5832249	-5.5084942	1.6491831
H	-1.3567542	-3.9351234	1.9209194
H	1.6064128	-3.1682401	4.0048652
H	3.2349278	-2.9626102	0.1153722
H	2.3080394	-6.7834881	5.7034284
C	3.2847391	-6.4432782	6.0167783
C	4.0261536	-5.6574718	4.9812708
N	4.1359438	-6.0962300	3.6813805
C	4.8078093	-4.5112808	5.0495430
C	4.9518565	-5.2428746	3.0179114
N	5.3917690	-4.2641663	3.8172049
H	3.6590092	-6.9879770	3.3457051
H	3.1775893	-5.8501562	6.9411762
H	3.8484450	-7.3661876	6.2568427
H	4.9603943	-3.8554099	5.9077753
H	5.1638835	-5.3641848	1.9557877
H	-0.5415698	-3.3970690	-4.4447588
C	-0.1808599	-4.1972088	-3.8231589

C	0.5270477	-3.7235428	-2.5546224
O	1.3614563	-2.7657130	-2.6503794
O	0.2636697	-4.3903103	-1.5110574
H	0.5457080	-4.7684382	-4.4320067
H	-1.0114105	-4.8757782	-3.5724025
H	1.8977495	-0.1090000	-5.5861584
C	2.6112593	0.2730699	-4.8524886
O	3.4468844	1.1165397	-5.1940085
N	2.4955955	-0.1821692	-3.5731699
C	3.4992329	0.2220316	-2.5763830
C	4.8772263	-0.3864772	-2.9748334
O	4.9411125	-1.3648330	-3.7367385
C	2.9586278	-0.1597504	-1.1665150
C	3.9391936	-0.0868677	0.0000708
O	3.7024459	0.6023205	1.0381860
O	5.0045422	-0.7907406	-0.0555772
H	1.8960338	-0.9961319	-3.3566713
H	3.6145066	1.3184033	-2.6331146
H	2.5929823	-1.2024702	-1.2142783
H	2.0932292	0.4848891	-0.9485082
H	8.0447577	0.1959599	-2.2970194
N	5.9909946	0.2005835	-2.4938995
O	8.3611147	-2.5058774	-3.1088828
C	7.5244693	-1.8376010	-2.4833354
H	5.9613514	0.9579183	-1.7753966
C	7.2898380	-0.3556099	-2.8349792
H	7.4905900	-0.2648610	-3.9166417

N	6.8568425	-2.2831933	-1.3951392
C	7.0961703	-3.5866930	-0.8255053
C	6.0299381	-4.5928127	-1.3174312
O	4.9448244	-4.2239616	-1.7869065
C	7.1254801	-3.4782711	0.7270167
C	8.1167255	-2.4501649	1.2028256
N	7.7620990	-1.1389838	1.5124695
C	9.4884089	-2.5766476	1.3416575
C	8.8851191	-0.4974946	1.8233467
N	9.9516694	-1.3338084	1.7347012
H	6.0980417	-1.7052020	-0.9841279
H	10.9210442	-1.0865092	1.9256528
H	8.0780358	-3.9279305	-1.2004601
H	7.3681208	-4.4639190	1.1585193
H	6.1275969	-3.1853969	1.0889045
H	10.1583994	-3.4216981	1.1914998
H	8.9436924	0.5536708	2.1058863
H	5.9194383	-7.8607478	-1.7443495
N	6.3412930	-5.9126056	-1.2472225
O	4.2005551	-6.9208979	0.3863183
C	4.1333179	-6.9558821	-0.8439214
H	7.2089555	-6.2137020	-0.8086808
C	5.3901585	-6.9192581	-1.7193995
H	5.1086642	-6.6778248	-2.7605385
N	2.9770825	-7.0350284	-1.5513251
C	1.7007800	-7.3040520	-0.8925559
C	1.6664842	-8.8059035	-0.5481372

O	1.3034486	-9.6594785	-1.3762825
H	3.0244146	-7.0035508	-2.5736649
H	0.8821675	-7.0365446	-1.5739204
H	1.6348536	-6.6715229	0.0078134
H	1.6063495	-10.6777270	1.8256595
N	2.1766927	-9.0685562	0.6698460
C	2.4445693	-10.3905971	1.2221497
C	3.7280220	-10.3774094	2.0651681
C	3.7109174	-9.3884819	3.2629681
O	4.4684101	-9.6084355	4.2282666
O	2.9263047	-8.3732383	3.1249266
H	2.4254663	-8.3187437	1.3376480
H	2.5129994	-11.1032696	0.3814042
H	4.5800747	-10.0796406	1.4226502
H	3.9459747	-11.3910247	2.4448508
H	1.7735995	-6.6652581	-6.1462083
C	2.0946394	-7.1276280	-5.2265485
O	3.1890584	-6.4142197	-4.6145358
H	4.0063891	-6.6767520	-5.0765048
H	2.3616188	-8.1833775	-5.4123464
H	1.2594751	-7.1055461	-4.5087580
H	4.6767987	-4.3759888	-7.9091178
C	5.2893085	-3.9249589	-7.1434580
O	4.8166330	-3.0631555	-6.3944426
H	7.0954880	-4.2313188	-6.2052983
N	6.5004182	-4.4741187	-7.0133780
H	6.8220257	-5.2076529	-7.6420349

H	7.4986079	4.4701787	-0.2574699
C	7.7040778	3.7413389	0.5462598
C	6.8399023	2.4724568	0.3680613
O	6.2914267	2.2909716	-0.7447294
O	6.7624211	1.6826511	1.3838574
H	8.7727251	3.4601482	0.5239222
H	7.4994281	4.2050222	1.5243685
H	0.7360998	2.7154592	2.9971992
C	1.6626895	3.1528591	3.3529491
O	1.6507769	4.0148846	4.2443046
H	4.2115188	4.2583588	2.7577492
N	2.7822709	2.7096909	2.7493453
C	4.0717589	3.2580991	3.1365391
H	2.7723540	1.9660537	2.0368874
H	4.8643445	2.6134049	2.7321577
H	4.1603425	3.2903653	4.2357978
H	8.8453875	7.5318479	4.9068386
C	9.5478473	6.7386481	5.0986086
C	8.8082513	5.4360313	5.2930870
C	8.5190305	4.5955071	4.2005501
C	8.3101707	5.0645913	6.5580563
C	7.7390326	3.4404011	4.3425345
C	7.5329240	3.9136028	6.7251541
C	7.2293282	3.1023012	5.6139094
O	6.4481387	2.0127052	5.8140554
H	6.2709168	1.5596217	4.9421560
H	10.2405164	6.6906408	4.2398898

H	10.1301742	7.0171835	5.9949378
H	8.9046574	4.8579462	3.2078964
H	8.5251986	5.6959381	7.4292815
H	7.5039624	2.8078968	3.4756694
H	7.1383317	3.6310791	7.7069769
H	8.8968375	-2.3677193	6.2765782
C	7.8532078	-2.1791494	6.4094082
O	5.7112049	0.4284250	3.8283841
C	5.1292939	-0.3352688	4.6756722
C	3.6350297	-0.2894956	4.8443653
N	5.8503433	-1.1061343	5.4830668
H	5.3899965	-1.6564274	6.2245205
C	7.3029537	-1.2554197	5.3271595
H	3.1730374	-0.4013198	3.8512868
H	3.2648091	-1.0726130	5.5236278
H	3.3606876	0.7056581	5.2361318
H	7.7668745	-0.2550182	5.3747927
H	7.4974170	-1.6607741	4.3174704
H	7.7039636	-1.7428722	7.4121863
H	7.3363067	-3.1535886	6.3772318
O	5.1340204	-1.8936593	2.6004590
H	4.1450790	-2.1838020	2.4876891
H	5.4908064	-2.6903869	3.1074542
Mg	5.8151256	-0.1228802	1.7844067
K	3.6195172	-3.5286447	-3.9231256
C	5.4339985	-3.2089991	8.7399975
O	4.9405283	-2.5469634	7.8085451

N	5.3349985	-2.8549992	10.0039972
H	5.7378889	-3.4495753	10.7282237
H	5.9559983	-4.1099988	8.5299976
H	4.8299986	-2.0139994	10.3499971

Mn^{2+}

H	-0.7091598	-4.8591886	3.2845591
C	-0.5230899	-4.5844887	2.2532494
C	0.8129946	-3.9367611	2.0648712
N	1.3914210	-3.8107703	0.8132734
C	1.7067507	-3.3417432	2.9435297
C	2.5730445	-3.1691934	0.9636290
N	2.7919920	-2.8604383	2.2405317
H	0.9566205	-4.0600312	-0.1331661
H	-0.5955474	-5.5046145	1.6447306
H	-1.3456927	-3.9200153	1.9240532
H	1.6426638	-3.2531886	4.0289272
H	3.2596175	-2.9685944	0.1404869
H	2.3080394	-6.7834881	5.7034284
C	3.2847391	-6.4432782	6.0167783
C	4.0461992	-5.6775230	4.9822394
N	4.1719022	-6.1301152	3.6887514
C	4.8414406	-4.5414220	5.0538906
C	5.0093303	-5.2939082	3.0309219
N	5.4493679	-4.3145879	3.8296145
H	3.6891334	-7.0195839	3.3545801
H	3.1728310	-5.8358629	6.9314265
H	3.8403348	-7.3654117	6.2786013
H	4.9864301	-3.8773818	5.9069595
H	5.2340360	-5.4242123	1.9725031
H	-0.5415698	-3.3970690	-4.4447588
C	-0.1808599	-4.1972088	-3.8231589

C	0.5284062	-3.7161534	-2.5580798
O	1.3701379	-2.7658314	-2.6668246
O	0.2592469	-4.3644279	-1.5050275
H	0.5445377	-4.7699742	-4.4316915
H	-1.0119826	-4.8737778	-3.5687789
H	1.8977495	-0.1090000	-5.5861584
C	2.6112593	0.2730699	-4.8524886
O	3.4418940	1.1246452	-5.1863311
N	2.4975879	-0.1984172	-3.5821918
C	3.4910707	0.2013357	-2.5778133
C	4.8755805	-0.3980989	-2.9710439
O	4.9486564	-1.3664681	-3.7443763
C	2.9288929	-0.1994865	-1.1848366
C	3.8909228	-0.1157372	-0.0085685
O	3.6603749	0.6216411	0.9914368
O	4.9340687	-0.8594218	-0.0375857
H	1.8967108	-1.0134736	-3.3725488
H	3.6028840	1.2988052	-2.6218119
H	2.5749553	-1.2450176	-1.2487962
H	2.0540658	0.4369115	-0.9802977
H	8.0447577	0.1959599	-2.2970194
N	5.9871593	0.1868509	-2.4788220
O	8.3690665	-2.5016032	-3.1097890
C	7.5288813	-1.8362478	-2.4863516
H	5.9551294	0.9357394	-1.7533254
C	7.2898380	-0.3556099	-2.8349792
H	7.4807714	-0.2581676	-3.9174749

N	6.8590979	-2.2856819	-1.3998813
C	7.1015334	-3.5929128	-0.8409334
C	6.0335367	-4.5950063	-1.3301525
O	4.9514752	-4.2266818	-1.8064264
C	7.1542177	-3.4987861	0.7112511
C	8.1621763	-2.4813497	1.1692477
N	7.8187737	-1.1664093	1.4697301
C	9.5350740	-2.6125265	1.2868935
C	8.9485711	-0.5257022	1.7586110
N	10.0095973	-1.3679235	1.6604860
H	6.0913431	-1.7188128	-0.9921669
H	10.9828065	-1.1238758	1.8363667
H	8.0789684	-3.9307658	-1.2300586
H	7.3976768	-4.4887456	1.1318355
H	6.1652673	-3.2010674	1.0939186
H	10.1993337	-3.4615088	1.1334108
H	9.0143251	0.5271757	2.0328342
H	5.9194383	-7.8607478	-1.7443495
N	6.3429747	-5.9142141	-1.2515684
O	4.2114608	-6.9225505	0.3919054
C	4.1372353	-6.9537088	-0.8381410
H	7.2071244	-6.2150023	-0.8060121
C	5.3901585	-6.9192581	-1.7193995
H	5.1051802	-6.6786500	-2.7596407
N	2.9763923	-7.0243516	-1.5386822
C	1.7031686	-7.2889748	-0.8719418
C	1.6638285	-8.7919869	-0.5333030

O	1.2901730	-9.6398232	-1.3625884
H	3.0176091	-6.9930210	-2.5612680
H	0.8811725	-7.0148334	-1.5466214
H	1.6470402	-6.6594667	0.0312576
H	1.6063495	-10.6777270	1.8256595
N	2.1824913	-9.0634540	0.6793098
C	2.4445693	-10.3905971	1.2221497
C	3.7312548	-10.3934166	2.0593822
C	3.7236931	-9.4159107	3.2662136
O	4.4827344	-9.6480056	4.2271698
O	2.9436649	-8.3954684	3.1392691
H	2.4377952	-8.3188545	1.3500416
H	2.5047698	-11.0973215	0.3756775
H	4.5823497	-10.0947542	1.4159651
H	3.9444172	-11.4121657	2.4279068
H	1.7735995	-6.6652581	-6.1462083
C	2.0946394	-7.1276280	-5.2265485
O	3.1854022	-6.4111987	-4.6114998
H	4.0051330	-6.6747652	-5.0684831
H	2.3660705	-8.1820227	-5.4134733
H	1.2580068	-7.1100315	-4.5103232
H	4.6767987	-4.3759888	-7.9091178
C	5.2893085	-3.9249589	-7.1434580
O	4.8174568	-3.0614241	-6.3960171
H	7.0954880	-4.2313188	-6.2052983
N	6.5004182	-4.4741187	-7.0133780
H	6.8209039	-5.2094473	-7.6405217

H	7.4986079	4.4701787	-0.2574699
C	7.7040778	3.7413389	0.5462598
C	6.8555446	2.4679137	0.3925851
O	6.2899876	2.2539781	-0.7033667
O	6.8156043	1.6979542	1.4291438
H	8.7740722	3.4650124	0.5199532
H	7.5072705	4.2100562	1.5239838
H	0.7360998	2.7154592	2.9971992
C	1.6626895	3.1528591	3.3529491
O	1.6505021	4.0128744	4.2464734
H	4.2115188	4.2583588	2.7577492
N	2.7818794	2.7107134	2.7485337
C	4.0717589	3.2580991	3.1365391
H	2.7696482	1.9727887	2.0306940
H	4.8632107	2.6108517	2.7342214
H	4.1588738	3.2897562	4.2357127
H	8.8453850	7.5318458	4.9068372
C	9.5478446	6.7386462	5.0986072
C	8.7945532	5.4458137	5.3085539
C	8.5238721	4.5763969	4.2342472
C	8.2529608	5.1207760	6.5685085
C	7.7189956	3.4392145	4.3870715
C	7.4507876	3.9891231	6.7462405
C	7.1636945	3.1494536	5.6514470
O	6.3537105	2.0836235	5.8636012
H	6.1830936	1.6073747	5.0021834
H	10.2330638	6.6770783	4.2351733

H	10.1378830	7.0195690	5.9891594
H	8.9421313	4.8014733	3.2455429
H	8.4507183	5.7756887	7.4262756
H	7.4995315	2.7876569	3.5306994
H	7.0207215	3.7447280	7.7232703
H	8.8968375	-2.3677193	6.2765782
C	7.8532078	-2.1791494	6.4094082
O	5.6473006	0.4567012	3.9092527
C	5.0912621	-0.3343747	4.7461423
C	3.5950561	-0.3453553	4.9050700
N	5.8375440	-1.1012255	5.5377008
H	5.3934070	-1.6766210	6.2702720
C	7.2914380	-1.2118556	5.3714315
H	3.1472883	-0.4819265	3.9081009
H	3.2497768	-1.1396019	5.5845118
H	3.2760015	0.6400461	5.2872420
H	7.7393295	-0.2068774	5.4668287
H	7.4914757	-1.5642942	4.3427506
H	7.7016837	-1.7880919	7.4305908
H	7.3365853	-3.1518926	6.3358884
O	5.1751821	-1.9426256	2.6749455
H	4.1891370	-2.2395245	2.5413785
H	5.5422665	-2.7552829	3.1566321
Mn	5.8542129	-0.1397228	1.7991159
K	3.6211494	-3.5297471	-3.9348052
C	5.4339985	-3.2089991	8.7399975
O	4.9274550	-2.5545853	7.8105118

N	5.3349985	-2.8549992	10.0039972
H	5.7511467	-3.4415819	10.7272350
H	5.9559983	-4.1099988	8.5299976
H	4.8299986	-2.0139994	10.3499971

Ni^{2+}

H	-0.7091598	-4.8591886	3.2845591
C	-0.5230899	-4.5844887	2.2532494
C	0.7717022	-3.8720859	2.0467229
N	1.3712196	-3.8091790	0.8011368
C	1.5978601	-3.1484156	2.8924435
C	2.5021949	-3.0783703	0.9229440
N	2.6636872	-2.6528853	2.1735963
H	0.9651396	-4.1187760	-0.1414455
H	-0.5462149	-5.5157400	1.6579076
H	-1.3802518	-3.9715525	1.9107682
H	1.4989205	-2.9767447	3.9651475
H	3.1988020	-2.9096144	0.1013423
H	2.3080394	-6.7834881	5.7034284
C	3.2847391	-6.4432782	6.0167783
C	3.9829943	-5.6263548	4.9716827
N	4.0485486	-6.0480102	3.6624718
C	4.7306099	-4.4559274	5.0177825
C	4.8060313	-5.1626207	2.9728130
N	5.2497625	-4.1776039	3.7619556
H	3.5879582	-6.9525562	3.3362255
H	3.1855848	-5.8683005	6.9530463
H	3.8645514	-7.3634298	6.2239746
H	4.9036733	-3.8059904	5.8767413
H	4.9773226	-5.2705935	1.9020900
H	-0.5415698	-3.3970690	-4.4447588
C	-0.1808599	-4.1972088	-3.8231589

C	0.5449729	-3.7355847	-2.5599959
O	1.3670437	-2.7669690	-2.6521966
O	0.3096417	-4.4276271	-1.5258398
H	0.5352752	-4.7759796	-4.4375023
H	-1.0135562	-4.8703153	-3.5647096
H	1.8977495	-0.1090000	-5.5861584
C	2.6112593	0.2730699	-4.8524886
O	3.4476111	1.1149468	-5.1960138
N	2.4958284	-0.1776634	-3.5714425
C	3.5047402	0.2303731	-2.5822503
C	4.8800048	-0.3779648	-2.9853423
O	4.9415589	-1.3607761	-3.7419927
C	2.9786444	-0.1556403	-1.1693148
C	3.9497277	-0.0330561	0.0005704
O	3.6928345	0.6605816	1.0224041
O	5.0373453	-0.7124421	-0.0390861
H	1.8996339	-0.9935808	-3.3521276
H	3.6170121	1.3269879	-2.6391640
H	2.6412202	-1.2077159	-1.2072897
H	2.0947568	0.4648476	-0.9551699
H	8.0447577	0.1959599	-2.2970194
N	5.9930232	0.2089611	-2.5047164
O	8.3334368	-2.5290434	-3.0888222
C	7.5080376	-1.8409029	-2.4712356
H	5.9655127	0.9642358	-1.7821757
C	7.2898380	-0.3556099	-2.8349792
H	7.4969206	-0.2754102	-3.9162474

N	6.8314247	-2.2664251	-1.3797821
C	7.0439295	-3.5614609	-0.7781194
C	6.0002151	-4.5832942	-1.2960193
O	4.9123560	-4.2220521	-1.7653936
C	6.9809342	-3.4117042	0.7696320
C	7.9633038	-2.3952583	1.2883910
N	7.6245177	-1.0806654	1.6053690
C	9.3254898	-2.5452071	1.4846000
C	8.7420967	-0.4616193	1.9758601
N	9.7934398	-1.3181047	1.9165584
H	6.0959344	-1.6611909	-0.9689141
H	10.7577989	-1.0923242	2.1538110
H	8.0453392	-3.9100614	-1.0894395
H	7.1720228	-4.3897837	1.2410299
H	5.9684463	-3.0910601	1.0591691
H	9.9848875	-3.4009995	1.3499702
H	8.7990716	0.5817444	2.2822212
H	5.9194383	-7.8607478	-1.7443495
N	6.3290042	-5.9008072	-1.2394913
O	4.1832788	-6.9271003	0.3762524
C	4.1261652	-6.9640347	-0.8546246
H	7.1974357	-6.1937903	-0.7966842
C	5.3901585	-6.9192581	-1.7193995
H	5.1144354	-6.6779233	-2.7622523
N	2.9771818	-7.0540515	-1.5715690
C	1.6966481	-7.3275928	-0.9232185
C	1.6695853	-8.8270871	-0.5686659

O	1.3213341	-9.6893637	-1.3939974
H	3.0321470	-7.0164689	-2.5933820
H	0.8829620	-7.0702222	-1.6140637
H	1.6172758	-6.6894254	-0.0279762
H	1.6063495	-10.6777270	1.8256595
N	2.1684417	-9.0761290	0.6565878
C	2.4445693	-10.3905971	1.2221497
C	3.7254234	-10.3587972	2.0683274
C	3.6988098	-9.3561046	3.2547141
O	4.4671899	-9.5524008	4.2166027
O	2.8948908	-8.3568219	3.1119160
H	2.4066159	-8.3185744	1.3191782
H	2.5205912	-11.1105298	0.3882830
H	4.5774118	-10.0634130	1.4246830
H	3.9490889	-11.3664180	2.4605554
H	1.7735995	-6.6652581	-6.1462083
C	2.0946394	-7.1276280	-5.2265485
O	3.1961288	-6.4199621	-4.6203622
H	4.0096044	-6.6868733	-5.0865578
H	2.3538153	-8.1856364	-5.4104731
H	1.2622557	-7.0977982	-4.5058704
H	4.6767987	-4.3759888	-7.9091178
C	5.2893085	-3.9249589	-7.1434580
O	4.8168503	-3.0628724	-6.3945671
H	7.0954880	-4.2313188	-6.2052983
N	6.5004182	-4.4741187	-7.0133780
H	6.8218036	-5.2080564	-7.6416889

H	7.4986079	4.4701787	-0.2574699
C	7.7040778	3.7413389	0.5462598
C	6.8347728	2.4704274	0.3618692
O	6.3084830	2.2794177	-0.7591026
O	6.7455938	1.6990578	1.3912914
H	8.7716385	3.4570793	0.5190600
H	7.5018055	4.2034074	1.5254711
H	0.7360998	2.7154592	2.9971992
C	1.6626895	3.1528591	3.3529491
O	1.6494754	4.0078153	4.2515190
H	4.2115188	4.2583588	2.7577492
N	2.7810400	2.7166175	2.7449336
C	4.0717589	3.2580991	3.1365391
H	2.7704616	1.9822220	2.0239558
H	4.8624573	2.6056239	2.7400068
H	4.1525792	3.2931772	4.2364863
H	8.8453875	7.5318479	4.9068386
C	9.5478473	6.7386481	5.0986086
C	8.8157590	5.4367727	5.3134831
C	8.4978257	4.5985596	4.2273103
C	8.3601901	5.0596348	6.5922646
C	7.7339206	3.4366888	4.3904800
C	7.5998340	3.9007308	6.7809701
C	7.2713730	3.0883772	5.6777867
O	6.5144858	1.9865921	5.8945501
H	6.3346515	1.5406767	5.0173568
H	10.2318749	6.6837755	4.2332218

H	10.1402434	7.0258768	5.9855944
H	8.8493289	4.8681923	3.2240513
H	8.5972253	5.6911872	7.4576193
H	7.4749088	2.8049234	3.5301874
H	7.2397796	3.6112895	7.7739403
H	8.8968375	-2.3677193	6.2765782
C	7.8532078	-2.1791494	6.4094082
O	5.8196489	0.4846556	3.8402340
C	5.1860546	-0.2798963	4.6469547
C	3.6938276	-0.1623660	4.8100591
N	5.8655490	-1.0968461	5.4522495
H	5.3809177	-1.6450645	6.1771946
C	7.3099955	-1.2935711	5.2950772
H	3.2445649	-0.0506081	3.8130384
H	3.2613066	-1.0317366	5.3288246
H	3.4799959	0.7549463	5.3881167
H	7.7999871	-0.3047833	5.2908957
H	7.4872089	-1.7494695	4.3039070
H	7.7067349	-1.7062171	7.3959300
H	7.3352864	-3.1534922	6.4141892
O	5.0099865	-1.7526355	2.6274027
H	4.0153422	-2.0201781	2.4768138
H	5.3749428	-2.5861273	3.0690629
Ni	5.8256083	-0.1142704	1.7502264
K	3.6246357	-3.5293963	-3.9270980
C	5.4339985	-3.2089991	8.7399975
O	4.9427023	-2.5474247	7.8076568

N	5.3349985	-2.8549992	10.0039972
H	5.7401707	-3.4479817	10.7281928
H	5.9559983	-4.1099988	8.5299976
H	4.8299986	-2.0139994	10.3499971

Zn^{2+}

H	-0.7091596	-4.8591872	3.2845582
C	-0.5230898	-4.5844874	2.2532488
C	0.8090100	-3.9282925	2.0564669
N	1.3945751	-3.8397606	0.8051358
C	1.6908951	-3.2913984	2.9185825
C	2.5673976	-3.1793857	0.9403274
N	2.7758361	-2.8225997	2.2058877
H	0.9636896	-4.1046089	-0.1389627
H	-0.5896700	-5.5065239	1.6471695
H	-1.3508113	-3.9268782	1.9227727
H	1.6175375	-3.1647088	3.9996824
H	3.2555302	-3.0047227	0.1127349
H	2.3080388	-6.7834862	5.7034268
C	3.2847382	-6.4432764	6.0167766
C	4.0276638	-5.6586299	4.9814380
N	4.1391412	-6.0998221	3.6825986
C	4.8090327	-4.5123337	5.0486479
C	4.9551958	-5.2479782	3.0183218
N	5.3942780	-4.2676122	3.8162910
H	3.6634484	-6.9934635	3.3483281
H	3.1769997	-5.8495066	6.9406497
H	3.8482087	-7.3659101	6.2584376
H	4.9623222	-3.8560829	5.9062725
H	5.1668909	-5.3705526	1.9562851
H	-0.5415696	-3.3970680	-4.4447576
C	-0.1808598	-4.1972076	-3.8231578

C	0.5405320	-3.7317561	-2.5606256
O	1.3854876	-2.7845751	-2.6647636
O	0.2779934	-4.3951829	-1.5149135
H	0.5367569	-4.7743123	-4.4371436
H	-1.0144940	-4.8699349	-3.5664152
H	1.8977490	-0.1090000	-5.5861568
C	2.6112586	0.2730698	-4.8524872
O	3.4390371	1.1282852	-5.1834652
N	2.5017705	-0.2065921	-3.5840564
C	3.4975398	0.1877933	-2.5782374
C	4.8818389	-0.3940295	-2.9843266
O	4.9596867	-1.3667438	-3.7527461
C	2.9592182	-0.2559938	-1.1909321
C	3.8755295	-0.0594732	0.0150824
O	3.5265634	0.6485321	0.9861650
O	4.9962277	-0.6991067	0.0099094
H	1.9008660	-1.0222941	-3.3766519
H	3.5936561	1.2874886	-2.5983986
H	2.6984702	-1.3277136	-1.2510642
H	2.0253958	0.2975327	-1.0065889
H	8.0447554	0.1959598	-2.2970188
N	5.9895191	0.1975748	-2.4935470
O	8.3339489	-2.5199783	-3.0992972
C	7.5102365	-1.8363018	-2.4744983
H	5.9540716	0.9372613	-1.7583264
C	7.2898360	-0.3556098	-2.8349784
H	7.4896672	-0.2706235	-3.9168044

N	6.8439066	-2.2732711	-1.3775951
C	7.0864559	-3.5842063	-0.8234939
C	6.0192927	-4.5896685	-1.3135561
O	4.9317441	-4.2211562	-1.7772858
C	7.1419033	-3.4987114	0.7275057
C	8.1589428	-2.4973542	1.2000499
N	7.8380907	-1.1748811	1.4939735
C	9.5252251	-2.6526619	1.3541771
C	8.9704525	-0.5536157	1.8134194
N	10.0136861	-1.4177427	1.7416733
H	6.0950924	-1.6894103	-0.9579173
H	10.9865722	-1.1933459	1.9433488
H	8.0627544	-3.9187642	-1.2180309
H	7.3741664	-4.4933745	1.1427279
H	6.1560926	-3.1914585	1.1103597
H	10.1759728	-3.5148635	1.2187902
H	9.0452688	0.4970805	2.0921683
H	5.9194366	-7.8607456	-1.7443490
N	6.3374528	-5.9081909	-1.2493271
O	4.2031187	-6.9182071	0.3873622
C	4.1339293	-6.9555917	-0.8427330
H	7.2073196	-6.2061191	-0.8128660
C	5.3901570	-6.9192562	-1.7193990
H	5.1072207	-6.6795672	-2.7606396
N	2.9769194	-7.0349077	-1.5483607
C	1.7013634	-7.3008627	-0.8867879
C	1.6648668	-8.8027319	-0.5438217

O	1.2980814	-9.6548437	-1.3717928
H	3.0231297	-7.0074402	-2.5707494
H	0.8816759	-7.0297838	-1.5653733
H	1.6402573	-6.6690867	0.0144165
H	1.6063490	-10.6777240	1.8256590
N	2.1786413	-9.0673874	0.6722196
C	2.4445686	-10.3905942	1.2221494
C	3.7292947	-10.3826056	2.0629476
C	3.7194116	-9.3924990	3.2598034
O	4.4821534	-9.6116782	4.2209616
O	2.9343977	-8.3769829	3.1251888
H	2.4292805	-8.3189822	1.3409102
H	2.5101399	-11.1019111	0.3799709
H	4.5819709	-10.0902946	1.4187211
H	3.9427224	-11.3969920	2.4432291
H	1.7735990	-6.6652562	-6.1462066
C	2.0946388	-7.1276260	-5.2265470
O	3.1962227	-6.4204549	-4.6199759
H	4.0100312	-6.6908504	-5.0834877
H	2.3536264	-8.1856419	-5.4107633
H	1.2621681	-7.0981450	-4.5059031
H	4.6767974	-4.3759876	-7.9091156
C	5.2893070	-3.9249578	-7.1434560
O	4.8178748	-3.0607116	-6.3966700
H	7.0954860	-4.2313176	-6.2052966
N	6.5004164	-4.4741174	-7.0133760
H	6.8200863	-5.2107200	-7.6394514

H	7.4986058	4.4701774	-0.2574698
C	7.7040756	3.7413378	0.5462596
C	6.8724319	2.4485358	0.3880675
O	6.2885967	2.2455212	-0.6989124
O	6.8688104	1.6640423	1.4165864
H	8.7783984	3.4818637	0.5384563
H	7.4852189	4.2072435	1.5207902
H	0.7360996	2.7154584	2.9971984
C	1.6626890	3.1528582	3.3529482
O	1.6485392	4.0079064	4.2520878
H	4.2115176	4.2583576	2.7577484
N	2.7805700	2.7156117	2.7458086
C	4.0717578	3.2580982	3.1365382
H	2.7644640	1.9815707	2.0233548
H	4.8616511	2.6050358	2.7386866
H	4.1543419	3.2925167	4.2363874
H	8.8453850	7.5318458	4.9068372
C	9.5478446	6.7386462	5.0986072
C	8.8145496	5.4327780	5.2885385
C	8.5498459	4.5834174	4.1967951
C	8.3018623	5.0642407	6.5483478
C	7.7798539	3.4214476	4.3344671
C	7.5341847	3.9063287	6.7112663
C	7.2552063	3.0858816	5.6003647
O	6.4833047	1.9886290	5.7929261
H	6.3256119	1.5352871	4.9164316
H	10.2434808	6.6944490	4.2420795

H	10.1272313	7.0168641	5.9970777
H	8.9474852	4.8439881	3.2084348
H	8.4980502	5.7033292	7.4184197
H	7.5629349	2.7812973	3.4686562
H	7.1284136	3.6251465	7.6888501
H	8.8968350	-2.3677186	6.2765764
C	7.8532056	-2.1791488	6.4094064
O	5.8010045	0.4260424	3.7734926
C	5.1754559	-0.3344161	4.5913806
C	3.6761020	-0.2776606	4.6954581
N	5.8644121	-1.0962701	5.4397715
H	5.3786869	-1.6420291	6.1650117
C	7.3213296	-1.2441479	5.3278576
H	3.2622295	-0.2382514	3.6771942
H	3.2611652	-1.1362978	5.2452108
H	3.4016867	0.6597862	5.2120154
H	7.7839396	-0.2446941	5.3991089
H	7.5435658	-1.6366749	4.3187933
H	7.6995953	-1.7513510	7.4149943
H	7.3341483	-3.1519629	6.3641841
O	5.1752338	-1.9108972	2.5886329
H	4.1828664	-2.1787583	2.4656286
H	5.5186742	-2.7124435	3.1039359
Zn	5.9845079	-0.1858163	1.7199748
K	3.6324410	-3.5344487	-3.9369009
C	5.4174575	-3.1939005	8.7232910
O	4.9405269	-2.5469627	7.8085429

N	5.3349970	-2.8549984	10.0039944
H	5.7378873	-3.4495743	10.7282207
H	5.9559966	-4.1099976	8.5299952
H	4.8384652	-2.0084016	10.2825828

Table 6.1. List of literature binding energies with regards to a series of chelating agents for all of the relevant metal ions studied in HDAC8 (in kcal/mol).

	DTPA Stability constant	**DTPA ΔG binding (kcal/mol)**	**GEDTA Stability Constant**	**GEDTA ΔG binding (kcal/mol)**
Co^{2+}	18.4	-25.	12.5	-17.
Fe^{2+}	16.55	-23.	11.92	-16.
Mg^{2+}	9.3	-13.	5.21	-7.1
Mn^{2+}	15.6	-21.	12.3	-17.
Ni^{2+}	20.32	-28.	13.6	-19.
Zn^{2+}	18.75	-26.	14.5	-20.

Table 6.2. Calculated ΔG (kcal/mol) - DTPA metal swapped with EGTA

	Co^{2+}	Fe^{2+}	Mg^{2+}	Mn^{2+}	Ni^{2+}	Zn^{2+}
Co^{2+}	0	-2.25	-3.41	-1.96	1.31	-2.37
Fe^{2+}		0	-1.16	0.29	3.56	-0.12
Mg^{2+}			0	1.45	4.72	1.03
Mn^{2+}				0	3.27	-0.41
Ni^{2+}					0	-3.68
Zn^{2+}						0

Table 6.3. Experimental ΔG (kcal/mol) - DTPA metal swapped with EGTA

	Co^{2+}	Fe^{2+}	Mg^{2+}	Mn^{2+}	Ni^{2+}	Zn^{2+}
Co^{2+}	0	-1.73	-2.47	-3.55	1.12	-2.25
Fe^{2+}		0	-0.74	-1.81	2.85	-0.52
Mg^{2+}			0	-1.08	3.59	0.22
Mn^{2+}				0	4.66	1.30
Ni^{2+}					0	-3.37
Zn^{2+}						0

Table 6.4. Calculated ΔG – Experimental ΔG (kcal/mol) - DTPA metal swapped with EGTA

	Co^{2+}	Fe^{2+}	Mg^{2+}	Mn^{2+}	Ni^{2+}	Zn^{2+}
Co^{2+}	0	-0.52	-0.94	1.59	0.19	-0.12
Fe^{2+}		0	-0.42	2.11	0.71	0.40
Mg^{2+}			0	2.53	1.13	0.82
Mn^{2+}				0	-1.40	-1.71
Ni^{2+}					0	-0.31
Zn^{2+}						0

Table 6.5. Comparison of energies at various dialectric constant values from interior of protein (dielectric = 4) to water (dielectric= 80)

Original (ε = 4) converged Energy + COSMO correction compared with recomputed single point Energy + COSMO correction at different values of □ (All energies relative to respective reactant state, kcal/mol)			
	ε = 4	ε = 8	ε = 20
TS1	10.43	10.96	11.32
Intermediate	7.07	7.73	8.17
TS2	13.81	14.20	14.45

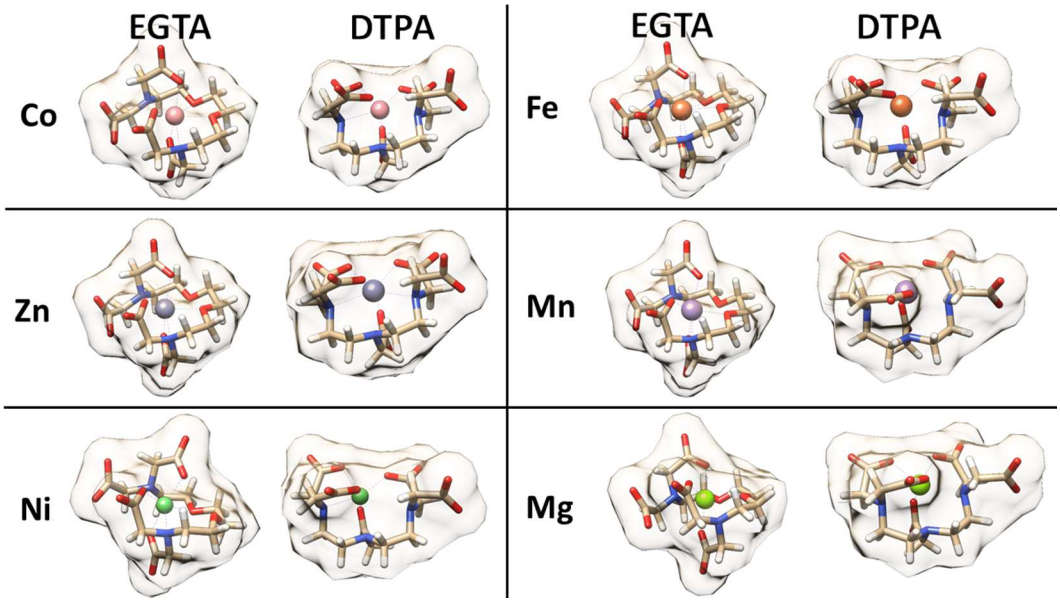


Figure 6.1. All divalent metals in this study coordinated by EGTA and DTPA, for use as a reference in determining metal binding affinities to HDAC8. Shaded areas surrounding each complex represent the solvent accessible surface area and highlight the solvent exclusion of these chelating complexes.

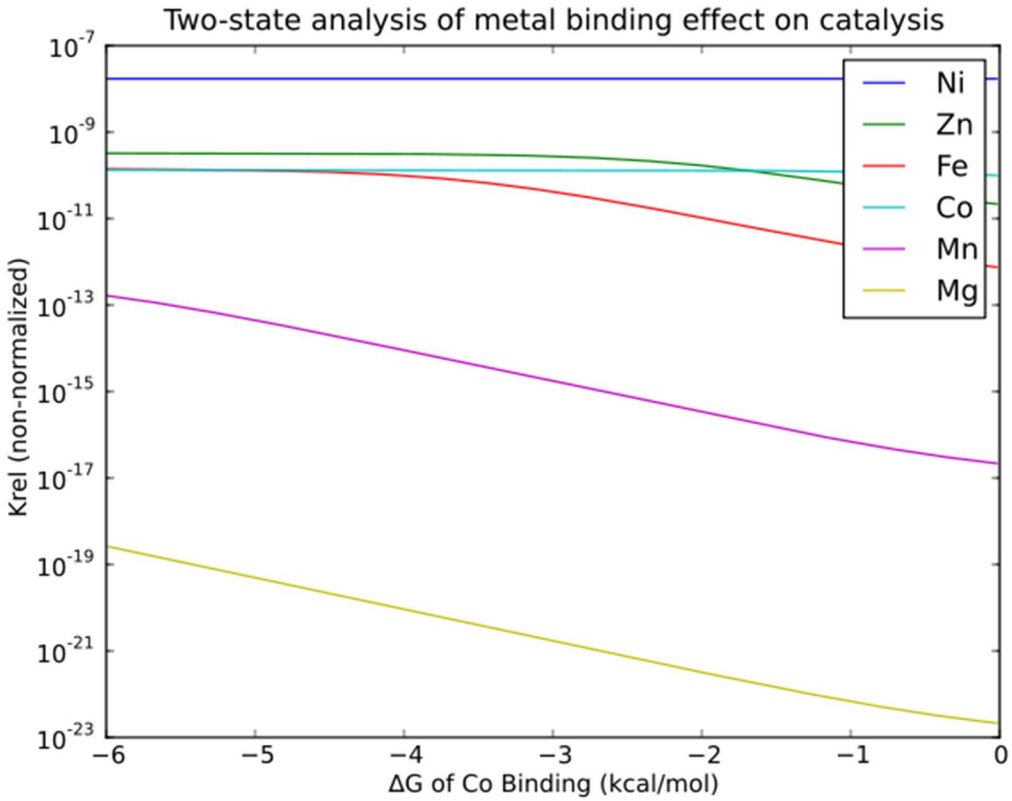


Figure 6.2. Catalytic model utilizing a two-state ensemble for $K_{binding}$ over a range of possible binding affinities for Co. In this model, $K_{binding}(M) = \frac{e^{-(\Delta G_{Co}+\Delta\Delta G_M)/RT}}{1+e^{-(\Delta G_{Co}+\Delta\Delta G_M)/RT}}$, where ΔG_{Co} is the binding affinity of Co and $\Delta\Delta G_M$ is the binding affinity of metal M relative to Co.

CHAPTER 7
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*, 19696-19701.

4 Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* 2006, *106*, 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*, 1087-1092.

7 Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 1970, *57*, 97-109.

8 Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* 2004, *303*, 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*, 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*, 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*, 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*, 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*, 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*, 41-64.

17 Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* 1985, *55*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 4062-4071.

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

39 Hammes-Schiffer, S.; Stuchebrukhov, A. A. Theory of Coupled Electron and Proton Transfer Reactions. *Chem. Rev.* 2010, *110*, 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*, 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*, 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*, 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*, 577-587.

44 Sparta, M.; Alexandrova, A. N. How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. *PLoS ONE* 2012, *7*, 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*, 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*, 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*, 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*, 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*, 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*, 59-73.

51 Kaplan, J.; DeGrado, W. F. De Novo Design of Catalytic Proteins. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 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*, 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*, 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*, 313-340.

55 Yin, S.; Ding, F.; Dokholyan, N. V. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* 2007, *15*, 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*, 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*, 1587-1590.

58 Shirvanyants, D.; Alexandrova, A. N.; Dokholyan, N. V. Rigid Substructure Search. *Bioinformatics* 2011, *27*, 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*, 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*, 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*, 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*, 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*, 481-485.

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

65 Kuhlman, B.; Baker, D. Native Protein Sequences Are Close to Optimal for Their Structures. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 10383-10388.

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

67 Yin, S.; Ding, F.; Dokholyan, N. V. Eris: An Automated Estimator of Protein Stability. *Nat. Methods* 2007, *4*, 466-467.

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

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

70 Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* 1981, *14*, 325-332.

71 Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* 2001, *115*, 6289-6292.

72 Schlitter, J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance Matrix. *Chem. Phys. Lett.* 1993, *215*, 617-621.

73 Masgrau, L.; Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* 2015, *48*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 593-628.

89 Guldberg, H. C.; Marsden, C. A. Catechol-O-Methyl Transferase: Pharmacological Aspects and Physiological Role. *Pharmacol. Rev.* 1975, *27*, 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*, 4202-4210.

91 Vidgren, J.; Svensson, L. A.; Liljas, A. Crystal Structure of Catechol O-Methyltransferase. *Nature* 1994, *368*, 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*, 120-130.

93 Axelrod, J.; Tomchick, R. Enzymatic O-Methylation of Epinephrine and Other Catechols. *J. Biol. Chem.* 1958, *233*, 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*, 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*, 954-958.

96 Valentine, J. S.; Hart, P. J. Misfolded CuZnSOD and Amyotrophic Lateral Sclerosis. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 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*, 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*, 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*, 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*, 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*, 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*, 8038-8047.

104 Dokholyan, N. V. Studies of Folding and Misfolding Using Simplified Models. *Curr. Opin. Struct. Biol.* 2006, *16*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 31-45.

113 Yang, X. J.; Seto, E. HATs and HDACs: From Structure, Function and Regulation to Novel Strategies for Therapy and Prevention. Oncogene 2007, 26, 5310-5318.

114 Wang, Z.; Zang, C.; Rosenfeld, J. A.; Schones, D. E.; Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Peng, W.; Zhang, M. Q., et al. Combinatorial Patterns of Histone Acetylations and Methylations in the Human Genome. Nat. Genet. 2008, 40, 897-903.

115 Kouzarides, T. Acetylation: A Regulatory Modification to Rival Phosphorylation? EMBO J. 2000, 19, 1176-9.

116 Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science 2009, 325, 834-40.

117 Kouzarides, T. Chromatin Modifications and Their Function. Cell 2007, 128, 693-705.

118 Iwabata, H.; Yoshida, M.; Komatsu, Y. Proteomic Analysis of Organ-Specific Post-Translational Lysine-Acetylation and -Methylation in Mice by Use of Anti-Acetyllysine and -Methyllysine Mouse Monoclonal Antibodies. Proteomics 2005, 5, 4653-64.

119 Kim, S. C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L., et al. Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. Mol. Cell 2006, 23, 607-18.

120 Xie, H.; Bandhakavi, S.; Roe, M. R.; Griffin, T. J. Preparative Peptide Isoelectric Focusing as a Tool for Improving the Identification of Lysine-Acetylated Peptides from Complex Mixtures. J. Proteome Res. 2007, 6, 2019-2026.

121 Gu, W.; Roeder, R. G. Activation of P53 Sequence-Specific DNA Binding by Acetylation of the P53 C-Terminal Domain. Cell 1997, 90, 595-606.

122 Imhof, A.; Yang, X. J.; Ogryzko, V. V.; Nakatani, Y.; Wolffe, A. P.; Ge, H. Acetylation of General Transcription Factors by Histone Acetyltransferases. Curr. Biol. 1997, 7, 689-92.

123 Yang, X. J.; Gregoire, S. Metabolism, Cytoskeleton and Cellular Signalling in the Grip of Protein Nepsilon - and O-Acetylation. EMBO Rep. 2007, 8, 556-62.

124 Starai, V. J.; Celic, I.; Cole, R. N.; Boeke, J. D.; Escalante-Semerena, J. C. Sir2-Dependent Activation of Acetyl-CoA Synthetase by Deacetylation of Active Lysine. Science 2002, 298, 2390-2.

125 Hallows, W. C.; Lee, S.; Denu, J. M. Sirtuins Deacetylate and Activate Mammalian Acetyl-CoA Synthetases. Proc. Natl. Acad. Sci. USA 2006, 103, 10230-5.

126 Schwer, B.; Bunkenborg, J.; Verdin, R. O.; Andersen, J. S.; Verdin, E. Reversible Lysine Acetylation Controls the Activity of the Mitochondrial Enzyme Acetyl-CoA Synthetase 2. Proc. Natl. Acad. Sci. USA 2006, 103, 10224-9.

127 L'Hernault, S. W.; Rosenbaum, J. L. Chlamydomonas A-Tubulin Is Posttranslationally Modified by Acetylation on the E-Amino Group of a Lysine. Biochemistry 1985, 24, 473-478.

128 Piperno, G.; Fuller, M. T. Monoclonal Antibodies Specific for an Acetylated Form of A-Tubulin Recognize the Antigen in Cilia and Flagella from a Variety of Organisms. J. Cell Biol. 1985, 101, 2085-94.

129 Kamine, J.; Elangovan, B.; Subramanian, T.; Coleman, D.; Chinnadurai, G. Identification of a Cellular Protein That Specifically Interacts with the Essential Cysteine Region of the HIV-1 Tat Transactivator. Virology 1996, 216, 357-66.

130 Reifsnyder, C.; Lowell, J.; Clarke, A.; Pillus, L. Yeast Sas Silencing Genes and Human Genes Associated with Aml and HIV-1 Tat Interactions Are Homologous with Acetyltransferases. Nat. Genet. 1996, 14, 42-9.

131 Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. Proc. Natl. Acad. Sci. USA 1964, 51, 786-94.

132 Phillips, D. M. The Presence of Acetyl Groups of Histones. Biochem. J. 1963, 87, 258-63.

133 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.

134 Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. Structures of a Histone Deacetylase Homologue Bound to the Tsa and Saha Inhibitors. Nature 1999, 401, 188-193.

135 Hodawadekar, S. C.; Marmorstein, R. Chemistry of Acetyl Transfer by Histone Modifying Enzymes: Structure, Mechanism and Implications for Effector Design. Oncogene 2007, 26, 5528-40.

136 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.

137 Smith, B. C.; Denu, J. M. Chemical Mechanisms of Histone Lysine and Arginine Modifications. Biochim. Biophys. Acta 2009, 1789, 45-57.

138 Yang, X. J.; Seto, E. The Rpd3/Hda1 Family of Lysine Deacetylases: From Bacteria and Yeast to Mice and Men. Nat. Rev. Mol. Cell Biol. 2008, 9, 206-18.

139 Bolden, J. E.; Peart, M. J.; Johnstone, R. W. Anticancer Activities of Histone Deacetylase Inhibitors. Nat. Rev. Drug Discov. 2006, 5, 769-84.

140 Paris, M.; Porcelloni, M.; Binaschi, M.; Fattori, D. Histone Deacetylase Inhibitors: From Bench to Clinic. J. Med. Chem. 2008, 51, 1505-29.

141 Marks, P. A.; Breslow, R. Dimethyl Sulfoxide to Vorinostat: Development of This Histone Deacetylase Inhibitor as an Anticancer Drug. Nat. Biotechnol. 2007, 25, 84-90.

142 Bradner, J. E.; West, N.; Grachan, M. L.; Greenberg, E. F.; Haggarty, S. J.; Warnow, T.; Mazitschek, R. Chemical Phylogenetics of Histone Deacetylases. Nat. Chem. Biol. 2010, 6, 238-243.

143 Suzuki, N.; Suzuki, T.; Ota, Y.; Nakano, T.; Kurihara, M.; Okuda, H.; Yamori, T.; Tsumoto, H.; Nakagawa, H.; Miyata, N. Design, Synthesis, and Biological Activity of Boronic Acid-Based Histone Deacetylase Inhibitors. J. Med. Chem. 2009, 52, 2909-2922.

144 Gregoretti, I. V.; Lee, Y. M.; Goodson, H. V. Molecular Evolution of the Histone Deacetylase Family: Functional Implications of Phylogenetic Analysis. J. Mol. Biol. 2004, 338, 17-31.

145 Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E., et al. Structural Snapshots of Human HDAC8 Provide Insights into the Class I Histone Deacetylases. Structure 2004, 12, 1325-34.

146 Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.; Gallinari, P., et al. Crystal Structure of a Eukaryotic Zinc-Dependent Histone Deacetylase, Human HDAC8, Complexed with a Hydroxamic Acid Inhibitor. Proc. Natl. Acad. Sci. USA 2004, 101, 15064-15069.

147 Vannini, A.; Volpari, C.; Gallinari, P.; Jones, P.; Mattu, M.; Carfi, A.; De Francesco, R.; Steinkuhler, C.; Di Marco, S. Substrate Binding to Histone Deacetylases as Shown by the Crystal Structure of the HDAC8-Substrate Complex. EMBO Rep. 2007, 8, 879-84.

148 Dowling, D. P.; Gantt, S. L.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W. Structural Studies of Human Histone Deacetylase 8 and Its Site-Specific Variants Complexed with Substrate and Inhibitors. Biochemistry 2008, 47, 13554-63.

149 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, 6170-8.

150 Gantt, S. L.; Joseph, C. G.; Fierke, C. A. Activation and Inhibition of Histone Deacetylase 8 by Monovalent Cations. J. Biol. Chem. 2010, 285, 6036-43.

151 de Ruijter, A. J.; van Gennip, A. H.; Caron, H. N.; Kemp, S.; van Kuilenburg, A. B. Histone Deacetylases (HDACs): Characterization of the Classical HDAC Family. Biochem. J. 2003, 370, 737-49.

152 Anzellotti, A. I.; Farrell, N. P. Zinc Metalloproteins as Medicinal Targets. Chem. Soc. Rev. 2008, 37, 1629-1651.

153 Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. A Proton-Shuttle Reaction Mechanism for Histone Deacetylase 8 and the Catalytic Role of Metal Ions. J. Am. Chem. Soc. 2010, 132, 9471-9.

154 Corminboeuf, C.; Hu, P.; Tuckerman, M. E.; Zhang, Y. Unexpected Deacetylation Mechanism Suggested by a Density Functional Theory QM/MM Study of Histone-Deacetylase-Like Protein. J. Am. Chem. Soc. 2006, 128, 4530-4531.

155 Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V. Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. J. Phys. Chem. B 2012, 116, 8375-8382.

156 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.

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

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

159 Nechay, M. R.; Valdez, C. E.; Alexandrova, A. N. Computational Treatment of Metalloproteins. J. Phys. Chem. B 2015, 119, 5945-5956.

160 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, 730-737.

161 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, 4112-4124.

162 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, 3007-3018.

163 Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. Chem. Phys. Lett. 1989, 162, 165-169.

164 Klamt, A.; Schuurmann, G. Cosmo: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. Perkin Trans. 2 1993, 799-805.

165 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.

166 Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. Acta Crystallogr. Sect. A 1976, 32, 922-923.

167 Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. Phys. Rev. B 1992, 45, 13244-13249.

168 Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. Phys. Rev. Lett. 2003, 91, 146401.

169 Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. Phys. Chem. Chem. Phys. 2005, 7, 3297-305.

170 Grimme, S. Accurate Description of Van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. J. Comput. Chem. 2004, 25, 1463-1473.

171 Wu, R.; Hu, P.; Wang, S.; Cao, Z.; Zhang, Y. Flexibility of Catalytic Zinc Coordination in Thermolysin and HDAC8: A Born−Oppenheimer Ab Initio QM/MM Molecular Dynamics Study. J. Chem. Theory Comput. 2010, 6, 337-343.

172 Chen, K.; Zhang, X.; Wu, Y.-D.; Wiest, O. Inhibition and Mechanism of HDAC8 Revisited. J. Am. Chem. Soc. 2014, 136, 11636-11643.

173 Bader, R. F. W. Atoms in Molecules: A Quantum Theory. Oxford University Press: Oxford, 1990.

174 Matta, C. F.; Boyd, R. J. The Quantum Theory of Atoms in Molecules. Wiley-VCH Verlag GmbH & Co. KGaA: 2007.

175 Boyd, R. J.; Choi, S. C. Hydrogen Bonding between Nitriles and Hydrogen Halides and the Topological Properties of Molecular Charge Distributions. Chem. Phys. Lett. 1986, 129, 62-65.

176 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. J. Comput. Chem. 2001, 22, 931-967.

177 Fonseca Guerra, C.; Snijders, G. J.; te Velde, G.; Baerends, J. E. Towards an Order-N Dft Method. Theor. Chem. Acc. 99, 391-403.

178 ADF2014 SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.

179 Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative Assessment of a New Nonempirical Density Functional: Molecules and Hydrogen-Bonded Complexes. J. Chem. Phys. 2003, 119, 12129-12137.

180 Van Lenthe, E.; Baerends, E. J. Optimized Slater-Type Basis Sets for the Elements 1-118. J. Comput. Chem. 2003, 24, 1142-56.

181 Tecplot. http://www.tecplot.com/ (2014).

182 McCall, K. A.; Huang, C.; Fierke, C. A. Function and Mechanism of Zinc Metalloenzymes. J. Nutr. 2000, 130, 1437s-46s.

183 Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. The Carboxylate Shift in Zinc Enzymes: A Computational Study. J. Am. Chem. Soc. 2007, 129, 1378-1385.

184 Bell, R. P. The Tunnel Effect in Chemistry. Springer US: 1980.

185 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. 1995, 117, 5179-5197.

186 Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? J. Comput. Chem. 2000, 21, 1049-1074.

187 Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins: Struct., Funct., Bioinf. 2006, 65, 712-725.

188 Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. J. Comput. Chem. 2004, 25, 1157-1174.

189 Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L., et al. Nwchem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. Comput. Phys. Commun. 2010, 181, 1477-1489.

190 Dojindo.com Metal Chelates - Analytical & Biological Products. pp 252-253.

191 Williams, J. & Moreton, K. The distribution of iron between the metal-binding sites of transferrin human serum. *Biochem. J.* **185,** 483–8 (1980).

192 Tinoco, A. D. & Valentine, A. M. Ti(IV) binds to human serum transferrin more tightly than does Fe(III). *J. Am. Chem. Soc.* **127,** 11218–11219 (2005).

193 Emsley, J. *The elements*. (Clarendon Press, 1998).

194 Lavi, N. & Alfassi, Z. B. Determination of trace amounts of cadmium, cobalt, chromium, iron, molybdenum, nickel, selenium, titanium, vanadium and zinc in blood and milk by neutron activation analysis. *Analyst* **115,** 817 (1990).

195 Hydrolytic precipitation reaction of titanium(IV) from (Na, H)Cl aqueous solution. *J. Inorg. Nucl. Chem.* **43,** 2443–2448 (1981).

196 Loza-Rosas, S. A. *et al.* A ubiquitous metal, difficult to track: towards an understanding of the regulation of titanium(iv) in humans. *Metallomics* **9,** 346–356 (2017).

197 Hafner, A. & Okuda, J. Titanium NMR data for some titanium half-sandwich complexes bearing substituted cyclopentadienyl ligands. *Organometallics* **12,** 949–950 (1993).

198 Koch, R. & Bruhn, T. Theoretical 49Ti NMR chemical shifts. *J. Mol. Model.* **12,** 723–729 (2006).

199 Mujika, J. I., Escribano, B., Akhmatskaya, E., Ugalde, J. M. & Lopez, X. Molecular Dynamics Simulations of Iron- and Aluminum-Loaded Serum Transferrin: Protonation of Tyr188 Is Necessary To Prompt Metal Release. *Biochemistry* **51,** 7017–7027 (2012).

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

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

202 Nechay, M. R., Valdez, C. E. & Alexandrova, A. N. Computational treatment of metalloproteins. *J. Phys. Chem. B* **119,** 5945–5956 (2015).

203 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.* **9,** 730–737 (2013).

204 Sparta, M. & Alexandrova, A. N. How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase. *PLoS One* **7,** e47172 (2012).

205 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.* **426,** 4112–24 (2014).

206 Mizutani, K., Yamashita, H., Kurokawa, H., Mikami, B. & Hirose, M. Alternative structural state of transferrin. The crystallographic analysis of iron-loaded but domain-opened ovotransferrin N-lobe. *J. Biol. Chem.* **274,** 10190–4 (1999).

207 Nechay, M. R. *et al.* Histone Deacetylase 8: Characterization of Physiological Divalent Metal Catalysis. *J. Phys. Chem. B* **120,** 5884–5895 (2016).

208 Tinoco, A. D., Eames, E. V & Valentine, A. M. Reconsideration of serum Ti(IV) transport: albumin and transferrin trafficking of Ti(IV) and its complexes. *J. Am. Chem. Soc.* **130,** 2262–70 (2008).

209 Wells, A. S., Finch, G. L., Michels, P. C. & Wong, J. W. Use of Enzymes in the Manufacture of Active Pharmaceutical Ingredients—A Science and Safety-Based

Approach To Ensure Patient Safety and Drug Quality. *Org. Process Res. Dev.* **16,** 1986–1993 (2012).

210 Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).

211 Li, X., Chianese, A. R., Vogel, T. & Crabtree, R. H. Intramolecular alkyne hydroalkoxylation and hydroamination catalyzed by iridium hydrides. *Org. Lett.* **7,** 5437–5440 (2005).

212 Shirvanyants, D., Alexandrova, A. N. & Dokholyan, N. V. Rigid substructure search. *Bioinformatics* **27,** 1327–1329 (2011).

213 Adamczyk, A. J., Cao, J., Kamerlin, S. C. L. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 14115–20 (2011).

214 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. *J. Chem. Phys.* **132,** 154104 (2010).

215 Weymuth, T., Couzijn, E. P. A., Chen, P. & Reiher, M. New benchmark set of transition-metal coordination reactions for the assessment of density functionals. *J. Chem. Theory Comput.* **10,** 3092–3103 (2014).

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

217 Cusanelli, A., Frey, U., Richens, D. T. & Merbach, E. The Slowest Water Exchange at a Homoleptic Mononuclear Metal Center : Variable-Temperature and Variable-Pressure 17 O NMR Study on [Ir(HO)]3+. *J. Am. Chem. Soc.* **118,** 5265–5271 (1996).

218 Poth, T., Paulus, H., Elias, H., Dücker-Benfer, C. & van Eldik, R. Kinetics and Mechanism of Water Substitution at Half-Sandwich Iridium(III) Aqua Cations Cp*Ir(A−B)(H2O)2+/+ in Aqueous Solution (Cp* = η5-Pentamethylcyclopentadienyl Anion; A−B = Bidentate N,N or N,O Ligand). *Eur. J. Inorg. Chem.* **2001,** 1361–1369 (2001).

219 Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angewandte Chemie - International Edition* **52,** 5700–5725 (2013).

220 Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **82,** 447–470 (2013).

221 Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **51,** 3933–3940 (2012).

222 Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **51,** 3933–3940 (2012).

223 Weber, P. C. & Salemme, F. R. Structural and functional diversity in 4-α-helical proteins. *Nature* **287,** 82–84 (1980)..

224 Morgan, A. A. & Rubenstein, E. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. *PLoS One* **8,** e53785 (2013).

225 Morgan, A. A. & Rubenstein, E. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. *PLoS One* **8,** e53785 (2013).

226 Lee, C.-W., Wang, H.-J., Hwang, J.-K. & Tseng, C.-P. Protein Thermal Stability Enhancement by Designing Salt Bridges: A Combined Computational and Experimental Study. *PLoS One* **9,** e112751 (2014).

227 Mayer, C. & Hilvert, D. A Genetically Encodable Ligand for Transfer Hydrogenation. *European J. Org. Chem.* **2013,** 3427–3431 (2013).