Title
Accounting for Chemical Change in Classical and QM-Hybrid Molecular Dynamics Simulations of Proteins and Metalloproteins

Author
Reilley, David

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Accounting for Chemical Change in Classical and QM-Hybrid Molecular Dynamics Simulations of Proteins and Metalloproteins

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemistry

by

David John Reilley

2021

ABSTRACT OF THE DISSERTATION

Accounting for Chemical Change in Classical and QM-Hybrid Molecular Dynamics Simulations of Proteins and Metalloproteins

by

David John Reilley

Doctor of Philosophy in Chemistry

University of California, Los Angeles 2021

Professor Anastassia N. Alexandrova, Chair

Molecular dynamics (MD) is a powerful tool to study atomic scale changes in proteins underpinning biological and catalytic pathways. It is routinely used to identify small molecule binding sites and study protein folding. MD simulations struggle to achieve quantitatively accurate energies and traditionally sample a fixed chemical state – they do not allow covalent bonds to form or break. However, chemical change underlies many important biochemical

processes, including shifts in metal ligation that occur during metalloenzyme activity or when metal ions bind to metalloproteins, or the protonation and deprotonation of amino acids that dictate pH-dependent activity and stability. Understanding these processes at the atomic scale demands computational tools that operate in conjunction with MD to calculate more accurate energies.

Hybrid quantum mechanical/molecular mechanics approaches (QM/MM) can capture the dynamics of chemical change as they provide accurate energies for small regions of interest, such as the active site. We have developed one such hybrid method, QM/DMD, which combines density functional theory (DFT) calculations with discrete molecular dynamics simulations for rapid metalloprotein conformational sampling that can even capture shifts in metal ligation. With QM/DMD we identified the orientational preferences of a cofactor in phenylalanine hydroxylase that distinguishes the disease state of phenylketonuria. To study some forms of chemical change further theory even beyond QM/DMD is necessary. We created a competitive metal affinity (CMA) method, a semi-empirical thermodynamic cycle, to calculate relative metal binding affinities to metalloproteins. With our CMA method and QM/DMD we assessed the ability of human serum transferrin to transport non-native metals and therefore its role in metal cytotoxicity. We also identified the different mechanisms by which Li^{+} and Be^{2+} inhibit glycogen synthase kinase 3β, which could inform future drug design targeting that protein.

In contrast to the detailed QM treatment of a small region needed to study metal behavior, pH-dependent behavior requires rapid sampling of chemical changes across the whole protein in the form of protonation and deprotonation reactions. We therefore developed a constant-pH molecular dynamics method, called titr-DMD, that stochastically updates protonation states based on the efficient DMD method and the semi-empirical electrostatic

method Propka. We successfully benchmarked titr-DMD on experimentally verified pH-dependent conformational changes in a staphylococcal nuclease mutant. Our work demonstrates the utility of properly modified molecular dynamics, and QM/DMD in particular, to study many forms of chemical change in proteins with good accuracy and speed.

The dissertation of David John Reilley is approved.

Todd O. Yeates

David S. Eisenberg

Mathieu Bauchy

Anastassia N. Alexandrova, Committee Chair

University of California, Los Angeles

2021

To my family, friends, mentors, and peers who all supported me on this journey

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I pause here on my journey to thank all those that helped me reach this point. My curiosity, love of learning, and perseverance, that blend which forms the bedrock of all scientific pursuits, began with my family in my youngest years. To my mother, Beth, and father, Kevin, I could not have hoped for better first teachers or more supportive parents. I would not have become the scholar I am today without the shared books, museum trips, camping, and, yes, auto-repair and woodshop chores. Whenever I found myself stuck, your encouragement always helped me find a way forward. To my brother, Andrew, I thank you for being my closest friend growing up – the future holds more concerts and roller coasters. To my sister, Alyssa, your laughter and joy taught me that there is always a lighter side to life. I also thank my extended family for comforting reunions and memories with plentiful food and plentiful conversation. I especially thank my grandmother Sharon, who nurtured my creativity and desire to explore the world around me. To my fiancee, Rose, since you came into my life I have never looked back. You have always been there for me, even when a whole continent separated us. Your love and understanding has paved our road together and I have truly found my best traveling companion in you.

I am grateful to all the teachers and mentors I have had over the years. I thank my high school chemistry teachers Mr. Glenn Black and Mr. David Dougan for nurturing my nascent interest in chemistry with colorful, engaging experiments. I thank Prof. Joshua Schrier for opening the road to computational chemistry for me at Haverford College. I will always remember that conversation I had in your office my freshman year to first learn about the subject, which lead to a new found passion, three years of research in your lab, and now a PhD. I

especially thank my academic advisor at UCLA, Professor Anastassia Alexandrova. You have helped me grow immensely as a scientist over these past five years. I value your candid and deep advice and intellectual flexibility. My conversations with you always leave me brimming with ideas to test and investigate, but also keep me grounded enough to see my research through. When my scientific interests took me to new areas of study, you were willing to learn and expand your scope with me. I feel privileged to share in a part of your scientific journey and I wish you and the lab all the best going forward.

I also shared my road to a PhD with many wonderful, talented friends and peers. I thank all of my fellow members of the Alexandrova group, including Jack Fuller, Matthew Hennefarth, PJ Robinson, Daniel Bim, Mike Nechay, Elisa Jimenez-Izal, Han Guo, Julen Munarriz, Addis Fuhr, Mai-Anh Ha, Huanchen Zhai, Zerina Mehmedovic, Kirill Shumilov, Patricia Poths, Derek Urwin, Borna Zandkarimi, Zisheng Zhang, Santiago Vargas, Claire Dickerson, Amy Lai, and Jiyuan Liu. I enjoyed the scientific conversations, good times, and minor pranks we all shared together (bio boiz and girls forever). I thank my good friends in the department from other labs: Sam Lilak, Matthew Voss, Robert 'Boots' Boutelle, Hammy Buditama, Erik Farr, Taylor Aubry, Will Narvaez, Andy Vong, and Ted Zee. Celebrating our successes together, and commiserating over our failures, made my path that much more memorable and enjoyable.

I thank all of the scientific collaborators with whom I have worked over the years. I thank my PhD committee members, Professor Todd Yeates, Professor Mathieu Bauchy, Professor Christopher Lee, and Professor David Eisenberg for taking the time to serve on my committee and for your advice. I also thank all those who have worked with me for research projects, especially Professor Nikolay Dokholyan, Professor Emily Que, Professor Alexander Spokoyny, Professor Cathleen Zeymer, Professor Andy Borovik, Professor Konstantin Popov, Dr. Radhika

Mehta, Jian Wang, and Morgan Hopp. Our discussions and research together have greatly broadened my knowledge.

As the road goes ever onward, so do my scientific pursuits. But no journey is worth taking or even possible without traveling companions. I express my deepest gratitude to those who shared my path with me so far.

This research was supported by the National Science Foundation through NSF-CAREER Award CHE-1351968 and NSF CHE-1903808, and by the Passan Foundation and National Institute of Health through NIGMS R01GM134047, R01GM114015, R01GM064803, R01GM123247, and R35GM134864. This work was also made possible through grants of supercomputer time from the Institute for Digital Research and Education (IDRE) at UCLA and the Extreme Science and Engineering Discovery Environment (XSEDE).

Chapter 1 is adapted with permission from the following publication:

1. **Reilley, D. J.**; Hennefarth, M. R.; Alexandrova, A. N. The Case for Enzymatic Competitive Metal Affinity Methods. *ACS Catal.* **2020**, *10*, 2298-2307. (DOI: 10.1021/acscatal.9b04831, link: https://pubs.acs.org/doi/full/10.1021/acscatal.9b04831)

Chapter 2 is adapted with permission from the following publication:

1. **Reilley, D. J.**; Fuller, J. T.; Nechay, M. R.; Victor, M.; Li, W.; Ruberry, J. D.; Mujika, J. I.; Lopez, X.; Alexandrova, A. N. Toxic and Physiological Metal Uptake and Release by

Human Serum Transferrin: Insight from QM/MM Dynamics Simulations. *Biophys. J.* **2020**, *118*, 2979-2988. (DOI: 10.1016/j.bpj.2020.05.006) Copyright 2020 Elsevier.

Chapter 3 is adapted with permission from the following publication:

1. **Reilley, D. J.**; Arraf, Z.; Alexandrova, A. N. Contrasting Effects of Inhibitors Li^{+} and Be^{2+} on Catalytic Cycle of Glycogen Synthase Kinase-3β. *submitted.*

Chapter 4 is adapted with permission from the following publication:

1. **Reilley, D. J.**; Wang, J.; Dokholyan, N. V.; Alexandrova, A. N. Titr-DMD – A Rapid, Coarse-Grained Quasi-All-Atom Constant *pH* Molecular Dynamics Framework. *submitted*.

Chapter 5 is adapted with permission from the following publication:

1. **Reilley, D. J.**; Popov, K.; Dokholyan, N. V.; Alexandrova, A. N. Uncovered Dynamic Coupling Resolves the Ambiguous Mechanism of Phenylalanine Hydroxylase Oxygen Binding. *J. Phys. Chem. B.* **2019**, *123*, 4534-4539. (DOI: 10.1021/acs.jpcb.9b02893, link: https://pubs.acs.org/doi/full/10.1021/acs.jpcb.9b02893) Copyright 2019 American Chemical Society.

VITA

Timeline

2020	Gave talk at Gordon Research Seminar: Metals in Biology
2018	Seaborg Symposium Poster Award
2018	Gave talk at International Conference on Theoretical Aspects of Catalysis
2016	UCLA Pavelsky Scholarship
2016	B.Sc. in Chemistry *Magna Cum Laude* with honors, Haverford College
2016	ACS Scholastic Achievement Award
2016	Elected to Phi Beta Kappa: Haverford Zeta Chapter

Publications

1. **Reilley, D. J.**; Wang, J.; Dokholyan, N. V.; Alexandrova, A. N. Titr-DMD – A Rapid, Coarse-Grained Quasi-All-Atom Constant *pH* Molecular Dynamics Framework. *submitted*. (ChemRxiv: 13717870)
2. **Reilley, D. J.**; Arraf, Z.; Alexandrova, A. N. Contrasting Effects of Inhibitors Li^{+} and Be^{2+} on Catalytic Cycle of Glycogen Synthase Kinase-3β. *submitted.* (ChemRxiv: 14637741)
3. Mehta, R.; Rivera, D.; **Reilley, D. J.**; Thomas, P.; Hinojosa, A.; Tan, D.; Stewart, A.; Cheng, Z.; Crowder, M., Alexandrova, A. N.; Fast, W.; Que, E. Visualizing the Dynamic Metallation State of NDM-1 in Bacteria Using a Reversible Fluorescent Probe. *submitted.*
4. **Reilley, D. J.**; Fuller, J. T.; Nechay, M. R.; Victor, M.; Li, W.; Ruberry, J. D.; Mujika, J. I.; Lopez, X.; Alexandrova, A. N. Toxic and Physiological Metal Uptake and Release by

Human Serum Transferrin: Insight from QM/MM Dynamics Simulations. *Biophys. J.* **2020**, *118*, 2979-2988.

5. **Reilley, D. J.**; Hennefarth, M. R.; Alexandrova, A. N. The Case for Enzymatic Competitive Metal Affinity Methods. *ACS Catal.* **2020**, *10*, 2298-2307.
6. **Reilley, D. J.**; Popov, K.; Dokholyan, N. V.; Alexandrova, A. N. Uncovered Dynamic Coupling Resolves the Ambiguous Mechanism of Phenylalanine Hydroxylase Oxygen Binding. *J. Phys. Chem. B.* **2019**, *123*, 4534-4539.
7. Atmuri, N. D. P.; **Reilley, D. J.**; Lubell, W. D. Peptidomimetic Synthesis by way of Diastereoselective lodoacetoxylation and Transannular Amidation of 7-9-Membered Lactams. *Org. Lett.* **2017**, *19*, 5066-5069.
8. Martin-Noble, G. C.; **Reilley, D. J.**; Rivas, L. M.; Smith, M. D.; Schrier, J. EQeq+C: An Empirical Bond-Order-Corrected Extended Charge Equilibration Method. *J. Chem. Theory Comput.* **2015**, *11*, 3364-3374.

Chapter 1

The Case for Enzymatic Competitive Metal Affinity Methods

1.1 Introduction: The Need for Accurate Metal Binding Affinity Calculations in Catalytic and Biomedical Contexts

We often want to know which metal will bind to a protein most readily, which metal or metals actually bind *in vivo*, and which one will be the best at enzymatic catalysis. It is not guaranteed that a single metal could satisfy all of the above for a given natural metalloenzyme. For artificial metalloenzymes (ArMs), we also want to know if the protein can bind the desired metal, and if the metal would then function as a catalyst with the desired activity and selectivity. Hence, being able to compute the metal binding affinities to proteins is desirable in the studies of enzymatic catalysis and enzyme design. Unfortunately, this goal is non-trivial. Efforts toward solving this problem are the focus of this chapter.

The questions of which metal is used in a natural enzyme, and which metal we want to employ in an artificial enzyme are not easily answerable because different forces drive the evolution of enzymes in nature than the priorities of man-made catalysts. Instead of maximizing enzyme activity, biology caps it to maintain the complex equilibria of homeostasis. Biology prioritizes the bio-availability of the starting materials and fold stability, but also ensures that enzymes can be readily destroyed when needed. These constraints also apply to the way in which metals are selected for natural metalloenzymes.[1,2] Furthermore, the catalytically relevant metals for many metalloproteins are not truly known. Many enzymes are assumed to be Zn(II)-dependent based on X-ray crystal structures, but this can be an artifact of experimental conditions.[3] Follow up studies on systems such as histone deacetylase[4,5] carbonic anhydrase,[3] S-ribosylhomocysteinase,[6] and peptide deformylase[7] show that sometimes other metals can bind and report significant activity. In some cases, the metal reported by crystallography is not even a particularly significant contributor to the protein's function. Without considering the binding

affinity of different metals, *in vitro* and computational studies of metalloproteins could be based on a false or incomplete picture of metal preferences.

A major goal in the design of artificial metalloenzymes is maximal catalytic performance, with less emphasis on stability in their simpler *in vitro* environment of operation. Previous efforts already found that while proteins provide powerful platforms for new catalysts, the reactions they can perform, and sometimes their catalytic rates, have hard limitations.[8,9] Recently, directed evolution has become an indispensable tool to develop new ArMs or refine existing ones.[10–14] However, directed evolution is constrained by the roles for which a given protein scaffold has evolved.[8,9,15] While there is promiscuity of function in many proteins, some reactions are simply out of reach of conventional methodologies. Metals that are not natively bioavailable can expand the space of accessible reactions. For example, recent efforts show that noble metals can expand the repertoire of porphyrin-dependent enzymes.[16] However, non-physiological metals must bind sufficiently strongly to their protein scaffolds, whose amino acids did not originally evolve to ligate non-physiological metals. Thus, determination of metal affinity is required. Additionally, as we will show shortly, the affinity of the metal to the protein (e.g. the stability gain upon metal binding) and the catalytic activity may follow a non-trivial and non-linear mutual dependence, via the Brønsted-Evans-Polanyi (BEP) relation.

Lastly, metal-protein affinity is of broader interest than biocatalysis. It is relevant to metal transport about the body, particularly the activity of metal chaperones, which, unlike many proteins, bind metals in a highly selective manner and in specific environments.[17–20] Chaperones help maintain the distinct metal concentrations in different organ systems, tissues, and even different subcellular organelles within cells.[21] Tracking the metal affinity of these proteins in different contexts is important for metal toxicology. A large number of transition and heavy

metals are now bioavailable with their use in modern industries, including industrial catalysis. Some metals, such as Cd(II), Hg(II), As(III), and Pb(II) are highly toxic and lead to non-specific syndromes.[22,23] The extent of cytotoxicity of other metals, such as Al(III), Ti(IV), and Ga(III), is unclear but demands investigation as they are introduced into the body both from the environment, and for medical purposes.[24–28] Metal binding may even play a role in neurodegenerative diseases, hypothetically facilitating the protein-protein aggregation and fibril formation.[29] Ultimately it is of high interest to know the metal-protein affinity, and have ways to calculate it.

1.2 Existing Methods and Their Limitations

Dedicated computational tools to investigate protein-metal binding, which we will refer to as competitive metal affinity (CMA) methods, are hard to come by. The ideal CMA would incorporate an accurate energy evaluation and significant dynamical sampling to capture configurational entropy in order to fully describe the thermodynamics of metal binding. Clearly, the expense of the accurate energy calculations severely limits the amount of sampling that can be afforded. While there are many methods to study metalloprotein behavior in general, not all are suited to form the basis of a CMA method.

Classical force field based methods can be parameterized to model some metalloenzyme structures, but are insufficient to obtain thermodynamic values. Force field parameters for metals are based on a point charge supplemented with various harmonic terms, and operate on the basis of a fixed metal coordination (e.g. octahedral, tetrahedral) that cannot change significantly as a function of protein dynamics. These potentials can contain bonding and non-bonding interactions, but are generally fitted to capture structure (within limits) rather than energy.[30–33] In

this respect they can be fairly successful for systems containing closed shell metals with ideal geometries (Zn(II), Mg(II), Mn(II)), remaining stable over long molecular dynamics (MD) simulations.[34,35] However, even the most successful applications of these methods do not obtain reliable energies for catalytic studies.

Electronic structure calculations are necessary to obtain accurate metal binding energies. One possible approach is to use a small cluster model of the active site and treat it quantum mechanically. However, this approach ignores the entropy of the protein scaffold and the impact of the protein dynamics on the energy and entropy of the active site. The only portion of the entropy in the free energy of the active site that this approach captures is the vibrational entropy - typically calculated within the harmonic approximation and subject to the constraints imposed by the rest of the protein structure. While cluster models are useful for catalytic mechanism mapping,[36,37] and as such can play a role in artificial metalloenzyme design,[38–40] these applications rely on the cancellation of errors when protein entropy is ignored equivalently throughout the reaction profile. On the other hand, many metal exchange phenomena are inaccessible to the approach, as enzymes frequently undergo some amount of restructuring when a new metal binds.

A more promising avenue to obtain metal binding free energies based on electronic structure calculations are mixed QM/MM simulations. This class of methods combines a quantum mechanical description of the metal center and its surrounding environment and molecular mechanical modeling of the rest of the protein (Figure 1.1). Statistical mechanical sampling of the protein becomes possible within QM/MM, and there has been intensive research into and development of these methods over the last two decades.[41–43] Sufficient sampling is still a problem, however, for most established QM/MM methods. Our group developed the QM/DMD method,[44] which combines DFT with discrete molecular dynamics (DMD)[45] for enhanced

sampling. Several features make QM/DMD significantly more affordable than other methods, permitting lengthy trajectories while treating the metal quantum mechanically. DMD is based on simplified square-well potentials, ballistic equations of motion, and slight coarse graining, which permit speedy sampling. The majority of the metalloprotein is sampled with DMD. Only the metal(s) and the atoms within the metal ligands that are immediately attached to the metal are never touched by DMD. Periodically, this tiny metal coordination region is expanded to a large cluster model, and that larger region undergoes a relaxation on a DFT potential energy surface. Then the DMD domain expands again, to include the majority of the (now optimized) active site. An overlapping region modeled with both theories allows passing the geometric and energetic information between QM and DMD. We have used QM/DMD to successfully study many aspects of metalloprotein behavior, including the effect of mutagenesis on structure and function,[44,46,47] metal-dependent catalytic activity,[5,48–50] redox functionality,[44,51] and recently, metal affinity.[5,52] Because of the sampling efficiency and capability of dynamically changing metal coordination sphere, QM/DMD is suitable for building a CMA technology.

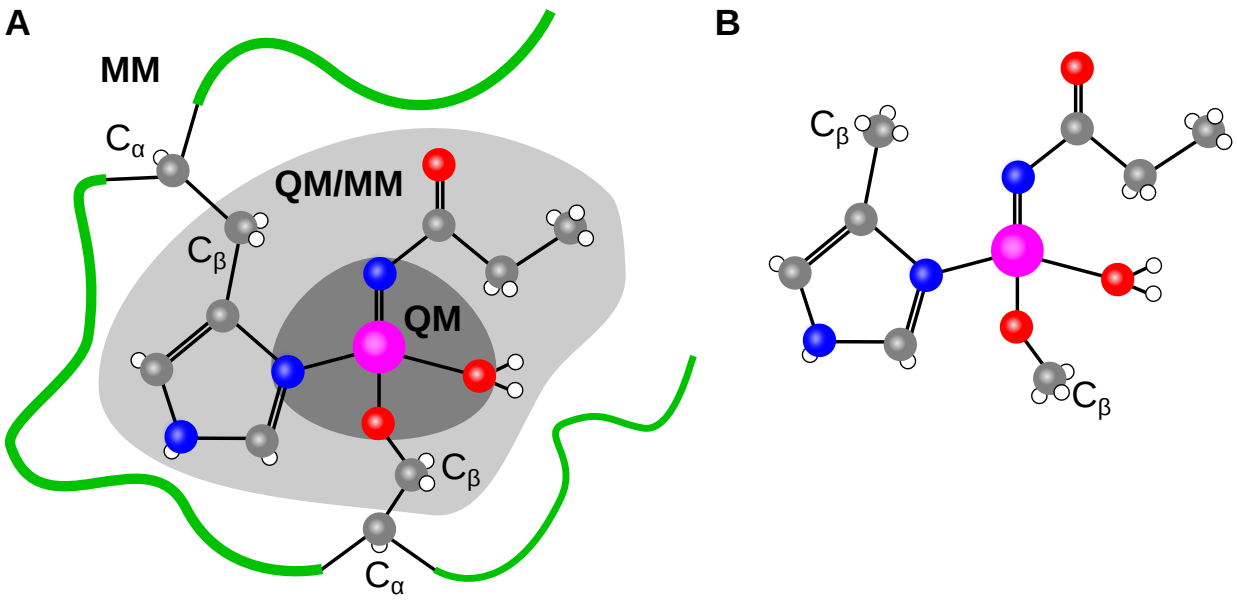


Figure 1.1. Diagrams demonstrating the active space of (A) QM/MM and (B) small cluster methods. QM/MM models the entire protein, with QM for the active site (the dark and light gray regions) and MM for the rest of the protein (the white region). In some forms of QM/MM, such as QM/DMD, there is an overlapping region treated with both QM and MM (in light gray) and MM modeling is only excluded from a small central region (in the case of this diagram, the metal and its first coordination sphere in dark gray). Small cluster methods, by contrast, only model the QM region.

The exact form of the necessary free energy terms is another major complication in CMA evaluation. One would think that metal affinities could be calculated as the difference between the free energies of the bound metalloprotein, and the apo-protein and the solvated metal ion. However, the accuracy of the free energies of metal binding obtained in this way will depend on the precision of evaluating the entropy change upon binding, which requires complete sampling of the conformational space of the protein, both with and without the metal. Such full equilibration is practically impossible.[53] Secondly, evaluations of the free energy of the solvated metal ion requires expensive and laborious quantum mechanical treatment, explicit solvent, and sufficient sampling of solvent configurations (on the order of 10^6). Furthermore, as solvated metals are charged ionic species their free energies cannot be directly obtained by experiment

either.[54–56] In what follows we describe our CMA method that avoids both complications described in this paragraph. We will discuss several diverse applications of the method, its current limitations, and propose further directions to improve upon it. To the best of our knowledge, this technique is unprecedented.

1.3 Thermodynamic CMA Method

Our method calculates the relative metal binding free energy, $\Delta\Delta G$, with respect to one metal chosen as a reference. For most applications, relative free energies are sufficient. The approach combines QM/DMD sampling with a semi-empirical thermodynamic cycle that avoids ill-defined terms. First, we employ QM/DMD simulations run to convergence (on the order of an approximate 10-100 ns) of the protein with each considered metal. Second, we determine the lowest energy QM region for each metal and calculate its Gibbs free energy using the harmonic approximation. Finally, we use these free energies in a thermodynamic cycle shown in Figure 1.2. The cycle consists of the metal ions going into the protein from a complex with a chelating agent (typically EDTA, which we exclusively used in all systems described in this article) rather than directly from solution. Hence, instead of using a dubious, calculated value for the free energy of a metal in solution, this cycle uses computationally tractable metal-chelator complexes. The free energies of metal complexation from solution to the chelator are readily available from experiment. The final step of the cycle cancels the chelator terms through the computed free energies of metal exchange in the protein (from QM/DMD) and in the chelator complex (from *ab initio* or DFT calculations and harmonic vibrational entropies). Closing the thermodynamic cycle yields the $\Delta\Delta G$ of one metal, M_a, binding to the protein relative to the other metal, M_b. This

means that when comparing the results of this method to experiment, only the trend can be reproduced, not the absolute free energies of metal binding.

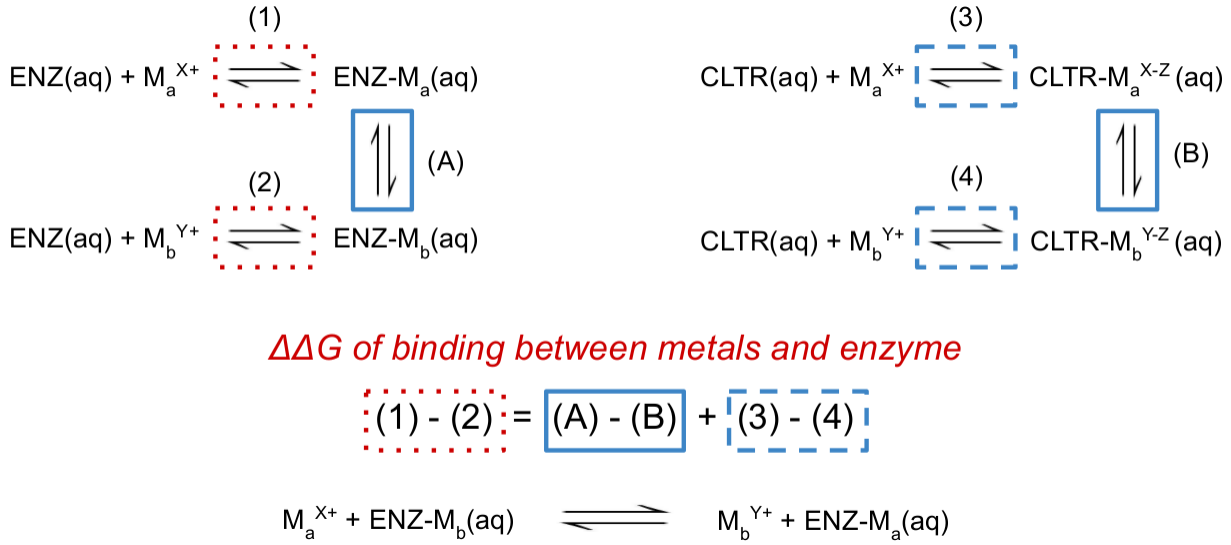


Figure 1.2. Thermodynamic cycles for the relative free energy of metal binding method. The left cycle is intractable as the structure of free metal ions in solution is not defined (dashed red boxes). The right cycle uses experimentally available data for chelator (CLTR) binding to avoid this problem (dashed blue boxes). The sum of this cycle and the easily calculated transition from CLTR to the protein (solid blue boxes) gives the free energy of exchanging metals in the protein by canceling all the CLTR terms.

1.4 Method Benchmarks: Acireductone Dioxygenase and Histone Deacetylase 8

We have successfully applied the described CMA method to a series of problems of catalytic and biological relevance. To illustrate the method's performance and accuracy, we now describe several diverse examples, each with principally different biological functionality and chemistry. We consider a mononuclear oxidase, a mononuclear metal-dependent hydrolase, and a metal transporter protein.

Acireductone dioxygenase (ARD) can tightly bind different metals and performs different reactions depending on which metal binds. The protein is involved in the methionine salvage

pathway and acts on the substrate 1,2-dihydroxy-3-keto-5-(methylthio)pentene, oxidizing it to two possible sets of products.[57,58] ARD bound with Ni(II) catalyzes the formation of methylthiopropinate, while ARD bound with Fe(II) catalyzes the formation of 2-keto-4-methylthiobutyric acid, a precursor of methionine (Figure 1.3).[59] The bound metal does not change the structure of the protein, or the way in which the substrate binds to it, as we showed with QM/DMD. This means that the properties of the metal itself dictate catalytic selectivity. As such, ARD is the subject of many mechanistic studies.[49,60,61] We showed that the mechanistic bifurcation relies on the differences in charge transfer from the metal ligands, through the metal, and to the dioxygen bound to the substrate. Experimental binding studies show that ARD has an appreciable affinity for both Ni(II) and Fe(II).[62,63] The measured activity and metal binding affinities together demonstrate that both ARD reactive pathways are meaningful. The ARD's preference for the metal should then be context-dependent. Hence, the relative affinity of ARD to Fe(II) versus Ni(II) in the absence of other environmental factors is of interest.

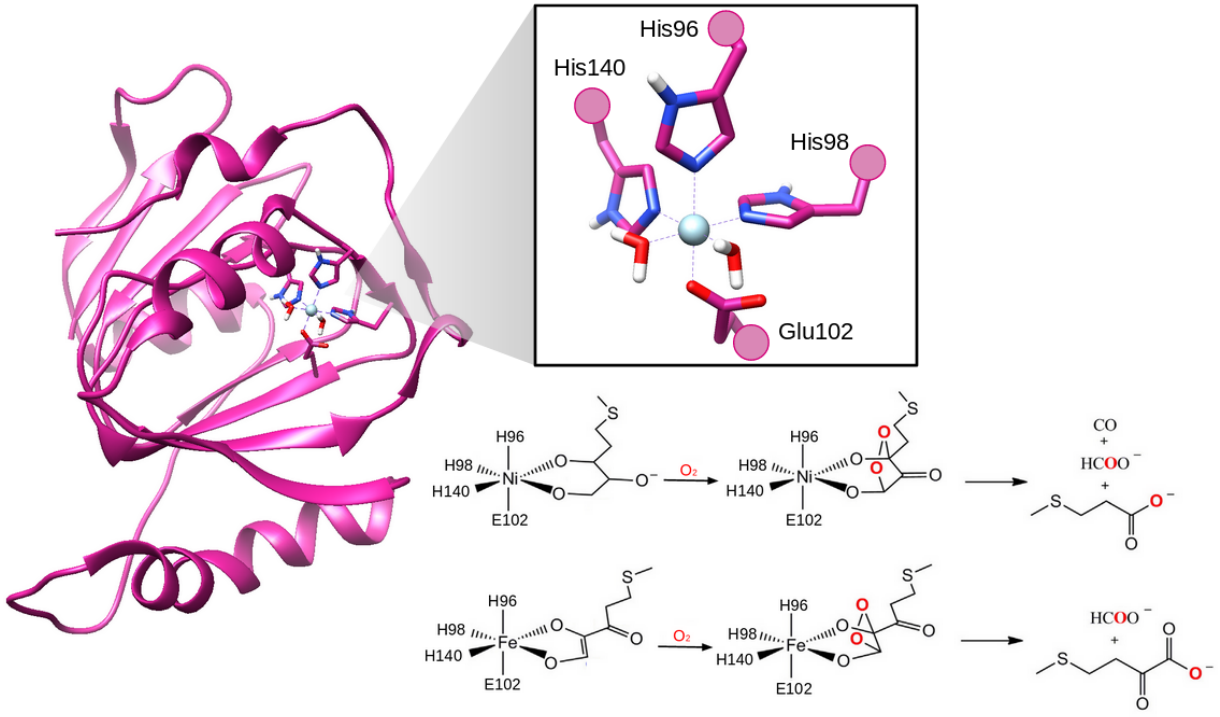


Figure 1.3. The structure of ARD (PDB ID: 1ZRR) and its active site, and the mechanisms of the metal-dependent reactions the protein can perform. The Ni(II) and Fe(II) bound forms of ARD preferentially bind different substrates and therefore perform different reactions.

The application of our CMA method to the catalytic metals in ARD, including Co(II), is illustrated in Table 1.1. To calculate the binding affinities of Fe(II), Ni(II), and Co(II) to ARD, we started with QM/DMD trajectories from our previous studies.[49] We selected the three lowest energy structures of the QM regions for each metal variant of ARD. We tested all feasible spin states of the metals with further geometry optimizations on these systems, looking for the multiplicity that minimizes the electronic energy. Our calculations showed that the multiplicity of Fe(II) was a singlet or quintet (depending on the structure), Ni(II) was a triplet, and Co(II) was a doublet. For each multiplicity we then performed frequency calculations and selected the lowest free energy among them. The calculations were done with Turbomole (version 6.6).[64] The pure meta-GGA TPSS DFT functional[65] with the D3 dispersion correction[66] was used. The metal was

treated with the triple-zeta basis set def2-TZVPP while all other atoms were treated with the double-zeta def2-SVP basis set.[67] The conductor-like Screen Model (COSMO) with a constant dielectric of 20 was used to approximate the screening and solvation effects in the partially buried active site of the protein.[68] These settings are consistent with the initial QM/DMD runs. The results correctly capture that the affinity of the protein for Ni(II) is stronger than for Fe(II) and that ARD's affinity for Co(II) is roughly the same as for Fe(II). The error from the exact values is on the order of 3 kcal/mol (Table 1.1). Note that this is at the limit of DFT accuracy. Given the many approximations needed along the way and despite the cancellation of errors in the relative calculations, the qualitative agreement with experiment we obtained is still satisfying.

Table 1.1. Table of the experimental (Dai and Chai) and calculated binding affinities to ARD. The energies are relative to Fe(II), which correspondingly has a value of 0 kcal/mol. The experimental values here are based on Boltzmann weighted ratios of molar metal content.

	Fe(II)	Ni(II)	Co(II)
Dai (kcal/mol)	0.0	-1.23	-0.65
Chai (kcal/mol)	0.0	-0.28	N/A
calc. (kcal/mol)	0.0	-3.76	0.38

Our next system is a histone deacetylase (HDAC), which is part of a class of enzymes that remove acetyl groups from histone lysines and potentially some nonhistone proteins.[69,70] Alongside histone acetyltransferases, which add acetyl groups, HDACs regulate how tightly histones bind to DNA and therefore gene regulation.[71–73] Overexpression of HDACs is associated with many pathologies, particularly cancer, while inhibition leads to the activation of genes related to growth arrest and tumor cells.[73,74] Consequently, many anti-cancer drugs are HDAC inhibitors.[75,76] Many of these bind to the transition metal center of their HDAC targets,

including FDA approved suberanilohydroxamic acid (Vorinostat)[77] and FK228 (Romidepsin).[78] To reliably develop tighter binding drugs with computational methods, knowledge of which metal or metals bind to HDAC is necessary.

The catalytically relevant metals for histone deacetylases are not well understood. Historically, researchers assumed that HDACs are Zn(II) enzymes on the basis of X-ray structures and kinetic studies.[79,80] While Zn(II) is clearly a catalytically active metal in HDACs, as discussed earlier in this article, the promiscuity of metalloproteins means that crystallographic data does not preclude the relevance of other metals. Indeed, kinetic studies report significant activity with Co(II), Fe(II), and Ni(II) in HDAC8 and Co(II) even shows higher activity than Zn(II).[4] This variety in metals that HDAC8 can use has important implications in traditional mechanistic studies.

Binding affinities from our method proved necessary to properly identify the catalytically relevant metals besides Zn(II) in HDAC8 and calculate their activities. Our group recently investigated the mechanism of HDAC8 and how it varies with physiologically abundant metals (Zn(II), Fe(II), Co(II), Mn(II), Ni(II), and Mg(II)) (Figure 1.4).[5] Pairing a traditional transition state search with QM/DMD simulations, we mapped the mechanism and calculated the activation barrier of the reaction for each metal. However, these results do not capture the experimental catalytic order and suggest that experimentally inactive Mn(II), Ni(II), and Mg(II) are reactive. We theorized that the binding affinities of these metals to HDAC8 contributes to their *in vitro* catalytic activity. We calculated the $\Delta\Delta G$ for each metal and combined this with our computed barriers ($\Delta G^{\ddagger}$) to get a series of K_{rel}:

$$K_{rel} = \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right)\exp\left(\frac{\Delta\Delta G_{binding}}{RT}\right) \quad (1.1)$$

which in contrast to the barriers, match the experimental catalytic order and identify Mn(II) and Mg(II) as inactive (Table 1.2). The K_{rel} of Ni(II) is the one outlier, which suggests that it is highly reactive, driven by predicted tight binding. Ultimately, our study of HDAC8 demonstrates the utility of our metal binding $\Delta\Delta G$ method when the catalytic metal or metals of a natural metalloenzyme are not known.

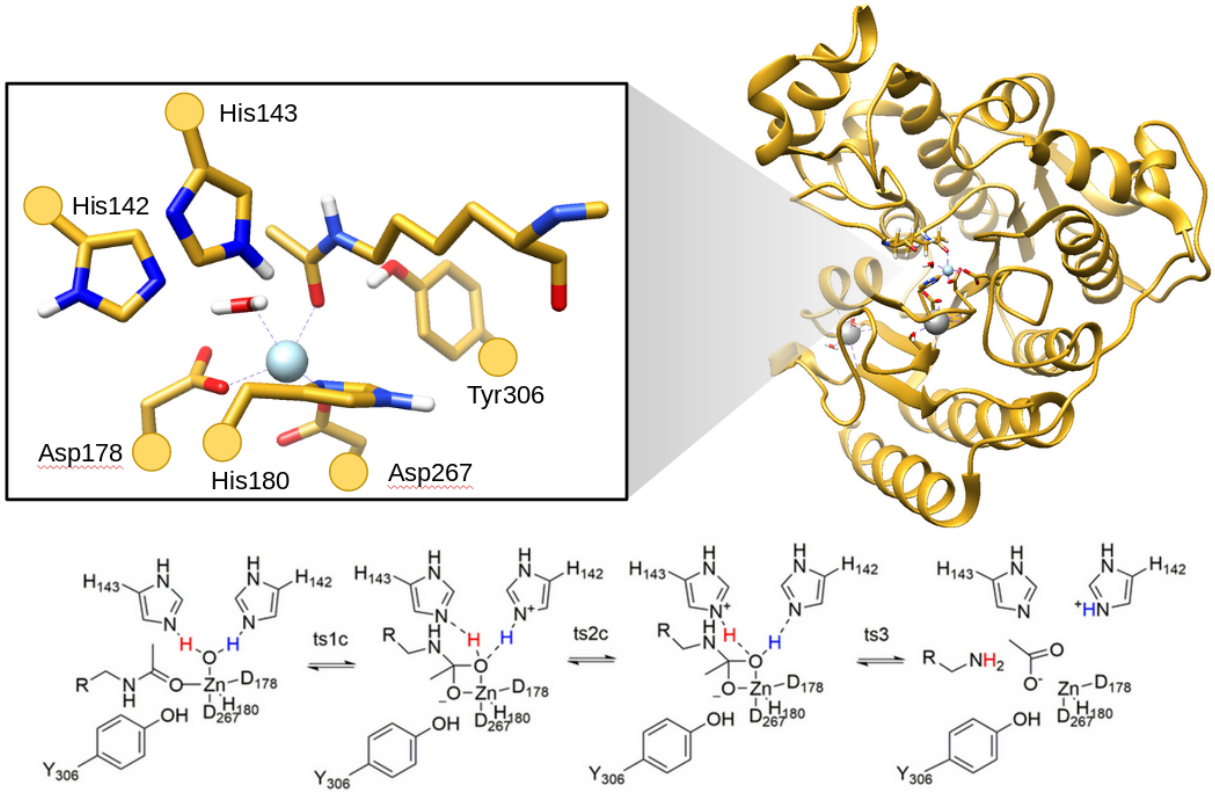


Figure 1.4. The structure of HDAC8 (PDB ID: 2V5W) and its active site with example substrate, and the most plausible mechanism of the deacetylation reaction it performs.

Table 1.2. Experimental k_{cat} and calculated K_{rel} values for HDAC8. While the exact values are not comparable, the qualitative order of the two catalytic measures match. Notice that Ni(II) is an exception, with the highest K_{rel} despite its experimental inactivity. Also note that Mg(II) and Mn(II) have K_{rel} that are many orders of magnitude lower than Co(II) meaning that they are consistent with their inactive experimental result.

	Co(II)	Zn(II)	Fe(II)	Ni(II)	Mn(II)	Mg(II)
exp. k_{cat}	1.2	0.90	0.48	N/A	N/A	N/A
calc. K_{rel}	7.64×10^{-11}	1.27×10^{-11}	1.75×10^{-13}	1.89×10^{-8}	1.37×10^{-17}	1.46×10^{-23}

As an aside, we further hypothesize that in some cases the metal binding affinity could be a descriptor of enzymatic catalytic activity. Specifically, by the BEP principle, the binding of the rate-determining intermediate to the active site should be neither too strong not too weak for the maximal catalytic activity to emerge. On the other hand, the stability of the active site itself and the metal ion in it should impact the stability of the intermediate of interest. That is because both the binding energy of the metal to its ligands, and the binding energy of the metal to the reaction intermediate depend on the energy and spatial extent of the orbitals of the metal. Therefore, there should be some relationship between the affinity of the protein to the metal and the catalytic activity of the metalloenzyme. We tested this conjecture using the computational data that we generated for the different metal variants of HDAC8, focusing just on the rate-determining, second step of the reaction (as shown in Figure 1.5). We excluded the Mg(II) form from the dataset, since it is known from the experiment to not bind appreciably to HDAC8. We correlate the $\Delta\Delta G$ of the metal ion binding to the protein to the Boltzmann weighted reaction barriers $e^{-E_a/RT}$ (normalized to remove the pre-exponential factor which we may assume to be approximately the same for all considered metals). The result is shown in Figure 1.5. We observe a classic volcano plot that all metals obey, even Ni(II), with Zn(II) at the peak. While we cannot assume that all metalloenzymes obey this sort of scaling relation, this demonstrates the utility of CMAs for yet another catalytic application.

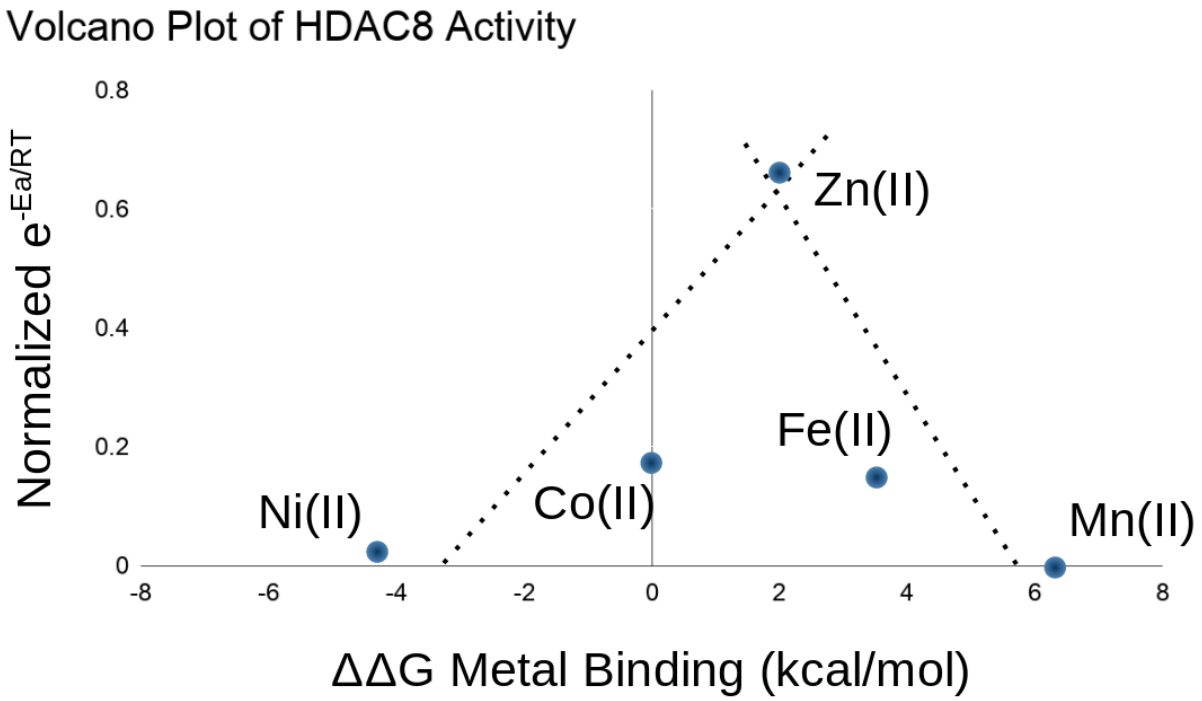


Figure 1.5. Volcano plot showing scaling relation of HDAC8 between binding $\Delta\Delta G$ and reaction rate. We calculated the reaction rates as the normalized, Boltzmann weighted ratios between each reaction barrier and the Co(II) reference. Notice how even Ni(II) is consistent with this trend.

Human serum transferrin (hTF) is an example of how CMAs could be used in a different context. This protein is not catalytic, but is interesting because it can uptake and also release metals through pH dependent protein conformations with potentially profound implications in metal toxicology. hTF is implicated in the promiscuous transport of many metals across the blood-brain barrier and the concentration of metals in tumor cells, including potentially cytotoxic metals like Ti(IV), Al(III), and Ga(III). We used our CMA method to get the first insight into the toxic metal transport abilities of hTF *in vivo* conformational states. Full consideration of hTF and the application of our CMA method to this system is covered in the following chapter.

1.5 Limitations and Outlook

Further research into CMA methods is important, especially as our method is not without limitations. Its reliance on chelating agents introduces other problems besides limiting calculations to referential $\Delta\Delta G$. The best way to calculate the thermodynamic terms involving the chelating agent is unclear. Experimental stability constants for EDTA and many related chelating agents are fortunately available for most metals in their common oxidation states.[81,82] Unfortunately, the corresponding structures of these metal complexes are not fully known, and they are necessary to accurately calculate the free energy associated with the transition from the chelator complex to the protein. In the studies we discuss above, we assume full chelation of each metal with no other ligands in the complexes. This makes most metals conform to an octahedral geometry. This is likely fine for large transition metals, but breaks down for small and low charge metals such as Li(I) and Mg(II). Indeed, crystallographic studies of Mg-EDTA binding show that a water molecule is also a ligand in the complex.[83] One way to mitigate these problems would be a benchmark study of a wide range of chelators on a system that has been experimentally well characterized for many metals. Calculating the set of $\Delta\Delta G$ for each chelator without varying any other parameters would reveal which chelator can be used most accurately for each metal.

Our method is also limited to proteins which undergo only minor conformational changes upon the binding of different metals. The first concern here is that the QM regions must share the same atoms besides the metal center to satisfy the thermodynamic cycle. Metals that bind entirely different sites on a protein are consequentially inaccessible to our current method. A second concern largely involves computational scaling, as significant rearrangement (like

refolding) upon metal binding requires even more expensive structural sampling in order to accurately assess the entropy component of $\Delta\Delta G$. While this is a general problem with protein and metalloprotein simulations, enhanced sampling for the specific purpose of metal binding affinities would be impactful. Solutions to both of these concerns would render many systems more accessible, particularly metal chaperones as these proteins can adopt different folds for different metals.[19]

Further advancements in CMA methods would greatly propel understanding of natural metalloenzymes and the design of new ArMs. Such techniques could determine the catalytically relevant metals in natural metalloenzymes, which cannot be taken for granted from crystal structures. CMA calculations would be indispensable in the effort to better understand metal transport pathways throughout the body, especially with regards to metal toxicology. In the design of ArMs, replacing the bound metal in an existing metalloprotein scaffold can introduce new functions, often inaccessible to current design methodologies like directed evolution. Placing a metal into a specifically designed artificial scaffold is also an attractive opportunity for ArMs catalysis. For all such design tasks, it is critical to assess the metal affinity and its ability to outperform other metals that might be present in the synthesis conditions. New tools such as CMAs will expand the catalytic space of metalloenzymes.

Chapter 2

An Application of Competitive Metal Affinity Methods: Toxic and Physiological Metal Uptake and Release by Human Serum Transferrin

2.1 Introduction: Ubiquitous Human Serum Transferrin and its Promiscuous Metal Binding

Human serum transferrin (hTF) is of medical interest for its unique role in toxic metal pathology. The protein natively carries atomic iron into the cell from blood by receptor-mediated endocytosis. However, it is believed to also competitively transport other transition metals based on *in vitro* binding studies, including Ti(IV), Co(III), Ga(III), Cr(III), and Al(III).[84–86] While some of these metals, namely Co(III), may be necessary for homeostasis in the appropriate concentrations, others (including Ti(IV) and Al(III)) are believed to exhibit some degree of cytotoxicity even at low concentrations.[24,28] Ti(IV) in particular can bind to DNA as well as inhibit various intracellular enzymes.[87,88] The ability of hTF to transport toxic metals is therefore of great concern with the increasing bioavailability of these metals due to modern industry and their medical applications.[27,28,89,90] This is acutely problematic as the protein could bring these metals to sensitive parts of the body by its ability to pass the blood-brain barrier as well as concentrate them in cancer cells due to the over expression of the transferrin receptor across a range of tumors.[91,92] Though troubling, these abilities have been providing unique opportunities to develop new anticancer drugs and drug delivery techniques to the brain based on hTF or its receptor.[93,94] More recently, this has included human serum albumin nanoparticles coupled to transferrin to facilitate targeted cancer drug delivery.[95] Ultimately, a thorough structural understanding of hTF and how it binds and transports a variety of transition metals can improve knowledge of their toxicology as well guide the development of drugs and new tools for drug delivery.

The basic biological activity and structure of hTF are well characterized. It is a member of the transferrin family of glycoproteins, which regulate free iron concentration in physiological

fluids such as blood by hTF, milk and tears by lactoferrin, and egg whites by ovotransferrin. hTF is a 80 kDa protein comprised of two domains, termed the N- and C-domains, each containing a pair of highly similar subdomain lobes connected by a hinge. Each domain can bind an iron ion with a synergistic anion (typically carbonate) near the hinge between its two lobes (Figure 2.1).[96] When hTF binds two iron atoms, it is recognized by transferrin receptor 1 and the entire protein is brought into the cell by endocytosis.[97] The change from the blood serum pH (7.4) to the lower pH of the endosomes (5.6) then triggers the release of iron from hTF.[98]

The mechanism of hTF binding and releasing iron and other metals is not fully understood despite extensive study, and further clarification poses particular challenges for experimental approaches. A large body of work, including native PAGE gels, small angle X-ray scattering, X-ray absorption fine structure spectroscopy (XAFS) and crystal structures, suggest the protein undergoes a hinging conformational change between the lobes in each domain from a closed state upon metal binding at physiological pH to an open one at endosomal pH immediately before metal release.[99,100] This is believed to facilitate iron capture and release by protecting and then exposing the metal binding site to solvent. There are a few possible chemical triggers for this action which may operate in conjunction or separately, including metal reduction[101] and protonation of specific residues. In the N-domain, the better studied of the two, the conformational change was long thought to be driven primarily by a dilysine bridge: a pair of interacting lysines (Lys206 and Lys296), one from each lobe, where the lower pH protonates and breaks this interaction.[102] However, while mutagenesis studies did indicate that the bridge is important for proper hTF activity,[103] there is no experimental confirmation of the exact role it plays. The necessary, direct, structural inspection of the open, endosomal forms of hTF is hampered by their embedding in the cellular machinery of endocytosis. Some crystal structures

have been obtained with hTF in a receptor-bound state or at low pH,[104,105] but it is difficult to capture the effect of both conditions on structure and dynamics.[106] Only two crystal structures (PDB ID: 5DYH, 5H52) demonstrate a truly open conformation of the holoprotein beyond a few degrees of interlobal twisting.[107,108] However, both use the larger citrate as a synergistic anion and were obtained in blood serum uptake, rather than endosomal release, conditions. These issues have made further study into the release activity of hTF difficult to pursue with experiment alone. The release mechanism is unknown at a structural level and even its basic kinetics (akin to references 84 and 85) across all metals is simply unstudied.

Past computational studies have filled in and revised the structural and mechanistic details of the hTF metal transport mechanism, but these are not yet complete. Crystal structures of two mutants to the dilysine bridge first suggested that breaking this interaction alone fails to trigger the conformational change.[109] Later molecular dynamics studies by the Lopez group,[110] based on a hypothesis from Rinaldo and Field,[111] showed that protonation of a tyrosine that coordinates the metal (Tyr188) prompts the conformational change regardless of the protonation state of the dilysine bridge. The simulations found this behavior present for both native Fe(III) and Al(III). These results, however, were based mainly on force field based molecular dynamics, which has severe limitations for the appropriate treatment of transition metals. Moreover, previous QM/MM MD simulations at the semi-empirical level coupled with DFT/MM minimizations[112] for Fe(III) and Al(III) suggested the importance of introducing a quantum method to fully characterize the interaction mode of these two cations with the surrounding ligands.

This study aims to investigate the atomistic root for the hinging phenomenon in the N-domain of hTF with extensive QM/DMD simulations and consider how this mode of binding and

release might vary for a range of transition metals. QM/DMD is a rapid sampling method for metalloenzymes (described thoroughly in the Theoretical Methods section of this chapter) that can treat the chemistry of the bound metal and its coordination quantum mechanically. In addition, we describe a new method to evaluate the relative binding affinities of the metals to hTF in all its forms at both uptake and release. Besides Fe(III), we performed simulations with metals necessary for the human body, Fe(II), Co(III), Cr(III), and Zn(II), and increasingly bioavailable non-native metals that may have acute cytotoxicity, Ti(IV) and Ga(III). We conducted the simulations on four different protonation states derived from reference 110: the physiological form found in blood serum (Phys), the protein with just the dilysine bridge protonated (Acid), and the protein with Tyr188 protonated and either one or two additional water molecules in the QM active space (PrTr and Double respectively) (Figure 2.1). The Phys and Acid forms are hypothesized to be closed, while the Double and Prtr forms are open and likely implicated in metal release into the endosome. The results show a remarkable structural similarity across all considered metals with some subtle differences in transient interactions about the binding site that may explain their relative affinities to hTF, ability to undergo uptake and release, and implications for metal toxicity.

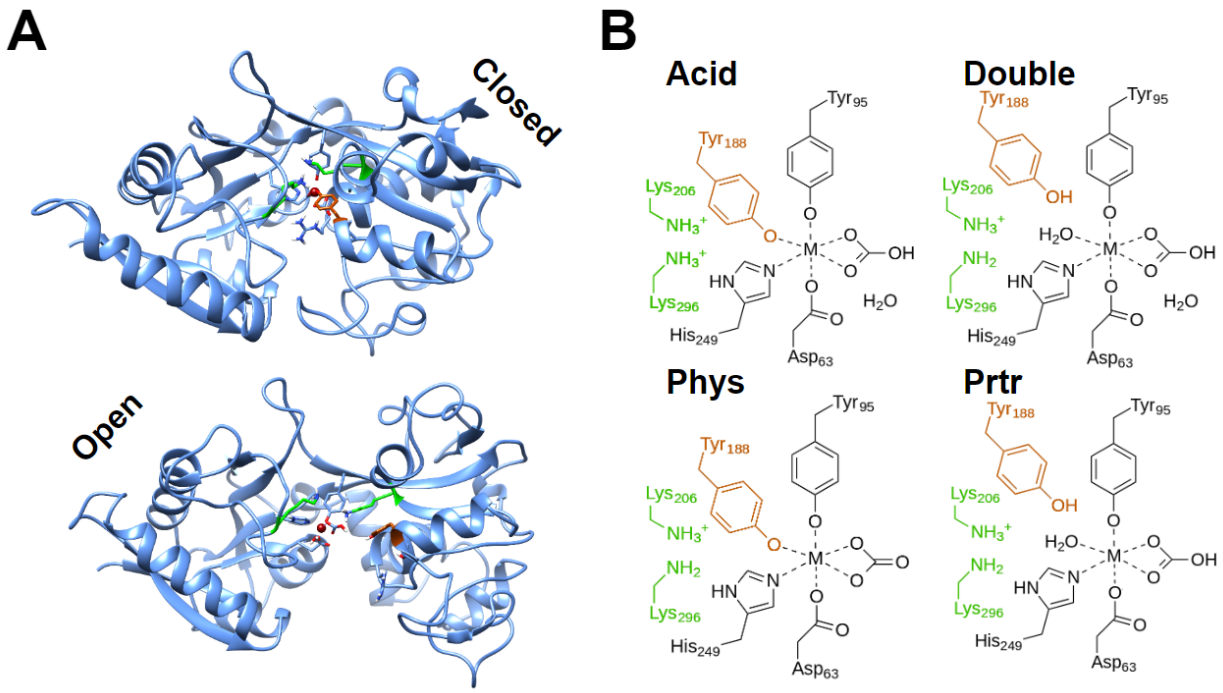


Figure 2.1. (A) Examples of open and closed hTF N-domain conformers with iron (in red) loaded in the metal binding site between two lobes. The closed form comes from a diferric bound crystal structure (PDB ID: 3V83) at blood serum pH, and the open form comes from a computationally generated structure from this study. (B) The hTF binding site in four different protonation states considered in our calculations based on structures from the reference 30. According to that study, the Phys and Acid forms yield closed conformers while the Double and Prtr forms, with Tyr188 protonated, become open. The difference between the Double and Prtr structures is the inclusion of an additional explicit water molecule in the binding site of the Double form. In both parts of this figure, the dilysine bridge is in bright green while Tyr188 is in dark orange.

2.2 Theoretical Methods

A total of 5 replicate QM/DMD trajectories were run for each metal for each form of the protein for a total of 140 simulations. Each trajectory corresponds roughly to 20 ns of simulation. Full rationalization and details about the preparation of each system can be found in Appendix A.

These simulations were performed with the established QM/DMD method.[44] This is a technique for sampling metalloprotein conformations using quantum mechanical (QM) electronic structure calculations necessary to model the metal and its coordination (referred to as

the 'QM region') and discrete molecular dynamics (DMD)[45] to describe the rest of the protein. Both methods treat an overlapping QM/DMD region, consisting of species participating in important, non-covalent interactions near the metal, to enable inter-region communication and mitigate discontinuity errors. QM/DMD has a strong record of successfully explaining a variety of metalloenzyme behaviors. A full description of QM/DMD's capabilities can be found in Chapter 1.2.

All QM calculations in this study were performed at the DFT level of theory with Turbomole (version 6.6).[64] The pure meta-GGA TPSS functional[65] was used with the D3 dispersion correction.[66] The metal was treated with the triple-zeta basis set def2-TZVPP and all other atoms with the double-zeta def2-SVP basis set.[67] While the small basis set may result in some degree of basis set superposition error, the large size of the QM regions have precluded the use of larger basis sets. Furthermore, the level of theory employed has proven effective in the past studies, cited above, including for quantitative free energy comparisons. Finally, the Conductor-like Screen Model (COSMO) with a constant dielectric of 4 was applied to approximate the screening and solvation effects in the relatively buried metal binding sites of the systems.[68] Water molecules which coordinate to the metal were modeled explicitly. The QM calculations were performed to convergence within 1.0×10^{-7} Hartree or at least 100 SCF cycles. This approach enhances sampling and plotting the energy trajectories shows most QM calculations are close to convergence by this point. All DMD phases in the iterative QM/DMD simulations in this study were performed for 10,000 steps per iteration (0.5 ns). DMD runs with an implicit solvent through appropriate potentials in its forcefield.

Convergence of the QM/DMD simulations was achieved according to a series of benchmarks. These consist of the protein backbone RMSD (calculated with respect to the alpha

carbon and amide carbon, nitrogen, and oxygen of each amino acid), the DMD energy, and QM energy. The RMSD values were calculated with the initial protein equilibrated for one QM/DMD iteration as the reference structure. The backbone RMSD trajectories of each system are included in this text as an example (Figure 2.2), while charts of the other two metrics can be found in Appendix A.

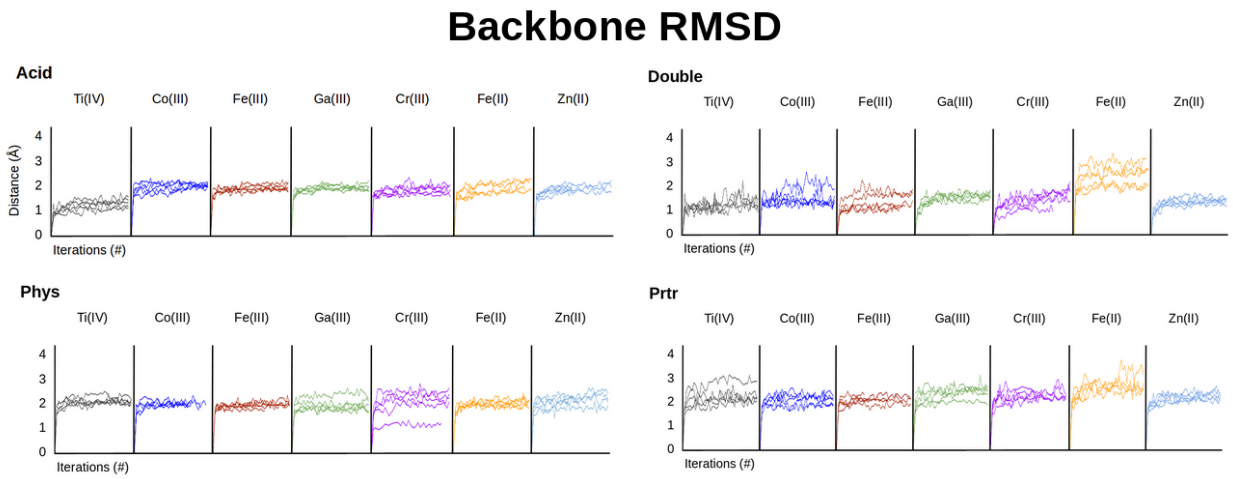


Figure 2.2. Plots of the backbone RMSD by timestep for every QM/DMD simulation in this study. The plots group all replicates by metal: Ti(IV), Co(III), Fe(III), Ga(III), Cr(III), Fe(II), Zn(II) for each protein protonation form: Acid, Phys, Double, Prtr. Most replicates oscillate around the value of 2 Å, indicating convergence. The open forms of the protein (Double and Prtr) report more variations from this value befitting their greater flexibility, especially with Fe(II), but still show convergence in each case.

The spin state for each metal was estimated and then verified so that the best was used to generate the data for this study. Geometry optimizations with DFT at the same level of theory as described for QM/DMD were performed on each feasible spin multiplicity for each metal in the acid form of hTF. The spin state of the optimized structure with the lowest electronic energy for each metal was then used for the appropriate QM/DMD simulations. After the simulations were completed, all reasonable spin multiplicities were tested for the lowest energy structure from

each trajectory of the acid forms of the most suspect metals: Fe(II), Fe(III), Cr(III), and Co(III). Fe(II), Fe(III), and Cr(III) were found to be high-spin, and Co(III) was found to be low-spin. The QM/DMD simulations for these metals were rerun with the corrected spin-states as necessary.

All computational binding affinities in this study were calculated by a new, relative approach.[5] A direct approach would involve computing free energies of the metal ion in solution, the apo-protein, and the metal-containing protein. However, the free energy of a metal cation in water is highly dependent on the local structure of water and therefore ill-defined. Furthermore, a metalloenzyme in its apo-form can be prone to unfolding or refolding and so assessing its free energy becomes prohibitively expensive. Instead, the new method uses EDTA-metal complexes as an intermediate step in a thermodynamic cycle (Figure 2.3). This cycle captures the energy of the unbound metal in the experimentally derived metal-EDTA binding energy and leaves the tractable terms of EDTA-protein transition to calculation. However, to fully cancel the EDTA terms, the approach can only calculate relative affinities between different metals within the same form of hTF.

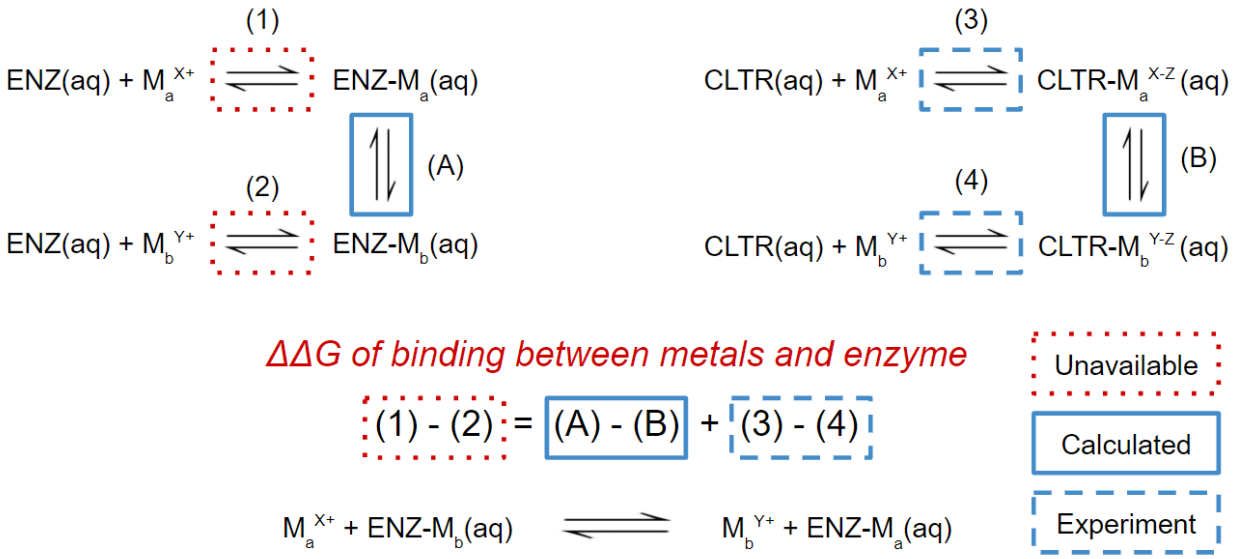


Figure 2.3. Thermodynamic cycle for the relative affinity of metal binding to hTF. The parenthesized label next to each box around an equilibrium process corresponds to the free energy of that transition. The desirable processes (1) and (2) are intractable as the structure of free metal ions in solution is not defined (dotted boxes). The new method in this study provides the free energy associated with the chemical reaction at the bottom. It is calculated as the sum of the difference between processes (3) and (4) which utilize available, experimental data for EDTA-metal binding (dashed boxes). This is summed with the difference between processes (A) and (B) which capture metal exchange (solid boxes). The result is a ΔΔG that reconstructs the difference between processes (1) and (2): the difference between the binding affinities of the metals.

The calculation of the relative binding affinities required just some data on EDTA besides information from the QM/DMD simulations. Experimental values for the free energy of binding of each metal to EDTA were obtained from tabulated data of beta110 stability constants.[82] The free energy calculation for each of the EDTA-metal complexes started with a DFT geometry optimization with the same software and at nearly the same level of theory used in QM/DMD. The only differences are that COSMO was given a dielectric of 84 corresponding to the water solvent and each optimization was run out to total convergence. The free energy was then

calculated from the optimized geometry with a harmonic frequency calculation with the same settings.

Further optimization of the QM region was done to generate the structures for free energy and metal angle variance calculations. This was done on the lowest lying unoptimized QM regions for each metal in each form of the protein. First, the three lowest electronic energy structures were optimized. The average deviation in the drop in electronic energy they experienced was taken. All unoptimized QM/DMD structures within two standard deviations of the lowest unoptimized structure were then selected for full optimization. Each set of structures were optimized to full convergence and free energies were calculated for them by a harmonic frequency calculation at the same level of theory and with the same software as above. Of these the structure with the lowest free energy was then selected as the representative minimum for its structure of the protein and metal.

2.3 Results and Discussion

Interlobal distances calculated from the QM/DMD simulations confirm the role of Tyr188 as the switch controlling the functional hTF conformational transition. The interlobal distance was calculated as the smallest distance between two sets of residue alpha carbons that define the two sides of the central binding pocket (Figure 2.4). One set consists of residues 12-14, 43-46, 290, and 291 while the other is comprised of residues 179-182. When the interlobal distance was determined for each iteration of all simulations, it shows that the Acid and Phys forms of the protein maintain a distance of about 5 Å, while the Double and Prtr forms vary over their trajectories but rarely shrink below 10 Å (Figure 2.5). The 5 Å distance corresponds to a closed conformation and any distance above that corresponds to an open conformation. This

demonstrates that the forms of the protein with Tyr188 deprotonated stay closed, regardless of the protonation state of the dilysine bridge (as this is protonated in the closed Acid form); the forms with Tyr188 protonated by contrast generally stay open but occasionally approach a closed state in Ti(IV) and Fe(II) Double. The simulations therefore corroborate the results of reference 110 on the role of Tyr188 over the dilysine bridge and shows that the open forms of the protein are more flexible.

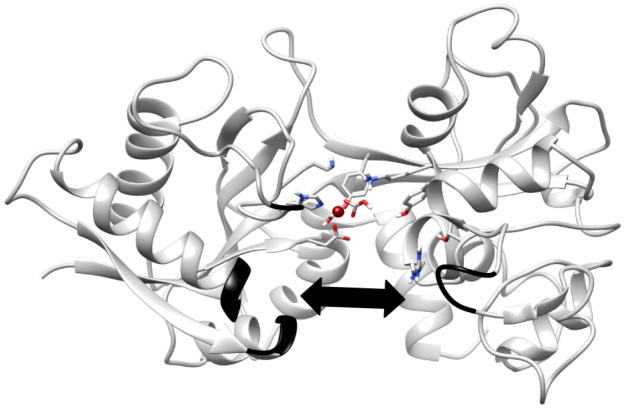


Figure 2.4. The interlobal distance was measured between the alpha carbon of the black loops opposite each other on the hTF cleft.

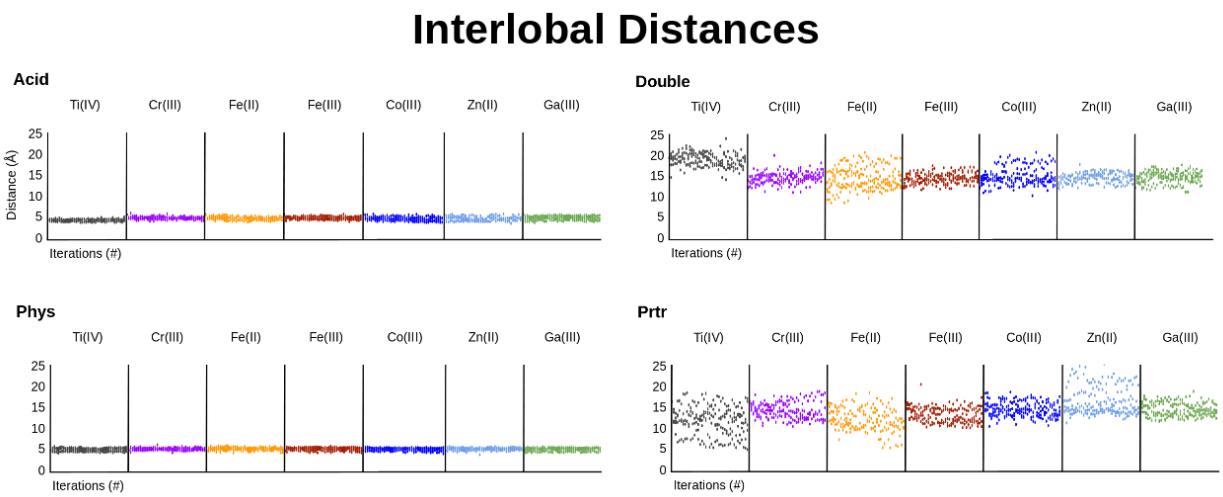


Figure 2.5. Plots of the interlobal distance for each form of the protein and each metal. The distance is recorded as a scatter plot with all replicates overlaid. These plots show that, regardless of metal, the Acid and Phys forms of the protein maintain a closed conformation for all metals with an interlobal distance of about 5 Å, while the Double and Prtr forms typically stay open with distances consistently higher.

The calculated relative free energies of metal-hTF binding qualitatively match experimental results. To calculate these free energy differences, we utilized the relative binding affinity approach described in the Theoretical Methods section. The most significant drawback to this approach is that the free energy of binding must be calculated relative to another metal. This is still a valuable and vetted technique: when applied to the lowest energy structures from the QM/DMD simulations with physiological Fe(III) as reference the results can be qualitatively compared to in vitro binding affinities. In fact, the experimental order determined and estimated in references 84 and 85 closely matches the order of the relative free binding energies calculated for the Phys form, the dominant structure in the experimental conditions (Table 2.1). The observed differences are for metals with similar enough experimental and estimated affinities and theoretical free energies to agree within an acceptable margin of error (due to aberrations in the experimental setup and computational techniques). Only Ti(IV), Co(III), Cr(III) deviate

significantly, and the experimental values for Co(III) and Cr(III) are suspect as they were extrapolated from binding constants to small molecules rather than directly measured.[84] This concurrence supports our methods for the calculation of the relative free energies of binding and what they say about hTF activity.

Table 2.1. The experimental[84,85] and calculated relative binding affinities to the N-terminal domain of hTF each sorted by metal in descending order. The experimental affinities reported as ranges (Co(III) and Cr(III)) were estimated. The energies are relative to their respective unscaled values of Fe(III), which correspondingly have values of 0 kcal/mol.

Exp. (kcal/mol)	Ti(IV): -5.8	Co(III): -2.0 to +1.9	Fe(III): 0	Ga(III): 2.6	Cr(III): 4.1 to 8.2	Fe(II): 20.1	Zn(II): 21.4
Calc. (kcal/mol)	Ti(IV): -37.5	Co(III): -14.8	Fe(III): 0	Ga(III): -3.2	Cr(III): 0.0	Fe(II): 26.5	Zn(II): 19.7

The order of the relative free energies of binding provides unique insight into the potential of various metals to compete with physiological Fe(III) for hTF - insight which could distinguish cytotoxic from healthy behavior. As the chemistry of the binding site is slightly different between the forms of the protein, relative free energies can't be calculated across the open to closed forms. However, comparisons can be made relative to Fe(III) for each state (Figure 2.6). As established above, the calculated binding free energies in the Phys form of the protein describe how well each metal is uptaken by hTF from solution akin to blood serum. Given the mechanistic role of the open conformers of hTF, orders for these forms of the protein correspondingly suggest how well each metal is released in conditions similar to the endosome. Our calculations therefore divide the metals into three categories based on how well they bind in the closed and open forms of hTF: (1) those that don't compete with Fe(III), (2) those that do, and (3) those that could outcompete Fe(III) for hTF. The first category comprises divalent metals Fe(II) and Zn(II). These metals bind worse than Fe(III) in both conformational states and

therefore do not compete for hTF as they are uptaken poorly and released readily. The second category is composed of Co(III), Cr(III). These metals, in contrast, can be transported by hTF as they bind better than Fe(III) in the Phys form but worse in the open forms of the protein. The final category may include Ti(IV) and Ga(III), which could interfere in natural hTF behavior. They bind better than Fe(III) in the Phys form, but according to the Double form of the protein are released about as easily as the physiological metal. Whether Ti(IV) and Ga(III) are in the second or third category is dependent on whether the Double of Prtr form is more mechanistically relevant as these disagree on the release potential of these metals. Note that these two metals are the non-essential and potentially cytotoxic ones in our study, so sequestering hTF could be related to their toxicity and merits future investigation. This is particularly important for Ti(IV) with its immense predicted affinity for the protein. As our study suggests that Co(III), Cr(III), and especially Ti(IV) and Ga(III) all compete with Fe(III) in hTF activity, it is important to understand the structural underpinning for this.

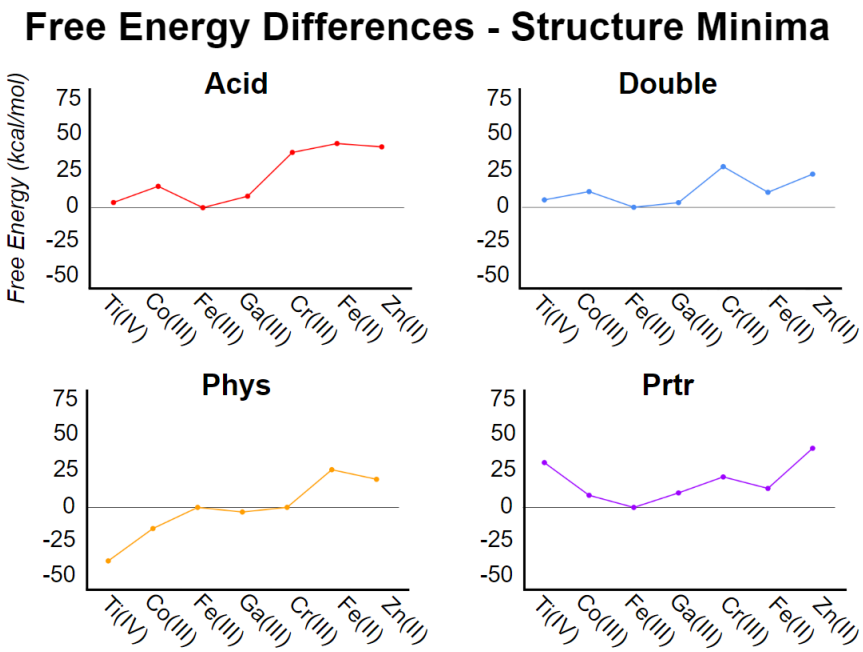


Figure 2.6. Free energies of binding relative to Fe(III) for each metal and form of hTF. Notice the marked difference in metal binding preferences across the four forms. These were all calculated for the fully optimized, lowest energy QM regions from the QM/DMD simulations.

The structural analysis of the QM/DMD simulations focuses on the binding site region, which varies the most with the generally subtle effect of different metals. Considered here are the metal angle variance and distance of critical hydrogen bonding interactions. Consistent with the observations from the interlobal distances, these analyses demonstrate that the closed Acid and Phys forms of hTF are rigid and do not significantly change over the course of the QM/DMD simulations, while the open Double and Prtr forms are flexible. However, the analyses do not identify a single interaction or property that distinguish individual metal binding preferences. Instead, the order of metal binding affinities arises from an ensemble of critical interactions implicated in hTF metal transport behavior.

The metal angle variance simply matches chemical intuition and does not correlate with the binding free energies. This is a measurement of how much the geometry of the metal deviates from the ideal octahedral. It arises from the equation: $\sigma_{oct}^2 = \frac{1}{11}\sum_{i=1}^{12}(\theta_i - 90°)^2$ which sums the difference of each of the 12 characteristic angles of an octahedral geometry from the ideal 90° (Figure 2.7). The metal angle variance was calculated for the optimized, lowest energy structure of each metal and protein form (Figure 2.8). The divalent metals Fe(II) and Zn(II) are the only ones to consistently deviate from octahedral, which makes sense as they generally prefer a tetrahedral geometry. Indeed, both of these metals typically reject a ligand to adopt a coordination closer to this geometry over the course of their QM/DMD simulations, in agreement with previous calculations by Sakajiri et al.[113] Otherwise, just Ti(IV) and Fe(III) vary a small amount from the baseline in open forms of the protein. Ultimately, the results further demonstrate the greater flexibility of the open forms of the protein, but do not identify a direct structural correlation to the free energy of binding besides the intuitive poor binding performance of the divalent metals.

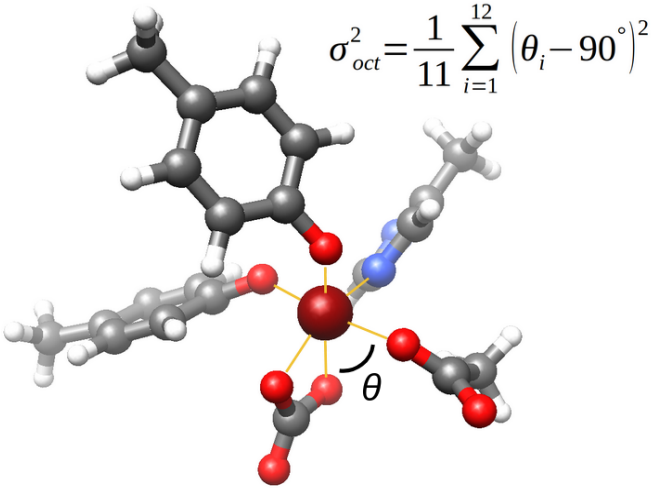


Figure 2.7. Example metal geometry in the hTF binding site; the twelve angles used to calculate the metal angle variance are defined between each pair of adjacent ligand bonds (solid lines).

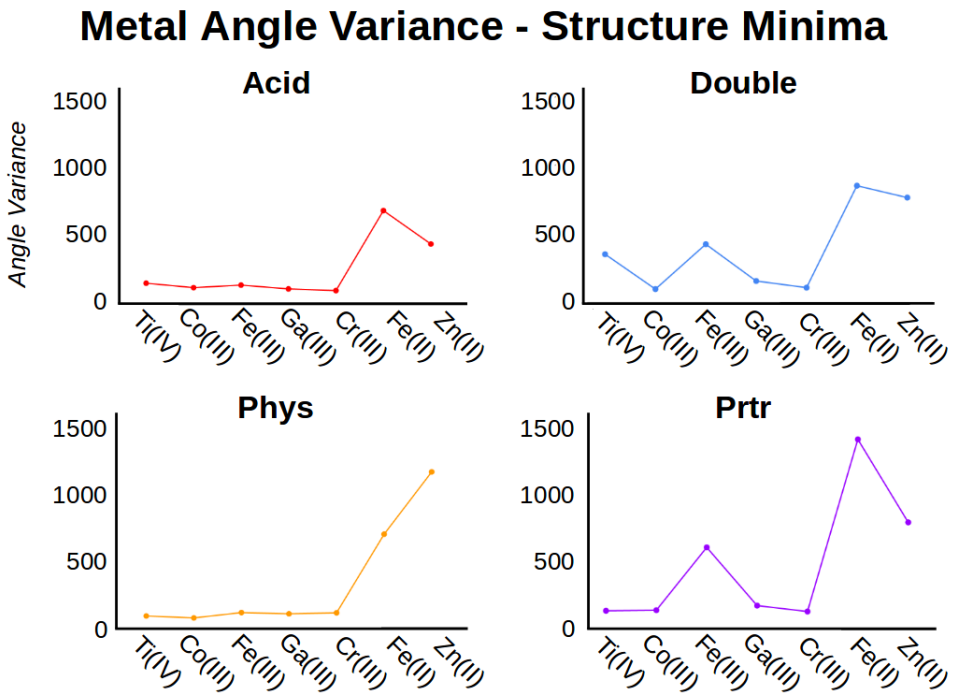


Figure 2.8. Metal angle variances calculated for each metal and form of the protein. Notice how only the divalent metals and occasionally Ti(IV) and Fe(III) deviate significantly from the low values and therefore octahedral geometries. As with the calculated relative binding free energies, these were calculated for just the fully optimized, lowest energy QM regions from the QM/DMD simulations.

There are a number of critical, hydrogen-bonding interactions around the binding site that vary significantly between metals and forms of the protein, but no single interaction directly correlates with the binding free energies across all systems (Figure 2.9). This makes sense if we consider that the roles of the secondary hydrogen-bonding interactions around the binding site are to keep the binding site residues in the optimum arrangement, while the electrostatic interactions of the residues coordinating to the metal most directly affect relative binding energy. Note that relatively good qualitative results were obtained in a previous study for the binding affinity in the Phys state by considering only first-shell residues provided that these are fixed at their corresponding binding site positions.[113] Regardless, the characterization of the hydrogen-bonding interactions around the binding site is important to at least understand the relevant interactions in the stabilization of the structure of the metal-binding site. Therefore, we have made a thorough analysis of the important, most varying interactions. For each of them, the smallest distance between potential atomic partners was recorded for each timestep of each relevant simulation. The data were summed over intervals of 1 Å and plotted as smoothed histograms by protein form with all metals color-coded and overlaid (Figure 2.10). The first set of distances considered here are between the synergistic carbonate anion and its hydrogen bonding partners on Arg124, Ser125, Tyr188. These graphs show that as the protein transitions from its closed to the open forms, Arg124 generally moves away from the characteristic hydrogen-bonding distance of 2.5 Å, while Ser125 and Tyr188 generally move towards it. These residues therefore take over the role of stabilizing the carbonate from Arg124. This is consistent with observations made in reference 110, which saw Arg124 as an indicator of protein conformational change. Another highly varying distance is between hydrogen bonding partners

on Asp292 and the water ligand found in the open conformers. The histograms show that this interaction is only present in the Prtr form, and is only preferred to a significant degree in the Co(III), Ti(IV), Fe(III), and Zn(II) forms of the protein. This interaction in particular has effects which extend out of the binding site, as Asp292 adopting it tugs on the loop comprised of residues 289-294 and changes its shape. Ultimately, these interactions vary significantly between metals, but none individually correlate with the calculated free energies. Instead, it is the composite of these interactions and the electrostatic interactions with first-shell residues which explains hTF behavior.

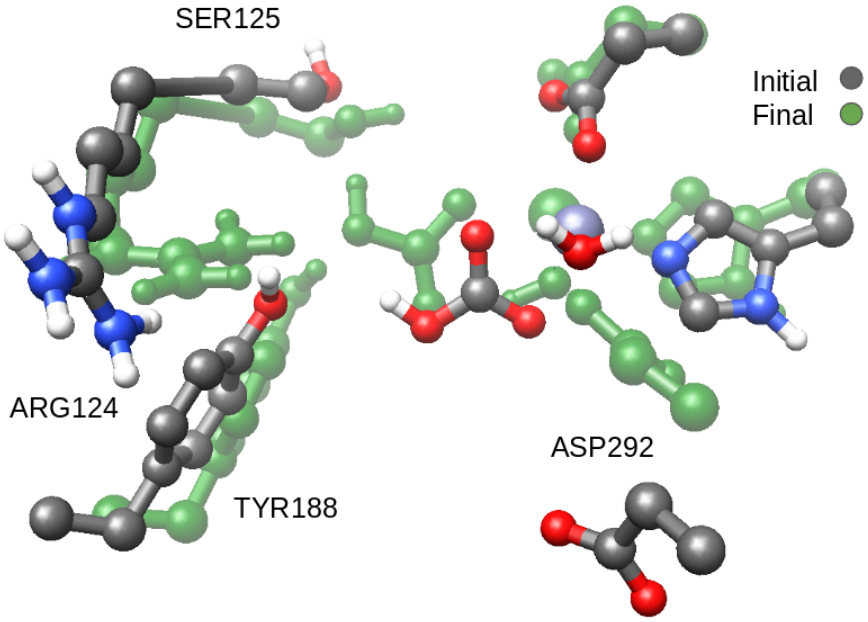


Figure 2.9. Overlay of two exemplary metal binding sites (both of the Prtr form, with gray from a briefly equilibrated structure and green from a structure toward the end of a simulation) from the QM/DMD simulations demonstrating the possible range of motion. The structures show how the Asp292-water distance and interactions between the carbonate anion and Arg124, Ser125, Tyr188 are flexible.

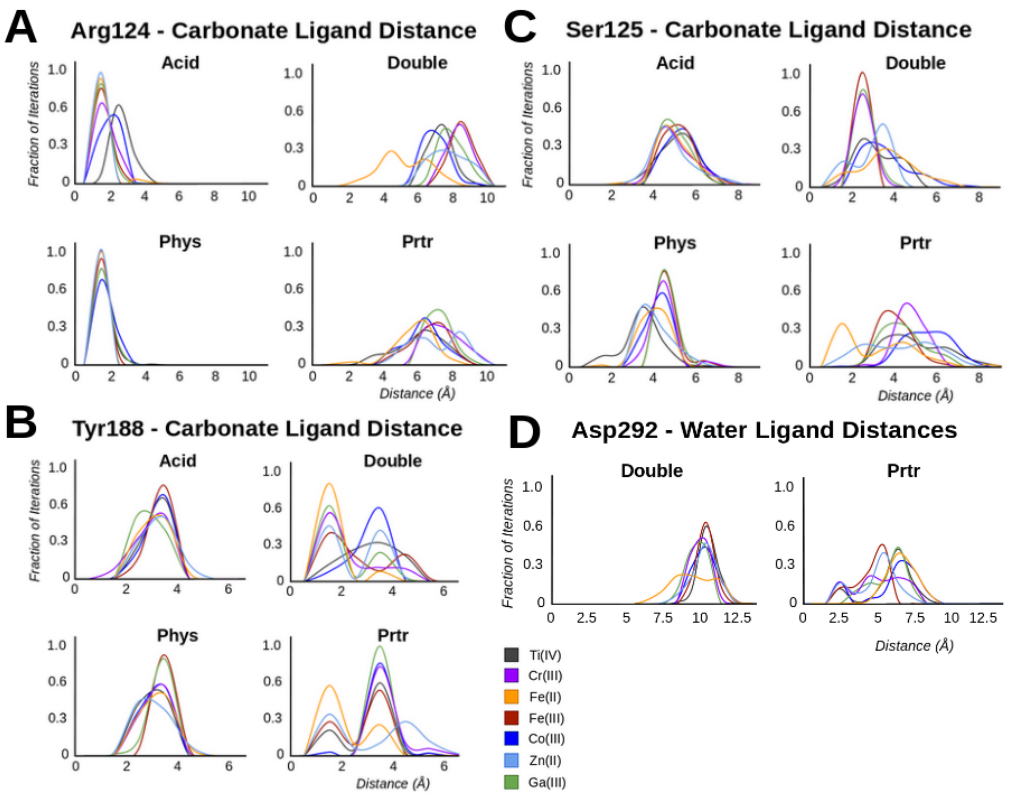


Figure 2.10. Histogram plots of critical interactions near the binding site, sorted by metal and form of protein. The histograms are constructed of the interaction distances calculated for all iterations across all replicates for each state. Included here are plots of the (A) Arg124-carbonate distance and it replacements in the (B) Tyr188-carbonate distance and (C) Ser125-carbonate distance. The gating effect of Arg124 is clearly visible as the interaction begins consistent with a characteristic hydrogen bonding distance of about 2.5 Å in the closed Acid and Phys forms, but generally disappears to a greater distance in the Double and Prtr forms as hTF opens; both Ser125 and Tyr188 see the opposite trend, greatly preferring hydrogen-bonding distances in the open Double and Prtr forms of the protein. (D) The stabilizing interaction between Asp292 and the metal water ligand is only present in the Prtr form.

2.4 Conclusion

The atomistic insight QM/DMD simulations provide into the hTF uptake and release process shows that cytotoxic metals can successfully compete with Fe(III) in transport and may get trapped in the protein, but are hard to target as the only observed differences are small and nuanced. The simulations confirm that the conformational hinging which drives hTF metal

transport is fundamental to each of the considered metals beyond just Fe(III). Furthermore, for all these metals the transition is dependent on the protonation of Tyr188, rather than that of the Lys206-Lys296 dilysine bridge. Orders of metal free energies of binding relative to physiological Fe(III) were calculated, suggesting how well different metals can be transported by hTF. The binding energies for the closed Phys form are consistent with experiment, while those for the open forms of the protein provide a unique, initial indication of hTF release preferences. Most importantly, the data from the Double form suggest that the root of Ti(IV) and Ga(III) cytotoxicity could arise from the difficulty by which they are released from hTF, which should be considered in future studies of their toxicity. Structural details of the simulations show that no single interaction explains the stability of the structure of the metal binding site, which instead arises from an aggregate of interactions largely about the metal binding site. Given the central role of the synergistic anion, future research should focus on this moiety for the purposes of drug development and protein redesign. Also of interest is Asp292 and the loop it rests on, which could be used to target the Prtr state specifically through the unique interaction it makes with the metals ligands in that form. The simulations demonstrate throughout that the open Double and Prtr states of hTF are flexible in both binding site interactions and protein conformation, while the closed Phys and Acid states are uniformly rigid. Ultimately, this study uncovered a nuanced network of interactions that could be modified to target hTF metal transport activity and address cytotoxic behavior and should be considered in any future development of drugs which use or target the human serum transferrin receptor.

Chapter 3

An Application of Competitive Metal Affinity Methods: Contrasting Effects of Inhibitors Li^{+} and Be^{2+} on the Catalytic Cycle of Glycogen Synthase Kinase-3β

3.1 Introduction: The Potential and Challenges of Li^+ as a Drug for Neurological Conditions

Glycogen synthase kinase-3β (GSK-3β) is an important therapeutic target for a wide range of neurological conditions. GSK-3β dysfunction is known or proposed to be implicated in bipolar disorder (BPD), Parkinson's disease, epilepsy, and Alzheimer's disease (AD) (according to the prominent tau hyperphosphorylation hypothesis).[114–116] Consequentially, GSK-3β is an attractive target for inhibitors to treat these conditions,[117–120] and some have even entered clinical trials.[121] However, no recent efforts have yet passed, encountering problems with toxicity and efficacy.[122] GSK-3β is difficult to target selectively as it is highly homologous with related kinases, particularly mitogen-activated protein kinases (MAPKs), cyclin-dependent kinases (CDKs), and protein kinase C (PKC). This family of proteins all bind an ATP molecule alongside two Mg^{2+} ions at a site adjacent to a peptide/protein substrate binding channel (Figure 3.1A).[123] As the most conserved region between these proteins is the ATP binding site, a well-studied potential solution to this are allosteric inhibitors.[124,125] Although ATP non-competitive compounds can improve selectivity, their more subtle modes of binding and inhibition make them difficult to develop into effective drugs. Study of a past, and rare, successful therapeutic agent would be especially helpful in the design of new ones.

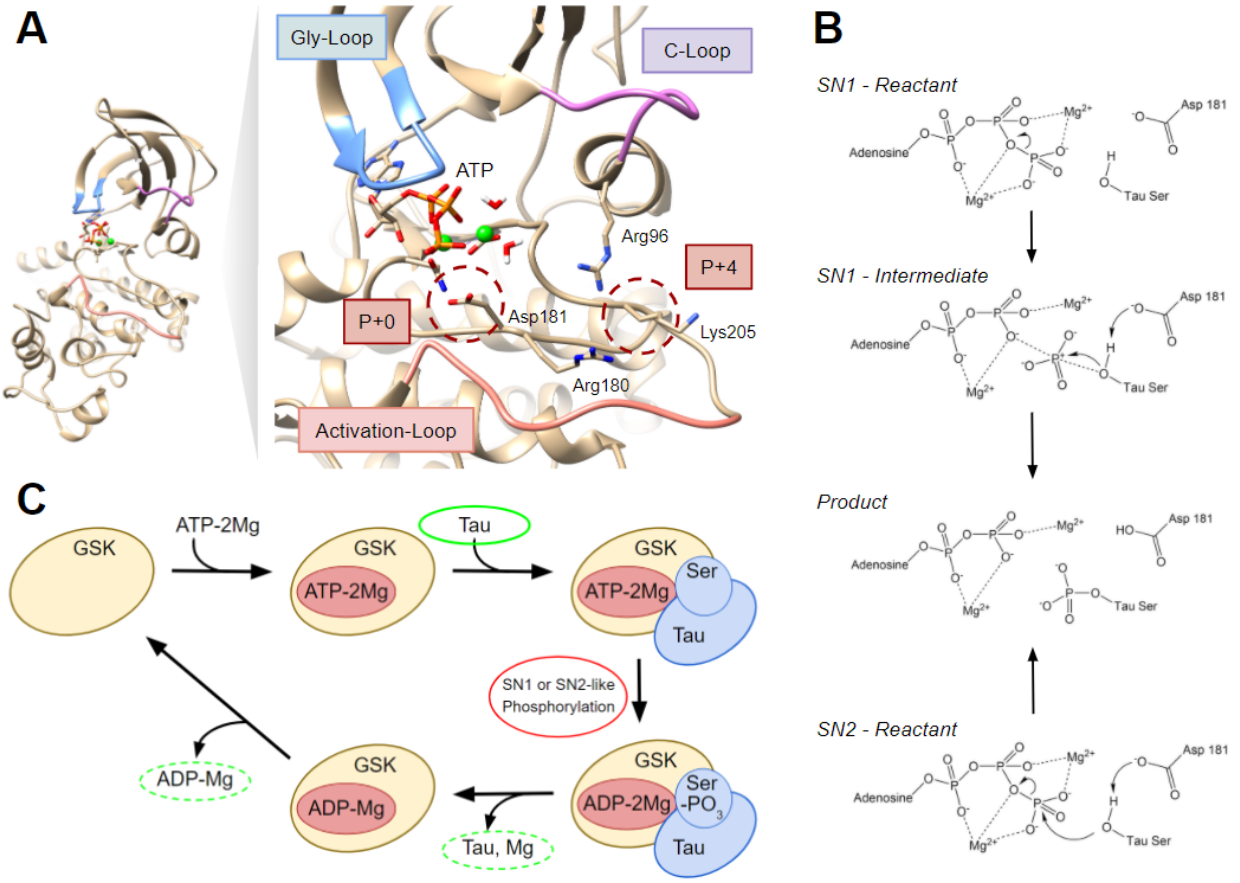


Figure 3.1. (A) Ribbon diagram of GSK-3β structure with insert showing the ATP binding site under the Gly-Loop (in blue) and the peptide/protein substrate binding channel, defined by the Gly-Loop, C-Loop (in purple) and Activation-Loop (in pink). The substrate serine or threonine residue binds at the P+0 site while a pre-phosphorylated residue binds at the P+4 site, held tightly by a series of charged residues (Arg96, Arg180, Lys205). (B) The dominant phosphorylation mechanism for GSK-3β based on studies of related kinases. The reaction proceeds by either an SN1 or SN2 path, with a nearby Asp181 residue acting as the base. (C) The catalytic cycle of GSK-3β. Li^{+} and Be^{2+} binding could inhibit activity by directly increasing the phosphorylation reaction barrier (circled red) or by disrupting the binding of the substrate (tau in this study, circled solid green) or either the unbinding of the substrate or ADP complex (circled dashed green).

Lithium is one of the most effective treatments for BD, and evidence for its beneficial effects on other neurological conditions have recently been gathering. The metal has been used for its mood stabilizing effect for BD since its serendipitous discovery in 1949.[126] *In-vivo* and even clinical studies report some beneficial effects on AD pathology, but with conflicting

results.[127,128] Li^+ therapeutic action is believed to be driven primarily through altering pro- and anti-apoptotic gene expression and GSK-3β inhibition.[129] Kinetic studies comparing Li^+ to Be^{2+}, another potent inhibitor, demonstrate that Li^+ binds non-competitively with ATP, just replacing one of the two native Mg^{2+} ions.[130,131] Remarkably, Li^+ has a limited effect on the activity of closely related CDK5, PKC, and MAPK.[114] Lithium does, however, have major drawbacks as a drug: a narrow therapeutic index, with therapeutic levels between 0.6 and 1.5 mEg/Lit,[132] and several side effects[133] – though it is specific, regarding altering activity of GSK-3β among other kinases, it clearly has other targets. An atomic-scale picture of GSK-3β activity and Li^+ inhibition would explain some of this behavior and aid the development of new, selective therapeutics without these drawbacks.

There have been many structural studies on related kinases and the biological effect of Li^+, but how the metal inhibits GSK-3β is not fully known. Experimental and computational research on the related protein kinase A (PKA) has established a likely SN1 or SN2-like mechanism of phosphoryl transfer facilitated by a nearby aspartate residue acting as a base (Figure 3.1B).[134,135] Quantum mechanical (QM) calculations on small mimic complexes of the GSK-3β metal binding site suggest that Li^+ can replace Mg^{2+} in only the ATP-bound forms of the protein,[136] but truncated model systems like these do not take the full effect of the protein environment into account. Later dynamics simulations on GSK-3β indicate that Li^+ preferentially binds to one Mg^{2+} site but not the other.[137,138] However, these studies did not confirm their results by calculating thermodynamic quantities such as binding energies or reaction barriers, nor did they consider the role Li^+ could play in other steps of the catalytic cycle.

This computational study investigated the atomic-scale effect of Li^+ and Be^{2+} on each relevant step of the GSK-3β catalytic cycle. Metal inhibition could arise from a direct increase to

the phosphorylation reaction barrier or disruption of the timely binding or detachment of other groups that could slow turnover: ATP, ADP, or the peptide substrate (Figure 3.1C). We employed the established QM/DMD method to study the structure of the ATP and ADP bound states of GSK-3β. This method combines QM calculations for appropriate treatment of the metal environment with discrete molecular dynamics simulations for rapid and extensive sampling of protein conformations. We also performed additional QM calculations to obtain binding affinities and phosphorylation reaction barriers based on the lowest energy structures identified by QM/DMD. We consider the effect of docking a tau protein fragment, a species implicated in AD, as an example substrate. We also identify structural differences between the native and Li^+ states that could serve as targets for future inhibitor development.

3.2 Theoretical Methods

This study began with dynamics simulations using the QM/DMD method.[44] This technique samples metalloprotein configurations using quantum mechanical (QM) electronic structure calculations (normally density functional theory) necessary to describe the metal and its coordination environment (referred to as the 'QM region') and discrete molecular dynamics (DMD)[45,139] to model the rest of the protein. Both methods treat an overlapping QM/DMD region which consists of species constituting the active site but not directly binding to the metal. This region enables inter-region geometric communication and mitigates discontinuity errors. QM/DMD has a strong record of successfully explaining a range of metalloenzyme behaviors. These include metal-dependent catalytic activity,[5,49] protein-metal binding affinity,[5,52] the effect of mutagenesis on structure,[46,47] and flexible docking of substrates.[140]

All of the QM calculations in this study were performed at the density functional theory (DFT) level using Turbomole (version 6.6).[64] The pure meta-GGA TPSS functional[65] was used with the D3 dispersion correction.[66] The metal was modeled with the triple-zeta basis set def2-TZVPP while all other atoms were treated with the double-zeta def2-SVP basis set.[67] A small basis set may result in some degree of basis set superposition error, but the large size of the QM regions (111-116 atoms) precluded any larger basis set. Regardless, this level of theory has proven effective in previous studies, including for quantitative free energy comparisons.[5,52] Lastly, the Conductor-like Screen Model (COSMO)[68] with a constant dielectric of 20 was applied to approximate the partial screening and solvation effects in the partly buried active site. Any water molecules that directly coordinate to the metals were modeled explicitly. All QM calculations were performed to convergence withtin 1.0 x 10^{-7} Hartree or at least 100 SCF cycles to afford more sampling during QM/DMD simulations. All DMD phases in the QM/DMD simulations in this study were performed for 10,000 steps per iteration (roughly 0.5 ns). DMD operates with an implicit solvent through the appropriate potentials in its forcefield.

QM/DMD simulations were performed on all the catalytic forms of GSK-3β for each metal binding state with one exception. Attempts to dock a short peptide proxy for tau protein were unsuccessful due to the poor potentials for the pre-phosphorylated residue in DMD. Thus, the simulations only capture the ATP and ADP bound states of the protein without tau, which proves to be sufficient for the purposes of this study.

QM/DMD simulations were performed on the 8 total metal bound states of GSK-3β-ATP and GSK-3β-ADP. For the ADP form these include the Mg^{2+}, Li^{+}, and Be^{2+} states. In the case of the ATP form, as there are two metal binding sites, these include the $2Mg^{2+}$, $Li^{+}Mg^{2+}$, $Mg^{2+}Li^{+}$, $Be^{2+}Mg^{2+}$, and $Mg^{2+}Be^{2+}$ states. The first denoted metal represents the site closer to the ATP

adenosine group while the second one represents the further site. Simulations were performed in 5 replicate QM/DMD trajectories for the native Mg^{2+} and Li^{+} containing states of the ATP form and 3 replicate trajectories for the other systems, i.e. a total of 30 simulations. Each trajectory was continued for 40 iterations, which roughly corresponds to 20 ns. Full rationalization of the construction of each system can be found in Appendix B.

The convergence of the QM/DMD simulations was achieved according to several benchmarks: the protein backbone RMSD, the DMD energy, and QM region energy. All QM/DMD simulations were converged by 40 iterations / ~ 20 ns by these metrics. The full convergence charts with respect to the three standards are shown in Appendix B.

Further optimization of the QM region was done to generate the structures for free energy and metal angular variance calculations. The ten QM regions for each system with the lowest unoptimized electronic energy were fully converged (when not already achieved) using the same level of theory as the QM calculations from QM/DMD. The free energy was then calculated for each with a harmonic frequency calculation. The structure with the lowest free energy was then selected as the representative minimum of each system.

Potential energy surface (PES) scans were performed to calculate the phosphorylation reaction barriers for the $2Mg^{2+}$, $Li^{+}Mg^{2+}$, and $Be^{2+}Mg^{2+}$ states of the protein. The two reaction coordinates were the distance between the O and H of the tau serine (from 0.8Å to 2.2Å), and the distance between O of the tau serine and the terminal P on ATP (from 1.4Å to 2.6Å), with a step of 0.1Å. The active spaces were truncated from the lowest energy QM optimized structures with a fragment of tau manually docked into these structured and truncated at the Cβ (capped with a hydrogen and with its position frozen in all calculations). Without full dynamics for the tau substrate, these constraints could introduce errors into the calculated energies and barriers, but

full dynamics could not be obtained as explained above. Full details of the system construction are found in Appendix B. Once the PES were plotted, the stationary points were fully optimized with the same level of DFT theory as the QM calculations from QM/DMD with accompanying harmonic frequency calculations to obtain free energies for the reactant and product states as well as any intermediates and transition states.

All metal binding affinities in this study were calculated through the recently developed competitive metal affinity (CMA) method,[141] which shows that it can obtain quantitatively accurate relative metal binding affinities compared to experiment.[52,141] This method determines the relative binding affinity to a protein compared to a reference metal using an indirect thermodynamic cycle dependent on experimental metal binding free energies to a chelator. The absolute binding affinities are computationally inaccessible (both resource intensive and error prone) because of the indeterminate structure of solvated metal ions. EDTA was used as a chelator complex for CMA calculations. The first benefit of this complex is that experimental binding affinities are available for nearly all metals, including Mg^{2+}, Li^{+}, and Be^{2+}.[142] Normally the second benefit of EDTA is that its metal-bound structure is known and well behaved: fully coordinating the metal in an octahedral geometry. However, this assumption may not hold for small, low charge metals like the alkali and alkali earth metals of this study. These metals can not satisfy the -4 charge of EDTA alone and so the complex is likely to coordinate additional water molecules – one of the complications that the CMA method seeks to avoid. Indeed, a crystal structure for a Ca^{2+} complex with EDTA shows only partial coordination of the metal with EDTA and a number of closely interacting waters.[143] Nonetheless, CMA binding $\Delta\Delta G$ were calculated using EDTA as the sole ligand in implicit solvent water, as no alternative complex with experimental binding affinities to the examined metals could be found. Free energies for the

octahedrally coordinated EDTA complexes were calculated in the same manner as the QM/DMD optimizations but with a dielectric constant of 84 for COSMO to represent the aqueous environment. To mitigate the shortcomings and verify the results using EDTA, an alternative approach using ATP and ADP as chelators was also pursued.

This study also relied on the free energy of exchange of Li^{+} and Be^{2+} from the solvated forms of ATP and ADP to the holo-protein. While they do not offer the same full coordination of EDTA, ATP and ADP are both chelators and recent research has gone into their structure and the position of the water molecules that complete their coordination shells.[144] The solution phase structures used in this study are based on the bimetallic complexes determined recently by Dudev et al. That study did not ascertain structures for Be^{2+}, so the input structures for Be^{2+} are the same as Li^{+}. The Li^{+} input structures were used as both metals prefer tetrahedral geometries while Mg^{2+} adopts an octahedral geometry. The solvated complexes were then fully optimized to account for structural differences and obtain their free energies in the same manner as EDTA, including the use of the COSMO dielectric of 84.

3.3 Results and Discussion

Analysis of the conformational ensembles we generated through QM/DMD for the GSK-3β-ATP form shows no evidence that the metal could alter the nature of peptide/protein substrate binding. To assess this, we calculated the variance of two important features of the protein binding channel from a crystal structure of GSK-3β bound to an axin-derived peptide (PDB ID: 4NU1).[123] While the axin peptide in 4NU1 is different from tau, the binding site should be the same due to the anchoring role the P+4 site plays for pre-phosphorylated substrates. We therefore selected the crystal structure as our reference to see if Li^{+} or Be^{2+} impede the adoption of a

structure predisposed to binding. The two geometric features of our analysis are the shape of the binding channel itself as defined by its loops and the structure of the critical P+4 site. If the channel defining loops or the positively charged residues of the P+4 site in the Li^{+} and Be^{2+} states are significantly more distant, this incurs an energetic penalty to pull them together. Furthermore, the greater distance would also make additional binding modes more likely. Together, these changes to the potential energy surface of this event would reduce tau binding affinity.

The variance in the conformation of the peptide channel shows no significant differences between the different metal bound states. We calculated this as the variance from the 4NU1 structure in the distance between the closest alpha carbon for each pair of loops. Our graph depicts the average of this value across all iterations for all replicates of QM/DMD for each metal bound form of GSK-3β (Figure 3.2). For all the metals, the average variance is small at about 2Å for the Gly-C distance and large, over 5Å, for the distances involving the Act loop. This simply represents the binding channel closing around the peptide upon binding. The variances for the Gly-C loop distance are 0.91 ± 0.57Å for $2Mg^{2+}$, 0.79 ± 0.55Å for $Li^{+}Mg^{2+}$, 1.03 ± 0.51Å for $Mg^{2+}Li^{+}$, 1.26 ± 0.87Å for $Be^{2+}Mg^{2+}$, and 0.66 ± 1.00Å for the $Mg^{2+}Be^{2+}$ form of the protein. The variances for the Gly-Act loop distance are 4.53 ± 1.79Å for $2Mg^{2+}$, 4.47 ± 1.32Å for $Li^{+}Mg^{2+}$, 5.09 ± 1.12Å for $Mg^{2+}Li^{+}$, 4.59 ± 1.03Å for $Be^{2+}Mg^{2+}$, and 5.98 ± 1.12Å for the $Mg^{2+}Be^{2+}$ form of the protein. Finally, the variances for the C-Act loop distance are 7.90 ± 1.39Å for $2Mg^{2+}$, 8.06 ± 1.41Å for $Li^{+}Mg^{2+}$, 9.33 ± 1.32Å for $Mg^{2+}Li^{+}$, 8.75 ± 1.94Å for $Be^{2+}Mg^{2+}$, and 8.64 ± 1.36Å for the $Mg^{2+}Be^{2+}$ form of the protein. However, the difference in the variance between the different metals is small, subsumed by their standard deviations. The shape of the binding channel does not look especially predisposed to protein binding with any metal over the others.

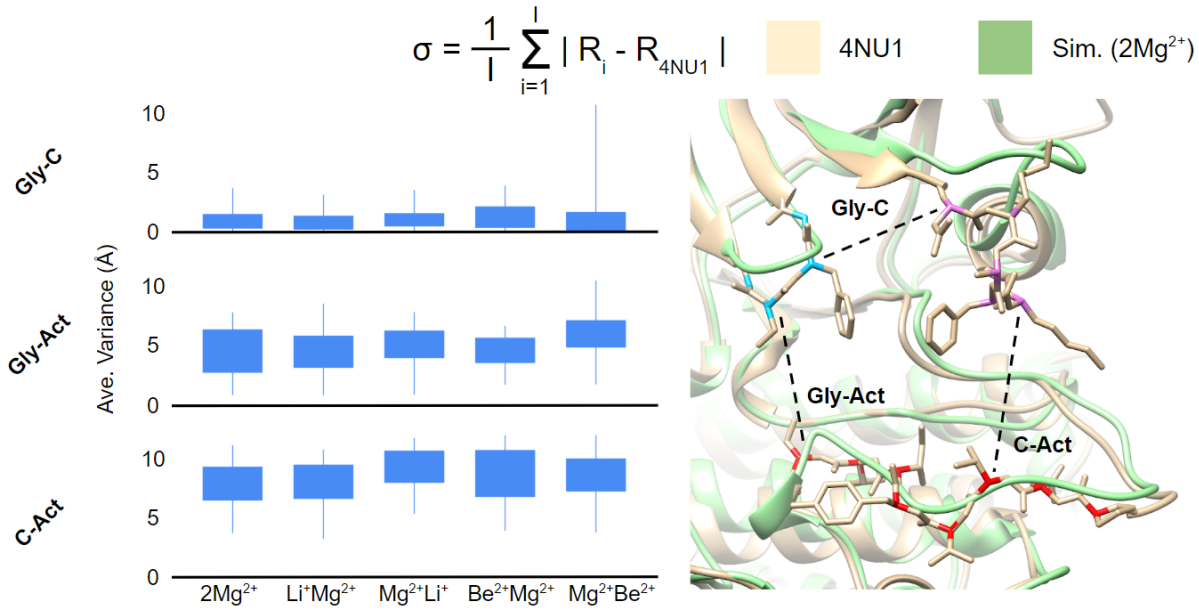


Figure 3.2. Variance of the peptide binding channel conformation in QM/DMD ensembles (green) from reference, peptide-bound crystal structure (4NU1). The green structure is an example from the $2Mg^{2+}$ simulations. The average variance for each metal-bound state is graphed on the left with the blue bar spanning one standard deviation above and below the average and the thin blue line showing the minimum and maximum loop distances. The variance was calculated by the equation at the top using the inscribed distances between loops in the picture on the right. The distances used were the minimum distance between any two alpha carbon on opposite loops. I is the number of QM/DMD iterations, R_i is the loop distance in the QM/DMD simulation, and R_{4NU1} is the loop distance in the crystal structure. Note that all the metal bound states report roughly the same variances.

The structure of the P+4 site is similarly invariant between the different metal bound states. As with the channel loops, we calculated the variance from the 4NU1 structure between alpha carbons, this time of Arg96, Arg180, and Lys205. In this case we averaged all three distances together for all iterations of all replicates of QM/DMD for each metal bound state (Figure 3.3). The variances are all around a small 0.5Å: 0.52 ± 0.25Å for $2Mg^{2+}$, 0.52 ± 0.25Å for $Li^{+}Mg^{2+}$, 0.52 ± 0.25Å for $Mg^{2+}Li^{+}$, 0.52 ± 0.25Å for $Be^{2+}Mg^{2+}$, and 0.57 ± 0.24Å for the $Mg^{2+}Be^{2+}$ form of the protein. As before the differences are minor between the metal states and well below the standard deviation of each ensemble.

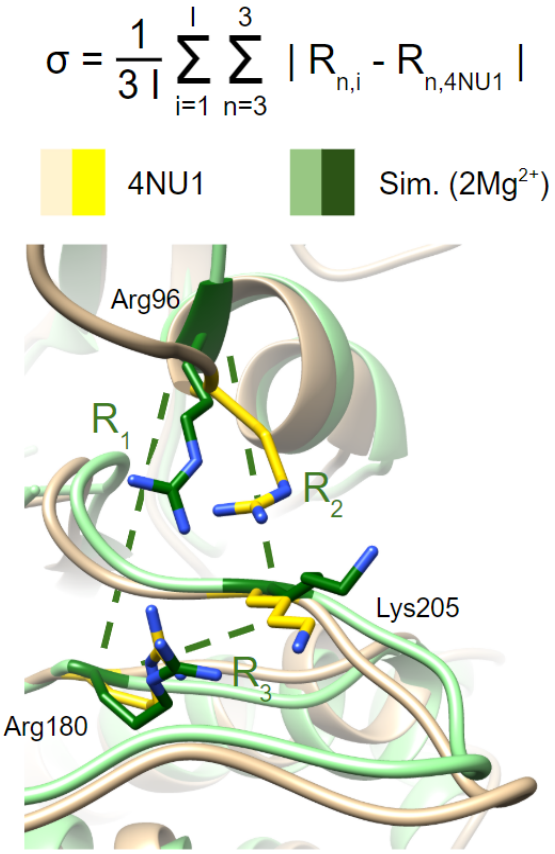


Figure 3.3. Variance of the P+4 peptide phosphate binding site geometry in QM/DMD ensembles (green) from reference, peptide-bound crystal structure (4NUI). The green structure is an example from a $2Mg^{2+}$ simulation. The variance was calculated using the equation at the top based on distances between the alpha carbon of three positively charged residues that define the P+4 site. I is the number of iterations, $R_{n,i}$ is the distance in the QM/DMD simulation, and $R_{n,4NU1}$ is the distance in the crystal structure.

The lowest energy active site structures from QM/DMD show significant differences when Li^{+} and Be^{2+} bind. The QM calculations suggest that the inhibitory metals preferentially bind to the first binding site closer to the adenosine group. The structures with the inhibitory metals occupying the second, further site are less stable by 10.9 kcal/mol for Li^{+} and 8.1 kcal/mol for Be^{2+}. For our future analyses we therefore consider metal substitutions only to the first, apparently dominant site.

Full investigation of the dominant active site structures immediately reveals significant geometric variations between the different metal bound states, dictated by the size and charge of the metal cations. The largest cation is Mg^{2+}, followed by Li^{+}, and Be^{2+}. This is demonstrated by the average ligand-metal distances which are 2.05 ± 0.06Å for Mg^{2+}, 1.94 ± 0.04Å for Li^{+}, and 1.64 ± 0.02Å for Be^{2+} in the ADP bound states, and 2.09 ± 0.1Å for Mg^{2+}, 1.95 ± 0.06Å for Li^{+}, and 1.63 ± 0.04Å for Be^{2+} in the dominant ATP-bound state. Furthermore, both Li^{+} and Be^{2+} prefer fewer ligands than Mg^{2+}; they feature tetrahedral coordination with both ATP and ADP, whereas Mg^{2+} is octahedral in both forms (Figure 3.4). Obviously, the doubly-charged cations attract nucleophilic ligands more strongly, resulting in more favorable electrostatics. On the other hand, smaller cations could favor fewer ligands, benefiting less from metal-ligand electrotactic attraction, but reducing the crowding of the coordination sphere and ligand-ligand repulsion. The total ligand-metal attraction, ligand-ligand repulsion, ligand and metal coordination strain, and binding site accommodation constitute a complex interplay that we aim to uncover. The resultant interactions in the binding pocket should also contribute to the site thermodynamics.

The joint effects of the metal charge, size, and ligand interactions are reflected in the differences in the strain that these complexes exhibit, judged by the metal angular variance. In the ADP bound state of GSK-3β, the phosphate tail of ADP can better accommodate the coordination of the smaller Li^{+} and especially Be^{2+} with less strain than the native Mg^{2+}. In addition, less crowded tetrahedral coordination should have weaker ligand-ligand repulsions and also tighter binding to the metal. In the ADP bound state of GSK-3β the metal angular variances are 14.3° for Mg^{2+}, a smaller 10.5° for Li^{+}, and only 3.7° for Be^{2+} (Figure 3.4). Smaller strain is expected to be associated with tighter binding, again featuring Be^{2+} as a strong binder, though the most valid comparison in this case is between Li^{+} and Be^{2+}, since both are tetrahedral. A similar

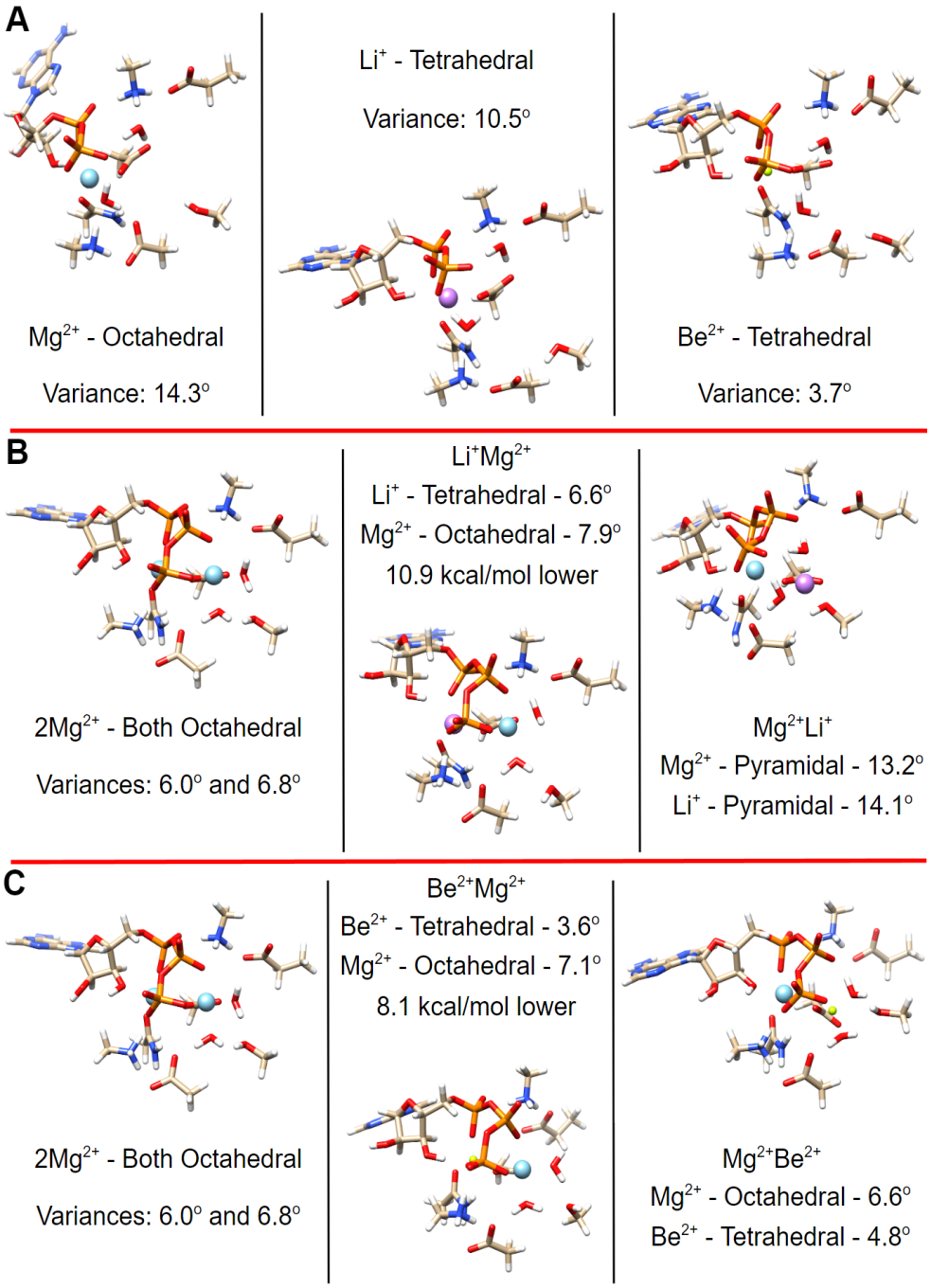


Figure 3.4. QM optimized geometries of the active site from the QM/DMD simulations for (A) ADP, (B) ATP with Li^+, and (C) ATP with Be^{2+}. Mg^{2+} is shown in cyan, Li^+ in pink, and Be^{2+} in yellow-green. Also reported for each structure are the preferred metal geometries with the average angular variance from ideal angles. The structures from the ATP-bound state also feature the free energy preference for metal substitution at the first binding site over the second. Note that the smaller Li^+ and especially Be^{2+} report much smaller angular variances.

trend emerges for the preferred ATP bound states. A notable difference in this case, however, is that Li^+ reports a comparable angular variance of 6.6° to the 6.0° and 6.8° of the native Mg^{2+} structure, while that of Be^{2+} continues to be significantly smaller at 3.6°. Relative to Mg^{2+}, Li^+ coordination appears less strained with ATP than with ADP, which (while being just one of the relevant parameters) can be expected to yield a greater barrier to the catalytic reaction step for Li^+ compared to Mg^{2+}. Be^{2+}, being the tightest binder of both ATP and ADP, should feature specific reactivity as well. Mg^{2+}, while having greater variance, benefits electrostatically from having more ligands in the octahedral sphere, and the exact energetic balance is thus far unclear. The compounding effects of the geometric trends on the thermodynamics of binding and reactivity are evaluated next.

Binding affinities that we calculated suggest that Li^+ and Be^{2+} can replace native Mg^{2+} in GSK at various points during its catalytic cycle. When Mg^{2+} is used as the reference metal in our CMA method, the ΔΔG of replacement with Li^+ or Be^{2+} are uniformly negative for the ATP and ADP bound forms of GSK-3β. These values are -16.7 kcal/mol for Li^+ and -38.1 kcal/mol for Be^{2+} in the ADP bound form of the protein and -17.6 kcal/mol for Li^+Mg^{2+}, -6.7 kcal/mol for $Mg^{2+}Li^+$, -23.0 kcal/mol for $Be^{2+}Mg^{2+}$, and -14.9 kcal/mol for $Mg^{2+}Be^{2+}$ in the ATP bound form of the protein. The ΔΔG correspond to the following reactions

$$GSK_{ADP,Mg} + M \rightarrow GSK_{ADP,M} + Mg^{2+} \quad (3.1)$$

$$GSK_{ATP,2Mg} + M \rightarrow GSK_{ATP,M,Mg} + Mg^{2+} \quad (3.2)$$

Where M can be either Li^+ or Be^{2+} and $GSK_{ADP,Mg}$ and $GSK_{ADP,M}$ are the ADP bound form of GSK-3β with Mg^{2+} and the substituted non-native metal in the active site, respectively. This notation carries for equation 3.2 and all that follow. The uniformly negative competitive metal affinities indicate that Li^+ and Be^{2+} readily replace Mg^{2+} and bind to GSK more tightly at all

considered points during the catalytic cycle. While ambiguities with the EDTA complexes discussed above might contribute to error, these would cancel out when comparing the ADP and ATP states, corresponding to reaction 3:

$$GSK_{ADP,Mg} + GSK_{ATP,M,Mg} \rightarrow GSK_{ADP,M} + GSK_{ATP,2Mg} \quad (3.3)$$

We see that the ΔΔG between the dominant ATP state and the ADP state is far larger and negative for Be2+ (-15.1 kcal/mol) compared to Li+ (0.9 kcal/mol). As a check, we also calculated the ADP-ATP ΔΔG based on the binding ΔG from the solvated forms of ATP and ADP. Each ΔG calculated this way corresponds to the reactions

$$GSK_{ADP,Mg} + ADP_{M} \rightarrow GSK_{ADP,M} + ADP_{Mg} \quad (3.4)$$

$$GSK_{ATP,2Mg} + ATP_{M,Mg} \rightarrow GSK_{ATP,M,Mg} + ATP_{2Mg} \quad (3.5)$$

Therefore, subtracting the binding ΔG for GSK-3β-ATP from GSK-3β-ADP yields energies consistent with the following reaction

$$\begin{array}{l} GSK_{ADP,Mg} + ADP_{M} + GSK_{ATP,M,Mg} + ATP_{2Mg} \rightarrow \\ GSK_{ADP,M} + ADP_{Mg} + GSK_{ATP,2Mg} + ATP_{M,Mg} \end{array} \quad (3.6)$$

The results for equation 3.6 are qualitatively consistent with the earlier EDTA-based CMA approach (equation 3.3), with +11.9 kcal/mol for Li^+ and -12.3 kcal/mol for Be^{2+}. Together, our results suggest that Li^+ has a preference for binding the ATP form of GSK-3β, while Be^{2+} prefers binding to the ADP form. This is consistent with experiment, showing that Li^+ can not easily bind alongside ADP while Be^{2+} does. The particularly small size and angular variance of Be^{2+}, as well as its greater charge relative to Li^+, and better placement of Be^{2+}-bound ADP in the binding pocket all likely contributes to this.

Analysis of the GSK-3β phosphorylation mechanism shows that Li^+ inhibition is driven by its direct reduction of the rate of the phosphorylation step itself, as inferred from the geometric differences between Li^+ and Mg^{2+}. The mechanism involves the deprotonation and

phosphorylation of a serine residue either in concert or sequentially. To assess this we calculated potential energy surfaces.

The PES for the tau phosphorylation are shown in Figure 3.5, and they exhibit significant differences for the three metals. The deep blue wells in the lower right correspond to the reactant states. In the upper left, a well for the product state can be found for Mg^{2+} and Li^{+}. However, the Mg^{2+} and Li^{+} mechanisms are slightly different. The Mg^{2+}-based PES reveals a small well corresponding to an intermediate, whereas for Li^{+} the reaction consists of a single step and has no intermediate. Therefore, the reaction proceeds as SN1 for Mg^{2+} and SN2 for Li^{+}. Furthermore, full optimizations of the stationary points on the PES and frequency calculations yield free energies that suggest a reduced rate of phosphorylation in the Li^{+} bound protein compared to Mg^{2+}. In the native protein the reaction is slightly exothermic, by 0.1 kcal/mol with our level of theory. The highest of its two reaction barriers is a reasonable 17.9 kcal/mol. By contrast in the Li+ case, the reaction is majorly endothermic by 19.6 kcal/mol, and the barrier is 26.0 kcal/mol. This is in line with the qualitative predictions based on the geometries and binding energies of ATP-Li^{+} and ADP-Li^{+} relative to Mg^{2+} forms, presented earlier, and is additionally related to geometric effects in the binding pocket discussed in the following paragraphs. Hence, the reaction with Li^{+} is both thermodynamically and kinetically unfavorable, enabling Li^{+} to take an inhibitory role on GSK-3β.

The PES for Be^{2+} shows no well for the product state within the investigated mechanism. This result seems surprising, given that the ADP-Be^{2+} state, the product of this reaction, is very stable and preferred over the ATP-Be^{2+} state (see the previous paragraphs). However, this is clarified in the next few paragraphs, where we discuss how structural misalignment of the ATP-Be^{2+} system impedes efficient phosphorylation.

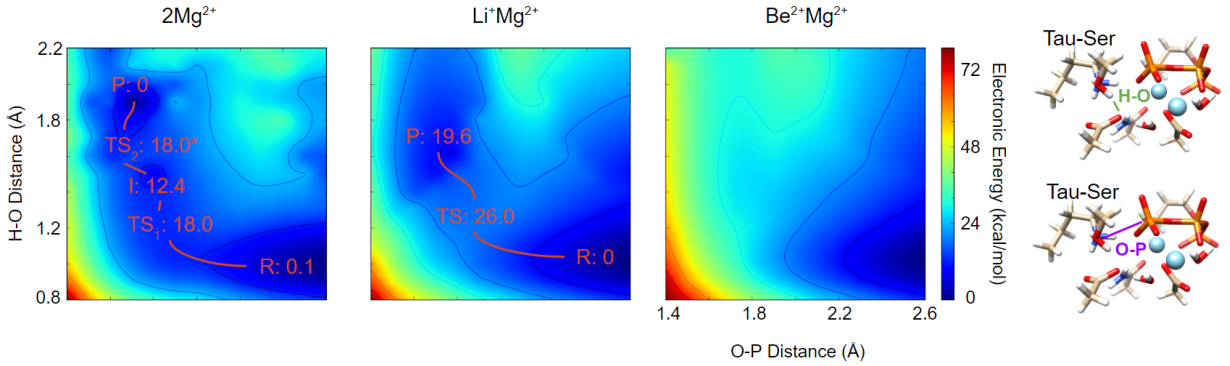


Figure 3.5. Potential energy surfaces for serine phosphorylation in the native form of GSK and the Li^+ and Be^{2+} bound forms. These plots vary the two reaction coordinates: serine hydroxyl hydrogen to aspartate carboxyl oxygen and serine hydroxyl oxygen to ATP terminal phosphorus. Free energies are shown as a heat map going from blue (low) to red (high). Stationary points along the reaction path are recorded in red for the native and Li^+ states. Note how $2Mg^{2+}$ reports two barriers, Li^+ reports one higher one, and Be^{2+} shows no product state and can't perform this reaction.

The interactions between the metals and the binding pocket of GSK-3β could reveal the structural root of the differences between phosphorylation mechanisms and inform future drug development. The structural features that show the greatest difference between the three considered metal-bound states of GSK-ATP are (i) the distance between the final ATP phosphate group and the oxygen on Asp181 that acts as the base in the reaction and (ii) the angle of the final phosphate group relative to metal centers (Figure 3.6A-B). We tracked each of these quantities both across the full ensemble of structures generated with QM/DMD and for the phosphorylation reactant structures based on the lowest energy structures from QM/DMD. The Asp181 to final phosphate distance can affect activity. The distance is much smaller for the Mg^{2+} form both on average and for the phosphorylation reactant states than in the Li^+ form. The Be^{2+} form shows a relatively low distance as well, so its product-less PES likely arises from other quantities. The correlation between activity and the Asp181 distance, at least for Mg^{2+} and Li^+, makes sense as Asp181 is the only negatively charged group near the active site that could accept

a proton: its greater distance from the substrate Ser and ATP phosphate incurs a greater reaction barrier as the proton or phosphate group must move more over the course of the reaction. In light of our results, we hypothesize that a potential inhibitor which can provide Asp181 with a competing non-covalent interaction would have a similar effect to Li^{+}.

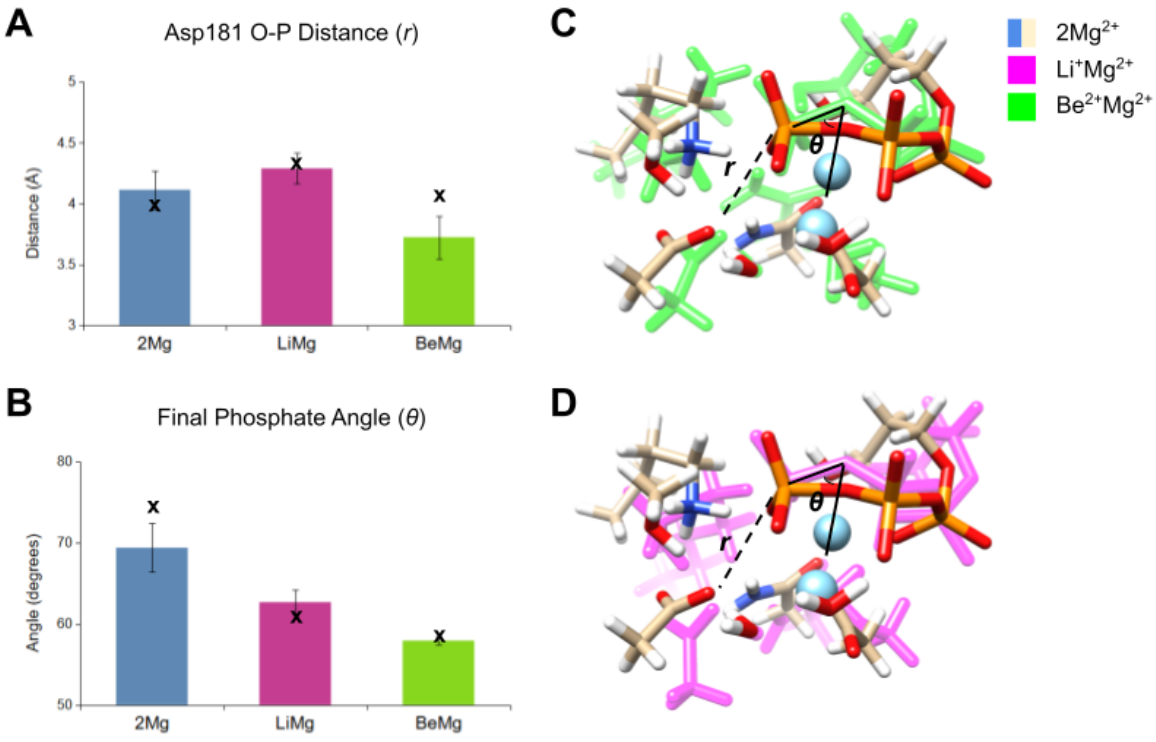


Figure 3.6. Bar charts of the average and standard deviation of the (A) Asp181 oxygen to final ATP phosphorous distance (r) and (B) final ATP phosphate angle relative to the metal centers (θ) across the full ensemble of structures from the QM/DMD simulations. The values from the phosphorylation reactants are plotted for each state as an 'x'. Blue is the native Mg^{2+} form, magenta the Li^{+} form, and green the Be^{2+} form. The distance *r* is based on whichever carboxylate oxygen in Asp181 is closest to the final ATP phosphorous in each given structure. The angle θ is inscribed by the vector of the last bridging ATP P-O bond and the vector between the last bridging O and midpoint between the two metals. Note that the values for the phosphorylation reactant Mg^{2+} are distinct from those of the Li^{+} and Be^{2+} states. Both quantities are marked on the structures on the right. Structural deviations of the phosphorylation reactant geometry in the (C) Li^{+} and (D) Be^{2+} bound states are also shown. In both, the native Mg^{2+} geometry is overlaid in tan while the metal substituted geometry is colored pink or green respectively. Note how for both Li^{+} and Be^{2+} the final phosphate group is angled down and away from the tau serine substrate.

The angle of the final phosphate group correlates fully with activity across the metal bound forms of GSK-3β and likewise yields opportunities for new inhibitors. Both for the phosphorylation reactant structures and the full QM/DMD ensemble, the Mg^{2+} form reports the largest angles, followed by Li^{+}, and finally Be^{2+}. The implications of the angle on the phosphorylation reaction barrier is clear from inspection of the reactant structures. While the final ATP phosphate faces the tau serine substrate with Mg^{2+}, it is shifted slightly down and away from it in the inhibitory structures (Figure 3.6C-D). This is particularly true of the Be^{2+} active site. This incurs an energetic penalty as the phosphate or tau must move and geometrically reorganize to perform the reaction. In the case of Be^{2+} this penalty appears too large for the reaction to occur at all according to the phosphorylation PES. This is most likely due to the smaller size of Li^{+} and especially Be^{2+}, as they pull the phosphate group toward them with their tighter coordination sphere. Though metal properties appear to drive the final phosphate angle, we suspect any structural changes which shrink that angle would reduce GSK-3β activity, given the resulting poor alignment of the reactant moieties. This might be achieved by crowding of the Gly-C loop, located directly above the ATP binding site.

Finally, based on these structural differences, we put forward a hypothesis regarding the mechanism of Be^{2+} inhibition of GSK-3β. In part, it is not principally different from that of Li^{+}: because Be^{2+} features an even tighter binding of ATP than Li^{+}, the misalignment with the tau serine is also greater, further making the reaction unfavorable. However, Be^{2+} can bind ADP, the product of phosphorylation, even tighter than ATP. While the PES for the reaction suggests that ADP cannot be reached in the Be^{2+} form of the protein, Be^{2+}-ADP could be recruited from solution. Hence, due to very strong binding, both ATP and ADP are predicted to stay bound and

inhibit the protein with Be^{2+}, reducing the turnover number of GSK-3β . The opportunity for ADP-based inhibition is accessible only to Be^{2+}, and not to Li^{+}, which constitutes the largest difference in their inhibitory mechanisms. One possible way to check this proposed mechanism would be to bind a small trivalent cation instead of Be^{2+}, to enforce an even tighter binding. No smaller cations than Be^{2+} exist, to the best of our knowledge, but higher valency might still lead to a similar effect.

3.4 Conclusion

GSK-3β is an important, but challenging, target for treating neurological and psychiatric disorders. Atomic-scale information about the effect of rare, successful inhibitors is helpful to design new inhibitory drugs. In this study, we identified how Li^{+} affects GSK-3β structure to inhibit its activity and how this contrasts with another metal ion inhibitor, Be^{2+}. Li^{+} binds most strongly to the ATP bound form of GSK-3β and directly increases the barrier for the phosphorylation reaction through poor orientation of the transferring final phosphate group. This differs from Be^{2+}, which binds ATP tighter than Li+ does and can also tightly bind ADP. The bound ATP is so strongly attracted to the compact and highly-charged Be^{2+} that it becomes too contorted and out of alignment with the rest of the binding site to undergo subsequent phosphorylation. If the protein recruits Be^{2+} alongside ADP, then this binding is even stronger. We therefore conclude that Be^{2+} stays bound to the protein either with unreacted ATP or ADP. In the case of both metals, their inhibitory effect arises from their small ionic radii relative to the native Mg^{2+}, exacerbated in Be^{2+} by its comparably high charge. While these are metallic properties, a similar effect might be achieved by new therapeutics that provide competing

interactions to an Asp181 residue that acts as the base during phosphorylation or crowding the final ATP phosphate group.

Chapter 4

Titr-DMD – A Rapid, Coarse-Grained Quasi-All-Atom Constant-*pH* Molecular Dynamics Framework

4.1 Introduction: Challenges in Modeling the Dynamics of Protein Protonations and Deprotonations

Solution *pH* is a chemical property with an immense effect on protein behaviors that are difficult to study at the atomic scale. Peak protein fold stability and catalytic activity are both dependent on an often narrow range of *pH*. Understanding the sequential and structural underpinning of these preferences contributes to the design and application of enzymes, particularly extremophile enzymes – which would allow for their use in harsher reaction conditions in industrial catalysis,[145–148] and answers a wide range of questions of medical interest as precise *pH* regulation is critical for cellular homeostatis.[149–151] However, this understanding demands atomistic information of fundamentally dynamic phenomena. *pH*-dependent dynamics is challenging to study experimentally, requiring a combination of techniques such as NMR monitored *pH*-titration, circular dichroism (CD) spectroscopy, and X-ray crystallography none of which alone provide the complete picture. Experimental complexity leaves computational investigation[152] as a critical tool to fill in the gaps.

Successful computational methods that assess *pH*-dependent protein behavior must accurately couple amino acid protonation state change with conformational dynamics. Typically, continuum electrostatic methods describe the protonation states of amino acids, assessing the free energy of protonation and deprotonation events or pK_a. Various solutions to the Poisson-Boltzmann equation can provide this,[153] especially the generalized Born model.[154,155] Simpler electrostatic methods are used as well. Tools such as UHBD,[156] H++,[157] and Propka[158,159] predict the pK_a of amino acid residues. Other tools,[160,161] including FPTS[162] rely on Monte Carlo simulations to sample protonation states as well as solvent and/or ion configurations in some cases. All of these methods are useful to study many *pH*-dependent protein properties, including

charge regulation during complexation and prediction of some titration curves.[163] However, these methods operate with largely static structures for the protein with little or no backbone motion, and so can not fully capture *pH* dependent dynamic behavior on their own. Molecular dynamics (MD) can provide the missing conformational sampling. Such combinations are known as constant *pH* molecular dynamics (CpHMD); these methods generally use electrostatic methods to model the protonation state changes of amino acids over the course of a molecular dynamics simulation.

The appropriate sampling of *pH*-coupled dynamics is difficult to achieve for all CpHMD methods and challenging to verify. The choice of solvation model is central to sampling and broadly breaks CpHMD methods into two categories: those using explicit solvation and those using implicit solvation. Explicit solvent based methods can provide greater accuracy through atomistic solvent treatment,[164,165] but sufficient sampling is difficult to achieve, as both conformational and protonation states need to be sampled. Furthermore, protonation sampling is affected by poor overlap between solvent configurations such that protonation state changes are often immediately rejected. To counter this, many groups have applied λ-dynamics, based off pioneering work by Brooks et al.[166] (in turn based on earlier work with other thermodynamic properties in mind),[167,168] which treats the protonation state of individual amino acid sites as continuous degrees of freedom rather than discrete ones sampled distinctly.[169–171] Other efforts focus on enhancing/accelerating conformational sampling through GPU processing[172] or replica exchange.[173–175] Implicit solvent-based methods offer increased sampling without acceleration techniques by treating the surrounding solution as a simple dielectric medium.[176–179] Implicit solvent therefore avoids the issue of solvent configuration sampling altogether.

Another approach to improve sampling is the use of coarse-grained (CG) molecular dynamics methods. CG can be implemented with either explicit or implicit solvent. Most often, CG models reduce the number of particles needed in a simulation by condensing atoms into supra-atomic beads. CpHMD methods based on these kinds of CG force fields, including Martini,[180] HiRE-RNA,[181] and OPEP6,[182] have been recently developed. Use of supra-atomic beads is, of course, more approximate that all-atom methods so this class of CpHMD methods is most attractive for particularly large systems, such as multiprotein complexes, or long timescales, such as those of protein refolding. Another form of CG is to simplify the MD force field potentials, reducing the number of calculations needed each timestep. This sort of CG is done in methods such as discrete molecular dynamics (DMD)[45,139,183] with square-well potentials used in place of continuous ones. Such CG potentials allow for quasi-all-atom simulations (with only some non-polar hydrogen excluded from full atomistic treatment), unlike other CG models. To our knowledge, no CpHMD methods based on the CG potential paradigm of DMD yet exist.

Ultimately, regardless of protonation scheme, solvation, and use of CG, verification of the generated ensemble of conformational and protonation states is of great important in all CpHMD methods. This is not trivial due to paucity of complementary experimental results. More plentiful indirect evidence, such as reconstruction of titration curves or estimation of experimental pK_a values, is not sufficient on its own for verification. Available results used in the past include limited helicity and secondary structural information from CD and NMR spectroscopy,[182,184–186] and occasionally X-ray crystal structures that demonstrate *pH*-dependent differences.[187] Such verification is of critical importance when first introducing a method.

We present here Titr-DMD as an undemanding method for the investigation of *pH* dependent protein behavior. Our method dynamically updates the protonation states of a DMD

simulation[45,139,183] using pK_a predicted for instantaneous structures along its trajectory as probabilities. In the current implementation Titr-DMD uses pK_a values generated through the semi-empirical electrostatics method Propka, but is not restricted to that specific tool. It follows a generally similar approach to the early CpHMD promulgated by Baptista et al.[164] However, DMD's CG square-well potentials and implicit solvation provide rapid conformational sampling at atomic resolution on limited resources, while periodic protonation state reassessment based on Propka confers extensive protonation state sampling. Our program is highly modular for easy modification as better approaches for instantaneous pK_a prediction develop. We benchmark Titr-DMD on both its ability to calculate ensemble pK_a compared to experiment and on its ability to recapitulate the *pH*-dependent conformational change found experimentally in staphylococcal nuclease (SNase), a rare, well-described system.[188,189] Titr-DMD proves to be an effective, affordable method to study *pH*-dependent protein dynamic behavior at atomic scale.

4.2 The Titr-DMD Algorithm

In contrast to traditional molecular dynamics, the interaction potentials of discrete molecular dynamics (DMD) are square-well step functions. Neighboring interactions (such as bonds, bond angles, and dihedrals) are modeled by infinitely high square well potentials. The square-well step functions result in 0 forces between atoms during a simulation. An atom's velocity remains constant until a potential step is encountered, when the velocity changes instantaneously according to the conservation of energy, momentum and angular momentum. Thus, the simulation can be viewed as a series of collision events. At an adequately small step size, the discrete step-function approaches the continuous potential function and DMD simulations become equivalent to traditional molecular dynamics. The sampling efficiency of

DMD compared to tradition MD arises mainly from rapid processing of collision events: only collided atoms are need to update at each collision. DMD uses a united-atom representation to model proteins, in which all heavy atoms and polar hydrogen atoms of each amino acid are included.[45] Covalent bonds and bond angles are modeled by single-well potentials, and dihedral interactions are modeled by multi-step potential functions of pair-wise distance. These parameters are obtained by sampling the corresponding distance distribution in a nonredundant database of high-resolution protein structures. The non- bonded interactions include the van der Waals (VDW), solvation, and hydrogen bond interactions. DMD uses the Lazaridis-Karplus solvation model and uses a standard 12-6 Lennard-Jones potential to model the Van der Waals interactions. The hydrogen bond interaction is modeled by the reaction algorithm.[190]

The Titr-DMD method combines rapid DMD[45] conformational sampling with a custom algorithm to resolve protonation based on Propka3.1[158] pK_a predictions. Simulations are performed iteratively, alternating between a short DMD simulation and a titration (Titr)-feature that discretely assigns protonation states. The algorithm for the Titr-feature itself comprises five steps: (1) titratable residues are identified, (2) contact networks are constructed from the identified residues, (3) the solvent accessibility of each network is determined, (4) the probability of protonation state change is determined for each network or residue, (5) protonation state changes are determined by a Monte Carlo step (Figure 4.1).

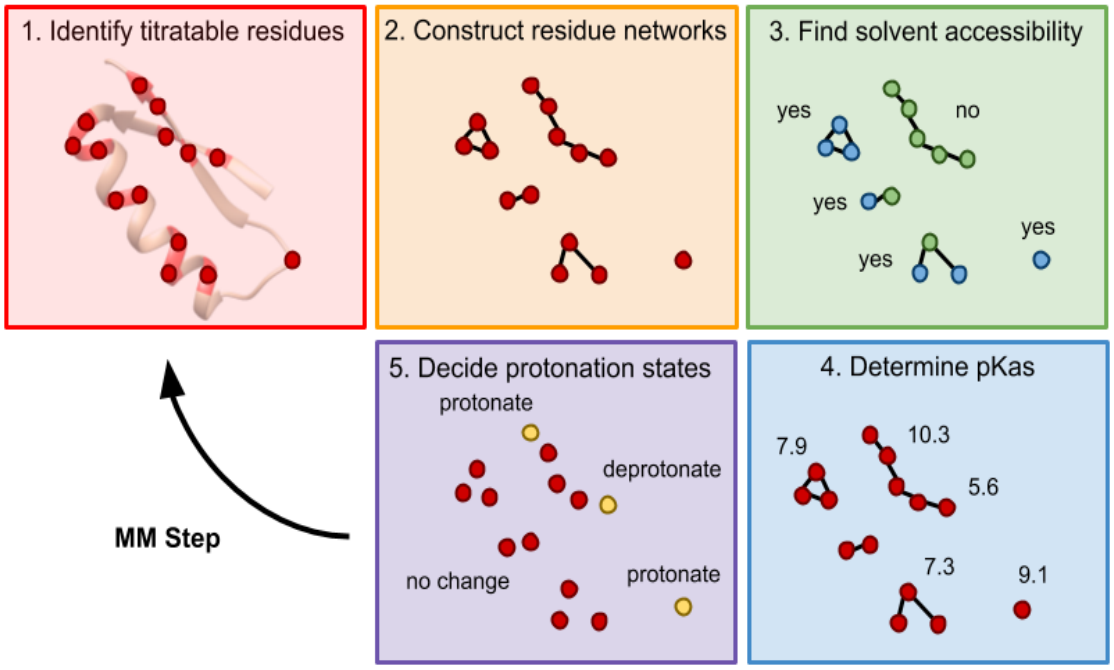


Figure 4.1. Schematic of the Titr-feature algorithm. The algorithm runs between short DMD (or any molecular mechanics) simulations to assign discrete protonation states.

The intervals between protonation state reassessment are run just long enough so that protonation and deprotonation are equilibrated over the DMD simulation timescale. As isolated proton transfer events, including many individual reaction steps in proteins,[191] generally occur on the femtosecond to picosecond timescale,[192] 200 DMD steps (which is ~10 ps) suffices – a comfortable separation of 1-3 orders of magnitude. The size of the timestep allows for both the consistent and meaningful application of theory, discussed more thoroughly throughout the rest of the description of the Titr-DMD method algorithm, and extensive sampling of a system's potential protonation states. A higher reassessment frequency is therefore unnecessary and computationally expensive; while additional time spent on the Titr-feature itself is minimal, a higher frequency requires more, shorter DMD simulations and thus more time overhead during the program initialization.

Selection of titratable residues is based on their solution pK_a values. The amino acids aspartate, glutamate, histidine, cysteine, tyrosine, lysine, and arginine are the only ones to have side-chain solution pK_a values in the physiological range of pH 1-13 and so are the only ones considered. While significant shifts in pK_a often occur when an amino acid is part of a protein, all other residues have side-chain pK_a that fall far enough out of the physiological range to be largely irrelevant in the vast majority of systems. For the same reason, only the first protonation/deprotonation event is considered for the included amino acids; states such as doubly deprotonated lysine or doubly protonated glutamine are inaccessible. The C-terminal carboxylate and N-terminal amine could be titrated as well, but are not currently implemented due to missing DMD potentials for their less preferred states.

Contact networks are constructed on the basis of the proximity of titratable residues. First, interacting pairs of residues are identified based on their (de)protonatable heteroatoms that are within a certain cutoff distance, r_p, of each other. The protonation contact distance r_p is selected as 3.5 Å to be consistent with the DMD definition of a long hydrogen bond. Each thus defined network represents a series of residues close enough that in the timeframe of the DMD phase of the Titr-DMD simulation the proton exchange is equilibrated between them and lies firmly under thermodynamic control.

Solvent accessibility of each residue contact network is determined in a manner consistent with Propka, which defines a specific residue as buried or exposed based on its contact number, $w(N)$. $w(N)$ is determined by the number of heavy atoms, N, within 15 Å of the residue's charge center according to

$$w(N)=\begin{cases} 0 & \text{if } N \le N_{min} \\ \dfrac{N-N_{min}}{N_{max}-N_{min}} & \text{if } N_{min} < N < N_{max} \\ 1 & \text{if } N \ge N_{max} \end{cases} \quad (4.1)$$

The residue is thus 0% buried if $N \le 280$ (N_{min}) and 100% buried if $N \ge 560$ (N_{max}).[158] In the Titr-feature, a network is considered solvent accessible if any residue in it is below a certain cutoff. As proton exchange is equilibrated within a network, so long as one residue is solvent accessible the rest of the network can freely exchange protons with solvent. The best value of the solvent access cutoff is a parameter in the model and is often system dependent. We find that the most appropriate value for the solvent access cutoff could range from 45% to 75% and matters most in systems with important, frequently buried residues. Alternative approaches to the solvent access cutoff are also possible. We discuss this fully in the future development of Titr-DMD section and within our test system simulations.

The probability of a protonation state change is assessed for each titratable residue based on instantaneous pK_a and the residue network information. In the current implementation, Propka3.1 is used for pK_a prediction, based on the latest structure from the preceding DMD trajectory. The protonation state change probability is then assessed for each residue. It is calculated differently depending on whether the residue is in a solvent accessible or inaccessible network. For a solvent accessible network the probability is based on the *pH* of the solution with which the residue can freely exchange protons (solvent is treated implicitly in DMD). This probability is based off the Henderson-Hasselbash equation

$$pH = pK_a + \log\left(\frac{[D]}{[P]}\right) \quad (4.2)$$

where $[D]$ is the concentration of the deprotonated state and $[P]$ is the concentration of the protonated state. Therefore, the probability of adopting the protonated state, P_P, can be defined as

$$P_P = \frac{[P]}{[P]+[D]} = \frac{10^{pK_a - pH}}{1+10^{pK_a - pH}} \quad (4.3)$$

In the solvent inaccessible case, only the titratable protons already present in the network can be exchanged. Buried residues not part of a network are therefore unable to change protonation state, unless, over the course of a simulation, they become solvent accessible or merge with another network. The probabilities of protonation state changes for the residues in a contact network are thus coupled; protonation state changes must be determined for the whole network at once, rather than residue by residue. Solvent inaccessible networks therefore require full enumeration of all proton configurations across the network. The preference of a proton to localize on any individual residue is determined by its pK_a, but with comparison to the competing residues in the network rather than the solution pH. To calculate the probability of a configuration, let R be the set of all residues in a network and n be the number of titratable protons in that network. Let $T|n(R)$ be the set of all possible proton configurations S, Q, ... such that $T|n(R) = \{S \in T(R) : |S| = n\}$. Then for every $S \in T|\mathrm{n}(R)$ the probability of adopting that proton configuration is

$$P_C(R,S) = \frac{\prod_{s \in S} 10^{pK_a(s)}}{\sum_{Q \in T|_n(R)} \prod_{q \in Q} 10^{pK_{a,T}(q)}} \quad (4.4)$$

The weighting term for each proton configuration is the product of 10 raised to the pK_a of each residue that holds a proton in that state ($s \in S$, $q \in Q$, ...). Equation 4.4 is used to calculate the probability of each possible configuration.

Finally, protonation state changes are decided discretely by a single Monte-Carlo step based on the probabilities generated for each network. As with the probabilities, the decision differs slightly between solvent exposed and buried networks. For solvent accessible networks, a

decimal between 0 and 1 is randomly generated for each residue and compared to its decimal probability. If it is above that probability the residue is unprotonated, and if below it is protonated. The solvent accessible approach holds regardless of what the previous protonation state was. For solvent inaccessible networks, the decimal probabilities of all potential protonation configurations are put in a sequential order. A probability range for each configuration, *S*, is then defined as from P_L up to $P_L + P_C(R,S)$, where P_L is the sum of all configuration probabilities already considered and $P_C(R,S)$ is that of the current configuration. A decimal between 0 and 1 is then randomly generated, and the configuration is decided based on which range the random number falls within. Probabilities are generated and protonation states are decided just once during the Titr-feature step before moving on to another DMD simulation step. Any changes from the previous structure are then made, with hydrogen removed when necessary and DMD placing any new hydrogen on the appropriate heteroatoms. The structure is then ready for the next DMD simulation.

A correction is needed to maintain consistency across DMD energies in a Titr-DMD trajectory. As the Titr-feature may add and remove hydrogen by exchange with implicit solvent, the chemical composition of the system can change. As protonation state changes are done through an external program, the energy associated with them are not directly taken into account in the DMD Hamiltonian which only sees the loss and gain of hydrogen. Consequentially, the correction does not affect how protonation state changes are made, but is simply for analysis of the energy trajectory. One approach for an energy correction would be to use a value for the solvation energy of a proton, but that can not be obtained directly from experiment and can only be determined by extrapolation.[56] Values that can be obtained for the solvation energy (-264.3[193] and -265.9 kcal/mol[56,194]) are large compared to the DMD energy changes associated

with structural fluctuations (ca. 100 kcal/mol). Unmodified use of the proton solvation free energy would result in unphysical behavior – Titr-DMD would always deprotonate any residue. Appropriate scaling of the solvation energy is one solution. For the current implementation of Titr-DMD an energy correction for each iteration is obtained instead based on the Propka pK_a of all residues with protonation states that deviate from the original structure. For each protonation state, take the following acid dissociation reaction

$$PRTN \rightleftharpoons PRTN^- + H^+ \quad (4.5)$$

where *PRTN* is the original protein and $PRTN^-$ is the new state. The free energy of reaction 5 can be written as

$$\Delta G_{deprot} = G(PRTN^-) + G(H^+) - G(PRTN) \quad (4.6)$$

Additionally, the K_a of the reaction is defined as

$$K_a = e^{-\Delta G_{deprot}/RT} \quad (4.7)$$

Hence

$$G(PRTN^-) + G(H^+) = G(PRTN) - RTln(10^{-pK_a}) \quad (4.8)$$

where *G*(*PRTN*) is the uncorrected DMD energy and the left-hand side of the reaction is a corrected energy for a comparable system with the same chemical composition. For the protonation reaction, casting *PRTN* as $PRTN^+$ and $PRTN^-$ as *PRTN* in the original reaction gives the equation

$$G(PRTN^+) - G(H^+) = G(PRTN) + RTln(10^{-pK_a}) \quad (4.9)$$

The energy associated with each protonation state change from the PRTN structure can therefore be written as $\pm RTln(10^{-pKa})$, positive for protonation and negative for deprotonation. The corrections are on the scale of 2-20 kcal/mol, consistent with DMD energy fluctuations.

Correction terms are calculated for each iteration and summed with its DMD energy for the corrected energy.

4.3 Current Limitations of Titr-DMD

The scope of Titr-DMD leaves it with a few limitations, which are worth describing here. Its reliance on Propka and DMD implicit solvent means that it does not take interactions with ions in solution into account. Other methods rely on Debye-Huckel theory to do this.[162] As covered in the Results and Discussion section, benchmarking suggests that Propka does not always provide accurate pK_a's for certain, specific residues, namely cysteine and aspartate residues with very acidic pK_a's (around 1.0). However, Propka does quite well with glutamate and asparate residues with pK_as' near to or higher than the solution value (above 4.5). The protonation of cysteine in disulfide bridges – and thus breaking of disulfide bridges – is not allowed in the current implementation. Titr-DMD does not assess protonation state changes to the C-terminal carboxylate or N-terminal amine. Titr-DMD has so far not been used to study catalytic protonation and deprotonation events, so it is unclear how well it can describe highly coupled, hydrogen bonding residues often involved in these processes.[195] However, Titr-DMD may obtain reasonably accurate pK_a's for catalytically coupled residues as both DMD and Propka contain hydrogen bonding terms. Investigations of such behavior are beyond the scope of this initial publication and left for future studies.

4.4 Future Development of Titr-DMD

The modularity of Titr-DMD allows for easy adaptation and refinement. Changes to the method do not require reparameterization of the forcefield. Future developments of Propka or

any other tool to calculate the instantaneous pK_a of a protein conformer can be exchanged to generate the probabilities of protonation state change and improve the quantitative accuracy of the feature. The Titr-feature could even be paired with another molecular mechanics method besides DMD, so long as it is in implicit solvent for consistency with the probabilities of protonation state change. Alternatives to the somewhat system dependent solvent access cutoff are also of interest. One is to use the Propka buried percentage as a scalar probability of solvent accessibility rather than assign a sharp cutoff, while another is based on the solvent-accessible surface (SAS) determined by reduced surface. The SAS method defines the contour of the protein that can be accessed by solvent by rolling a sphere with the van der Waals radius of the solvent (the 'probe') across the protein, avoiding the van der Waals radii of the other atoms.[196] The solvent accessibility of any residue can be determined by measuring the distance of its titratable group to the nearest vertex of the water SAS. If the vertex is within the van der Waals radius of the titratable group, it is solvent accessible. We are currently investigating a SAS approach for future developments.

4.5 Benchmark Systems and Settings

Most of the systems considered for pK_a prediction have been studied extensively both experimentally and with other computational methods (Figure 4.2). Hen egg-white lysozyme (HEWL) was used as it is a prototypical system for CpHMD benchmarking. The input structure of the protein was taken from the Protein Data Bank (PDB ID 1LZN).[197] All solvent molecules were removed for the simulation – water, nitrogen trioxide, and the sodium ion. As HEWL only reports experimental pK_a for GLU, ASP, LYS, TYR, and a single HIS residue, both human thioredoxin (HTRX) and human muscle creatine kinase (HMCK) were simulated as well. HTRX

brought in another HIS residue to the dataset alongside many more GLU and ASP. Two CYS residues in HTRX were not considered, as they are involved in a disulfide bridge. HMCK only added one CYS residue to the dataset, but was included as it is one of the largest proteins with an experimentally identified amino acid pK_a at 381 residues (compared to 105 residues for HTRX and 129 for HEWL). A Staphylococcal nuclease mutant, V66K (SNase V66K), was included as it contains a buried LYS66 residue that is deprotonated at neutral *pH*. The initial structure used for HTRX was PDB ID 1ERT,[198] with all water molecules removed and the rotamers labeled 'A' used when more than one was recorded. As it is unclear whether the 320-331 loop of HMCK is unstructured or an alpha helix, two structures were used. The unstructured case was based on the A chain of PDB ID 1U6R,[199] mutated back to the WT sequence with the substrate ADP, inhibitor (diaminomethyl-methyl-amino)-acetic acid, all water, nitrogen trioxide, and magnesium ions removed. The alpha loop structure was the same except the 320-331 loop was replaced with the 321-332 loop of the A-chain from PDB ID 3B6R.[200] The structure used for SNase V66K was PDB ID 2SNM[201] with thymidine-3',5'-diphosphate, water molecules, and the calcium ion removed. All experimental reference pK_a were drawn from the PKAD database.[202] The pK_a values used ultimately come from Bartik et al.[203] and Webb et al.[204] for HEWL, from Forman-Kay et al.[205] and Qin et al.[206] for HTRX, Wang et al.[207] for HMCK, and Fitch et al. for SNase V66K.[208] The pK_a predictions from our simulations measure error and deviations to the average of these datasets for each residue with more than one reported value. Titr-DMD pK_a predictions were also compared to existing methods. For HTRX, values were obtained from Harris et al.,[172] an explicit solvent replica exchange CpHMD method. Comparisons for HEWL were made based on a truncated set of residues that was also assessed by the explicit solvent Vila-Viçosa et al.[174] and Goh et al.[169] replica exchange CpHMD methods, the implicit solvent implementation of

the Wallace et al.[173] replica exchange CpHMD method, a CG CpHMD method using supra-atomic beads called OPEP6,[182] and the Monte-Carlo method FPTS.[178] The dataset includes mostly asparatate and glutamate residues as well as one histidine residue. Comparisons for the buried LYS66 of SNase V66K were made with the Wallace et al. implicit solvent-based CpHMD[173] and FPTS.[178] There was no other result to compare to for HMCK, but assessment of cysteine residues is unusual. Finally, for all the benchmark systems results were also compared to the NULL model. The NULL model does not involve any simulation but is used to calculate error with solution pK_a values assigned to each amino acid. In our case, we assign amino acids the reference solution pK_a values used by Propka. Beating the NULL model is important for any pK_a prediction tool as failure to do so means that the tool does not even qualitatively capture the pK_a shifting effect of the protein environment on residues.

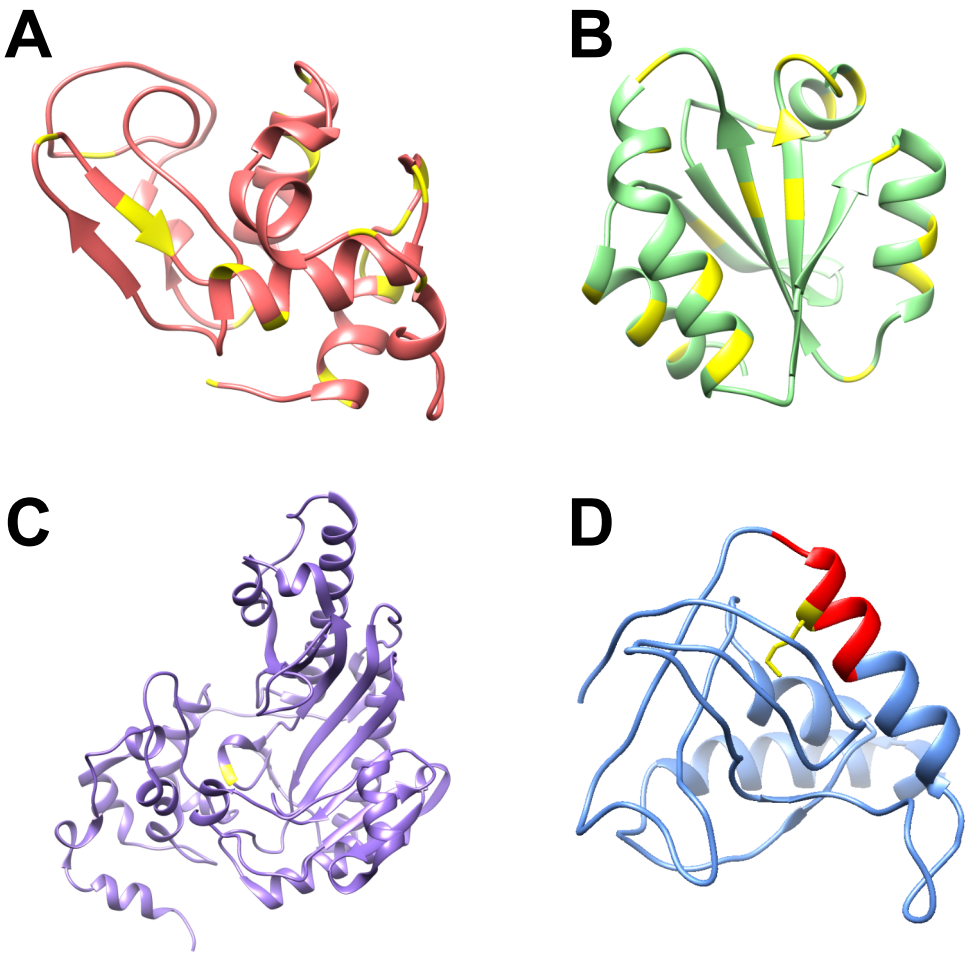


Figure 4.2. Ribbon diagrams of protein test systems for Titr-DMD benchmarking: HEWL (A), HTRX (B), HMCK (C), and SNase mutant V66K (D). The residues whose pK_a's are considered and compared to experiment are highlighted in yellow. In the case of the SNase mutant, this is the buried LYS66 residue. The alpha helical loop to which it belongs and that unravels is shown in red.

The system used to assess *pH*-conformational coupling was SNase V66K, a well characterized system (Figure 4.2D). Experimental information about protein conformational dynamics, including in the context of *pH* change, is difficult to obtain. As discussed in the introduction, the study of SNase mutants is a rare example with available experimental data on

dynamics. A combination of NMR, CD, and titration suggests that the protonation of LYS66 is concurrent with and may be coupled to the unraveling of the first loop of the alpha helix on which it is located.[188,209–211] The V66K mutant was selected as it demonstrates an extreme pK_a shift of 10.5 down to 6.4 – which alongside the conformational coupling is a real challenge for any CpHMD method.

Benchmarking simulations differ slightly between those done to estimate pK_a values and those that assess *pH*-conformational coupling. The *pH*-conformational coupling simulations were longer and hotter to achieve the necessary sampling. DMD simulations without the Titr-feature were also run for the *pH*-conformation coupling system as a control – to make sure conformational changes are *pH* dependent. The pK_a estimating simulations were run for 2,000,000 DMD timesteps (roughly 100 ns, defined empirically) at 50 K (note that temperature in DMD is defined specifically, and does not directly correspond to the physical temperature).[45] A high heat exchange of 10.0 was used for thermal stability because Titr-DMD consists of many short DMD simulations – a more typical, low value has a destabilizing effect. As discussed with the description of the method, a standard protonation contact distance of 3.5 Å was used, as well as the standard protonation state reassessment frequency of 200 steps. The solvent access cutoff was 75%, which is discussed in more detail in Appendix C. The *pH*-conformation coupling simulations were run for a longer 4,000,000 DMD timesteps (roughly 200 ns) with solvent access cutoff values of 65% and 45% ultimately selected and a temperature of 150K for increased mobility. The other settings were the same as for the pK_a prediction. The DMD control simulations without the Titr-feature were performed for the same time and temperature as the *pH*-conformational coupling simulations.

A total of 45 Titr-DMD and 4 DMD simulations were performed for benchmarking. Simulations were done for HEWL at *pH* 3, 5, 7, and 9, for HTRX at *pH* 3, 5, and 7, and for HMCK both with the unstructured and alpha helical 320-331 loop at *pH* 9. The *pH* values were selected to straddle the pK_a of residues with experimentally reported values. Simulations were run for SNase at *pH* 4.6, 5.7, and 7. These values are much below, slightly below, and above the experimental pK_a of the LYS66 residue and its coupled dynamic behavior. All are above the denaturing point of the protein. Three replicates were performed for each system and *pH*. The four DMD simulations were run for SNase to provide a point of comparison. Two were run with LYS66 permanently deprotonated and two with it permanently protonated.

Convergence of the Titr-DMD simulations was attained according to a series of metrics. This is comprised chiefly of the backbone RMSD and the corrected Titr-DMD energy (Figure 4.3). The RMSD was calculated with the initial structure as the reference and with respect to the alpha carbon and amide nitrogen, carbon, and oxygen of each amino acid. All trajectories come to oscillate around fixed values, indicating convergence of the overall protein structures. Convergence for HMCK and SNase V66K, systems with just one amino acid of interest, was tracked by additional metrics covered in Appendix C: the average pK_a and the average protonation state of the titratable residue of interest.

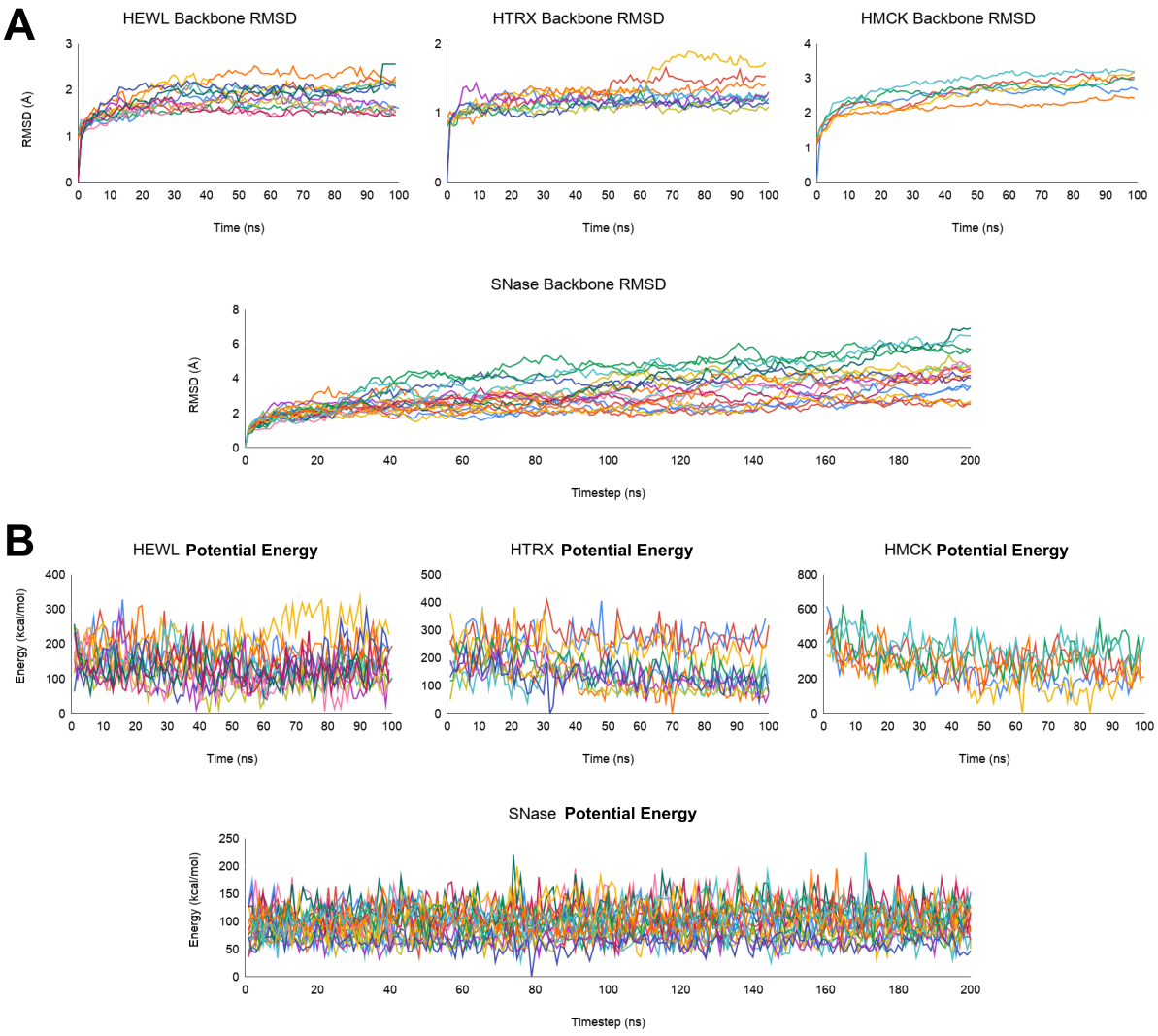
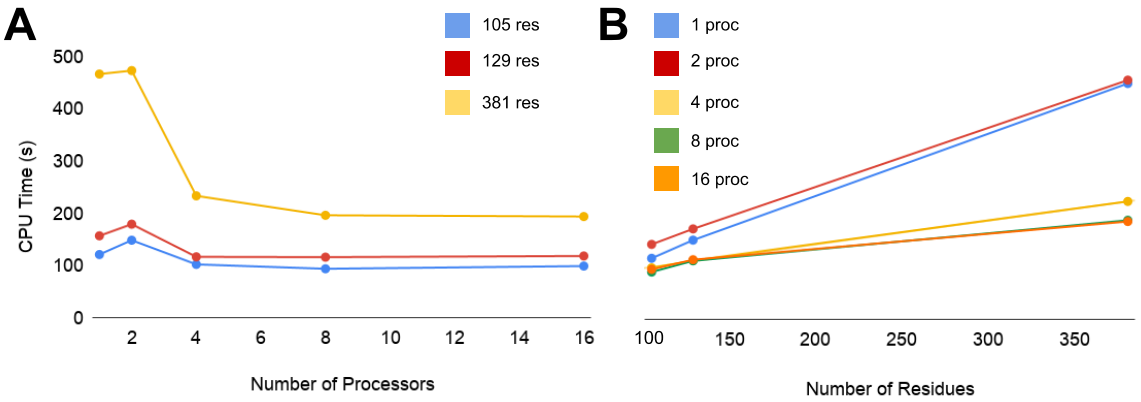


Figure 4.3. Convergence of Titr-DMD simulations tracked by (A) the backbone RMSD and (B) corrected DMD potential energy. Note that by both metrics the results come to oscillate around fixed values by the end of the simulations, indicating convergence for the overall protein structures. The average RMSDs across all trajectories are 1.77 ± 0.29 Å for HEWL, 1.20 ± 0.18 Å for HTRX, 2.51 ± 0.48 Å for HMCK, and 3.67 ± 1.57 Å for SNase. The average energies across all trajectories are 140.18 ± 57.79 kcal/mol for HEWL, 174.93 ± 77.49 kcal/mol for HTRX, 393.52 ± 104.51 kcal/mol for HMCK, and 95.93 ± 29.94 kcal/mol for SNase.

4.6 Benchmark Simulation Results

Titr-DMD offers rapid sampling on limited resources. The combination of DMD and Propka in an implicit solvent makes it a fast and affordable method. We assessed the scaling of

Titr-DMD through 1000 step (5 protonation assessments, about 50 ps) simulations of HEWL, HTRX, and HMCK executed with 1, 2, 4, 8, and 16 processors both with and without the Titr-feature. All simulations were run on the same node sequentially during a single submission to reduce the impact of the variability of other demands on the supercomputing cluster. Simulations were performed on AMD Opteron 2380 (2.5 GHz) cores on Hoffman2 at UCLA IDRE. This process was replicated five times, with the average of these results taken (Figure 4.4). Titr-DMD scales roughly linearly with the number of residues, and scales favorably out to four processors, with additional resources giving diminished returns. The Titr-feature does modestly increase the computational expense of DMD simulations, with the increase in relative runtime over base DMD growing some with the number of processors used. The increase largely derives from the need to initialize many short DMD simulations. However, Titr-DMD still runs quite well on limited resources; the CPU time for the four processor tests scales up to 500-1300 CPU hours (or 3-5.5 CPU hours per residue) to reach a 1 ms simulation.



C

	1 proc	2 proc	4 proc	8 proc	16 proc
105 res	4.50%	25.33%	39.16%	29.17%	40.61%
109 res	2.63%	16.64%	34.92%	33.71%	36.03%
381 res	3.38%	7.23%	18.97%	23.39%	22.13%

Figure 4.4. Computational resource scaling benchmark of Titr-DMD, plotted by (A) the number of processors and (B) the number of residues. Note the linear scaling with number of residues and that good performance is reached with four processors. (C) The percent increase of time for Titr-DMD over unmodified DMD. Note that the increase is relatively small and only becomes significant with many processors as the time DMD takes shortens.

With a couple exceptions, our Titr-DMD method successfully recapitulated the experimental pK_a of the test system residues with reasonable error. We calculated the average RMSE between the predicted and experimental values both by type of amino acid and by protein test system (Tables 4.1 – 4.4). The pK_a can be calculated two ways from Titr-DMD, therefore we calculated two average RMSE for each case. Propka-averaged pK_a is simply the average of the Propka predicted values from each timestep. The DMD-averaged pK_a for a residue is the natural logarithm of the fraction of timesteps in which the residue is protonated. That fraction is analogous to the K_a: the relative concentration of the protonated form of the residue. For solvent exposed residues (those that can freely change protonation state just based on their instantaneous

pK_a) the two pK_a predictions should converge to the same values with appropriate sampling. The results show that indeed the pK_a are in good agreement between the two methods for each system, with the notable exception of SNase V66K LYS66 – a deeply buried residue. The overall maximum absolute error, mean average error, and root mean square error across the full dataset (excluding an outlier HMCK CYS283, discussed later) were 3.25, 0.77, and 1.03 for the DMD-averaged pK_a's and 2.43, 0.81, and 1.05 for the Propka-averaged pK_a's respectively, which is decent agreement with experiment for a CpHMD method.

Table 4.1. Experimental and calculated pK_a values of HTRX. *Proximity of these residues meant that the exact experimental value in the Qin et al. study was unclear. **A series of possible pK_a were found for this residue in the Forman-Kay et al. study, the value of 9.0 was selected due to its consistency with the Qin et al. result. Experimental data from ref 176 (Qin) and ref 175 (Forman-Kay). Other calculated data from [a]ref 141.

	Experiment			Titr-DMD			
Residue	Qin	Forman-Kay	Ave.	Harris[a]	DMD-Ave.	Propka-Ave.	NULL
GLU13	4.4	4.8	4.6	4.4	4.7	4.7	4.5
ASP16	4.0	3.7	3.9	4.0	3.9	3.8	3.8
ASP20	3.8	3.6	3.7	2.9	3.0	2.9	3.8
ASP26**	9.9	9.0	9.5	6.2	7.6	7.6	3.8
HIS43		5.5	5.5	6.1	6.2	6.2	6.5
GLU47	4.1		4.1	4.3	4.5	4.5	4.5
GLU56*	4.1		4.1	4.5	4.7	4.7	4.5
ASP58*	2.8	3.1	3.0	3.8	4.4	4.4	3.8
ASP61*	5.3		5.3	4.6	4.0	4.0	3.8
GLU68	4.9	4.2	4.6	4.3	3.9	3.9	4.5
GLU70	4.6		4.6	5.0	3.8	3.9	4.5
GLU88	3.7	3.9	3.8	3.8	4.0	4.0	4.5
GLU95	4.1		4.1	3.5	4.6	4.6	4.5
GLU98	3.9		3.9	3.9	4.0	4.0	4.5
GLU103	4.4	4.9	4.7	4.7	4.6	4.6	4.5
MAX				3.3	1.9	1.9	5.7
MAE				0.6	0.6	0.6	0.8
RMSE				0.95	0.83	0.82	1.58

Table 4.2. Experimental and calculated pK_a values of HEWL. *Maximum absolute error, mean average error, and root mean square error were also calculated for a truncated set of ASP, GLU, and HIS residues so that Titr-DMD could be compared to referenced methods that only report those. Experimental data from ref 173 (Bartik) and ref 174 (Webb). Other calculated data from [a]ref 142, [b]ref 138, [c]ref 143, [d]ref 151, and [e]ref 147.

	Experiment								Titr-DMD		
Residue	Bartik	Webb	Ave.	Wallace[a]	Goh[b]	Vila-Viçosa[c]	OPEP6[d]	FPTS[e]	DMD-Ave.	Propka-Ave.	NULL
GLU7	2.85	2.60	2.73	2.58	2.70	3.34	3.36	3.30	4.26	4.21	4.5
LYS13		9.88	9.88						10.87	10.93	10.5
HIS15	5.36	5.50	5.43	5.34	6.00	5.43	5.99	5.60	5.74	5.76	6.5
ASP18	2.66	2.80	2.73	2.94	2.10	3.57	3.01	2.80	3.50	3.36	3.8
TYR20		10.30	10.30						9.45	9.55	10
TYR23		9.80	9.80						10.18	10.24	10
LYS33		9.92	9.92						9.91	10.01	10.5
GLU35	6.20	6.10	6.15	4.35	7.00	5.52	3.60	3.50	6.13	6.27	4.5
ASP48	1.60	1.40	1.50	2.84	1.30	1.95	3.31	3.40	3.52	3.61	3.8
ASP52	3.68	3.60	3.64	4.56	4.50	3.85	3.33	3.30	6.00	6.07	3.8
TYR53		12.10	12.10						11.39	11.82	10
ASP66	0.90	1.20	1.05	1.15	1.50	3.15	3.07	3.00	3.51	3.35	3.8
ASP87	2.07	2.20	2.14	2.03	1.30	2.31	3.22	3.20	3.57	3.50	3.8
LYS96		10.20	10.20						10.24	10.23	10.5
LYS97		9.64	9.64						10.88	11.10	10.5
ASP101	4.09	4.50	4.30	3.27	5.10	3.77	3.03	2.90	3.63	3.46	3.8
LYS116		9.76	9.76						10.40	10.38	10.5
ASP119	3.20	3.50	3.35	2.45	1.60	2.80	3.26	3.20	3.68	3.64	3.8
MAX									2.46	2.43	2.75
MAE									0.93	0.92	1.06
RMSE									1.19	1.19	1.31
MAX*				1.80	1.75	2.10	2.55	2.65	2.46	2.43	2.75
MAE*				0.66	0.70	0.61	1.06	1.03	1.19	1.19	1.34
RMSE*				0.89	0.83	0.82	1.32	1.34	1.46	1.45	1.56

Table 4.3. Experimental and calculated pK_a for CYS283 in HMCK. Experimental data from ref 177.

		Titr-DMD		
Residue	Experiment	DMD-Ave.	Propka-Ave.	NULL
CYS283	5.6	-	11.3	9
Abs. Error		-	5.7	3.4

Table 4.4. Experimental and calculate pK_a for LYS66 in SNase V66K. Experimental data from ref 65. Other calculated data from [a]ref 142 and [b]ref 147.

				Titr-DMD		
Residue	Experiment	Wallace[a]	FPTS[b]	DMD-Ave.	Propka-Ave.	NULL
LYS66	6.4	7.5	11	6.4	7.9	10.5
Abs. Error		1.1	4.6	0	1.5	4.1

Titr-DMD predictions of pK_a values are competitive with other CpHMD methods. Across nearly all of the benchmark systems, Titr-DMD outperforms the NULL model. The RMSEs by protein are lower at 0.82-0.83 versus 1.58 for HTRX and 1.19 versus 1.31 for HEWL. The absolute error for SNase V66K LYS66 is 0-1.5 versus 4.1 for the NULL model. HMCK is the one exception, as Propka predicts the pK_a of the single CYS283 residue poorly: 5.7 in absolute error from the experimental value versus 3.4. There was no DMD-averaged value for the residue as the Propka predicted pK_a's were too high and so the residue was rarely deprotonated in our simulations (all conducted at *pH* well below 11). Titr-DMD matches or outperforms more expensive CpHMD methods with HTRX and SNase V66K. For HTRX, Titr-DMD reports a smaller RMSE of 0.82-0.83 versus 0.95 for the Harris method, while the absolute error in the SNase V66K LYS66 pK_a is 0-1.5 compared to the 1.1 of the Wallace method. Titr-DMD performs worse than more expensive CpHMD methods with HEWL, but is comparable to another CG method. Its RMSE for a truncated dataset (comprised mostly of ASP and GLU residues) is 1.45-1.46, above the 0.82-0.89 of the more expensive Wallace, Goh, and Vila-Viçosa methods, but close to the 1.32 of the CG OPEP6.

The performance of Titr-DMD arises from its ability to accurately predict many large pK_a shifts. Our method generally does well with ASP and GLU residues that report pK_a's shifted to

more basic values, but struggles to provide accurate pK_a's for those shifted to very acidic values (around 2.0 or below). Titr-DMD outperforms the more expensive method with HTRX largely through its more accurate prediction of the pK_a of ASP26, shifted according to experiment to the very basic 9.5. Conversely, the poorer performance of Titr-DMD with HEWL is due to the large number of ASP residues shifted to highly acidic values in that system. Titr-DMD struggles to predict the large pK_a shift of the cysteine residue in HMCK, the one case where it fails to beat the NULL model. However, as this was the one CYS residue considered in the test set, Titr-DMD may do better with other examples. Titr-DMD does quite well predicting the shifts of LYS and TYR residues, including the buried and highly shifted LYS66 of the SNase system. The Propka-averaged and DMD-averaged values give qualitative agreement with experiment, but the latter is quantitatively more accurate. Conformational sampling frequents solvent inaccessible states for this residue, with the result of LYS66 spending more time deprotonated than the Propka pK_a would suggest and correcting it toward the experimental value. Furthermore, for LYS66, both our CpHMD method and the Wallace et al. method outperform FPTS, which doesn't perform extensive backbone dynamics and does worse than the NULL model. This demonstrates how important *pH* dependent conformational dynamics are for particular residues and proteins.

Titr-DMD holds promise for the study of the effect of solution *pH* on protein structure. Simulations of SNase V66K are qualitatively consistent with rare, experimentally studied dynamics. With the Titr-feature, we observe partial unraveling of the first turn of the alpha helix on which K66 is localized on (residues 65-69), which is not apparent in DMD without titration (Table 4.5). Unraveling is only observed in 0.002-0.015% of structures in base DMD, while Titr-DMD simulations show it occurs in 3-8% of structures. We define an unraveled state as one where the ALA69-LYS66 and ASN68-MET65 hydrogen bonds are broken or breaking and the

backbone RMSD of the loop is large relative to that of the full protein, indicative of significant, localized structural change (Figure 4.5). The criterion is

$$\frac{S_{RMSD} RMSD_T / RMSD_L}{((R_1 - R_{HB}) + (R_2 - R_{HB}))/S_{HB}} \leq 2 \quad (4.10)$$

where $RMSD_T$ is for the total protein and $RMSD_L$ is for the loop (residues 65-69), R_1 and R_2 are the backbone amide H to carbonyl O distances in Å of ALA69-LYS66 and ASN68-MET65 respectively, S_{RMSD} is 2, R_{HB} is 2.5 Å (for a long hydrogen bond length), and S_{HB} is 2 Å. We only consider structures where the ALA69-LYS66 and ASN68-MET65 backbone hydrogen bond distances are both at least 3 Å. While unraveling according to our criterion occurs in 3-8% of all states at the appropriate *pH*, it is not typically sustained for longer than about 1 ns at any one time. We surmise that our simulations do not have enough sampling to capture sustained loop unraveling, but do show the rare events that could lead to it.

Table 4.5. Frequency of SNase mutant V66K alpha helical loop 65-69 unraveling over the course of Titr-DMD (upper) and DMD (lower) simulations. Note that the frequency is much higher in the Titr-DMD simulations. The highest frequency that occurs a bit below the experimental pK_a of LYS66 (6.4) is when the solvent access cutoff is 45%.

Titr-DMD	*pH 4.6*	*pH 5.7*	*pH 7*
45% cutoff	5.68%	7.06%	4.68%
65% cutoff	3.84%	5.64%	7.93%
	Prot. Start	*Deprot. Start*	
DMD	0.002%	0.015%	

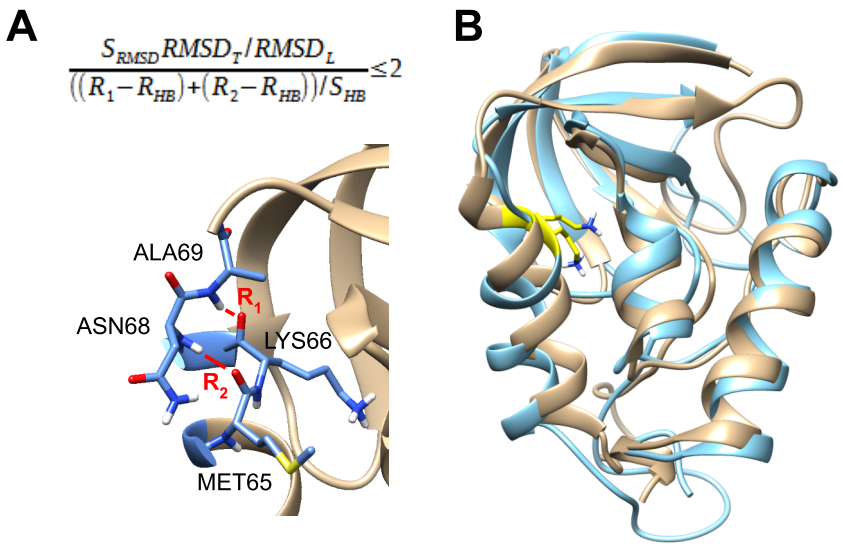


Figure 4.5. (A) Criterion for an unraveled 65-69 loop structure in our SNase simulations. This compares the RMSD of the loop (RMSDL) to the RMSD of the full protein (RMSDT) and compares the distances of important hydrogen bonding contacts (R1, R2) to standard values (RHB) to determine structures where the conformation of the loop varies significantly from the original structure. We give the values of the other variables in the main text. (B) Example of a SNase conformation with an unraveled 65-69 loop by our criterion (light blue) overlaid on a structure where it is not unraveled (tan). LYS66 is colored yellow here.

Protonation and deprotonation of LYS66 is coupled with loop unraveling according to Titr-DMD. At *pH* 5.7, the percentage of unraveled states is significantly higher around LYS66 protonation state changes than the total simulation average (Table 4.6). Moreover, few events at *pH* 5.7 occur without contemporaneous unraveling. The coupling we observe in our simulations is thus consistent with the experimental hypothesis.[188]

Table 4.6. Frequency of unraveling of the SNase mutant 65-69 loop around LYS66 protonation and deprotonation events. 'Near event' refers to the percentage of structures within 25 timesteps (before and after) of an event that are unraveled. This value is roughly on par with the total simulation average except at pH 5.7, particularly during the simulation with a 45% solvent access cutoff. 'By event' refers to the percent of events that have at least one unraveled structure within 25 timesteps. Again, note that the pH 5.7 simulations show high coupling where protonation state changes nearly always occur alongside some contemporaneous unraveling.

Near Event	*pH 4.6*	*pH 5.7*	*pH 7*
45% cutoff	1.11%	30.43%	0.54%
65% cutoff	3.20%	9.45%	10.97%
By Event			
45% cutoff	25%	85%	20%
65% cutoff	20%	80%	29%

Titr-DMD dynamics can predict the *pH* at which loop unraveling occurs. The Propka-averaged pK_a value of LYS66 is uniformly higher than the experimental 6.4, at an average of 7-8, but still shows a qualitatively correct large drop from the solution value of 10.5. However, as discussed before, the DMD-averaged pK_a is generally lower and close to the experimental value, representing the frequent solvent inaccessibility of the residue. The values in Table 4 are averages across the 45% cutoff and 65% cutoff simulations; their individual DMD-averaged pK_a are both close to the experimental result at 5.79 and 7.04 respectively. At these two cutoffs, unraveling is generally most common in the *pH* 5.7 simulations and nearly all protonation state changes occur alongside some unraveling. Titr-DMD can qualitatively model coupling between *pH* and protein structure, and when well calibrated can do so with more quantitative accuracy.

4.7 Conclusions

In this paper we demonstrate Titr-DMD as an effective new method to study *pH*-coupled protein dynamics. The challenges that face any CpHMD method are appropriate conformational and protonation state sampling, accuracy of protonation state changes, and whether the generated conformational ensemble is physically meaningful. Titr-DMD offers great sampling on just a few processors through atomic collision event calculations, implicit solvation, and semi-empirical pK_a prediction with Propka. Our method obtains reasonably accurate pK_a predictions for its computational expense. Titr-DMD was successfully benchmarked on the partial unraveling of SNase mutant V66K: one of the few experimentally studied *pH* coupled conformational changes. Titr-DMD generates a conformational ensemble consistent with experiment, and this ensemble even reflects the experimental *pH* value of the conformational change. Our method is also modular to further improve sampling and accurate assignment of protonation states. Titr-DMD stands as a promising method to address questions of *pH* dynamics in industrial catalysis and medicine.

Chapter 5

Uncovered Dynamic Coupling Resolves the Ambiguous Mechanism of Phenylalanine Hydroxylase Oxygen Binding

5.1 Introduction: The Role of Phenylalanine Hydroxylase (PAH) in Phenylketonuria and Hypothesized Mechanisms

While phenylalanine hydroxylase (PAH) is critically implicated in a range of problems of medical and biological interest, its activity is not fully understood. It is an important metabolic protein responsible for phenylalanine catabolism[212–214] and over 280 different mutants of the protein have been genetically linked to the well-studied condition phenylketonuria (PKU),[215] a candidate for gene therapy.[216] Two mutants of particular interest are R158Q, which is a common mutant as it comprises 40% of PKU haplotype 4 alleles in Europe,[217] and E280K, which is unusual among PKU mutants as it found on multiple haplotypes.[218] The protein is highly similar to both tryptophan hydroxylase and tyrosine hydroxylase, which perform the same function on their respective amino acids.[212] PAH is found as a highly conserved homodimer or homotetramer of 51.7 kDa subunits,[219,220] with an iron center in each monomeric unit that binds and cleaves diatomic oxygen.[221,222] Activated oxygen then participates in one of two competing pathways: the oxidation of phenylalanine to tyrosine, which dominates in wild type PAH, or the formation of hydrogen peroxide, which dominates in the PKU-inducing mutants.[223] Despite all the interest in PAH, two critical features in this mechanism of O_2 binding and activation are poorly understood. First, both PKU-inducing mutations, Arg158Gln and Glu280Lys, occur far from the active site. X-ray crystallography shows that in the wild type these two residues form a salt bridge around 20 Å away from the iron center (Figure 5.1). Why the disruption of this distant interaction affects PAH activity has not been firmly established. The second is the role of a (6*R*)-L-erythro-5,6,7,8-tetrahydrobiopterin (BH_4) cofactor. The binding of BH_4 to the active site is necessary for PAH function, where it presumably assists the iron center in binding and cleaving O_2. Despite extensive study, including on tryptophan hydroxylase and tyrosine hydroxylase which suggest

they operate by the same general pathway,[224–226] no research to date fully resolves either of these specific questions.

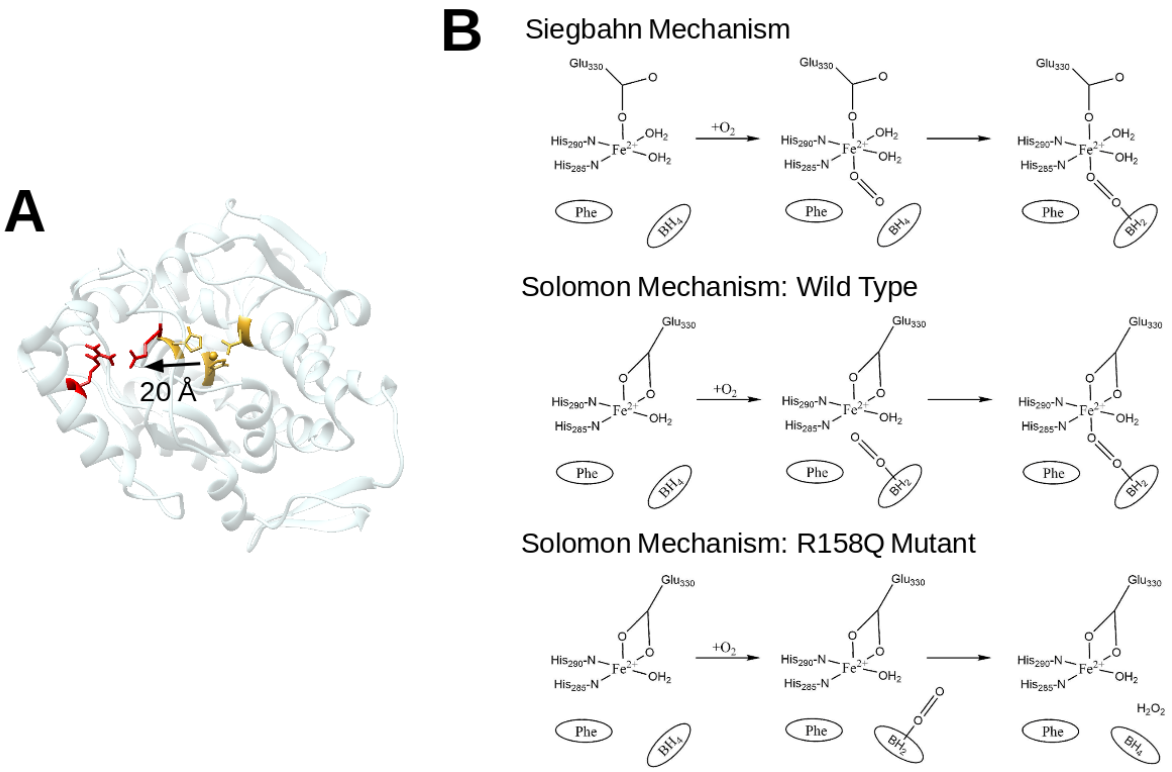


Figure 5.1. The phenylketonuria inducing mutants (red) of phenylalanine hydroxylase (PAH) are surprisingly far from its active site (yellow). (B) The oxygen activating and cleaving activity of PAH is not well understood structurally. In the Siegbahn mechanism, the metal has an additional water ligand and oxygen binds to the metal before the cofactor. According to the Solomon mechanism a change in orientation of the cofactor could explain the preferential H_2O_2 forming activity of the mutant.

Our study starts with structures implicated in a prominent mechanism put forward by the laboratory of Prof. Solomon based on spectroscopic data,[223] which posits solid initial hypotheses on our two mechanistic questions but lacks sufficient atomistic evidence to support them. This mechanism hypothesizes that O_2 binds first to BH_4 and then to the metal in the natural, tyrosine forming pathway. It then conjectures that the cofactor is positioned improperly in the active site

of the mutants, so that O_2 instead binds just to BH_4 and forms H_2O_2 in the pathological pathway. The Solomon study also notes the proximity of Arg158 to two loops around the active site identified in a previous study: one comprising residues 131-155 and another comprising residues 247-251 which forms some hydrogen-bonding contacts with the cofactor.[220,223] However, neither of these contacts fully explain the coupling between the mutant residues and the cofactor for a number of reasons. First, they hinge just on the simple proximity of Arg158 rather than specific interactions. Second, they are based on a single crystal structure of the PAH wild type and contain no information about the mutant state. Finally, as they are based on a single structure they consequentially ignore any dynamic aspects of long-range coupling. These gaps in knowledge necessitate an all-atom investigation of PAH O_2 binding activity.

We also consider structures from a potentially problematic mechanism suggested by the laboratory of Prof. Seigbahn based on DFT studies of small models of the PAH active site.[227] For the O_2 binding step of interest to us the model this mechanism utilized lacks the necessary presence of the phenylalanine substrate and has an inappropriate additional water ligand bound to the iron center rendering it inconsistent with previous kinetic and spectroscopic findings.[223] It thus corresponds more closely to an inactive form of PAH. Our consideration of this mechanism therefore acts as a negative control to see if any dynamic behavior we identify is truly unique to the O_2 binding step or a general property of the protein.

In this study, we performed extensive mixed quantum-classical dynamics simulations on the whole protein and its two pathogenic mutants at full atomistic resolution with QM/DMD, a rapid sampling method for metalloproteins (described fully in the methods section). The active sites are specified by both the Solomon and Seigbahn mechanisms with appropriate modifications. Simulations capture the wild type (WT), R158Q, and E280K variants for the

Seigbahn structures and the WT and R158Q variant for the Solomon structures, with the E280K variant excluded as its different metal coordination precludes direct comparison of energy with the wild type. Our simulations of the Seigbahn construction show it to be highly rigid, in contrast to the flexibility of the Solomon system; the difference demonstrates that the dynamic behaviors of the Solomon system are inherent to the O_2 binding step. These behaviors consist of a significant difference in cofactor orientational preference between the native and mutant forms of the protein and a robust coupling between the cofactor and mutant residues distinct from the static hydrogen-bonding networks suggested by Solomon. The obtained structural data corroborate and expand both Solomon hypotheses with necessary atomistic detail.

5.2 Theoretical Methods

A total of 7 replicates for each of the Siegbahn systems and 10 replicates for each of the Solomon systems were run. Simulations were run for 40 to 80 ns. Details about the preparation of each system and the results of the undiscussed systems can be found in Appendix D.

The established QM/DMD method was used for these simulations.[44] This is a technique for sampling metalloprotein conformations which uses discrete molecular dynamics (DMD)[45] to describe the protein and quantum mechanical (QM) electronic structure calculations necessary to model the metal and its environs. Both methods treat an overlapping region, consisting of species participating in important, non-covalent interactions near the metal, to mitigate discontinuity errors. QM/DMD has a strong record of successfully explaining a variety of metalloenzyme behaviors. A full description of QM/DMD's capabilities can be found in Chapter 1.2.

All QM calculations in this study were performed at the DFT level of theory with Turbomole (version 6.6).[64] The pure meta-GGA TPSS functional[65] with the D3 dispersion

correction[66] was used. The metal was treated with the triple-zeta basis set def2-TZVPP and all other atoms with the double-zeta def2-SVP basis set.[67] While the small basis set may result in some degree of basis set superposition error, the large size of our QM regions have precluded the use of larger basis sets. Furthermore, the level of theory employed without correction has proven effective in our cited past studies including for quantitative free energy comparisons. Finally, the Conductor-like Screen Model (COSMO) with a constant dielectric of 4 was applied to approximate the screening and solvation effects in the relatively buried active sites of the systems.[68] All DMD simulations in this study were performed for 10,000 steps per iteration (0.5 ns).

Convergence of the QM/DMD simulations was achieved according to a series of metrics. These consist of the full protein all-atom RMSD (calculated with respect to all atoms in our models), active site all-atom RMSD (calculated for just the atoms in our QM region) and the DMD energy and QM energy. The RMSD values were calculated with the initial protein equilibrated for one QM/DMD iteration as the reference structure. The full protein all-atom RMSD trajectories of each system are included in this text as an example (Figure 5.2), while plots of the other three standards can be found in Appendix D.

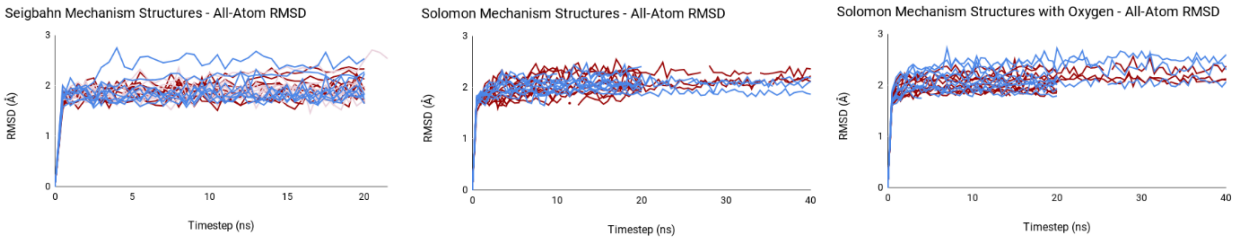


Figure 5.2 Plots of the full protein, all-atom RMSD by timestep for every QM/DMD simulation in this study. The plots include all replicates of the wild type (dark red), R158Q mutant (medium blue), and E280K mutant (light violet, only in Seigbahn simulations). Notice how all replicates oscillate around the value of 2 Å, indicating convergence.

5.3 Results and Discussion

The QM/DMD simulations show the Seigbahn system to be conformationally stiff, with little difference between the wild type and mutant structures. The full protein RMSDs of the PAH variants indicate well converged trajectories: 1.89 ± 0.19 Å for the WT versus 1.93 ± 0.26 Å for R158Q and 1.89 ± 0.19 Å for E280K. There are no significant differences between the structures of the active site, with RMSDs of 0.70 ± 0.23 Å for the WT versus 0.70 ± 0.24 Å for R158Q and 0.69 ± 0.27 Å for E280K (Figure 5.3). The average distances between the Fe center and the central (average) point of BH_4 are all nearly the same: 4.64 ± 0.10 Å for the WT versus 4.72 ± 0.17 Å for R158Q and 4.66 ± 0.09 Å for E280K. Even the orientation of the cofactor relative to the metal does not change with the mutations. The angle between the metal and BH_4 is 48.4 ± 2.1° for the WT versus 46.9 ± 5.5° for R158Q and 47.8 ± 1.6° for E280K. Furthermore, given the low standard deviations of these values, there are no significant populations of other angles that might characterize the position of BH_4 during the competing, H_2O_2 forming pathway. As the Seigbahn structures correspond to an inactive form of PAH, the lack of dynamic behavior in our simulations of them suggests that the behavior found for the Solomon systems is not inherent to PAH overall, but specific to its active form.

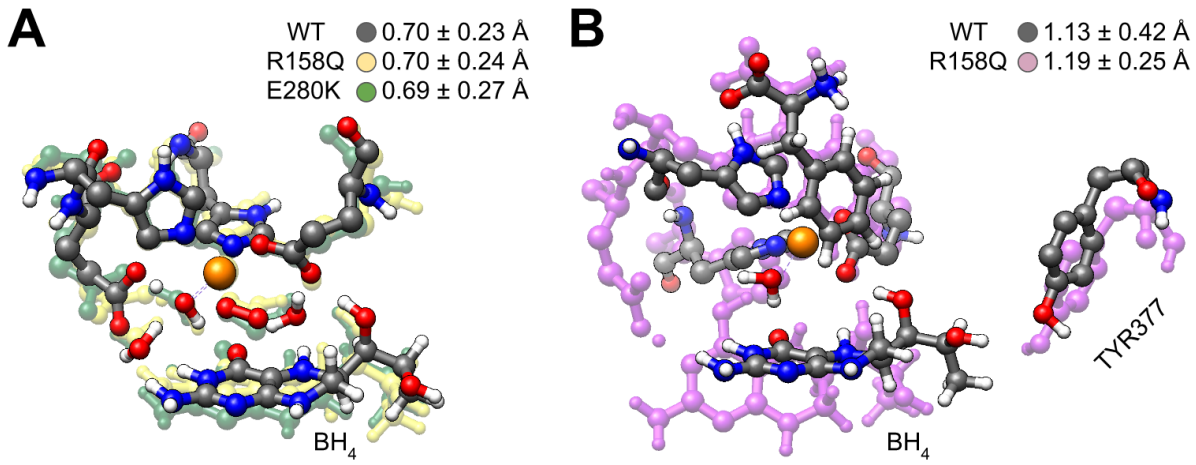


Figure 5.3. Equilibrated active site geometries for different forms of PAH, with reported all-atom RMSDs in the top right. (A) The wild type and mutants from the Siegbahn simulations show no significant differences in their geometries, matching their similar RMSD values. (B) On the other hand, the geometries from the Solomon mechanism without O_2 demonstrate the distinct angular preferences of BH_4. This pterin-like cofactor is angled further away from the metal center in the mutant relative to the wild type.

The results for the Solomon system with O_2 absent are consistent with the Solomon postulation of the role of the BH_4 cofactor. As before, the full RMSD indicates well-converged stable structures at 2.04 ± 0.20 Å for the WT versus 2.02 ± 0.14 Å for R158Q. The active site all-atom RMSD reports some difference between the WT at 1.13 ± 0.42 Å and the R158Q mutant at 1.19 ± 0.25 Å. The most important difference between the wild type and mutant here, clearly visible in Figure 5.3, is the metal to cofactor angles at 55.0 ± 24.3 degrees and 86.8 ± 17.3 degrees respectively. A histogram shows that the wild type has a much larger preference for the lower angle while the mutant greatly prefers the higher angle facing away from the metal (Figure 5.4). Furthermore, both angles are energetically accessible, as the average QM energy across both forms of the protein for the three most populated histogram bins of the lower angle peak is just 1.26 kcal/mol less than those of higher angle peak. Given the difference in activity between the wild type and mutant, these data suggests that the peaks about the lower and higher angles

correspond to the natural and H_2O_2 forming pathways respectively. This supports the Solomon hypothesis that the orientation of the cofactor dictates PAH activity as the active site in the wild type is structurally predisposed to both BH_4 and the metal binding O_2, but the active site in the mutant is not (further supported by simulations with O_2 present in Appendix D). Furthermore, just like both variants can still perform their unpreferred pathway they each report a small fraction of the unpreferred angle.

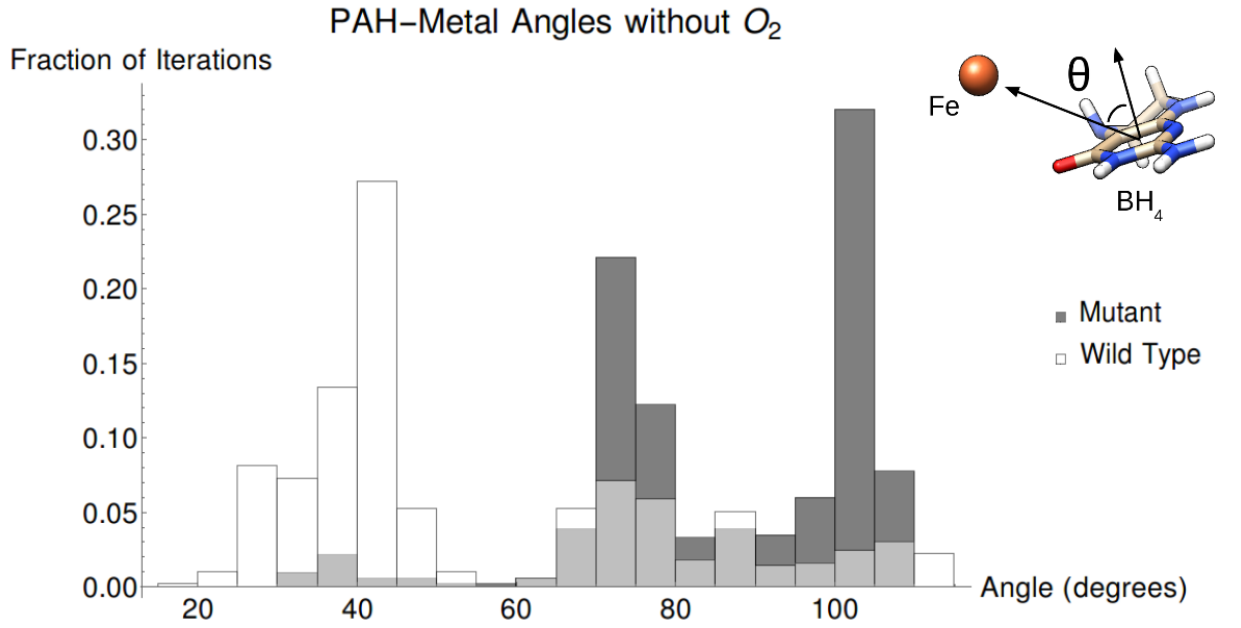


Figure 5.4. Histogram of the metal to cofactor angle in the Solomon mechanism system with O_2 absent for the wild type (white) and the R158Q mutant (dark gray). The height of each bar represents the percentage of iterations across all replicates with an angle in that five degree threshold. The color visible above a bar with the blended medium gray represents the form of the protein in excess for that threshold. Notice that the wild type and mutant overwhelmingly select different angles, consistent with the two pathways reported experimentally.

The QM/DMD simulations of the Solomon structures also explain how the distant mutation affects such a change in active site structure and dynamics. They identify a previously

unknown chain of tightly dynamically coupled residues connecting the mutant residues and the cofactor that controls its orientation. The position of BH_4 in the wild type is partially coupled to the position of the Tyr377 residue which frequently forms a hydrogen bond to it. The distances between the Tyr377 phenolic oxygen or hydrogen and potential hydrogen bonding partners on BH_4 were calculated and the minimum distance selected for each iteration of all replicates for both the wild type and mutant. A histogram of this data shows that in the wild type Tyr377 typically forms a hydrogen bond to the cofactor, with a maximum of population of this length around the characteristic distance of 2.5 Å. In contrast, the mutant generally does not form this H-bond, with a maximum of population ca. 6 Å (Figure 5.5). Furthermore, the position of Tyr377 is tethered to the residue 158 through a series of flexible loops. This suggests that the breaking of the residue 158 to 280 salt bridge by mutation communicates directly with the position of Tyr377 and affects its interaction with BH_4. This is not a large conformational change, and so does not significantly increase the all-atom RMSD. The results indicate that in the wild type the Tyr377 residue holds the cofactor in place to control proper activity, while the mutation breaks this interaction giving BH_4 greater orientational flexibility with respect to the metal center. This then results in the prominence of the H_2O_2 forming pathway.

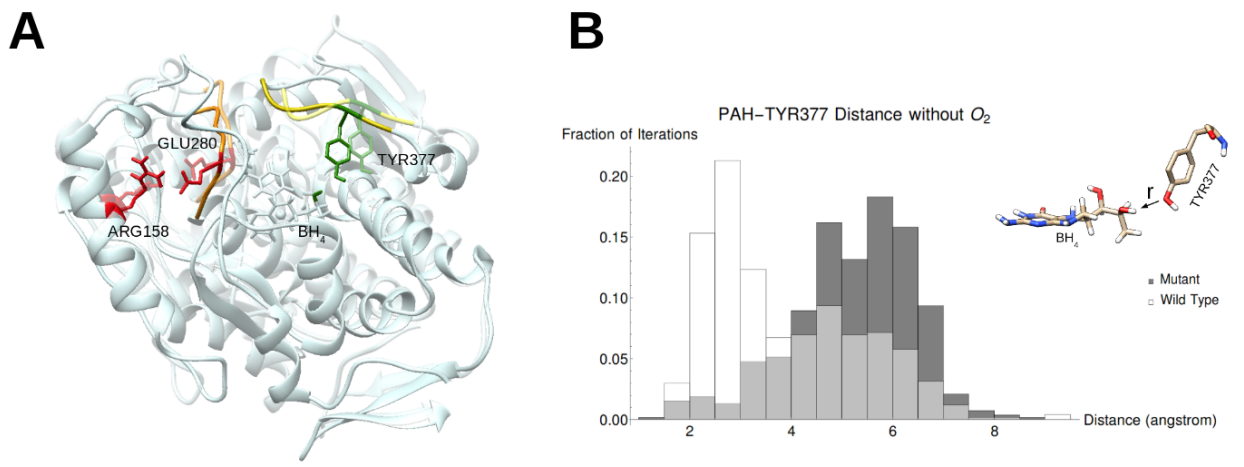


Figure 5.5. (A) Picture which illustrates the structural connection between the site of mutation and the active site in PAH with WT overlay on a faint R158Q structure. The mutation of one of the red residues breaks a salt bridge which in turn makes the orange loop (residues 276-281) more flexible. The orange loop then pushes the yellow loop (residues 376-381) out of its wild type position, affecting the cofactor and active site through Tyr377 (green). (B) Histogram of the *BH₄* to Tyr377 minimum distance for the wild type (white) and the R158Q mutant (dark gray). The height of each bar represents the percentage of iterations across all replicates with an angle in that 0.5 Å threshold. The color visible above a bar with the blended medium gray represents the form of the protein in excess for that threshold. This chart shows that the frequency of a hydrogen-bonding interaction between these two species, which occurs around 2 to 4 angstroms, is much higher in the wild type.

Additional confirmation for the discovered dynamic coupling was found through a dedicated coupled-dynamics method developed by the Dokholyan group. The method calculates correlated motions between residues and constructs pathways with Dijkstra's algorithm. It was used to great effect to study allosteric behavior in pathogenic cystic fibrosis regulator ion channel mutants.[228,229] A simplified protocol was applied to the lowest energy structure of each of the studied systems (results for the Siegbahn structures can be found in Appendix D). In the Solomon wild type, the method finds a strong coupled-dynamics pathway which connects the site of mutation to Tyr377 through the same series of loops identified in QM/DMD simulations

(Figure 5.6). This coupled-dynamics pathway is diverted in the R158Q mutant, consistent with its looser control of the position of Tyr377.

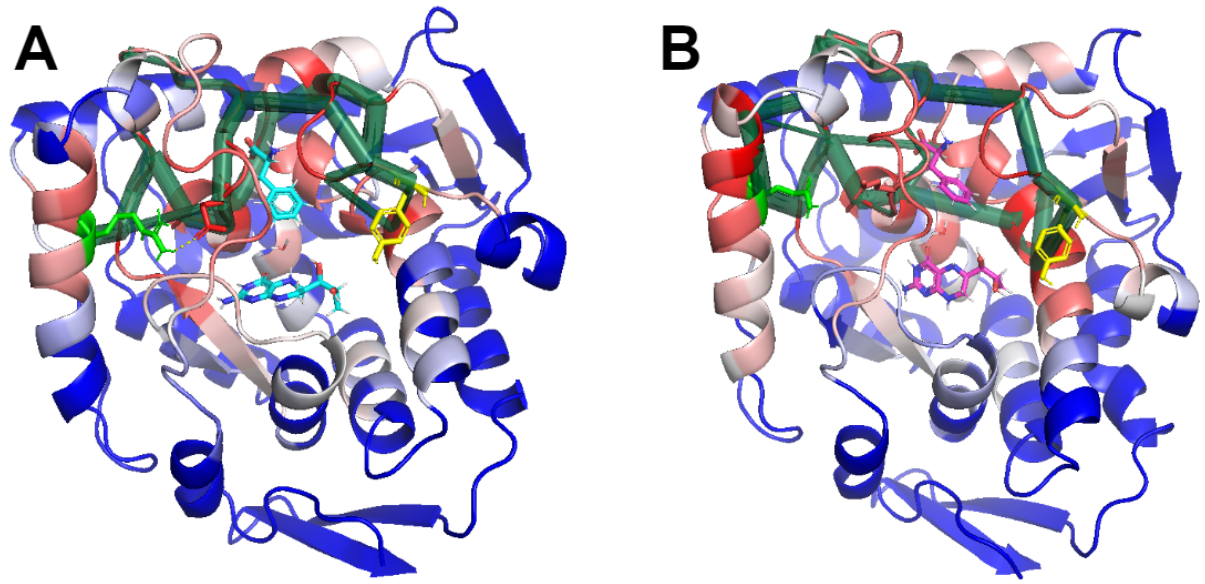


Figure 5.6. Coupled-dynamics pathways (deep green bars) between the site of mutation (bright green and red) and Tyr377 (yellow) identified by coupled-dynamics in the exemplary structures from QM/DMD for the Solomon system WT (A) and R158Q mutant (B). Residues are colored in a heat map going from red to blue indicating motion more to less correlated with the site of mutation. Notice how the pathways in the WT closely match the pathway identified with QM/DMD depicted in Figure 4, while those in the mutant do not.

The identified coupling between the cofactor, Tyr377, is a unique, directing interaction necessary to understand the BH_4 orientational preferences. The cofactor forms other hydrogen-bonding contacts, namely with the loop of residues 147-151. These are the most obvious interactions, as they are present in the original crystal structure (PDB: 1KW0) while the hydroxyl oxygen of the cofactor and Tyr377 are 6.39 Å apart. However, the contacts on the 147-151 loop can't explain the cofactor angle as they change little between the wild type and mutant. On this loop, the Ser151 side chain alcohol and the Gly130 and Leu132 backbone carbonyls all form

hydrogen bonds with the cofactor (Figure 5.7). As with Tyr377, we measured the distance between the atoms in these groups that can hydrogen bond and their nearest partner on the cofactor for all iterations of our simulations. Unlike Tyr377, the average distance of each of these contacts varies little between the wild type and mutant, at just 3.12 ± 1.48 Å (WT) vs 2.79 ± 1.10 Å (R158Q) for Ser151, 3.24 ± 1.09 Å (WT) vs 3.65 ± 1.26 Å (R158Q) for Leu132, and 3.26 ±1.09 Å (WT) vs 2.79 ± 1.10 Å (R158Q) for Gly130, compared to a significant 3.87 ± 1.48 Å (WT) vs 5.23 ± 1.23 Å (R158Q) for Tyr377. These differences suggest that the angular preferences of the cofactor correlate with its interaction with Tyr377, but not with the nearly constant hydrogen-bonds between BH_4 and the 147-151 loop. Ultimately, it is the subtle, shifting interaction that proves most important in understanding PAH activity at an atomic level.

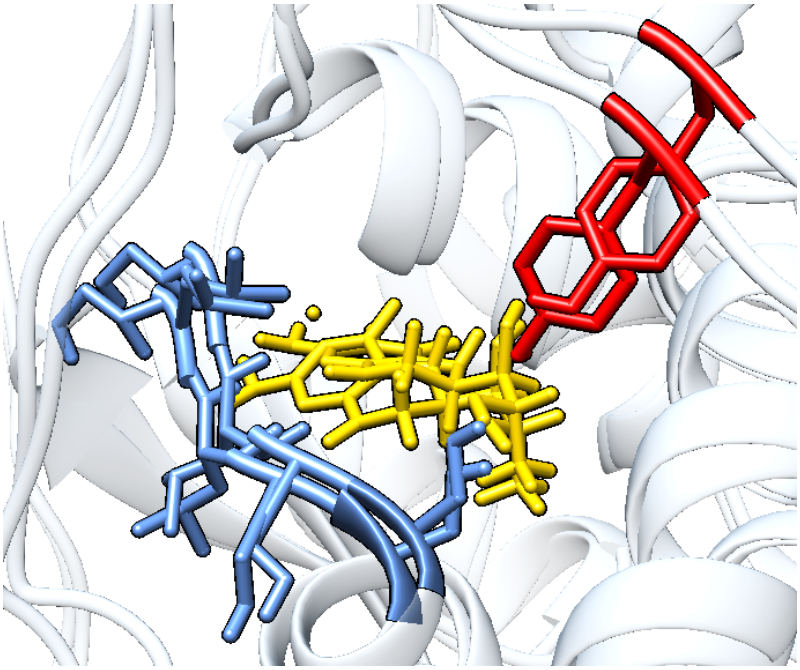


Figure 5.7. Overlay of structures indicative of the range of motion of the cofactor hydrogen bonding partners. Notice how much Tyr377 (in red) swings about the cofactor (in yellow) relative to the residues 147-151 (in blue).

5.4 Conclusion

In summary, QM/DMD simulations assessed whether the Solomon mechanism for the catabolism of phenylalanine by PAH could explain the sharp drop in healthy activity in the case of two disease-inducing mutants. The Solomon structure simulations capture the predicted role of both the cofactor and site of mutation on PAH activity through a dynamic, long-range pathway of communication. The results are reinforced by negative-control simulations of structures implicated in the Seigbahn mechanism and coupled-dynamics analysis. Together, they show that the orientation of BH_4 directs PAH to follow one of two pathways and is modified in the mutants by way of an interaction with Tyr377. Our discovery that a dynamic interaction correlates with PAH activity, especially in contrast to other hydrogen-bonding contacts with the cofactor that do not, demonstrates the importance of considering proteins as flexible objects comprised of an ensemble of important structures. Given the critical role of PAH and related amino acid hydroxylases in bodily amino acid regulation and their implications in disease we hope our findings and perspective will assist future biological and medical research.

Appendix A: Supporting Information for Chapter 2

System Construction

All QM/DMD simulations were based on structures used in Mujika et al.,[110] themselves derived from a crystal structure elucidated by MacGillivray et al. (PDB ID: 1A8E).[100] The QM regions were then selected to include the most important interactions around the metal center. Chemical differences between the four protonation states of hTF required different QM regions, though they were kept as consistent as possible. For the Phys form this consisted of the carbonate ligand and metal, the side chains of Asp63, Glu83, Tyr85, Tyr95, Thr120, Tyr188, Lys206, His249, Lys296, and Asp292 truncated to the functional end and the last side chain carbon bound to it, the side chain of Arg124 truncated to just the guanidino group, the entirety (including the backbone) of Ser125, and just the backbone atoms of Ala64. In this form, Tyr95, Tyr188, and His249 are all deprotonated to coordinate the metal and the terminal amine of Lys206 is deprotonated for the dilysine bridge. Whether Lys206 or Lys296 is initially deprotonated does not matter given the quantum mechanical treatment of the dilysine bridge, which allows for interchange. Indeed, during our simulations the DFT optimizations often remove the additional proton on Lys296 to other residues. The Acid form used the same QM region except for an additional water included near the metal, an additional proton on the carbonate ligand, and Lys206 was fully protonated to break the dilysine bridge. The Double and Prtr form QM regions were both the same as the Phys form, but with one or two additional water ligands included respectively and an additional proton on both the carbonate ligand and Tyr188. Our use of an implicit solvent in the both the QM and DMD regions necessitates the modeling of both Prtr and Double, as the additional water molecule forms interactions that the implicit solvent can not replicate like strong hydrogen bonds and sometimes even coordination to the metal.

Our quantum mechanical treatment of the active site also renders the protonation state of His249 a moot point. Previous studies, including that of Rinaldo and Field,[111] postulated that the His249 ligand to the metal is doubly deprotonated based on geometric comparisons to crystal structures. We do observe some evidence for this in the His249 N to metal distance difference between an applicable crystal structure (PDB ID: 1A8E)[100] and our comparable Fe(III) Phys simulations. This distance could provide information on the protonation state; if His249 is doubly deprotonated this distance should shrink because of the greater negative charge pulling on the positively charged metal. The average distance from our simulations with His249 singly protonated is 2.110 Å versus a smaller 2.036 Å in the crystal structure, which may indeed indicate a doubly deprotonated state for His249. However, we find this outcome unlikely as the exact reason for the smaller crystal structure distance can not be determined (it could even be due to crystallization conditions). Furthermore, His249 often forms a hydrogen bond with Glu81 in our QM/DMD simulations, but despite this chance is never deprotonated.

Simulation Convergence

Besides the protein backbone RMSD (included in the main text), both the DMD energy (Figure A.1) and QM region electronic energy (Figure A.2) were used to monitor convergence. These two energies were calculated for every iteration relative to the lowest energy structure for the given metal and protein form. The plots for both are recorded here and show convergence across all states.

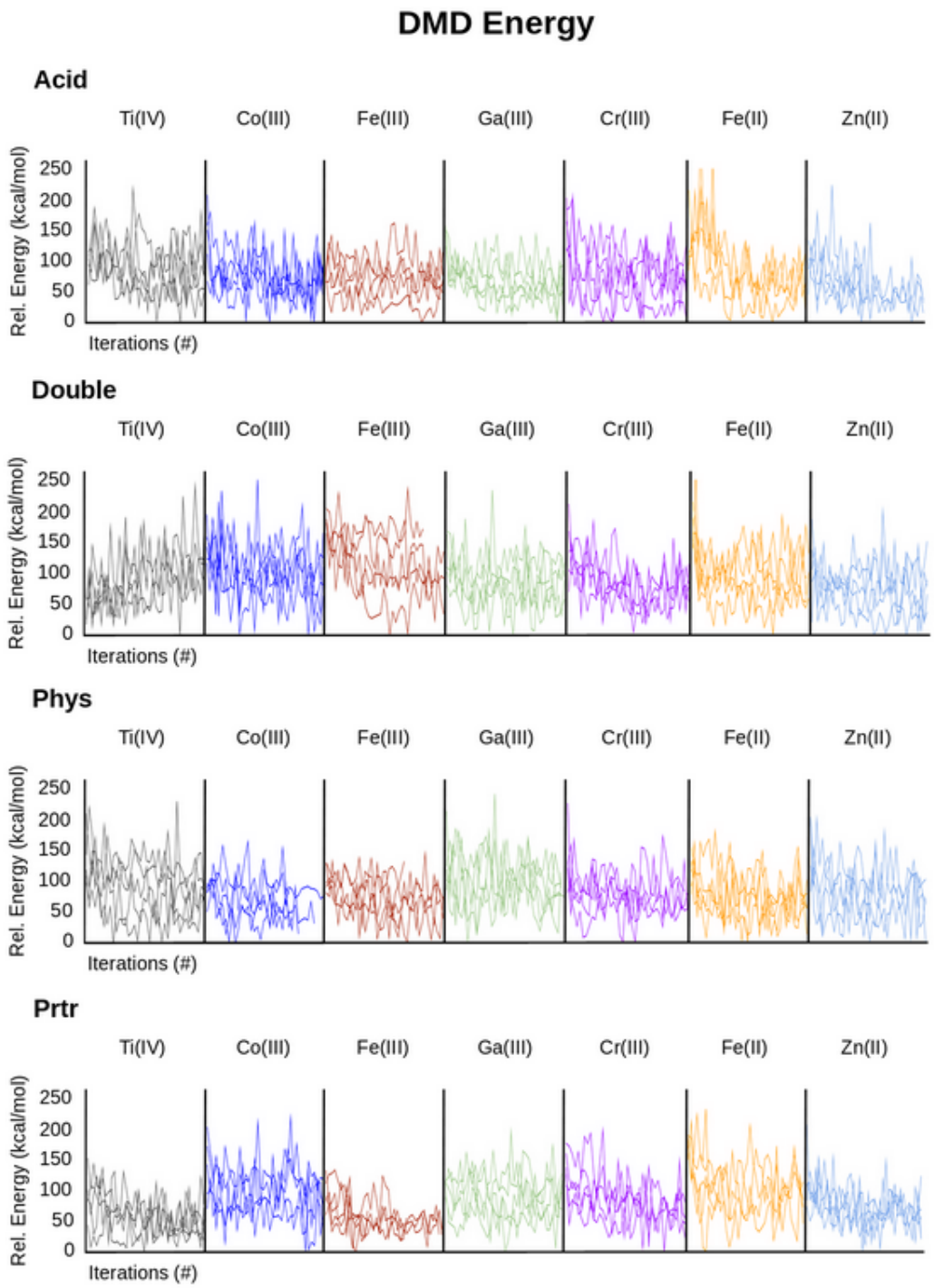


Figure A.1. Plots of the DMD energy by timestep for every QM/DMD simulation in this study. The plots group all replicates by metal for each protein form. Most replicates come to oscillate around values of about 50 kcal/mol, demonstrating convergence.

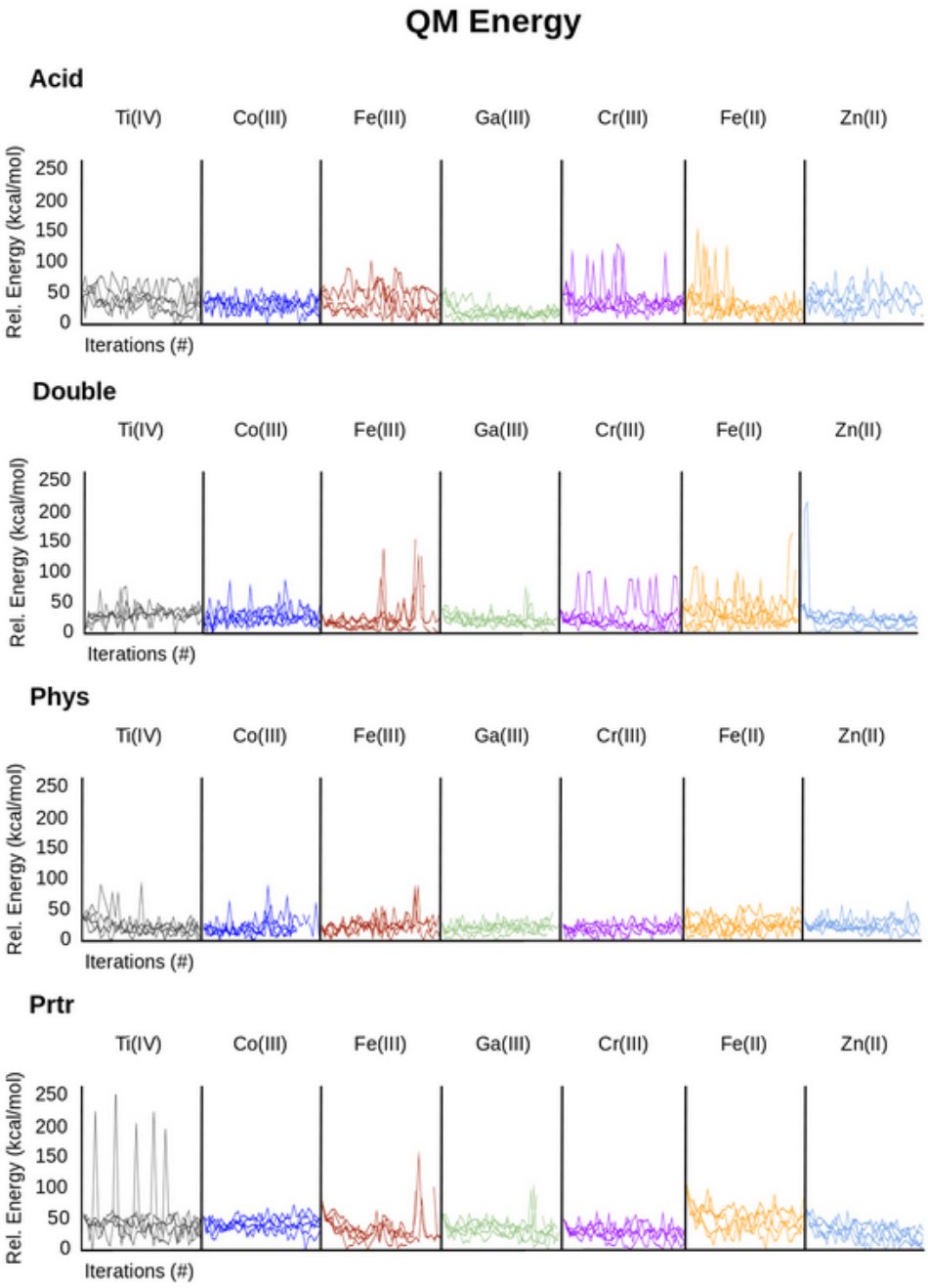


Figure A.2. Plots of the QM region electronic energy by timestep for every QM/DMD simulation in this study. The plots group all replicates by metal for each protein form. Despite occasional spikes, most replicates show convergence as they oscillate about the same 10-25 kcal/mol.

Experimental EDTA-Metal Stability Constants

For the convenience of the reader, the experimental EDTA-metal stability constants used to calculate the binding affinities of this study are recorded here (Table A.1).[81,82] These values are the beta110 constants, corresponding to cation binding to $EDTA^{-4}$.

Table A.1. Experimental EDTA-metal stability constants.

	Ti(IV)	Co(III)	Fe(III)	Ga(III)	Cr(III)	Fe(II)	Zn(II)
K	17.5	40.6	25.1	20.2	23.4	14.3	16.7

Appendix B: Supporting Information for Chapter 3

Simulation System Construction

The input structures for all QM/DMD simulations were based on one crystal structure of GSK-3β (PDB ID: 4NU1). The structure was truncated to residues 35-383 of chain A. The sequence was extended on the C-terminal end by three residues (IQA). These residues and any missing side chains were added using UCSF Chimera.[230] The glycerol, nitrate, and aluminum fluoride molecules were removed from the structure. All water molecules besides the two acting as ligands to the second magnesium ion were removed. For simulations in the ATP bound form of the protein, the ADP molecule was extended to ATP. For simulations in the ADP bound form, the second magnesium ion was removed (the one further from the adenosine group) along with its coordinating water.

The QM region comprised the full ADP/ATP molecule, metal(s) with retained water molecules, the guanidino group of Arg141, and the side chains of Lys85 from the epsilon carbon, Glu97 from the beta carbon, Asp181 from the beta carbon, Lys183 from the epsilon carbon, Asn186 from the beta carbon, Asp200 from the beta carbon, and Ser203 from the beta carbon. For the reaction coordinate scans of the phosphorylation reaction a truncated QM region was used, with only Asp181, Lys183, Asn186, and Asp200 included for amino acid residues. The ATP molecule was chopped to only include the phosphate groups and the portion of the 5-membered ring that links them to the hydroxyl group on that ring. Lys183 was extended to include the full side chain based on the corresponding DMD structures. A serine substrate was added including the side chain alcohol and beta carbon. This was based on manual docking with a short peptide fragment based on tau protein (residues 387-410) with a sequence of

DHGAEIVYK***S***PVV<u>S</u>GDTSPRHLSN.[231–233] The bolded Ser396 is the one at the P+0 site while the underlined Ser400 occupies the P+4 site.

Simulation Convergence

Convergence of QM/DMD was tracked based on metrics of backbone RMSD, DMD energy, and QM region energy (Figures B.1 – B.3). The simulations in this study all converged by about 20 ns, as the three quantities oscillate around fixed values for each trajectory by this point. Note that the definition of DMD and therefore QM/DMD timescale is approximate.

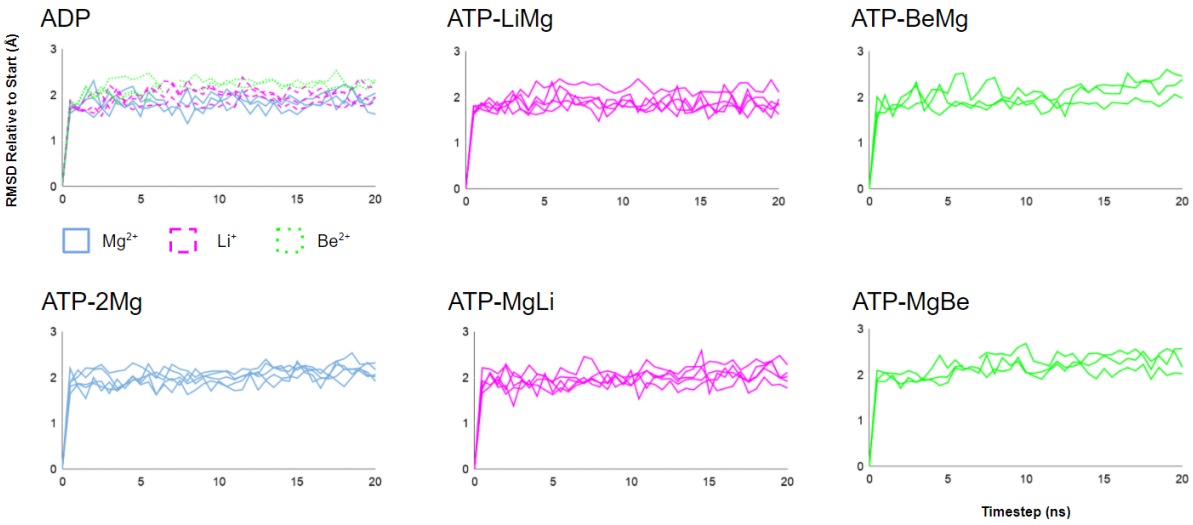


Figure B.1. Convergence plots of backbone RMSD over time for the QM/DMD simulations. This was calculated relative to the first, equilibrated timestep for each trajectory based on the alpha carbon and amide carbon, nitrogen, and oxygen of each amino acid. Note how each simulation oscillates around a fixed value of about 2Å.

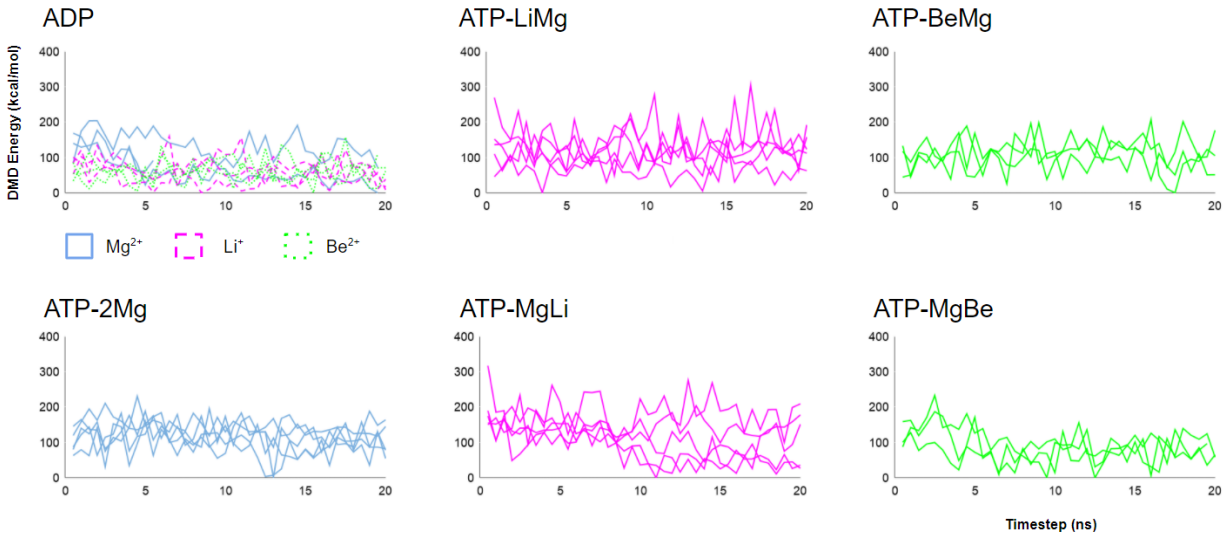


Figure B.2. Convergence plots of DMD energy over time for the QM/DMD simulations. The reported energies are relative to the lowest energy structure for each metal-bound form, which is given a value of 0 kcal/mol in each plot. While DMD trajectories are relatively unstable compared to other forms of molecular dynamics, they show oscillation around fixed values of about 100 kcal/mol.

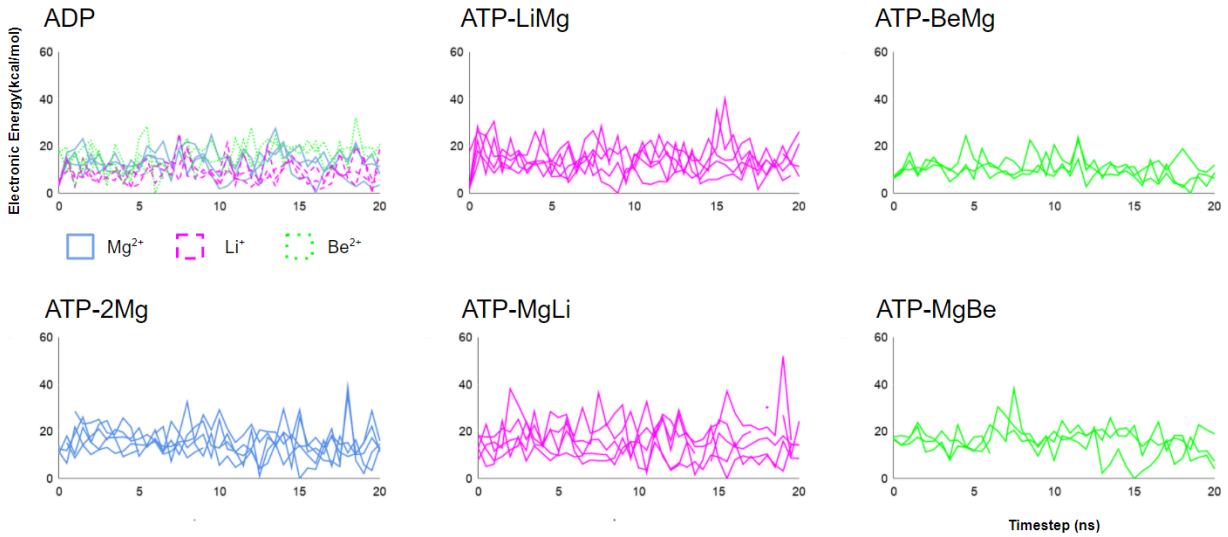


Figure B.3. Convergence plots of QM electronic energy over time for the QM/DMD simulations. The reported energies are relative to the lowest energy structure for each metal-bound form, which is given a value of 0 kcal/mol in each plot. All trajectories show oscillation around fixed values of about 20 kcal/mol.

Appendix C: Supporting Information for Chapter 4

Selection of Solvent Cutoff

The exact value of the solvent access cutoff is system dependent and requires special attention/calibration. A cutoff of 75% was originally selected for all of the simulations conducted in this study. A Propka calculation on the initial structure for each system yielded buried values above ~75% for only the entirely buried residues and those nestled in internal folds, while the external surface residues were all well below this percentage. Over the course of the Titr-DMD simulations, only 2 of the 40 residues with experimentally available pK_a in the benchmark set stayed consistently buried (ASP26 in HTRX and CYS282 in HMCK), so these simulations were ultimately not too sensitive to the choice of solvent buried cutoff. This choice was, however, very critical for simulations of *pH*-conformational coupling, as they involved a study of whether the protonation state of a deeply buried residue (LYS66 in a staphylococcal nuclease mutant V66K) is coupled to the dynamics of the protein. In this case, we ran full simulations at a series of different solvent access cutoff (25%, 35%, 45%, 55%, and 65%) and determined the most physical cutoff based on a comparison to the reduced surface SAS. LYS66 was then defined as solvent accessible if its amine group is within its van der Waals radius of three vertices (one face) of the SES. The comparison to the solvent access cutoff was done for a 20 ns test Titr-DMD simulation of SNase V66K with a solvent access cutoff of 75%. The solvent accessibility was then calculated for each iteration with a SAS generated by Chimera (v 1.13.1).[230] The iterations were binned based on their Propka predicted percent buried, and the percent of states deemed to be solvent accessible was calculated for each bin (Figure C.1). The results show that for SNase V66K there is a threshold around 65-75% buried where LYS66 is generally buried above that point and not buried below that point. Structures do not, however, become entirely solvent

accessible until below 55%. It is important to note that such a low Propka buried % is rare for LYS66 in our short test simulation. We therefore selected for analysis the SNase simulations with solvent access cutoffs of 65% and 45%, one around the inflection point (in Figure C.1) and one where LYS66 is only solvent accessible when it is assured based on our SES test. The results of the other cutoffs are reported in later sections (Figure C.3, Tables C.1 – C.3).

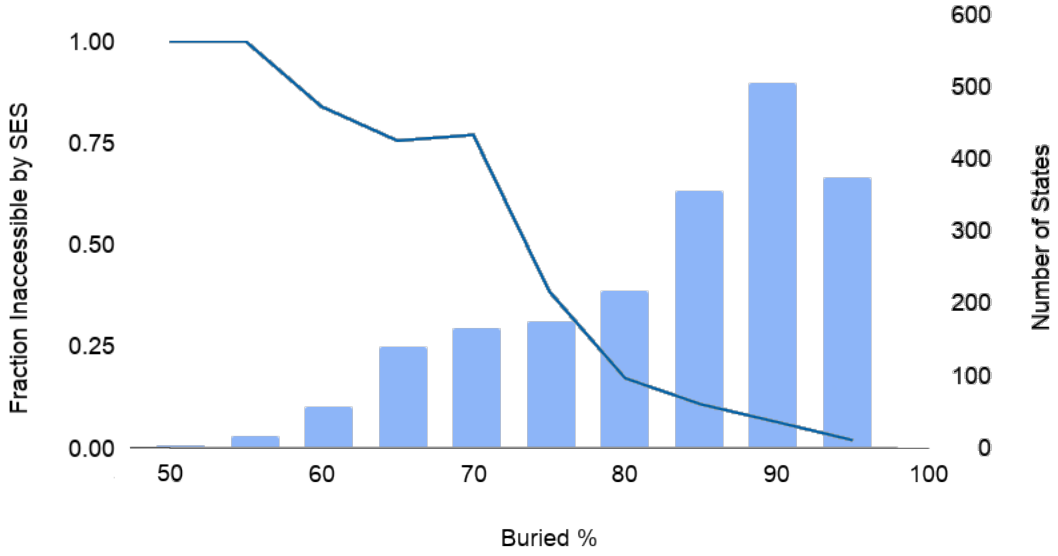


Figure C.1. Comparison of the Propka buried percentage to solvent accessibility according the reduced surface SES. The dark blue line is the fraction of states that are solvent inaccessible according to the SES, where 1 is totally inaccessible and 0 is entirely accessible (left axis). The light blue histogram is the number of states in that bin (right axis). States are binned in units of 5 buried percentage points. Note that most states below 65-70% buried are solvent inaccessible according to the SES and that those below 55-60% buried are entirely inaccessible according to the SES.

Additional Convergence Criterion for HMCK and SNase V66K

Besides backbone RMSD and energy, convergence was also assessed in the HMCK and SNase V66K simulations by the pK_a and frequency of (de)protonation of select residues (Figure C.2). For HMCK, this residue was CYS283. A plot of the average Propka-predicted pK_a over

time for each replicate shows convergence on fixed values. Convergence was not checked based on the frequency of deprotonation of this residue, as the Propka pK_a of this residue ended up consistently too high for deprotonation to occur at the *pH* of the performed simulations. For SNase V66K, the tracked residue was LYS66. As with the RMSD and energy in the main text, these results represent the simulations at cutoffs of 65% and 45%. The plot of the average Propka-predicted pK_a over time for each replicate simulation shows convergence in a tight band of fixed values largely below 8. The plot of the frequency of LYS66 protonation shows distinct behavior between the 65% and 45% cutoff simulations. The upper band of trajectories that converge on full protonation are the 65% cutoff simulations, while the others are the 45% cutoff simulations. As the Propka pK_a is higher than the *pH* of all the simulations, the more solvent accessible 65% cutoff simulations converge on full protonation. However, the 45% cutoff simulations show some amount of protonation and deprotonation as well as divergence not resolved by the end of the simulation in some cases. Ultimately, this suggests that these simulations have captured an initial state of unraveling and the full, sustained unraveling occurs on a (likely significantly) longer timescale. The results can still be used to study the coupling between protonation of LYS66 and structures that lead to unraveling, which is discussed in the Results and Discussion section of the main text.

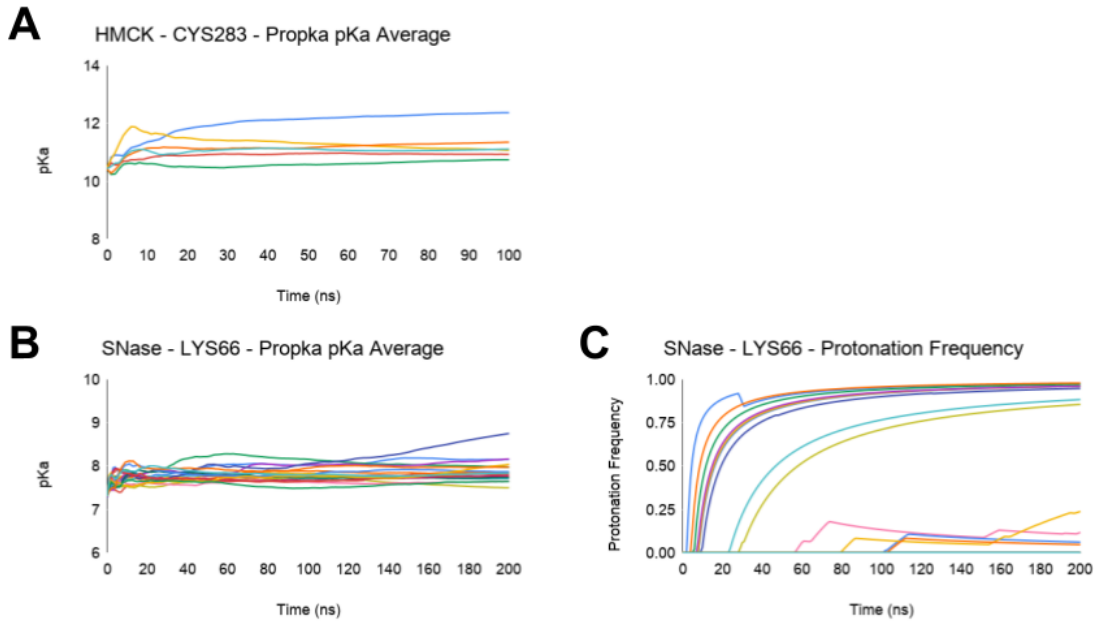


Figure C.2. Average Propka-predicted *pKa* of (A) CYS283 in HMCK and (B) LYS66 in SNase V66K. The value at each timepoint is the average up to that point. The frequency of protonation of LYS66 (C) is also plotted. The protonation state is 0 for deprotonation to 1 for protonation. The protonation frequency at each timestep is the average of the protonation state up to that point. All HMCK simulations are plotted while only the 65% and 45% buried cutoff simulations are plotted for SNase V66K.

Titr-DMD Simulations of SNase V66K at Other Cutoffs

Titr-DMD simulations of the staphylococcal nuclease mutant V66K were also performed with solvent access cutoff values of 55%, 35%, and 25%. All of the other settings for these simulations were the same as for the 45% and 65% cutoff simulations, as reported in the main text. These simulations achieved converged structures just like our others according to the metrics of backbone RMSD and corrected DMD potential energy (Figure C.3A-B). The results of the other convergence criterion, average pK_a and protonation frequency, are likewise similar to those of the 45% and 65% cutoff simulations (Figure C.3C-D).

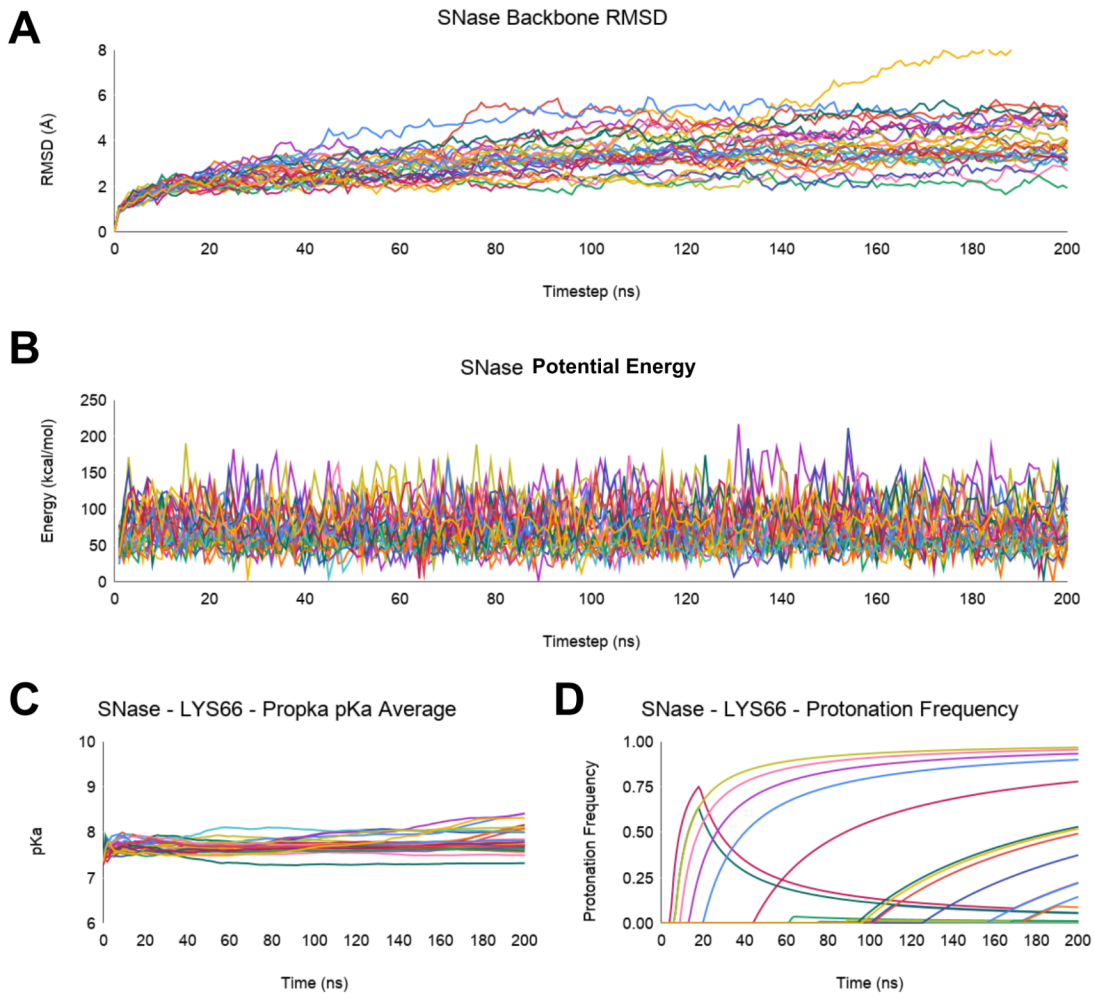


Figure C.3. Convergence of Titr-DMD simulations of SNase V66K with solvent access cutoffs of 25%, 35%, and 55%. Convergence is tracked by (A) the backbone RMSD, (B) the corrected DMD potential energy, (C) the average Propka-predicted *pKa* of LYS66, and (D) the protonation frequency of LYS66. For the protonation frequency, 0 represents deprotonation and 1 represents protonation. In both (C) and (D) the plotted point for each timestep is the average value up to that point. Nearly all simulations come to oscillate around fixed values by ~200 ns for metrics A-C. The unconverged trajectories in metric D are for the 55% solvent cutoff, showing intermediate behavior between the 65% and 45% simulations as the initial protonation takes a while. The average RMSD is 3.54 ± 1.26 Å, while the average energy is 73.05 ± 28.88 kcal/mol across all trajectories.

These simulations reinforce the 45% solvent access cutoff simulation as the most physically meaningful for SNase V66K *pH*-conformational dynamic coupling. As with the 45% and 65% cutoff simulations, the total percentage of iterations reporting an unraveled 65-69 loop across these simulations is much larger than the DMD simulations without titration and does not show any particular *pH* dependent trend (Table C.1). The iterations around protonation state changes show a consistently higher frequency of unraveling at *pH* 5.7 with most events at *pH* 5.7 coupled to some contemporaneous unraveling, consistent with the 45% and 65% cutoff simulations (Table C.2). The only exception are the 25% cutoff simulations, which report 0% unraveling by event at *pH* 5.7 as few protonation and deprotonation events occur at all. The calculated pK_a for LYS66 in these simulations are likewise consistent with the trend of the 45% and 65% cutoff simulations (Table C.3). The Propka pK_a are all off by the same amount, while the DMD pK_a drop down to near the experimental value of 6.4. Note that for the 55% cutoff simulation the DMD pK_a is intermediate between those of the 65% and 45% simulations, which then steadies out in the 35% and 25% cutoff simulations. This suggests that a solvent access cutoff of 45% is the first to capture the experimental pK_a of LYS66, making it the most physically meaningful solvent access cutoff for the SNase *pH*-dependent dynamic behavior.

Table C.1. Frequency of SNase V66K alpha helical loop 65-69 unraveling over the course of Titr-DMD simulations with solvent access cutoffs of 25%, 35%, and 55%. These all report a much higher percentage of unraveled states than the 0.002% - 0.015% for DMD simulations without titration (found in Table 3 in the main text).

Simulation	pH 4.6	pH 5.7	pH 7
25% cutoff	5.99%	4.15%	5.19%
35% cutoff	6.78%	6.53%	8.58%
55% cutoff	7.88%	6.43%	7.01%

Table C.2. Frequency of unraveling of the SNase V66K 65-69 loop around LYS66 protonation and deprotonation events for simulations with solvent access cutoffs of 25%, 35%, and 55%. As in Table 4 in the main text, 'Near event' refers to the percentage of iterations within 25 timesteps (before and after) of an event that are unraveled. This is only consistently elevated above the total simulation average in the *pH* 5.7 simulation. 'By event' refers to the percent of (de)protonation events that have at least one unraveled structure within 25 timesteps. This is also only consistently high for the *pH* 5.7 simulations. The 25% cutoff simulation is the one outlier to this trend, reporting 0% for both forms of coupling, but that is likely due to the infrequency of (de)protonation events in those simulations.

Near Event	pH 4.6	pH 5.7	pH 7
25% cutoff	10.98%	0.00%	7.25%
35% cutoff	5.34%	15.49%	6.13%
55% cutoff	31.33%	10.67%	24.09%
By Event			
25% cutoff	67%	0%	56%
35% cutoff	71%	93%	63%
55% cutoff	67%	67%	100%

Table C.3. Titr-DMD predicted pK_a for LYS66 in the SNase mutant for simulations with 25%, 35%, and 55% solvent access cutoffs. Note that the Propka pK_a is consistently high across the cutoffs, while the DMD pK_a drops down to a value close to the experimental pK_a of 6.4 between the cutoffs of 55% and 35%.

Simulation	DMD	Propka
25% cutoff	5.78	7.78
35% cutoff	5.80	7.86
55% cutoff	6.32	7.89

Appendix D: Supporting Information for Chapter 5

Intervening Research

Little research into the mechanism of PAH oxygen binding has been published over the intervening years between the introduction of the Solomon and Seigbahn mechanisms and this paper. To our knowledge at the point of this publication all but 2 of the 112 articles that cite the Seigbahn or Solomon mechanisms are reviews, concern study into different attributes of PAH, or are mechanistic studies of other biological systems. The two exceptions only cite Seigbahn and their findings do not explain the malfunction of PAH. One is another DFT analysis which considers the potential of side-on O_2 binding to the metal, but does not include any water ligands and is thus suspect given EXAFS data.[234] The other is a QM/MM study of PAH, which while thorough only considers the reaction of the activated iron(IV)-oxo complex with phenylalanine rather than the steps involving the binding of BH_4 and activation of O_2.[235] The best mechanism to explain the malfunction of PAH is therefore unclear and demanded this study.

System Construction

The initial structures of the enzymes were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB code 1J8U[219] for the Seigbahn mechanism and 1KW0[220] for the Solomon mechanism). The 1J8U form of PAH contains BH_4 as well as water molecules around the active site, including three bound to the metal. To prepare our active site for the Seigbahn mechanism all water was removed except for two bound to the metal and one which forms a critical hydrogen bond with the cofactor and added a diatomic oxygen molecule between the metal and the cofactor. The 1KW0 form of PAH contains BH_4 and thienylalanine, an analog of the phenyalanine substrate, as well as water molecules around the

active site. Two systems were prepared based on this PDB: one with diatomic oxygen between the metal and cofactor and one without. In both of these systems the thienylalanine was replaced with the zwitterion of phenylalanine and all of the water was removed except for the one bound to the metal. In all simulations iron was modeled with an oxidation state of II which corresponds to the active site before O_2 cleavage.[236] These active sites were constructed to approximate the rate determining step as described by each mechanism without having to expend the extensive resources to determine transition state structures. In principle, modeling with bound transition states is possible, but it would require expensive, continuous reevaluation of the Hessian to ensure the system remains on a saddle point.

Solomon Structures with Oxygen Results

The results of the Solomon system with O_2 present may suggest that the activity of PAH is dictated before O_2 binds. As with all of our simulations, the full RMSD indicate converged trajectories: 2.06 ± 0.17 Å for the wild type versus 2.14 ± 0.24 Å for R158Q. Unlike the simulations with O_2 absent, the difference in the average cofactor to metal angle is somewhat small at 44.8 ± 35.4 degrees for the wild type and 55.9 ± 34.8 degrees for R158Q. However, note the massive standard deviations. A histogram of the angle at the end of each iteration of all replicates is bimodal for both the wild type and mutant (Figure D.1). Consistent with the results without O_2, the mutant, in which the H_2O_2 forming pathway dominates, populates the higher angle peak slightly more. Again, the difference in the average QM energy between the two angles across both forms of the protein is small enough to suggest that both are energetically accessible, with the three most populous histogram bins of the higher angle peak 4.62 kcal/mol less than those of the lower angle peak. However, the protein in this state does not explain the

experimentally observed large difference in activity due to the nearly equal population of these geometries. Instead, in the framework of the Solomon mechanism this data suggests that when O_2 binds it selects one of the two competing reactions of PAH. Each of the simulations that ran with O_2 bound did not deviate significantly from the cofactor angle it selected after equilibration, indicative of a high barrier between the two geometries at this point (Figure D.2). Once a reaction is selected by O_2 binding, the protein is effectively incapable of deviating to the other reaction. Furthermore, in the Solomon mechanism the significant populations of each geometry for both forms of the protein seem consistent with experimental results which show that while each form of the protein prefers a different reaction, both forms are capable of both reactions. Note that this hypothesis is not conclusive, as the metal is often too close to the cofactor at the low angle for it to easily incorporate O_2 into its coordination sphere as would be necessary for proper O_2 cleavage (Figure D.3). This strange behavior likely derives from the fixed oxidation state of the metal in our simulations; to represent the active site with O_2 strongly bound to iron there may be a change to its oxidation state as discussed in the Seigbahn mechanism. Ultimately, a full mechanistic analysis of the Solomon system would be necessary to assess this hypothesis.

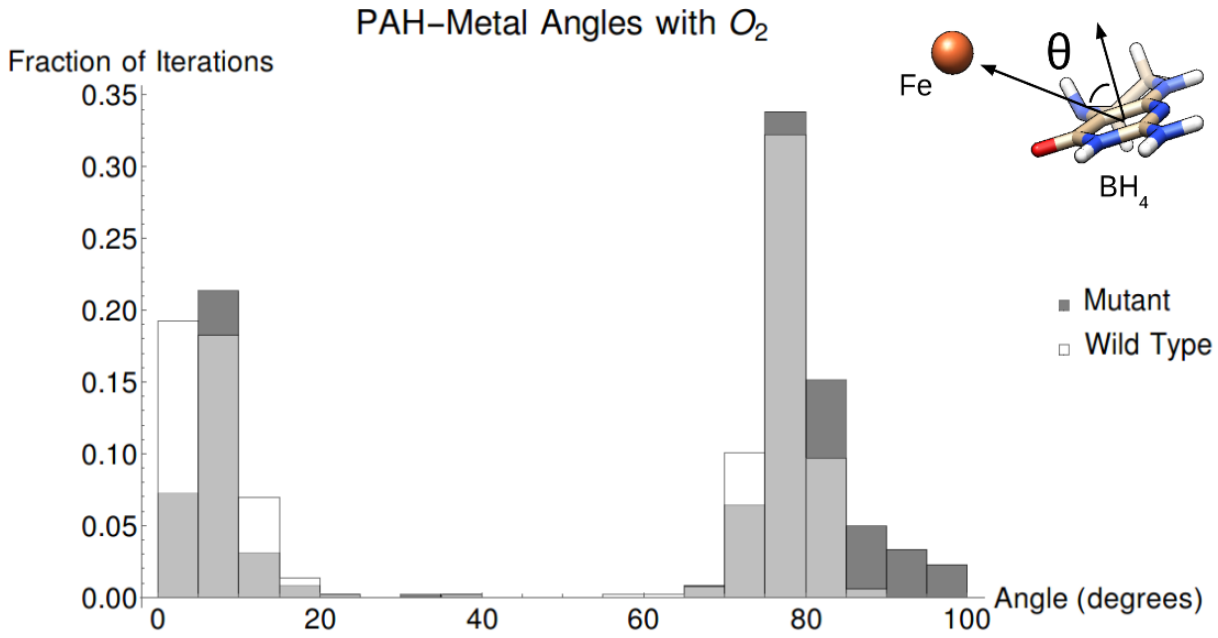


Figure D.1. Histogram of the metal to cofactor angle in the Solomon mechanism system with O_2 present for the wild type (red) and the R158Q mutant (blue). The height of each bar represents the number of iterations across all replicates with an angle in that five degree threshold. The color visible above a bar with blended colors represents the form of the protein in excess for that threshold.

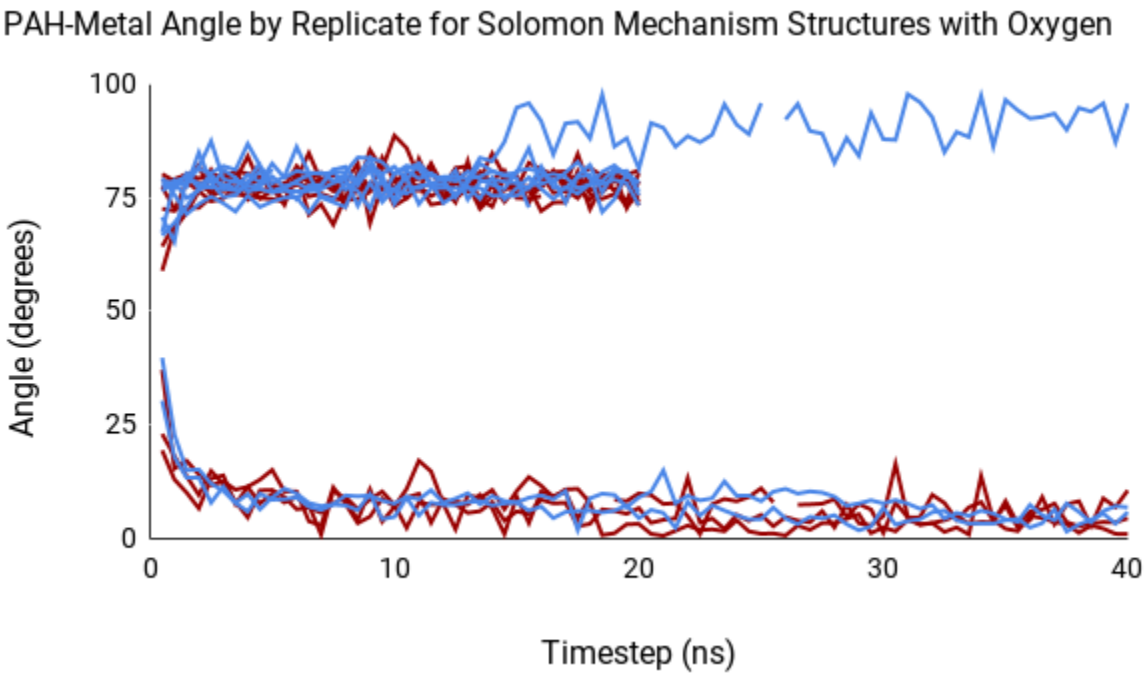


Figure D.2. Plot of wild type (dark red) and mutant (medium blue) cofactor to metal center angles by replicate of the Solomon mechanism system with oxygen present. Notice how no replicates cross over from one geometry to the other after equilibration. This suggests that the binding of O_2 decides the reaction PAH will perform, with the observed differences in activity dictated by the angular preferences of the protein before O_2 binds.

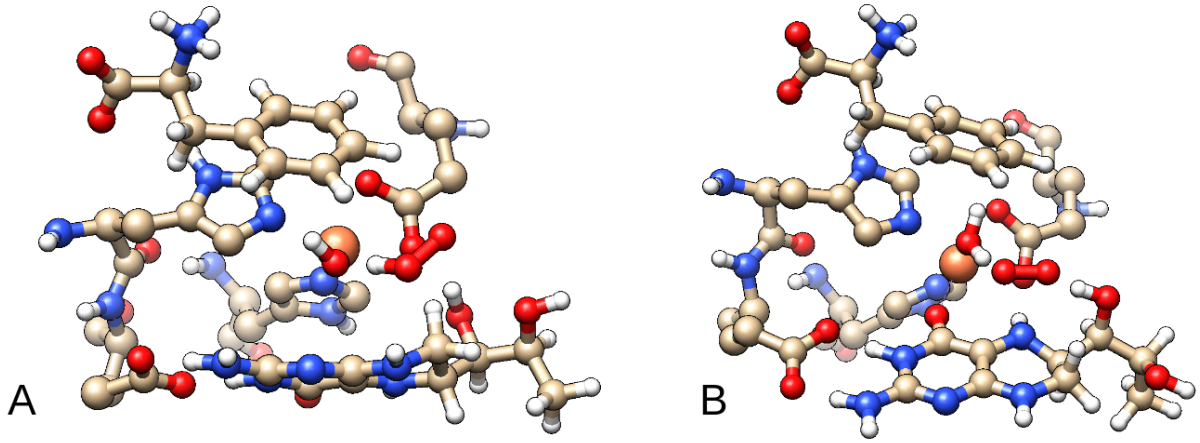


Figure D.3. Exemplary QM regions of the Solomon mechanism system with O_2 with metal to cofactor angles of 10 (A) and 80 (B) degrees. These geometries are the same for both the mutant and wild type. Note that the existence of two geometries may explain the two pathways observed experimentally, but that the distance between BH_4 and the metal in the 10 degree geometry may not be consistent with this.

Coupled Dynamics Analysis of Siegbahn Structures

Coupled dynamics simulations with the Dokholyan method of the Siegbahn structures show long, diffuse coupled-dynamics pathways between the site of mutation and Tyr377 (Figure D.4). None of these pass through the series of loops identified in the Solomon structures. These results further discredit the capability of the Siegbahn mechanism to explain the effect of the distant mutations and their role in manipulating PAH activity.

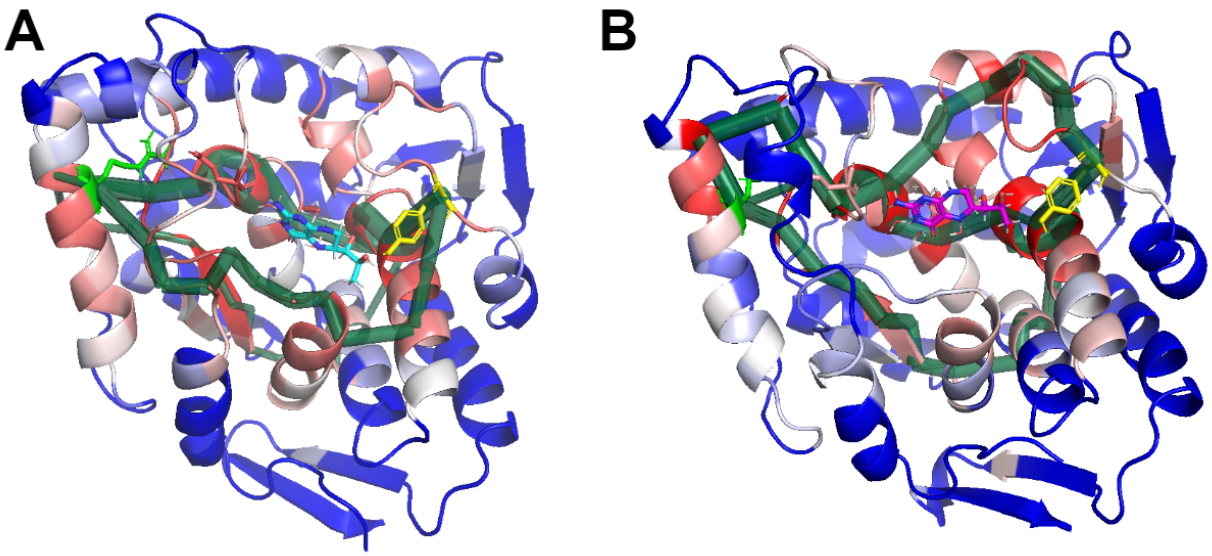


Figure D.4. Coupled-dynamics pathways (deep green bars) between the site of mutation (in bright green and red) and Tyr377 (in yellow) identified by coupled-dynamics in the exemplary structures from QM/DMD for the Siegbahn system WT (A) and R158Q mutant (B). Residues are colored in a heat map going from red to blue indicating motion more to less correlated with the site of mutation. Notice how the pathways go deep into the protein and avoid the surface loops identified in the Solomon structures.

QM/DMD Simulation Convergence

All conducted forms of analysis indicate that the QM/DMD simulations performed for this study are converged. The full protein all-atom RMSD is found in the methods section of the main text. Included here are plots for each system of the active site all-atom RMSD (Figure D.5) and trajectories of the DMD energy (Figure D.6) and QM energy (Figure D.7).

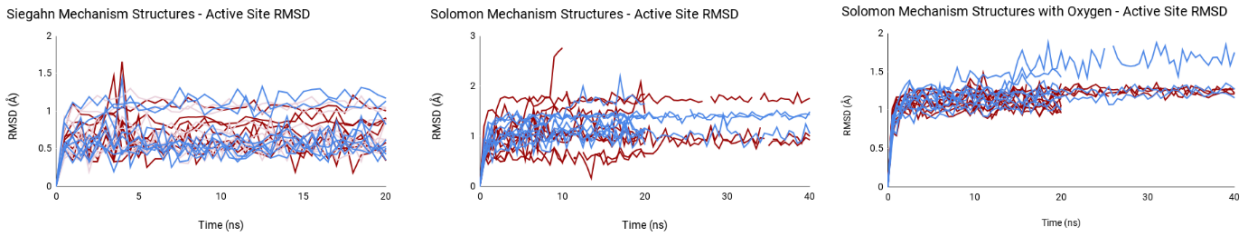


Figure D.5. Plots of the active site, all-atom RMSD by timestep for every QM/DMD simulation in this study. The plots include all replicates of the wild type (dark red), R158Q mutant (medium blue), and E280K mutant (light violet, only in Seigbahn simulations). Notice how all replicates oscillate around the value of 1 Å, indicating convergence.

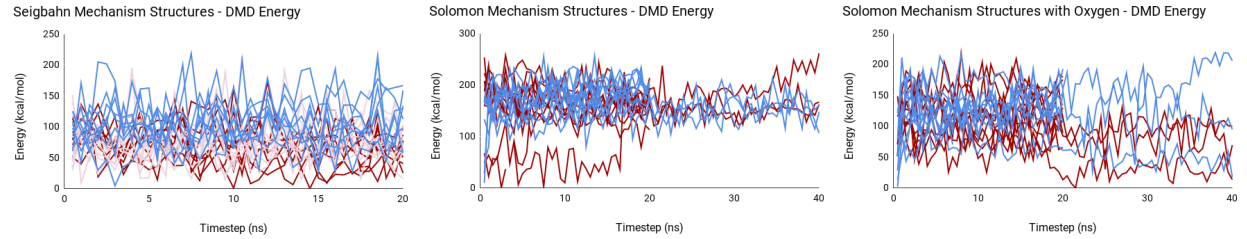


Figure D.6. Plots of the DMD energy relative to the lowest energy structure by timestep for every QM/DMD simulation in this study. The plots include all replicates of the wild type (dark red), R158Q mutant (medium blue), and E280K mutant (light violet, only in Seigbahn simulations).

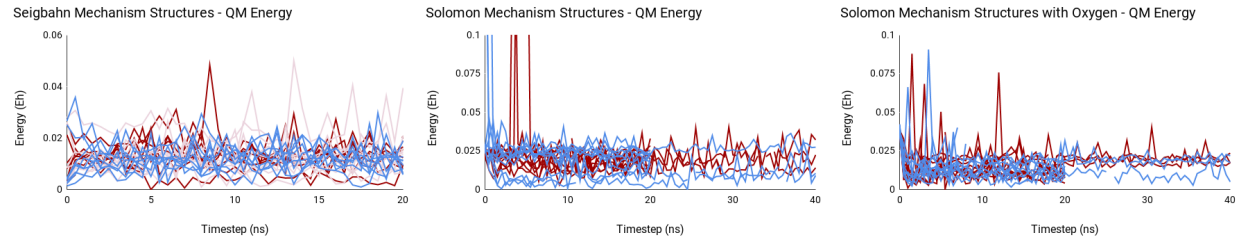


Figure D.7. Plots of the QM energy relative to the lowest energy structure by timestep for every QM/DMD simulation in this study. The plots include all replicates of the wild type (dark red), R158Q mutant (medium blue), and E280K mutant (light violet, only in Seigbahn simulations).

References

(1) Kepp, K. P. Heme: From Quantum Spin Crossover to Oxygen Manager of Life. *Coord. Chem. Rev.* **2017**, *344*, 363–374.

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

(3) Tripp, B. C.; Bell, C. B.; Cruz, F.; Krebs, C.; Ferry, J. G. A Role for Iron in an Ancient Carbonic Anhydrase. *J. Biol. Chem.* **2004**, *279* (8), 6683–6687.

(4) 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* (19), 6170–6178.

(5) Nechay, M. R.; Gallup, N. M.; Morgenstern, A.; Smith, Q. A.; Eberhart, M. E.; Alexandrova, A. N. Histone Deacetylase 8: Characterization of Physiological Divalent Metal Catalysis. *J. Phys. Chem. B* **2016**, *120* (26), 5884–5895.

(6) Zhu, J.; Dizin, E.; Hu, X.; Wavreille, A.-S.; Park, J.; Pei, D. S-Ribosylhomocysteinase (LuxS) Is a Mononuclear Iron Protein. *Biochemistry* **2003**, *42* (16), 4717–4726.

(7) Rajagopalan, P. T. R.; Yu, X. C.; Pei, D. Peptide Deformylase: A New Type of Mononuclear Iron Protein. *J. Am. Chem. Soc.* **1997**, *119* (50), 12418–12419.

(8) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chemie Int. Ed.* **2015**, *54* (11), 3351–3367.

(9) Hyster, T. K.; Ward, T. R. Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chemie Int. Ed.* **2016**, *55* (26), 7344–7357.

(10) Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* **2008**, *383* (5), 1069–1080.

(11) Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C.-H.; Arnold, F. H. Chemoenzymatic Elaboration of Monosaccharides Using Engineered Cytochrome P450BM3 Demethylases. *Proc. Natl. Acad. Sci.* **2009**, *106* (39), 16550–16555.

(12) Rentmeister, A.; Brown, T. R.; Snow, C. D.; Carbone, M. N.; Arnold, F. H. Engineered Bacterial Mimics of Human Drug Metabolizing Enzyme CYP2C9. *Chem. Cat. Chem.* **2011**, *3* (6), 1065–1071.

(13) Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L. Protein Design: Toward Functional Metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3495–3578.

(14) Reetz, M. T. Directed Evolution of Artificial Metalloenzymes: A Universal Means to Tune the Selectivity of Transition Metal Catalysts? *Acc. Chem. Res.* **2019**, *52* (2), 336–344.

(15) Prier, C. K.; Arnold, F. H. Chemomimetic Biocatalysis: Exploiting the Synthetic Potential of Cofactor-Dependent Enzymes to Create New Catalysts. *J. Am. Chem. Soc.* **2015**, *137* (44), 13992–14006.

(16) Natoli, S. N.; Hartwig, J. F. Noble- Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* **2019**, *52* (2), 326–335.

(17) Finney, L. A.; O'Halloran, T. V. Transition Metal Speciation in the Cell: Insights from the Chemistry of Metal Ion Receptors. *Science* **2003**, *300* (5621), 931–936.

(18) Tottey, S.; Harvie, D. R.; Robinson, N. J. Understanding How Cells Allocate Metals Using Metal Sensors and Metallochaperones. *Acc. Chem. Res.* **2005**, *38* (10), 775–783.

(19) Tottey, S.; Waldron, K. J.; Firbank, S. J.; Reale, B.; Bessant, C.; Sato, K.; Cheek, T. R.; Gray, J.; Banfield, M. J.; Dennison, C.; others. Protein-Folding Location Can Regulate Manganese-Binding versus Copper-or Zinc-Binding. *Nature* **2008**, *455* (7216), 1138.

(20) Foster, A. W.; Osman, D.; Robinson, N. J. Metal Preferences and Metallation. *J. Biol. Chem.* **2014**, *289* (41), 28095–28103.

(21) Xiao, Z.; Wedd, A. G. The Challenges of Determining Metal–Protein Affinities. *Nat. Prod. Rep.* **2010**, *27* (5), 768–789.

(22) Pieczenik, S. R.; Neustadt, J. Mitochondrial Dysfunction and Molecular Pathways of Disease. *Exp. Mol. Pathol.* **2007**, *83* (1), 84–92.

(23) Ibrahim, D.; Froberg, B.; Wolf, A.; Rusyniak, D. E. Heavy Metal Poisoning: Clinical Presentations and Pathophysiology. *Clin. Lab. Med.* **2006**, *26* (1), 67–97.

(24) Exley, C.; Burgess, E.; Day, J. P.; Jeffery, E. H.; Yokel, R. A. Aluminum Toxicokinetics. *J. Toxicol. Environ. Heal. Part A* **1996**, *48* (6), 569–584.

(25) Tinoco, A. D.; Thomas, H. R.; Incarvito, C. D.; Saghatelian, A.; Valentine, A. M. Cytotoxicity of a Ti (IV) Compound Is Independent of Serum Proteins. *Proc. Natl. Acad. Sci.* **2012**, *109* (13), 5016–5021.

(26) Guo, M.; Sun, H.; McArdle, H. J.; Gambling, L.; Sadler, P. J. TiIV Uptake and Release by Human Serum Transferrin and Recognition of Ti(IV)-Transferrin by Cancer Cells: Understanding the Mechanism of Action of the Anticancer Drug Titanocene Dichloride. *Biochemistry* **2000**, *39* (33), 10023–10033.

(27) Jakupec, M. A.; Keppler, B. K. Gallium in Cancer Treatment. *Curr. Top. Med. Chem.* **2004**, *4* (15), 1575–1583.

(28) Exley, C. Human Exposure to Aluminium. *Environ. Sci. Process. & Impacts* **2013**, *15* (10), 1807–1816.

(29) Gaggelli, E.; Kozlowski, H.; Valensin, D.; Valensin, G. Copper Homeostasis and Neurodegenerative Disorders (Alzheimer's, Prion, and Parkinson's Diseases and Amyotrophic Lateral Sclerosis). *Chem. Rev.* **2006**, *106* (6), 1995–2044.

(30) Hoops, S. C.; Anderson, K. W.; Merz Jr, K. M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113* (22), 8262–8270.

(31) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. Modeling the Charge Distribution at Metal Sites in Proteins for Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157* (3), 444–453.

(32) Neves, R. P. P.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Parameters for Molecular Dynamics Simulations of Manganese-Containing Metalloproteins. *J. Chem. Theory Comput.* **2013**, *9* (6), 2718–2732.

(33) Cho, A. E.; Goddard III, W. A. *Metalloproteins: Theory, Calculations, and Experiments*; CRC Press, 2015.

(34) Dal Peraro, M.; Vila, A. J.; Carloni, P.; Klein, M. L. Role of Zinc Content on the Catalytic Efficiency of B1 Metallo β-Lactamases. *J. Am. Chem. Soc.* **2007**, *129* (10), 2808–2816.

(35) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8* (4), 1314–1324.

(36) Rydberg, P.; Sigfridsson, E.; Ryde, U. On the Role of the Axial Ligand in Heme Proteins: A Theoretical Study. *JBIC J. Biol. Inorg. Chem.* **2004**, *9* (2), 203–223.

(37) Tantillo, D. J. How an Enzyme Might Accelerate an Intramolecular Diels- Alder Reaction: Theozymes for the Formation of Salvileucalin B. *Org. Lett.* **2010**, *12* (6), 1164–1167.

(38) Kries, H.; Blomberg, R.; Hilvert, D. De Novo Enzymes by Computational Design. *Curr. Opin. Chem. Biol.* **2013**, *17* (2), 221–228.

(39) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chemie Int. Ed.* **2013**, *52* (22), 5700–5725.

(40) Vaissier Welborn, V.; Head-Gordon, T. Computational Design of Synthetic Enzymes. *Chem. Rev.* **2018**, *119* (11), 6613–6630.

(41) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3601–3658.

(42) Ryde, U. QM/MM Calculations on Proteins. In *Methods in enzymology*; Elsevier, 2016; Vol. 577, pp 119–158.

(43) Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale Modeling of Enzymes: QM-Cluster, QM/MM, and QM/MM/MD: A Tutorial Review. *Int. J. Quantum Chem.* **2018**, *118* (9), e25558.

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

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

(46) Valdez, C. E.; Morgenstern, A.; Eberhart, M. E.; Alexandrova, A. N. Predictive Methods for Computational Metalloenzyme Redesign–a Test Case with Carboxypeptidase A. *Phys. Chem. Chem. Phys.* **2016**, *18* (46), 31744–31756.

(47) Reilley, D. J.; Popov, K. I.; Dokholyan, N. V; Alexandrova, A. N. Uncovered Dynamic Coupling Resolves the Ambiguous Mechanism of Phenylalanine Hydroxylase Oxygen Binding. *J. Phys. Chem. B* **2019**, *123* (21), 4534–4539.

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

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

(50) 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.

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

(52) Reilley, D. J.; Fuller III, J. T.; Nechay, M. R.; Victor, M.; Li, W.; Ruberry, J. D.; Mujika, J. I.; Lopez, X.; Alexandrova, A. N. Toxic and Physiological Metal Uptake and Release by Human Serum Transferrin. *Biophys. J.* **2020**, *118* (12), 2979–2988.

(53) Genheden, S.; Ryde, U. Will Molecular Dynamics Simulations of Proteins Ever Reach Equilibrium? *Phys. Chem. Chem. Phys.* **2012**, *14* (24), 8662–8677.

(54) Guggenheim, E. A. The Conceptions of Electrical Potential Difference between Two Phases and the Individual Activities of Ions. *J. Phys. Chem.* **1929**, *33* (6), 842–849.

(55) Klotz, I. M.; Rosenberg, R. M. *Chemical Thermodynamics*; Wiley, 1994.

(56) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous Solvation Free Energies of Ions and Ion− Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006**, *110* (32), 16066–16081.

(57) Myers, R. W.; Wray, J. W.; Fish, S.; Abeles, R. H. Purification and Characterization of an Enzyme Involved in Oxidative Carbon-Carbon Bond Cleavage Reactions in the Methionine Salvage Pathway of Klebsiella Pneumoniae. *J. Biol. Chem.* **1993**, *268* (33), 24785–24791.

(58) Oram, S. W.; Ai, J.; Pagani, G. M.; Hitchens, M. R.; Stern, J. A.; Eggener, S.; Pins, M.; Xiao, W.; Cai, X.; Haleem, R.; others. Expression and Function of the Human Androgen-Responsive Gene ADI1 in Prostate Cancer. *Neoplasia (New York, NY)* **2007**, *9* (8), 643.

(59) Dai, Y.; Wensink, P. C.; Abeles, R. H. One Protein, Two Enzymes. *J. Biol. Chem.* **1999**, *274* (3), 1193–1195.

(60) Wray, J. W.; Abeles, R. H. The Methionine Salvage Pathway in Klebsiella Pneumoniae and Rat Liver IDENTIFICATION AND CHARACTERIZATION OF TWO NOVEL DIOXYGENASES. *J. Biol. Chem.* **1995**, *270* (7), 3147–3153.

(61) Borowski, T.; Bassan, A.; Siegbahn, P. E. M. DFT Study of the Uncatalyzed Dioxygenation of Acireductone. *J. Mol. Struct.* **2006**, *772* (1–3), 89–92.

(62) Dai, Y.; Pochapsky, T. C.; Abeles, R. H. Mechanistic Studies of Two Dioxygenases in the Methionine Salvage Pathway of Klebsiella Pneumoniae. *Biochemistry* **2001**, *40* (21), 6379–6387.

(63) Chai, S. C.; Ju, T.; Dang, M.; Goldsmith, R. B.; Maroney, M. J.; Pochapsky, T. C. Characterization of Metal Binding in the Active Sites of Acireductone Dioxygenase Isoforms from Klebsiella ATCC 8724. *Biochemistry* **2008**, *47* (8), 2428–2438.

(64) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (2), 91–100.

(65) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative Assessment of a New Nonempirical Density Functional: Molecules and Hydrogen-Bonded Complexes. *J. Chem. Phys.* **2003**, *119* (23), 12129–12137.

(66) 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.* **2010**, *132* (15), 154104.

(67) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.

(68) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–2235.

(69) Kouzarides, T. Acetylation: A Regulatory Modification to Rival Phosphorylation? *EMBO J.* **2000**, *19* (6), 1176–1179.

(70) 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* (5942), 834–840.

(71) Phillips, D. M. P. The Presence of Acetyl Groups in Histones. *Biochem. J.* **1963**, *87* (2), 258.

(72) 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.* **1964**, *51* (5), 786–794.

(73) Gallinari, P.; Di Marco, S.; Jones, P.; Pallaoro, M.; Steinkühler, C. HDACs, Histone Deacetylation and Gene Transcription: From Molecular Biology to Cancer Therapeutics. *Cell Res.* **2007**, *17* (3), 195.

(74) 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* (1), 32.

(75) Marks, P. A.; Breslow, R. Dimethyl Sulfoxide to Vorinostat: Development of This Histone Deacetylase Inhibitor as an Anticancer Drug. *Nat. Biotechnol.* **2007**, *25* (1), 84.

(76) West, A. C.; Johnstone, R. W. New and Emerging HDAC Inhibitors for Cancer Treatment. *J. Clin. Invest.* **2014**, *124* (1), 30–39.

(77) Lobera, M.; Madauss, K. P.; Pohlhaus, D. T.; Wright, Q. G.; Trocha, M.; Schmidt, D. R.; Baloglu, E.; Trump, R. P.; Head, M. S.; Hofmann, G. A.; others. Selective Class IIa Histone Deacetylase Inhibition via a Nonchelating Zinc-Binding Group. *Nat. Chem. Biol.* **2013**, *9* (5), 319.

(78) Furumai, R.; Matsuyama, A.; Kobashi, N.; Lee, K.-H.; Nishiyama, M.; Nakajima, H.; Tanaka, A.; Komatsu, Y.; Nishino, N.; Yoshida, M.; others. FK228 (Depsipeptide) as a Natural Prodrug That Inhibits Class I Histone Deacetylases. *Cancer Res.* **2002**, *62* (17), 4916–4921.

(79) 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* (6749), 188.

(80) Drummond, D. C.; Noble, C. O.; Kirpotin, D. B.; Guo, Z.; Scott, G. K.; Benz, C. C. Clinical Development of Histone Deacetylase Inhibitors as Anticancer Agents. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 495–528.

(81) Dojindo Molecular Technologies, I. Metal Chelates.

(82) Anderegg, G. *Critical Survey of Stability Constants of EDTA Complexes*; Pergamon Press: New York, 1977.

(83) Passer, E.; White, J. G.; Cheng, K. L. The Crystal Structure of $Mg_2EDTA \cdot 9H_2O$. *Inorganica Chim. Acta* **1977**, *24*, 13–23.

(84) Li, H.; Sadler, P. J.; Sun, H. Rationalization of the Strength of Metal Binding to Human Serum Transferrin. *Eur. J. Biochem.* **1996**, *242* (2), 387–393.

(85) Tinoco, A. D.; Valentine, A. M. Ti (IV) Binds to Human Serum Transferrin More Tightly than Does Fe (III). *J. Am. Chem. Soc.* **2005**, *127* (32), 11218–11219.

(86) Tinoco, A. D.; Incarvito, C. D.; Valentine, A. M. Calorimetric, Spectroscopic, and Model Studies Provide Insight into the Transport of Ti (IV) by Human Serum Transferrin. *J. Am. Chem. Soc.* **2007**, *129* (11), 3444–3454.

(87) Vera, J. L.; Román, F. R.; Meléndez, E. Study of Titanocene--DNA and Molybdenocene--DNA Interactions by Inductively Coupled Plasma--Atomic Emission Spectroscopy. *Anal. Bioanal. Chem.* **2004**, *379* (3), 399–403.

(88) Mokdsi, G.; Harding, M. M. Inhibition of Human Topoisomerase II by the Antitumor Metallocenes. *J. Inorg. Biochem.* **2001**, *83* (2–3), 205–209.

(89) Cini, M.; Bradshaw, T. D.; Woodward, S. Using Titanium Complexes to Defeat Cancer: The View from the Shoulders of Titans. *Chem. Soc. Rev.* **2017**, *46* (4), 1040–1051.

(90) Sun, H.; Li, H.; Sadler, P. J. Transferrin as a Metal Ion Mediator. *Chem. Rev.* **1999**, *99* (9), 2817–2842.

(91) Li, H.; Qian, Z. M. Transferrin/Transferrin Receptor-Mediated Drug Delivery. *Med. Res. Rev.* **2002**, *22* (3), 225–250.

(92) Gupta, Y.; Jain, A.; Jain, S. K. Transferrin-Conjugated Solid Lipid Nanoparticles for Enhanced Delivery of Quinine Dihydrochloride to the Brain. *J. Pharm. Pharmacol.* **2007**, *59* (7), 935–940.

(93) Lesley, J.; Schulte, R.; Woods, J. Modulation of Transferrin Receptor Expression and Function by Anti-Transferrin Receptor Antibodies and Antibody Fragments. *Exp. Cell Res.* **1989**, *182* (1), 215–233.

(94) Huwyler, J.; Wu, D.; Pardridge, W. M. Brain Drug Delivery of Small Molecules Using Immunoliposomes. *Proc. Natl. Acad. Sci.* **1996**, *93* (24), 14164–14169.

(95) Ulbrich, K.; Hekmatara, T.; Herbert, E.; Kreuter, J. Transferrin-and Transferrin-Receptor-Antibody-Modified Nanoparticles Enable Drug Delivery across the Blood--Brain Barrier (BBB). *Eur. J. Pharm. Biopharm.* **2009**, *71* (2), 251–256.

(96) Noinaj, N.; Easley, N. C.; Oke, M.; Mizuno, N.; Gumbart, J.; Boura, E.; Steere, A. N.; Zak, O.; Aisen, P.; Tajkhorshid, E.; Evans, R. W.; Gorringe, A. R.; Mason, A. B.; Steven, A. C.; Buchanan, S. K. Structural Basis for Iron Piracy by Pathogenic Neisseria. *Nature* **2012**, *483* (7387), 53.

(97) Dautry-Varsat, A.; Ciechanover, A.; Lodish, H. F. PH and the Recycling of Transferrin during Receptor-Mediated Endocytosis. *Proc. Natl. Acad. Sci.* **1983**, *80* (8), 2258–2262.

(98) Sipe, D. M.; Murphy, R. F. Binding to Cellular Receptors Results in Increased Iron Release from Transferrin at Mildly Acidic PH. *J. Biol. Chem.* **1991**, *266* (13), 8002–8007.

(99) Grossman, J. G.; Crawley, J. B.; Strange, R. W.; Patel, K. J.; Murphy, L. M.; Neu, M.; Evans, R. W.; Hasnain, S. S. The Nature of Ligand-Induced Conformational Change in Transferrin in Solution. *J. Mol. Biol.* **1998**, *279*, 461–472.

(100) MacGillivray, R. T. A.; Moore, S. A.; Chen, J.; Anderson, B. F.; Baker, H.; Luo, Y.; Bewley, M.; Smith, C. A.; Murphy, M. E. P.; Wang, Y.; Mason, A. B.; Woodworth, R. C.; Brayer, G. D.; Baker, E. N. Two High-Resolution Crystal Structures of the Recombinant N-Lobe of Human Transferrin Reveal a Structural Change Implicated in Iron Release. *Biochemistry* **1998**, *37* (22), 7919–7928.

(101) Dhungana, S.; Taboy, C. H.; Zak, O.; Larvie, M.; Crumbliss, A. L.; Aisen, P. Redox Properties of Human Transferrin Bound to Its Receptor. *Biochemistry* **2004**, *43* (1), 205–209.

(102) He, Q.-Y.; Mason, A. B.; Tam, B. M.; MacGillivray, R. T. A.; Woodworth, R. C. Dual Role of Lys206- Lys296 Interaction in Human Transferrin N-Lobe: Iron-Release Trigger and Anion-Binding Site. *Biochemistry* **1999**, *38* (30), 9704–9711.

(103) Steinlein, L. M.; Ligman, C. M.; Kessler, S.; Ikeda, R. A. Iron Release Is Reduced by Mutations of Lysines 206 and 296 in Recombinant N-Terminal Half-Transferrin. *Biochemistry* **1998**, *37* (39), 13696–13703.

(104) Eckenroth, B. E.; Steere, A. N.; Chasteen, N. D.; Everse, S. J.; Mason, A. B. How the Binding of Human Transferrin Primes the Transferrin Receptor Potentiating Iron Release at Endosomal PH. *Proc. Natl. Acad. Sci.* **2011**, *108* (32), 13089–13094.

(105) Jeffrey, P. D.; Bewley, M. C.; MacGillivray, R. T. A.; Mason, A. B.; Woodworth, R. C.; Baker, E. N. Ligand-Induced Conformational Change in Transferrins: Crystal Structure of the Open Form of the N-Terminal Half-Molecule of Human Transferrin. *Biochemistry* **1998**, *37* (40), 13978–13986.

(106) Steere, A. N.; Byrne, S. L.; Chasteen, N. D.; Mason, A. B. Kinetics of Iron Release from Transferrin Bound to the Transferrin Receptor at Endosomal PH. *Biochim. Biophys. Acta (BBA)-General Subj.* **2012**, *1820* (3), 326–333.

(107) Tinoco, A. D.; Saxena, M.; Sharma, S.; Noinaj, N.; Delgado, Y.; Quiñones González, E. P.; Conklin, S. E.; Zambrana, N.; Loza-Rosas, S. A.; Parks, T. B. Unusual Synergism of Transferrin and Citrate in the Regulation of Ti (IV) Speciation, Transport, and Toxicity. *J. Am. Chem. Soc.* **2016**, *138* (17), 5659–5665.

(108) Curtin, J. P.; Wang, M.; Cheng, T.; Jin, L.; Sun, H. The Role of Citrate, Lactate and Transferrin in Determining Titanium Release from Surgical Devices into Human Serum. *JBIC J. Biol. Inorg. Chem.* **2018**, *23* (3), 471–480.

(109) Baker, H. M.; Nurizzo, D.; Mason, A. B.; Baker, E. N. Structures of Two Mutants That Probe the Role in Iron Release of the Dilysine Pair in the N-Lobe of Human Transferrin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63* (3), 408–414.

(110) 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* **2012**, *51* (35), 7017–7027. https://doi.org/10.1021/bi300584p.

(111) Rinaldo, D.; Field, M. J. A Computational Study of the Open and Closed Forms of the N-Lobe Human Serum Transferrin Apoprotein. *Biophys. J.* **2003**, *85* (6), 3485–3501.

(112) Mujika, J. I.; López, X.; Rezabal, E.; Castillo, R.; Marti, S.; Moliner, V.; Ugalde, J. M. A QM/MM Study of the Complexes Formed by Aluminum and Iron with Serum Transferrin at Neutral and Acidic PH. *J. Inorg. Biochem.* **2011**, *105* (11), 1446–1456.

(113) Sakajiri, T.; Yajima, H.; Yamamura, T. Density Functional Theory Study on Metal-Binding Energies for Human Serum Transferrin-Metal Complexes. *ISRN Biophys.* **2012**, *2012* (Iii), 1–5. https://doi.org/10.5402/2012/124803.

(114) Lei, P.; Ayton, S.; Bush, A. I.; Adlard, P. A. GSK-3 in Neurodegenerative Diseases. *Int. J. Alzheimer's Dis.* **2011**, *2011*.

(115) Lin, R.; Jones, N. C.; Kwan, P. Unravelling the Role of Glycogen Synthase Kinase-3 in Alzheimer's Disease-Related Epileptic Seizures. *Int. J. Mol. Sci.* **2020**, *21* (10), 3676.

(116) Hooper, C.; Killick, R.; Lovestone, S. The GSK3 Hypothesis of Alzheimer's Disease. *J. Neurochem.* **2008**, *104* (6), 1433–1439.

(117) King, M. K.; Pardo, M.; Cheng, Y.; Downey, K.; Jope, R. S.; Beurel, E. Glycogen Synthase Kinase-3 Inhibitors: Rescuers of Cognitive Impairments. *Pharmacol. Ther.* **2014**, *141* (1), 1–12.

(118) Aourz, N.; Serruys, A.-S. K.; Chabwine, J. N.; Balegamire, P. B.; Afrikanova, T.; Edrada-Ebel, R.; Grey, A. I.; Kamuhabwa, A. R.; Walrave, L.; Esguerra, C. V. Identification of GSK-3 as a Potential Therapeutic Entry Point for Epilepsy. *ACS Chem. Neurosci.* **2018**, *10* (4), 1992–2003.

(119) Morales-García, J. A.; Susín, C.; Alonso-Gil, S.; Pérez, D. I.; Palomo, V.; Pérez, C.; Conde, S.; Santos, A.; Gil, C.; Martínez, A. Glycogen Synthase Kinase-3 Inhibitors as Potent Therapeutic Agents for the Treatment of Parkinson Disease. *ACS Chem. Neurosci.* **2013**, *4* (2), 350–360.

(120) Davies, M. P.; Benitez, R.; Perez, C.; Jakupovic, S.; Welsby, P.; Rzepecka, K.; Alder, J.; Davidson, C.; Martinez, A.; Hayes, J. M. Structure-Based Design of Potent Selective Nanomolar Type-II Inhibitors of Glycogen Synthase Kinase-3β. *J. Med. Chem.* **2021**, *64* (3), 1497–1509.

(121) Eldar-Finkelman, H.; Martinez, A. GSK-3 Inhibitors: Preclinical and Clinical Focus on CNS. *Front. Mol. Neurosci.* **2011**, *4*, 32.

(122) Bhat, R. V; Andersson, U.; Andersson, S.; Knerr, L.; Bauer, U.; Sundgren-Andersson, A. K. The Conundrum of GSK3 Inhibitors: Is It the Dawn of a New Beginning? *J. Alzheimer's Dis.* **2018**, *64* (s1), S547–S554.

(123) Stamos, J. L.; Chu, M. L.-H.; Enos, M. D.; Shah, N.; Weis, W. I. Structural Basis of GSK-3 Inhibition by N-Terminal Phosphorylation and by the Wnt Receptor LRP6. *Elife* **2014**, *3*, e01998.

(124) Palomo, V.; Soteras, I.; Perez, D. I.; Perez, C.; Gil, C.; Campillo, N. E.; Martinez, A. Exploring the Binding Sites of Glycogen Synthase Kinase 3. Identification and Characterization of Allosteric Modulation Cavities. *J. Med. Chem.* **2011**, *54* (24), 8461–8470.

(125) Arfeen, M.; Bharatam, P. V. Design of Glycogen Synthase Kinase-3 Inhibitors: An Overview on Recent Advancements. *Curr. Pharm. Des.* **2013**, *19* (26), 4755–4775.

(126) Cade, J. F. J. Lithium Salts in the Treatment of Psychotic Excitement. *Med. J. Aust.* **1949**.

(127) Hampel, H.; Ewers, M.; Bürger, K.; Annas, P.; Mörtberg, A.; Bogstedt, A.; Frölich, L.; Schröder, J.; Schönknecht, P.; Riepe, M. W. Lithium Trial in Alzheimer's Disease: A Randomized, Single-Blind, Placebo-Controlled, Multicenter 10-Week Study. *J. Clin. Psychiatry* **2009**, *70* (6), 922–931.

(128) Leyhe, T.; Eschweiler, G. W.; Stransky, E.; Gasser, T.; Annas, P.; Basun, H.; Laske, C. Increase of BDNF Serum Concentration in Lithium Treated Patients with Early Alzheimer's Disease. *J. Alzheimer's Dis.* **2009**, *16* (3), 649–656.

(129) Breen, M. S.; White, C. H.; Shekhtman, T.; Lin, K.; Looney, D.; Woelk, C. H.; Kelsoe, J. R. Lithium-Responsive Genes and Gene Networks in Bipolar Disorder Patient-Derived Lymphoblastoid Cell Lines. *Pharmacogenomics J.* **2016**, *16* (5), 446–453.

(130) Ryves, W. J.; Harwood, A. J. Lithium Inhibits Glycogen Synthase Kinase-3 by Competition for Magnesium. *Biochem. Biophys. Res. Commun.* **2001**, *280* (3), 720–725.

(131) Ryves, W. J.; Dajani, R.; Pearl, L.; Harwood, A. J. Glycogen Synthase Kinase-3 Inhibition by Lithium and Beryllium Suggests the Presence of Two Magnesium Binding Sites. *Biochem. Biophys. Res. Commun.* **2002**, *290* (3), 967–972.

(132) Fountoulakis, K. N.; Vieta, E.; Bouras, C.; Notaridis, G.; Giannakopoulos, P.; Kaprinis, G.; Akiskal, H. A Systematic Review of Existing Data on Long-Term Lithium Therapy: Neuroprotective or Neurotoxic? *Int. J. Neuropsychopharmacol.* **2008**, *11* (2), 269–287.

(133) Schou, M. Lithium Treatment at 52. *J. Affect. Disord.* **2001**, *67* (1–3), 21–32.

(134) Cheng, Y.; Zhang, Y.; McCammon, J. A. How Does the CAMP-Dependent Protein Kinase Catalyze the Phosphorylation Reaction: An Ab Initio QM/MM Study. *J. Am. Chem. Soc.* **2005**, *127* (5), 1553–1562.

(135) Gerlits, O.; Tian, J.; Das, A.; Langan, P.; Heller, W. T.; Kovalevsky, A. Phosphoryl Transfer Reaction Snapshots in Crystals: Insights into the Mechanism of Protein Kinase a Catalytic Subunit. *J. Biol. Chem.* **2015**, *290* (25), 15538–15548.

(136) Dudev, T.; Lim, C. Competition between Li+ and Mg^{2+} in Metalloproteins. Implications for Lithium Therapy. *J. Am. Chem. Soc.* **2011**, *133* (24), 9506–9515.

(137) Sun, H.; Jiang, Y.; Yu, Q.; Luo, C.; Zou, J. The Effect of Li^{+} on GSK-3 Inhibition: Molecular Dynamics Simulation. *J. Mol. Model.* **2011**, *17* (2), 377–381.

(138) Lu, S.-Y.; Jiang, Y.-J.; Zou, J.-W.; Wu, T.-X. Dissection of the Difference between the Group I Metal Ions in Inhibiting GSK3β: A Computational Study. *Phys. Chem. Chem. Phys.* **2011**, *13* (15), 7014–7023.

(139) Proctor, E. A.; Ding, F.; Dokholyan, N. V. Discrete Molecular Dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (1), 80–92.

(140) 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.* **2013**, *9* (1), 730–737.

(141) Reilley, D. J.; Hennefarth, M. R.; Alexandrova, A. N. The Case for Enzymatic Competitive Metal Affinity Methods. *ACS Catal.* **2020**, *10* (3), 2298–2307.

(142) Anderegg, G. *Critical Survey of Stability Constants of EDTA Complexes: Critical Evaluation of Equilibrium Constants in Solution: Stability Constants of Metal Complexes*; Elsevier, 2013.

(143) Barnett, B. L.; Uchtman, V. A. Structural Investigations of Calcium-Binding Molecules. 4. Calcium Binding to Aminocarboxylates. Crystal Structures of Ca (CaEDTA). $7H_2O$ and Na (CaNTA). *Inorg. Chem.* **1979**, *18* (10), 2674–2678.

(144) Dudev, T.; Grauffel, C.; Lim, C. How Native and Alien Metal Cations Bind ATP: Implications for Lithium as a Therapeutic Agent. *Sci. Rep.* **2017**, *7* (1), 1–10.

(145) Dumorné, K.; Córdova, D. C.; Astorga-Eló, M.; Renganathan, P. Extremozymes: A Potential Source for Industrial Applications. *J Microbiol Biotechnol* **2017**, *27* (4), 649–659.

(146) Sarmiento, F.; Peralta, R.; Blamey, J. M. Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Front. Bioeng. Biotechnol.* **2015**, *3*, 148.

(147) Chen, G.-Q.; Jiang, X.-R. Next Generation Industrial Biotechnology Based on Extremophilic Bacteria. *Curr. Opin. Biotechnol.* **2018**, *50*, 94–100.

(148) Fernandes, P. Enzymes in Food and Feed Industries: Where Tradition Meets Innovation. In *Biocatalysis*; Springer, 2019; pp 233–253.

(149) Rivinoja, A.; Pujol, F. M.; Hassinen, A.; Kellokumpu, S. Golgi pH, Its Regulation and Roles in Human Disease. *Ann. Med.* **2012**, *44* (6), 542–554.

(150) Webb, B. A.; Chimenti, M.; Jacobson, M. P.; Barber, D. L. Dysregulated pH: A Perfect Storm for Cancer Progression. *Nat. Rev. Cancer* **2011**, *11* (9), 671–677.

(151) Fang, B.; Wang, D.; Huang, M.; Yu, G.; Li, H. Hypothesis on the Relationship between the Change in Intracellular pH and Incidence of Sporadic Alzheimer's Disease or Vascular Dementia. *Int. J. Neurosci.* **2010**, *120* (9), 591–595.

(152) Dokholyan, N. V. Experimentally-Driven Protein Structure Modeling. *J. Proteomics* **2020**, *220*, 103777.

(153) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson–Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–392.

(154) Bashford, D.; Gerwert, K. Electrostatic Calculations of the pK_a Values of Ionizable Groups in Bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224* (2), 473–486.

(155) Onufriev, A. V; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.

(156) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R. Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91* (1–3), 57–95.

(157) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537–W541.

(158) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

(159) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.

(160) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. Protonation of Interacting Residues in a Protein by a Monte Carlo Method: Application to Lysozyme and the Photosynthetic

Reaction Center of Rhodobacter Sphaeroides. *Proc. Natl. Acad. Sci.* **1991**, *88* (13), 5804–5808.

(161) Kesvatera, T.; Jönsson, B.; Thulin, E.; Linse, S. Focusing of the Electrostatic Potential at EF-hands of Calbindin D9k: Titration of Acidic Residues. *Proteins Struct. Funct. Bioinforma.* **2001**, *45* (2), 129–135.

(162) Teixeira, A. A. R.; Lund, M.; Barroso da Silva, F. L. Fast Proton Titration Scheme for Multiscale Modeling of Protein Solutions. *J. Chem. Theory Comput.* **2010**, *6* (10), 3259–3266.

(163) Barroso daSilva, F. L.; Dias, L. G. Development of Constant-pH Simulation Methods in Implicit Solvent and Applications in Biomolecular Systems. *Biophys. Rev.* **2017**, *9* (5), 699–728.

(164) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-p H Molecular Dynamics Using Stochastic Titration. *J. Chem. Phys.* **2002**, *117* (9), 4184–4200.

(165) Fuzo, C. A.; Degrève, L. The pH Dependence of Flavivirus Envelope Protein Structure: Insights from Molecular Dynamics Simulations. *J. Biomol. Struct. Dyn.* **2014**, *32* (10), 1563–1574.

(166) Lee, M. S.; Salsbury Jr, F. R.; Brooks III, C. L. Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins Struct. Funct. Bioinforma.* **2004**, *56* (4), 738–752.

(167) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(168) Tidor, B. Simulated Annealing on Free Energy Surfaces by a Combined Molecular Dynamics and Monte Carlo Approach. *J. Phys. Chem.* **1993**, *97* (5), 1069–1073.

(169) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks III, C. L. Constant pH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (7), 1319–1331.

(170) Chen, Y.; Roux, B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **2015**, *11* (8), 3919–3931.

(171) Huang, Y.; Harris, R. C.; Shen, J. Generalized Born Based Continuous Constant pH Molecular Dynamics in AMBER: Implementation, Benchmarking and Analysis. *J. Chem. Inf. Model.* **2018**, *58* (7), 1372–1383.

(172) Harris, R. C.; Shen, J. GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: pK_a Predictions with Single-pH Simulations. *J. Chem. Inf. Model.* **2019**, *59* (11), 4821–4832.

(173) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7* (8), 2617–2629.

(174) Vila-Viçosa, D.; Reis, P. B. P. S.; Baptista, A. M.; Oostenbrink, C.; Machuqueiro, M. A pH Replica Exchange Scheme in the Stochastic Titration Constant-pH MD Method. *J. Chem. Theory Comput.* **2019**, *15* (5), 3108–3116.

(175) Kim, M. O.; Blachly, P. G.; McCammon, J. A. Conformational Dynamics and Binding Free Energies of Inhibitors of BACE-1: From the Perspective of Protonation Equilibria. *PLoS Comput Biol* **2015**, *11* (10), e1004341.

(176) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating Proteins at Constant pH: An Approach Combining Molecular Dynamics and Monte Carlo Simulation. *Proteins Struct. Funct. Bioinforma.* **2002**, *47* (4), 469–480.

(177) Mongan, J.; Case, D. A.; McCAMMON, J. A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048.

(178) Barroso da Silva, F. L.; MacKernan, D. Benchmarking a Fast Proton Titration Scheme in Implicit Solvent for Biomolecular Simulations. *J. Chem. Theory Comput.* **2017**, *13* (6), 2915–2929.

(179) Harris, R. C.; Liu, R.; Shen, J. Predicting Reactive Cysteines with Implicit-Solvent-Based Continuous Constant pH Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2020**, *16* (6), 3689–3698.

(180) Grünewald, F.; Souza, P. C. T.; Abdizadeh, H.; Barnoud, J.; de Vries, A. H.; Marrink, S. J. Titratable Martini Model for Constant pH Simulations. *J. Chem. Phys.* **2020**, *153* (2), 24118.

(181) Pasquali, S.; Frezza, E.; Barroso da Silva, F. L. Coarse-Grained Dynamic RNA Titration Simulations. *Interface Focus* **2019**, *9* (3), 20180066.

(182) Barroso da Silva, F. L.; Sterpone, F.; Derreumaux, P. OPEP6: A New Constant-pH Molecular Dynamics Simulation Scheme with OPEP Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2019**, *15* (6), 3875–3888.

(183) 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* (29), 8375–8382.

(184) Khandogin, J.; Chen, J.; Brooks, C. L. Exploring Atomistic Details of pH-Dependent Peptide Folding. *Proc. Natl. Acad. Sci.* **2006**, *103* (49), 18546–18550.

(185) Vila-Viçosa, D.; Campos, S. R. R.; Baptista, A. M.; Machuqueiro, M. Reversibility of Prion Misfolding: Insights from Constant-pH Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116* (30), 8812–8821.

(186) Shi, C.; Wallace, J. A.; Shen, J. K. Thermodynamic Coupling of Protonation and Conformational Equilibria in Proteins: Theory and Simulation. *Biophys. J.* **2012**, *102* (7), 1590–1597.

(187) Chen, W.; Huang, Y.; Shen, J. Conformational Activation of a Transmembrane Proton Channel from Constant pH Molecular Dynamics. *J. Phys. Chem. Lett.* **2016**, *7* (19), 3961–3966.

(188) Isom, D. G.; Castañeda, C. A.; Cannon, B. R. Large Shifts in pK_a Values of Lysine Residues Buried inside a Protein. *Proc. Natl. Acad. Sci.* **2011**, *108* (13), 5260–5265.

(189) Peck, M. T.; Ortega, G.; De Luca-Johnson, J. N.; Schlessman, J. L.; Robinson, A. C.; García-Moreno E, B. Local Backbone Flexibility as a Determinant of the Apparent pK_a Values of Buried Ionizable Groups in Proteins. *Biochemistry* **2017**, *56* (40), 5338–5346.

(190) Ding, F.; Borreguero, J. M.; Buldyrey, S. V; Stanley, H. E.; Dokholyan, N. V. Mechanism for the A-helix to B-hairpin Transition. *Proteins Struct. Funct. Bioinforma.* **2003**, *53* (2), 220–228.

(191) Schramm, V. L.; Schwartz, S. D. Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* **2018**, *57* (24), 3299–3308.

(192) Fang, C.; Frontiera, R. R.; Tran, R.; Mathies, R. A. Mapping GFP Structure Evolution during Proton Transfer with Femtosecond Raman Spectroscopy. *Nature* **2009**, *462* (7270), 200–204.

(193) Zhan, C.-G.; Dixon, D. A. Absolute Hydration Free Energy of the Proton from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A* **2001**, *105* (51), 11534–11540.

(194) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V; Tuttle, T. R. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A* **1998**, *102* (40), 7787–7794.

(195) Huang, Y.; Yue, Z.; Tsai, C.-C.; Henderson, J. A.; Shen, J. Predicting Catalytic Proton Donors and Nucleophiles in Enzymes: How Adding Dynamics Helps Elucidate the Structure–Function Relationships. *J. Phys. Chem. Lett.* **2018**, *9* (6), 1179–1184.

(196) Sanner, M. F.; Olson, A. J.; Spehner, J. Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, *38* (3), 305–320.

(197) Bon, C.; Lehmann, M. S.; Wilkinson, C. Quasi-Laue Neutron-Diffraction Study of the Water Arrangement in Crystals of Triclinic Hen Egg-White Lysozyme. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55* (5), 978–987.

(198) Weichsel, A.; Gasdaska, J. R.; Powis, G.; Montfort, W. R. Crystal Structures of Reduced, Oxidized, and Mutated Human Thioredoxins: Evidence for a Regulatory Homodimer. *Structure* **1996**, *4* (6), 735–751.

(199) Ohren, J. F.; Kundracik, M. L.; Borders, C. L.; Edmiston, P.; Viola, R. E. Structural Asymmetry and Intersubunit Communication in Muscle Creatine Kinase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63* (3), 381–389.

(200) Bong, S. M.; Moon, J. H.; Nam, K. H.; Lee, K. S.; Chi, Y. M.; Hwang, K. Y. Structural Studies of Human Brain-Type Creatine Kinase Complexed with the ADP–Mg2+–NO3−–Creatine Transition-State Analogue Complex. *FEBS Lett.* **2008**, *582* (28), 3959–3965.

(201) Stites, W. E.; Gittis, A. G.; Lattman, E. E.; Shortle, D. In a Staphylococcal Nuclease Mutant the Side-Chain of a Lysine Replacing Valine 66 Is Fully Buried in the Hydrophobic Core. *J. Mol. Biol.* **1991**, *221* (1), 7–14.

(202) Pahari, S.; Sun, L.; Alexov, E. PKAD: A Database of Experimentally Measured pK_a Values of Ionizable Groups in Proteins. *Database* **2019**, *2019*.

(203) Bartik, K.; Redfield, C.; Dobson, C. M. Measurement of the Individual pK_a Values of Acidic Residues of Hen and Turkey Lysozymes by Two-Dimensional 1H NMR. *Biophys. J.* **1994**, *66* (4), 1180–1184.

(204) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. Remeasuring HEWL pK_a Values by NMR Spectroscopy: Methods, Analysis, Accuracy, and Implications for Theoretical pK_a Calculations. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (3), 685–702.

(205) Forman-Kay, J. D.; Clore, G. M.; Gronenborn, A. M. Relationship between Electrostatics and Redox Function in Human Thioredoxin: Characterization of pH Titration Shifts Using Two-Dimensional Homo-and Heteronuclear NMR. *Biochemistry* **1992**, *31* (13), 3442–3452.

(206) Qin, J.; Clore, G. M.; Gronenborn, A. M. Ionization Equilibria for Side-Chain Carboxyl Groups in Oxidized and Reduced Human Thioredoxin and in the Complex with Its Target Peptide from the Transcription Factor NFκB. *Biochemistry* **1996**, *35* (1), 7–13.

(207) Wang, P.-F.; McLeish, M. J.; Kneen, M. M.; Lee, G.; Kenyon, G. L. An Unusually Low pK_a for Cys282 in the Active Site of Human Muscle Creatine Kinase. *Biochemistry* **2001**, *40* (39), 11698–11705.

(208) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B. Experimental pK_a Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. *Biophys. J.* **2002**, *82* (6), 3289–3304.

(209) Takayama, Y.; Castaneda, C. A.; Chimenti, M.; García-Moreno, B.; Iwahara, J. Direct Evidence for Deprotonation of a Lysine Side Chain Buried in the Hydrophobic Core of a Protein. *J. Am. Chem. Soc.* **2008**, *130* (21), 6714–6715.

(210) Karp, D. A.; Stahley, M. R.; García-Moreno E, B. Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. *Biochemistry* **2010**, *49* (19), 4138–4146.

(211) Chimenti, M. S.; Khangulov, V. S.; Robinson, A. C.; Heroux, A.; Majumdar, A.; Schlessman, J. L.; García-Moreno, B. Structural Reorganization Triggered by Charging of Lys Residues in the Hydrophobic Interior of a Protein. *Structure* **2012**, *20* (6), 1071–1085.

(212) Fitzpatrick, P. F. Tetrahydropterin-Dependent Amino Acid Hydroxylases. *Annu. Rev. Biochem.* **1999**, *68* (1), 355–381.

(213) Kappock, T. J.; Caradonna, J. P. Pterin-Dependent Amino Acid Hydroxylases. *Chem. Rev.* **1996**, *96* (7), 2659–2756.

(214) Kaufman, S. The Phenylalanine Hydroxylating System. *Adv. Enzymol. Relat. Areas Mol. Biol. Vol. 67* **1993**, 77–264.

(215) Mitchell, J. J.; Trakadis, Y. J.; Scriver, C. R. Phenylalanine Hydroxylase Deficiency. *Genet. Med.* **2011**, *13* (8), 697.

(216) Camp, K. M.; Parisi, M. A.; Acosta, P. B.; Berry, G. T.; Bilder, D. A.; Blau, N.; Bodamer, O. A.; Brosco, J. P.; Brown, C. S.; Burlina, A. B. Phenylketonuria Scientific Review Conference: State of the Science and Future Research Needs. *Mol. Genet. Metab.* **2014**, *112* (2), 87–122.

(217) Okano, Y.; Wang, T.; Eisensmith, R. C.; Steinmann, B.; Gitzelmann, R.; Woo, S. L. Missense Mutations Associated with RFLP Haplotypes 1 and 4 of the Human Phenylalanine Hydroxylase Gene. *Am. J. Hum. Genet.* **1990**, *46* (1), 18.

(218) Byck, S.; Tyfield, L.; Carter, K.; Scriver, C. R. Prediction of Multiple Hypermutable Codons in the Human PAH Gene: Codon 280 Contains Recurrent Mutations in Quebec and Other Populations. *Hum. Mutat.* **1997**, *9* (4), 316–321.

(219) Andersen, O. A.; Flatmark, T.; Hough, E. High Resolution Crystal Structures of the Catalytic Domain of Human Phenylalanine Hydroxylase in Its Catalytically Active Fe (II) Form and Binary Complex with Tetrahydrobiopterin1. *J. Mol. Biol.* **2001**, *314* (2), 279–291.

(220) Andersen, O. A.; Flatmark, T.; Hough, E. Crystal Structure of the Ternary Complex of the Catalytic Domain of Human Phenylalanine Hydroxylase with Tetrahydrobiopterin and 3-(2-Thienyl)-L-Alanine, and Its Implications for the Mechanism of Catalysis and Substrate Activation. *J. Mol. Biol.* **2002**, *320* (5), 1095–1108.

(221) Lange, S. J.; Miyake, H.; Que, L. Evidence for a Nonheme Fe(IV) O Species in the Intramolecular Hydroxylation of a Phenyl Moiety. *J. Am. Chem. Soc.* **1999**, *121* (26), 6330–6331.

(222) Panay, A. J.; Lee, M.; Krebs, C.; Bollinger Jr, J. M.; Fitzpatrick, P. F. Evidence for a High-Spin Fe(IV) Species in the Catalytic Cycle of a Bacterial Phenylalanine Hydroxylase. *Biochemistry* **2011**, *50* (11), 1928–1933.

(223) Kemsley, J. N.; Wasinger, E. C.; Datta, S.; Mitić, N.; Acharya, T.; Hedman, B.; Caradonna, J. P.; Hodgson, K. O.; Solomon, E. I. Spectroscopic and Kinetic Studies of PKU-Inducing Mutants of Phenylalanine Hydroxylase: Arg158Gln and Glu280Lys. *J. Am. Chem. Soc.* **2003**, *125* (19), 5677–5686.

(224) McKinney, J.; Teigen, K.; Frøystein, N. Å.; Salaün, C.; Knappskog, P. M.; Haavik, J.; Martinez, A. Conformation of the Substrate and Pterin Cofactor Bound to Human Tryptophan Hydroxylase. Important Role of Phe313 in Substrate Specificity. *Biochemistry* **2001**, *40* (51), 15591–15601.

(225) Fitzpatrick, P. F. Steady-State Kinetic Mechanism of Rat Tyrosine Hydroxylase. *Biochemistry* **1991**, *30* (15), 3658–3662.

(226) Chow, M. S.; Eser, B. E.; Wilson, S. A.; Hodgson, K. O.; Hedman, B.; Fitzpatrick, P. F.; Solomon, E. I. Spectroscopy and Kinetics of Wild-Type and Mutant Tyrosine Hydroxylase: Mechanistic Insight into O2 Activation. *J. Am. Chem. Soc.* **2009**, *131* (22), 7685–7698.

(227) Bassan, A.; Blomberg, M. R. A.; Siegbahn, P. E. M. Mechanism of Dioxygen Cleavage in Tetrahydrobiopterin-Dependent Amino Acid Hydroxylases. *Chem. - A Eur. J.* **2003**, *9* (1), 106–115.

(228) Proctor, E. A.; Kota, P.; Aleksandrov, A. A.; He, L.; Riordan, J. R.; Dokholyan, N. V. Rational Coupled Dynamics Network Manipulation Rescues Disease-Relevant Mutant

Cystic Fibrosis Transmembrane Conductance Regulator. *Chem. Sci.* **2015**, *6* (2), 1237–1246.

(229) Dokholyan, N. V. Controlling Allosteric Networks in Proteins. *Chem. Rev.* **2016**, *116* (11), 6463–6487.

(230) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(231) Goedert, M.; Spillantini, M. G.; Jakes, R.; Rutherford, D.; Crowther, R. A. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. *Neuron* **1989**, *3* (4), 519–526.

(232) Imahori, K.; Uchida, T. Physiology and Pathology of Tau Protein Kinases in Relation to Alzheimer's Disease. *J. Biochem.* **1997**, *121* (2), 179–188.

(233) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. *PLoS Biol* **2009**, *7* (2), e1000034.

(234) Teigen, K.; Jensen, V. R.; Martinez, A. The Reaction Mechanism of Phenylalanine Hydroxylase. -- A Question of Coordination. *Pteridines* **2005**, *16* (1), 27–34.

(235) Shiota, Y.; Yoshizawa, K. QM/MM Study of the Mononuclear Non-Heme Iron Active Site of Phenylalanine Hydroxylase. *J. Phys. Chem. B* **2004**, *108* (44), 17226–17237.

(236) Gottschall, D. W.; Dietrich, R. F.; Benkovic, S. J.; Shiman, R. Phenylalanine Hydroxylase. Correlation of the Iron Content with Activity and the Preparation and Reconstitution of the Apoenzyme. *J. Biol. Chem.* **1982**, *257* (2), 845–849.