

CHARACTERIZATION OF ALCOHOL DEHYDROGENASE YQHD, AND EXPLORATION
OF THE ORIGINS OF FLUORINE NUCLEAR MAGNETIC RESONANCE CHEMICAL
SHIFTS IN PROTEINS

A Thesis by

Jonathan M. Ellis

Bachelor of Science, Wichita State University, 2014

Submitted to the Department of Chemistry
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

July 2018

© Copyright 2018 by Jonathan M. Ellis
All Rights Reserved

CHARACTERIZATION OF ALCOHOL DEHYDROGENASE YQHD, AND EXPLORATION
OF THE ORIGINS OF FLUORINE NUCLEAR MAGNETIC RESONANCE CHEMICAL
SHIFTS IN PROTEINS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science, with a major in Chemistry.

Katie Mitchell-Koch, Committee Chair

James Bann, Committee Member

Douglas English, Committee Member

Moriah Beck, Committee Member

William Hendry, Committee Member

Remember to look up at the stars and not down at your feet. Try to make sense of what you see and wonder about what makes the universe exist.

Be curious.

-Stephen Hawking

ACKNOWLEDGEMENTS

I would like to thank my family, including my parents, for supporting me in my pursuits of education and chemical inquiry. I would not be where I am today without their dedication, support, and love. I would like to thank my lab mates Dr. Rajni Verma, Dr. Jayangika Dehanayake, Dr. Chandana Kasireddy, Dr. Sreesha Mamillapalli, and Ryan Steinert for their wealth of knowledge and unique perspectives on all my research conducted, and for keeping me sane during the graduate school process. I would like to thank my friend Bryan Gant for his incredible friendship, sharing his wealth of programming knowledge, teaching me everything I know about CUDA, and for co-authoring the PFNMR program. I would also like to thank my advisors, Dr. Katie Mitchell-Koch and Dr. James Bann, for welcoming me into their lab groups and for instilling upon me a love and passion for the pursuit of chemical knowledge, both theoretical and biophysical.

I would like to thank my committee members, including Dr. Moriah Beck, Dr. Douglas English, Dr. William Hendry, Dr. Katie Mitchell-Koch, and Dr. James Bann. I would like to thank Dr. Kevin Langenwalter for his assistance and tutelage in the use of all the instrumentation available at Wichita State. I would like to thank Mary Cambridge and Susan McCoy for their assistance in acquiring research resources, and for the wealth of knowledge in teaching undergraduate laboratory courses. I would also like to thank Dr. Scott Lovell at University of Kansas for his efforts and success in crystallizing YqhD and producing a novel crystal structure solution.

My deepest thanks goes out to our funding sources which enabled this research, including Wichita State University and grants from the NSF (EIA-0216178, EPS-0236913), KINBRE (P20 GM103418), and COBRE-PSF (P30 GM110761).

ABSTRACT

Alcohol Dehydrogenase YqhD acts as a useful enzyme for the purpose of conversion of alcohols to aldehydes, and visa-versa. It acts on a wide range of substrates, and as such provides a potentially useful platform for production of a variety of compounds used in synthesis, biofuels, and more. To aid in future engineering efforts, biophysical characterization of YqhD was conducted, including exploration of the role of the metal center by Michaelis-Menten kinetics and FRET-based analysis of cofactor binding. To help validate previously conducted molecular dynamics simulations, ¹⁹F-NMR was employed to observe potential changes in protein structure and dynamics upon binding of zinc, cofactor, and substrate. However, this labeling technique suffers from so-called “The Assignment Problem” of NMR, which describes determining which resonance in a NMR spectrum corresponds to which residue in a labeled protein. In an effort to expand techniques available to those interested in ¹⁹F-NMR, a site-specific labeling technique was further developed and shown to be usable for a single resonance assignment of the acquired YqhD spectra.

One of the “holy grails” of computation chemistry in regards to NMR is to perform accurate assignments of NMR chemical shifts in macromolecular systems. To this end, a reasonably exhaustive set of QM calculation methods were tested to determine their accuracy in predicting the ¹⁹F-NMR spectra of small molecule sidechain analogues of commonly employed labeling amino acids. This information was then used to generate a large dataset describing the influence of dielectric and external electric fields on a ¹⁹F-NMR resonance. With this information, the PFNMR software package was developed in an effort to perform resonance assignments on the IFABP and PapD proteins in a fully computational fashion, with promising initial success.

PREFACE

This thesis is comprised of three unique projects in the realms of biophysics, computational chemistry, and programming. While more work still needs to be done for publication of the biophysical characterization of YqhD and development of the Protein Fluorine NMR (PFNMR) package, the computational characterization and prediction of fluorine NMR chemical shifts of small molecules has been published. The first of these is *The Biophysical Probes 2-fluorohistidine and 4-fluorohistidine: Spectroscopic Signatures and Molecular Properties*, Sci Rep. 2017 Feb 15; 7:42651. The second is *Tautomeric stabilities of 4-fluorohistidine shed new light on mechanistic experiments with labeled ribonuclease A*, Chem Phys Lett. 2016 Dec 1; 666:58-61. The third is *Evaluating electronic structure methods for accurate calculation of ¹⁹F chemical shifts in fluorinated amino acids*, J Comp. Chem. 2017 Nov 15;38(30):2605-2617. The final is a chapter published in Annual Reports on NMR Spectroscopy, entitled *Progress in Our Understanding of ¹⁹F Chemical Shifts*, 2017 Oct 21; 93:281-365.

TABLE OF CONTENTS

Chapter		Page
1.	INTRODUCION.....	1
1.1	Background on Alcohol Dehydrogenase YqhD	1
1.2	Computational Methods and Software Development	3
2.	BIOPHYSICAL CHARACTERIZATION OF YQHD	6
2.1	Experimental	6
2.2	Generation of an Apo Crystal Structure.....	11
2.3	ICP-OES Analysis of Zinc Occupancy.....	14
2.4	Kinetic Characterization	16
2.5	FRET-based Cofactor Binding Kinetics	19
2.6	Conclusions.....	24
3.	5F-TRP LABELING OF YQHD	26
3.1	Experimental	26
3.2	Uniform Labeling of YqhD	28
3.3	Site-Specific Labeling of YqhD.....	34
3.4	Conclusions.....	37
4.	COMPUTATIONAL STUDIES OF SMALL MOLECULE NMR PROPERTIES.....	39
4.1	Methodology	39
4.2	Analysis of Computational Methodology	40
4.3	Simulated Deprotonation of Fluorinated Hydrocarbons.....	42
4.4	Electric Field Effect Results	52
4.5	Conclusions.....	58
5.	PRODUCTION OF THE PFNMR SOFTWARE PACKAGE.....	59
5.1	Background	59
5.2	Methodology	60
5.3	Dielectric Environment Results	64
5.4	Preliminary Electric Field and NMR Prediction Results	66
5.5	Conclusions.....	70
6.	FUTURE DIRECTION	71
6.1	Exploration of the Fundamental Origins of ^{19}F -NMR Chemical Shifts	71
6.2	Generation of Diverse Datasets for Electric Field Analysis	71
6.3	Expansion of the Biochemical NMR “Toolbox”	72

TABLE OF CONTENTS (continued)

Chapter	Page
REFERENCES	73

LIST OF TABLES

	Page
Table 1. Crystallographic data of fully apo YqhD structure.....	12
Table 2. Metal occupancy data by ICP-OES analysis	16
Table 3. Kinetic parameterization of YqhD with butyraldehyde as a substrate	19
Table 4. Comparison of kinetic parameters for unlabeled (WT) and labeled (5FW) YqhD	34
Table 5. Error in ppm of various ¹⁹ F-NMR calculation methods.....	41
Table 6. Average error, in ppm, of NMR calculation methods with various solvents	42
Table 7. Change in ¹⁹ F-NMR chemical shifts (in ppm) through various deprotonation events.....	47

LIST OF FIGURES

Figure	Page
1. Examples of fluorinated molecules that exhibit reverse chemical shift behavior	4
2. Comparison of electron density maps for the holo and apo crystal structure metal binding pockets of YqhD	13
3. Structural alignment of the apo and holo YqhD Crystal Structures	14
4. Measurement of zinc occupancy of freshly purified YqhD by ICP-OES spectroscopy....	15
5. Michaelis-Menten kinetic analysis of YqhD converting butyraldehyde to butanol by measuring consumption of NADPH	17
6. Preliminary evidence of catalysis with non-zinc metals bound to YqhD.....	18
7. NADPH binding affinity analysis by FRET	21
8. NADP ⁺ binding affinity by FRET-based competitive binding assay	22
9. Cofactor-dependent distribution analysis of binding cleft distances by MD.....	23
10. Structure of YqhD showing 5-fluorotryptophan positions	29
11. ¹⁹ F-NMR comparison of 5-fluorotryptophan labeled YqhD in various conditions.....	30
12. Structure of cofactor/metal binding pocket depicting the close cofactor contact W264 and cofactor binding loop W117.....	31
13. ¹⁹ F-NMR of fully labeled YqhD incubated with 50 eq. butyraldehyde overnight	32
14. Thermal denaturation curves of labeled and unlabeled YqhD by CD spectroscopy	33
15. ¹⁹ F-NMR comparison of site-specifically labeled YqhD to fully labeled YqhD	35
16. ¹⁹ F-NMR comparison of site-specifically labeled YqhD and fully labeled YqhD in the presence of 5 eq. NADP ⁺	36
17. Depiction of the different tautomeric states of fluorohistidines, as well as naming conventions for each heavy atom.....	43
18. Energies and ¹⁹ F-NMR values of 2F-MetImid, 4F-MetImid, and 2,4F-MetImid during simulated deprotonation	45

LIST OF FIGURES (continued)

Figure		Page
19.	Boltzmann factor analysis of 2F-MetImid, 4F-MetImid, and 2,4F-MetImid during simulated deprotonation.....	48
20.	Contributions of N1 to the electron density of fluorines in 24F-MetImid.....	49
21.	Sample frame of a molecular orbital evolution animation for the deprotonation of 4F-MetImid	50
22.	Evolution of MO-20, MO-23, and MO-24 of 4F-MetImid from the cation to π -tautomer.....	51
23.	Predicted influence of applied electric field on 19F-NMR chemical shift of fluorobenzene, with a depiction of the x-axis aligned structure	54
24.	Left: C-F region of predicted IR spectra of fluorobenzene.....	56
25.	Histogram analysis of the 4-fluorotoluene relative shift perturbation dataset	57
26.	Illustration comparing integration method distance fittings	63
27.	Dielectric environment visualization of YqhD	64
28.	Dielectric environment visualization of GroES	65
29.	Complexity v.s error curve from Eureqa fitting data	66
30.	Comparison of IFABP experimental 19F-NMR spectrum with optimized simple and complex calculation methods.....	68
31.	Comparison of PapD experimental 19F-NMR spectrum to a PFNMR predicted spectrum with parameters optimized from IFABP	69

LIST OF ABBREVIATIONS

NMR	Nuclear Magnetic Resonance
HSQC	Heteronuclear Single Quantum Coherence experiment
NOE	Nuclear Overhauser Effect
FID	Free Induction Decay
TFA	Trifluoroacetic Acid
QM	Quantum Mechanical
MM	Molecular Modeling
DFT	Density Functional Theory
MO	Molecular Orbital
PFNMR	Protein Fluorine NMR prediction software package
GPU	Graphics Processing Unit
VSE	Vibrational Stark Effect spectroscopy
LB	Lysogeny Broth
PCR	Polymerase Chain Reaction
DMF	N,N-Dimethylformamide
FPLC	Fast Protein Liquid Chromatography
OE-PCR	Overlap Extension Polymerase Chain Reaction
ICP-OES	Inductively Coupled Plasma Optical Emission Spectroscopy
FRET	Förster Resonance Energy Transfer
EPR	Electron Paramagnetic Resonance spectroscopy
RMSD	Root Mean Square Deviation
PDB	Protein Data Bank

LIST OF ABBREVIATIONS (continued)

yPheRS	Yeast Phenylalanine tRNA Synthetase
DHFR	Dihydrofolate Reductase
IFABP	Intestinal Fatty Acid Binding Protein
BONCAT	Bio-Orthogonal Non-Canonical Amino Acid Tagging

LIST OF SYMBOLS

Int	Intensity
eq.	Equivalents
mAU	milli-Absorbance Units
Abs	Absorbance
a.u	Atomic or Arbitrary units
V/m	Volts per Meter
ppm	Parts per million
rpm	Revolutions per minute
RMS	Root Mean Square power
K _M	Michaelis constant
V _{max}	Maximum reaction velocity
k _{on}	Rate of binding
k _{off}	Rate of release
K _d	Dissociation constant
K _H	Hill constant
n	Hill coefficient

CHAPTER 1

INTRODUCTION

1.1 Background on Alcohol Dehydrogenase YqhD

YqhD is an alcohol/aldehyde oxidoreductase found in yeast which is able to convert a diverse array of alcohols into aldehydes, and visa-versa. It shows reasonable activity against primary alcohols and aldehydes greater than two carbons in length¹, and as such has promising applications for biofuel and commodity production. Its high degree of expression (~ 80 mg/L LB) also makes it an interesting candidate for potential large-scale bioreactor applications, as a large amount of catalyst is readily prepared and easily purified with little overall effort. These properties were the primary reasons that YqhD was chosen for study, in particular for future efforts of protein engineering as well as initial biophysical characterization for the purpose of validating computational simulations carried out previously by the Mitchell-Koch lab. Efforts have already been made in order to evolve YqhD to exhibit a greater catalytic activity towards 1,3-propanediol², an important industry precursor for polymer synthesis. The large degree of protein expression also makes YqhD a great candidate for Nuclear Magnetic Resonance (NMR) spectroscopy based studies, where protein concentrations on the order of high μM to mM range are necessary when a cryoprobe is not available. In order to work around some of the difficulties associated with instrumentation limitations, ¹⁹F-NMR was chosen for studying how the structural dynamics of YqhD change upon cofactor, metal, and substrate binding. ¹⁹F-NMR is a well-documented method for studying protein structure and dynamics^{3–6}, and affords many advantages compared to more traditional biomolecular NMR techniques. For one, canonical amino acids do not contain fluorine, so there is no background signal contamination from outside sources, and the ¹⁹F-nucleus is 100% naturally abundant. There is also no need for solvent suppression methods, so experiments can be carried out in H₂O. Because labeling only occurs at

specific amino acid sites (for example, if 4-fluorophenylalanine is used, the 8 phenylalanine sites in Intenstinal Fatty Acid Binding Protein would be labeled), ¹⁹F-NMR spectra are much less cluttered than standard ¹H spectra and cover a wide spectral range (800 ppm). This allows for simple 1D-NMR experiments to provide a rich amount of information, without the need for more complex multidimensional NMR techniques. However, these techniques (Such as HSQC, NOE, etc.) can still be employed if a high degree of overlap is observed between resonances, or if more information is needed.

As with all NMR based techniques, the issue of the so-called “Assignment Problem” arises. This problem describes the fact that, from a single spectrum alone, it is difficult to attribute which resonance belongs to which labeled amino acid in a protein. This is typically addressed by performing site-directed mutagenesis, for example swapping a previously labeled tryptophan for phenylalanine, determining which resonance is absent in the mutant protein spectrum, and attributing the missing resonance to the mutated residue. However, mutations are not always well tolerated, and protein misfolding can occur. An additional challenge to using fluorinated amino acids is, while a proton to fluorine swap is sterically innocuous in nature (the C-F bond is ~1.5 times larger than the C-H bond), the same swap is not at all similar from an electronics perspective. Fluorine is a much more electronegative atom compared to hydrogen, and certain fluorotryptophans (6-fluorotryptophan, in particular) are not well tolerated, resulting in a decrease in protein stability⁷. To address both of these issues simultaneously, efforts were made to develop a site-specific incorporation method for fluorotryptophans. Previous studies indicated that it was possible to site-specifically incorporate tryptophan analogues such as chloro and bromotryptophan, but fluorotryptophan was not reported⁸. Contained within this report are, to our knowledge, the first example of the site-specific incorporation of 5-fluorotryptophan for

the purpose of resonance assignment in a global labeling experiment. Efforts were also made to determine how labeling effects protein dynamics and catalytic activity of YqhD. It is our hope that these initial studies will lay the foundation for a more precise protein engineering effort of YqhD, as well as providing new and useful tools for studying proteins using ¹⁹F-NMR.

1.2 Computational Methods and Software Development

Fluorinated amino acids provide a rich and diverse wealth of labeling molecules for the purposes of ¹⁹F-NMR. However, assignment of resonances in fluorinated biomolecules and determining implications of chemical shift changes upon binding of docked fluorinated drug molecules can be cumbersome in nature. In particular, resonance assignments of fluorine-labeled proteins requires the use of site-directed mutagenesis techniques, requiring time (~1 week per resonance assignment) and money (~\$40 per assignment) without the guarantee of a successful assignment if a labeling mutation results in protein misfolding. In an effort to develop computational techniques to help understand the implications of various fluorine chemical shift perturbations, a fairly exhaustive study was conducted to test the accuracy of QM techniques for calculating ¹⁹F-NMR spectra. All QM studies were conducted in the Gaussian 09 software suite, where a wide variety of calculation methods, DFT methods, and basis sets were tested to see which combinations were the best at predicting experimentally observed NMR spectra. In addition, some fluorinated amino acids and molecules exhibit interesting properties that are not easy to tease out from an experimental perspective. Of interest to the Mitchell-Koch lab are molecules that exhibit so-called “reverse chemical shift behavior,” where an event that should cause shielding of the fluorine nuclei results in deshielding.

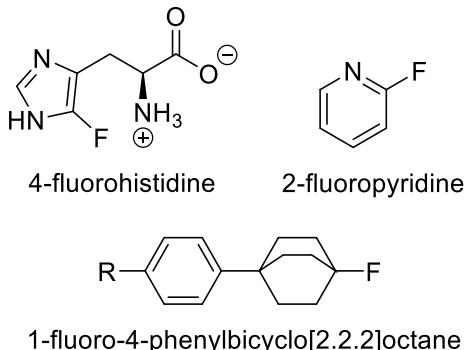


Figure 1: Examples of fluorinated molecules that exhibit reverse chemical shift behavior

An archetypical example of this is 4-fluoro-L-histidine, which exhibits deshielding behavior upon deprotonation of the imidazole ring by pH titration⁹. While previous efforts had been made within the group to explain this behavior, more molecules have since been found in literature that exhibit a reverse chemical shift behavior, such as 2-fluoropyridine¹⁰ and various substituted derivatives of 1-fluoro-4-phenylbicyclo[2.2.2]octanes¹¹. In order to explore the implications of the molecular structures on the chemical shift behavior of the fluorine nuclei, tools were developed to help visualize how the molecular orbitals of molecules “evolve” under a given change, in this case protonation/deprotonation. Fluorohistidines were chosen as test molecules for this study, as they were the most well understood candidates at the time. 2,4-difluoro-L-histidine was also chosen to see if both normal and reverse chemical shift behavior are predicted to be observed within the same system, as previous experimental studies indicated that 2-fluoro-L-histidine exhibits normal chemical shift behavior. These visualization tools provide an intuitive method for studying how electron density changes upon simple molecular changes to a given system.

While traditional computational methods could potentially afford chemical shift assignments, they are not amiable to increasingly large molecules. Even combined QM:MM techniques such as ONIOM, which allows for approximation of distant atomic interactions as purely classical in nature¹², can require unreasonable computational cost and time for particularly

large systems. With the rise of GPU-based techniques in computer science (such as NVIDIA’s CUDA programming language¹³) comes the possibilities of performing massive quantities of calculations in parallel, affording alluring improvements in computational speed and subsequent reductions in computational cost. With this in mind, work has been conducted to construct a rapid ¹⁹F-NMR resonance program leveraging CUDA-based programming techniques. These techniques allow for rapid calculations of parameters such as non-uniform dielectrics throughout proteins, as described in prior literature¹⁴. As more experimental data probing the electric environment of protein interiors, as well as more ¹⁹F-NMR spectra of labeled proteins, the program can be modified and expanded to become a versatile and useful tool for biophysical characterizations and predictions of proteins.

CHAPTER 2

BIOPHYSICAL CHARACTERIZATION OF YQHD

2.1 Experimental

Production of DNA and Protein Expression Plasmids of YqhD:

The original YqhD plasmid was a generous gift from Prof. Christian Cambillau (Marseille University/CNRS, France). The YqhD gene in this plasmid was PCR amplified out and subcloned into the PCR-XL-TOPO vector (Invitrogen) for sequence validation and future experimentation. This plasmid was then used to subclone the YqhD gene into pASK-IBA37plus (Invitrogen) using the BsaI restriction enzyme (NEB). This plasmid, denoted as pASK-YqhD, encodes for a 6x His-Tagged version of YqhD with a Factor Xa proteolysis site for tag removal.

Expression and Isolation of Wild-Type YqhD:

The pASK-YqhD vector was transformed into competent TunerTM (DE3)pLysS cells (Novagen) using the standard heat-shock transformation technique. An isolated colony of these cells (either from a fresh transformation, or from a re-plating of a glycerol stock) was used to inoculate 50 mL of LB broth with 100 mg/L ampicillin. This flask was shaken in an incubator overnight at 37°C. 10 mL of the overnight stock was used to inoculate 1 L of LB broth with 100 mg/L ampicillin in a 2.8 L Fernbach flask, up to 4 flasks per growth. These flasks were allowed to shake at 37°C until an OD₆₀₀ of 0.5 was achieved. Each flask was then induced for protein expression by addition of 100 uL of a 2 mg/mL anhydrotetracycline in DMF solution. The induced flasks were shaken at 37°C for 4-5 hours to permit adequate expression of YqhD. After this period, the cells were harvested by centrifugation at 3500 rpm and 4 °C. The harvested pellets were resuspended in 15 mL of His-column binding buffer (20 mM sodium phosphate buffer, 0.5 M sodium chloride, 5 mM imidazole, pH 7.4) per liter of growth. The resuspended

pellets were then frozen overnight. On the next day, the frozen pellets were thawed and sonicated for 1 hour (~20W RMS power, 5 second on/off cycles) to free the expressed protein from the cells. The sonicated slurry was then spun down in an ultracentrifuge at 20,000 rpm and 4°C for 45 minutes. While this was occurring, an AKTA FPLC purification system was washed, using the standard built-in system wash program, using His-column binding and elution buffer (binding buffer as above, elution buffer consisting of binding buffer containing 0.5 M imidazole). 5 mL HiTrap IMAC FF columns (GE Lifesciences, 1 per liter growth) charged with Ni were attached to the AKTA system and washed with 10 column volumes of binding buffer. Once the previous centrifugation has completed, the supernatant was isolated and loaded onto the columns using a peristaltic pump. The loaded columns were then re-attached to the AKTA system, and washed with binding buffer until an $\text{Abs}_{280} < 100$ mAU was achieved. A gradient elution (250 mL, 0% to 100% elution buffer) was then initiated, and fractionation (5 mL fractions) was started. The gradient elution was allowed to proceed until all apparent protein had eluted from the column. Relevant fractions were then pooled, concentrated to a volume no greater than 30 mL, and transferred to a dialysis bag for buffer exchange into a storage buffer containing 50 mM potassium phosphate buffer and 0.15 M sodium chloride at pH 7.5.

Generation of pASK-TEV-IBA37plus plasmid:

Overlap Extension PCR (OE-PCR) was employed to swap the Factor Xa cleavage site in pASK-IBA37plus for a TEV cleavage site. The following primers were used for the OE-PCR process: Primer A-CCATCGAACATGGCCAGATG, Primer B-GGATTGGAAGTACAGGTTCTCGTGATGGTGATGGTGATGCG, Primer C-GAGAACCTGTACTCCAATCCGAGACCGCGGTCCCC, Primer D-CCCTTATAAATCAAAAG. pASK-IBA37plus was incubated with both Primer A + Primer B

and Primer C + Primer D in two separate reactions, using the standard PrimeSTAR HS DNA polymerase (Takara) protocol for amplification. The amplified fragments from both of these PCR reactions were then purified via agarose gel electrophoresis, extracted using a Qiagen gel extraction kit, added to one another in a 1:1 ratio, and PCRed to form the complete replacement fragment. Both the replacement fragment and pASK-IBA37plus were then incubated with XbaI (NEB) and HindIII (NEB), gel purified, and ligated together to form the complete pASK-TEV-IBA37plus plasmid, hereby referred to as pASKT.

Crystallization of YqhD:

All crystallization and structure solution efforts were carried out by Dr. Scott Lovell through a COBRE Protein Structure Function Center grant at University of Kansas. What follows was copied from the crystallization report (unpublished) generated by Dr. Scott Lovell of the COBRE Protein Structure Laboratory:

Purified YqhD was concentrated to 15 mg/mL in 150mM NaCl, 50mM Tris 7.5 for crystallization screening. All crystallization experiments were conducted Compact Jr. (Rigaku Reagents) sitting drop vapor diffusion plates at 18 °C using equal volumes (0.7 μL) of protein and crystallization solution equilibrated against 75 μL of the latter. Plate clusters were initially obtained in 2-3 days from the Crystal Screen HT screen (Hampton Research) condition D4 (20% (w/v) PEG 4000, 20% (v/v) 2-propanol, 100 mM sodium citrate pH 5.6). Subsequent reproduction of the initial crystals using 2 μL of the Crystal Screen HT screen D4 condition in Clover Jr. plates (Rigaku Reagents) and 2 μL of protein equilibrated against 100 μL of reservoir solution, yielded well-defined single plates. Cryoprotection was carried out by transferring crystals to solutions containing 80% crystallant and 20% glycerol and storing in liquid nitrogen for data collection. X-

ray diffraction data were collected at the Advanced Photon Source IMCA-CAT beamline 17-ID using a Dectris Pilatus 6M pixel array detector.

Analysis of Metal Occupancy by ICP-OES:

Samples were prepared in 50 mM potassium phosphate buffer and 0.15 M sodium chloride at pH 7.5. All YqhD used was prepared from *E. coli* bearing the pASKT-YqhD plasmid using the standard purification procedure above. TEV protease bearing a 6xHis purification tag was used to cleave off the 6xHis tag. This was done by incubating a purified batch of YqhD with 50x concentration TEV protease overnight at 4 °C. Untagged protein was then purified by passing the cleaved mixture through 5 mL HiTrap IMAC FF columns and collecting the flow-through. This stock was then divided into 5 different test sets. For one set, nothing was done in addition to the purification. For the other sets, purified protein was incubated with 5 equivalents (with respect to enzyme concentration) β-mercaptoethanol, followed by 5x equivalents of a 1M divalent metal salt (ZnCl₂, NiSO₄, CoCl₂, or CuSO₄). Unbound metal was removed by centrifuge buffer exchange with sample buffer. Samples of the final flow-through were kept and analyzed in the same fashion as the protein samples in order to validate no excess metal was left in solution. 5 mL ICP-OES samples were prepared at a concentration of ~ 5 μM YqhD, and verified by measuring protein absorbance at 290 nm. A standard addition curve was created by spiking in 2, 4, 6, and 8 μM of a given metal before diluting to 5 mL. The quantity of bound metal was calculated by extrapolating and solving for the x-intercept of each curve, and the percentage occupancy was found by dividing this value by the concentration of YqhD measured by absorption spectroscopy. All ICP-OES measurements were carried out on a Varian 720 ICP-OES spectrometer using standard acquisition procedures for the instrument.

Absorbance Based Kinetic Characterization of YqhD:

For absorbance measurements, all experiments were carried out on either a Cary UV-vis Spectrometer or an Applied Photophysics SX-20 Stopped Flow apparatus in 1:10 mixing mode and equipped with a single source monochromator and absorbance detector. All reactions were carried out at 25 °C. The enzymatic reaction consisted of 50 mM Tris buffer, pH 7.0, with 100 mM NaCl. All YqhD stocks were treated with 5 equivalents (with respect to enzyme concentration) of EDTA to remove any non-specifically bound metals, followed by centrifuge filtration to remove excess EDTA and any chelated metals. Supplemental metals were introduced to purified stocks of YqhD by first treating the solution with 5 equivalents of β-mercaptoethanol, followed by addition of 5 equivalents of a 1M metal salt solution. For enzyme kinetics studies, 100 μM NADPH was present in the reaction buffer along with varying concentrations of butyraldehyde. 0.25 μM of YqhD was added in order to initiate catalysis, and the conversion of NADPH to NADP⁺ was monitored by measuring the decrease of absorbance at 340 nm, attributed to the presence of NADPH.

FRET-based Binding Analysis of NADPH and NADP⁺:

All fluorescence experiments were carried out on a Cary Eclipse Fluorescence Spectrometer. For NADPH binding studies, enzyme reaction buffer (as described above) containing 1 μM YqhD and 5 eq. ZnCl₂ was titrated with varying concentrations of NADPH. FRET emission was measured by exciting tryptophan at 295 nm and measuring the total fluorescence emission past 450 nm. For NADP⁺ binding studies, a competition assay was done by incubating 1 μM YqhD with 5 μM of NADPH for 1 hour to achieve saturation. Various quantities NADP⁺ were then titrated in, and the FRET emission of the remaining bound NADPH was observed as in the NADPH binding assay.

2.2 Generation of an Apo Crystal Structure

Through a NIH-COBRE grant, a newly solved crystal structure of fully apo YqhD was obtained. All crystallography efforts were carried out at The University of Kansas by Dr. Scott Lovell. Surprisingly, the structure solved showed little to no electron density in the metal binding center for either subunit.

Table 1: Crystallographic data of fully apo YqhD structure
(Credit Dr. Scott Lovell, University of Kansas)

	YqhD
Data Collection	
Unit-cell parameters (\AA , $^\circ$)	$a=60.45, b=135.74, c=116.49,$ $\beta=93.1$
Space group	$P2_1$
Resolution (\AA) ¹	48.91-2.05 (2.09-2.05)
Wavelength (\AA)	1.0000
Temperature (K)	100
Observed reflections	401,248
Unique reflections	117,033
$\langle I/\sigma(I) \rangle^1$	10.4 (2.0)
Completeness (%) ¹	99.7 (99.9)
Multiplicity ¹	3.4 (3.4)
$R_{\text{merge}} (\%)^{1,2}$	9.6 (65.2)
$R_{\text{meas}} (\%)^{1,4}$	11.4 (77.3)
$R_{\text{pim}} (\%)^{1,4}$	6.1 (41.1)
$CC_{1/2}^{1,5}$	0.996 (0.695)
Refinement	
Resolution (\AA) ¹	34.25-2.05
Reflections (working/test) ¹	111,058/5,904
$R_{\text{factor}} / R_{\text{free}} (\%)^{1,3}$	15.8/19.9
No. of atoms (Protein/ Na^+ /Water)	11,596/2/887
Model Quality	
R.m.s deviations	
Bond lengths (\AA)	0.011
Bond angles ($^\circ$)	1.012
Average B -factor (\AA^2)	
All Atoms	25.1
Protein	24.6
Na^+	30.7
Water	30.8
Coordinate error(maximum likelihood) (\AA)	0.21
Ramachandran Plot	
Most favored (%)	98.2
Additionally allowed (%)	1.7

1) Values in parenthesis are for the highest resolution shell

2) $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the i th reflection and $\langle I(hkl) \rangle$ is the average intensity of all reflections with indices hkl .

3) $R_{\text{factor}} = \sum_{hkl} ||F_{\text{obs}}(hkl) - |F_{\text{calc}}(hkl)||| / \sum_{hkl} |F_{\text{obs}}(hkl)|$; R_{free} is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

4) $R_{\text{meas}} = \text{redundancy-independent (multiplicity-weighted)} R_{\text{merge}}^{15,16}$. $R_{\text{pim}} = \text{precision-indicating (multiplicity-weighted)} R_{\text{merge}}^{17,18}$.

5) $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data^{18,19}.

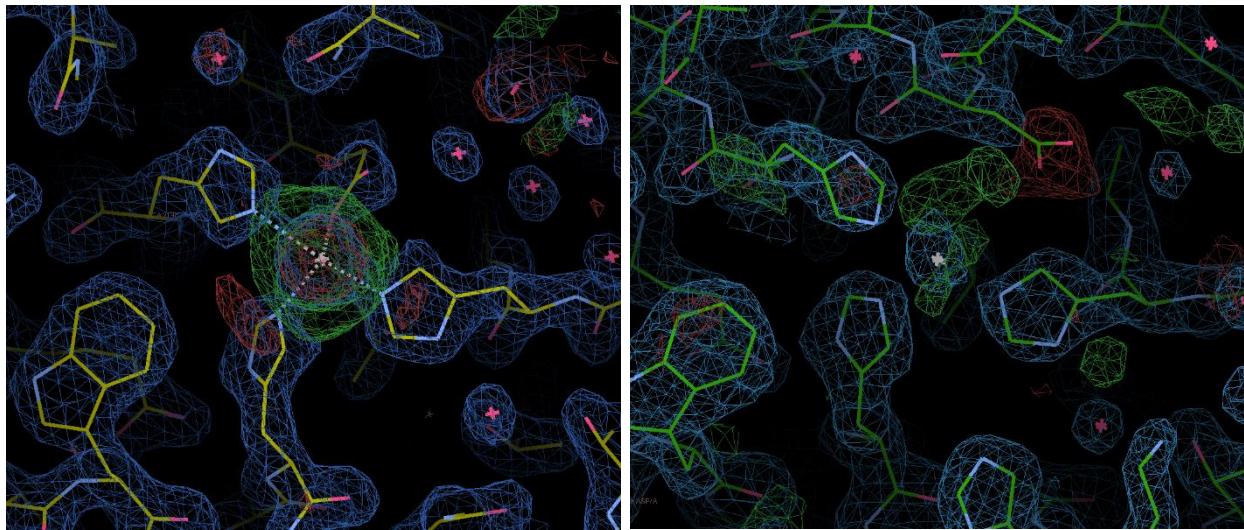


Figure 2: Comparison of electron density maps for the holo (left, PDB: 1OJ7) and apo (right) metal binding pockets (Software: WinCoot)

As the purified stock solutions had not been treated with any sort of metal chelating agent prior to shipment for crystallography, this result prompted future ICP-OES studies to determine what the percent occupancy of zinc was for purified YqhD. Aligning the newly solved apo crystal structure with the previously published holo structure modified where the zinc-bearing monomer is mirrored and replacing its metal-bound counterpart (PDB: 1OJ7) leads to an RMSD of 0.394 Å, indicating a high degree of similarity between the two structures. This provides tenuous evidence that the zinc metal center may not play a major role in the structure or conformation of YqhD, but rather serves a more catalytic role.

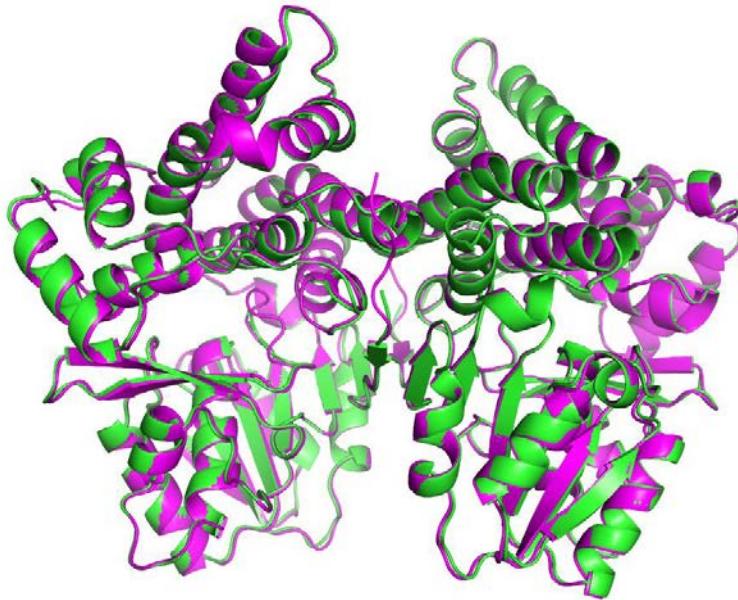


Figure 3: Structural alignment of the apo (green) and holo (magenta, PDB: 1OJ7, chains A and D)

2.3 ICP-OES Analysis of Zinc Occupancy

Analysis of the metal occupancy was conducted in order to study how much zinc metal was retained by YqhD during the purification process. This information is important, as it is crucial that the enzyme have an adequate metal occupancy for catalysis to occur. In addition, crystallography of purified protein had resulted in a solved structure bearing no significant electron density in the metal binding pocket. While this alone is strong evidence for the absence of a bound metal, it was still deemed important to validate metal binding using an alternate technique. All studies conducted for determining metal-binding occupancy by ICP-OES were done using protein which had it's polyhistidine cleaved by TEV protease, followed by re-purification and treatment with EDTA.

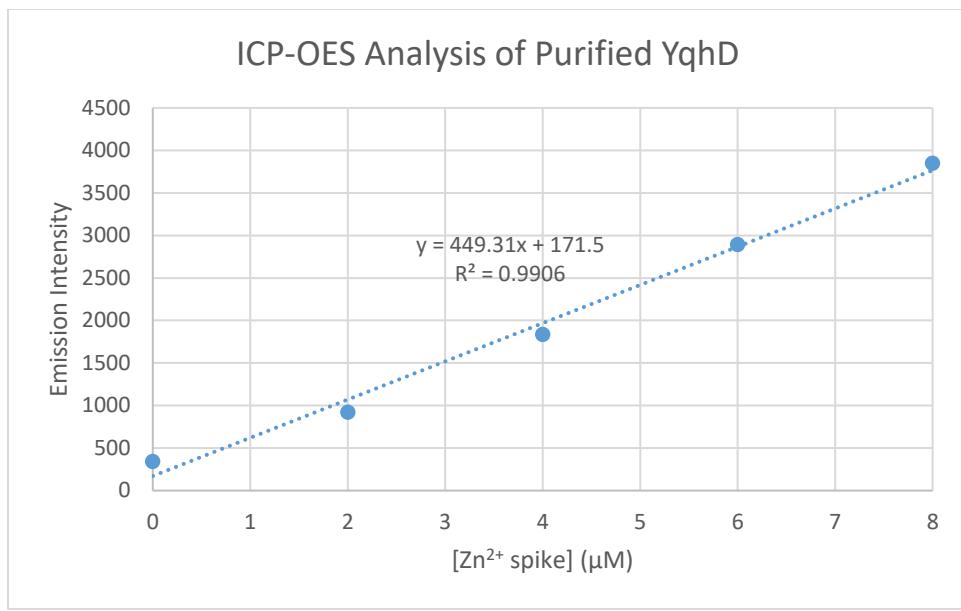


Figure 4: Measurement of zinc occupancy of freshly purified YqhD by ICP-OES spectroscopy. Standard additions were done using a 1 mg/mL ICP-OES standard ZnCl₂ solution.

Solving for the percent occupancy zinc in the untreated solution gave a value of ~7.66%, meaning that the vast majority of isolated YqhD, directly after purification, is in a fully apo state. This helps validate that supplemental zinc must be provided in order to observe proper catalysis from a purified protein stock. EDTA treated YqhD stocks were then incubated with various metals to explore if the metal coordination environment in the active site could support other metals. Care was taken to ensure that any unbound metal was filtered away, and the average protein concentration of each sample set was measured by UV-vis spectroscopy of each individual sample.

Table 2: Metal occupancy data by ICP-OES analysis

Metal Added:	Zn ²⁺	Ni ²⁺	Co ²⁺	Cu ²⁺
Slope (Int/ μ M)	604.07	102.2	220.19	304.96
Y-Intercept (Int)	1697.81	474.56	1020.9	3554.21
[M ²⁺] (μ M)	2.81	4.64	4.64	11.65
[YqhD] (μ M)	3.5	4.52	4.54	4.28
% Occupancy	80.21	102.71	102.17	272.39

Surprisingly, all 3 non-native metals chosen appear to have higher occupancy than the native zinc metal binding. This may be potentially indicative of a weaker binding interaction with zinc. The large discrepancy in copper binding is likely due to the use of an old copper metal stock solution, while the other metal solutions were spiked from a new standard metal stock purchased from a reputable supplier (Fisher Sci.) However, even overcompensating for a 25% degradation of the copper standard solution results in a 204% metal occupancy, strongly implying that there may be a non-specific metal binding location. Utilization of more sophisticated techniques such as EPR may be able to help elucidate if there is indeed a second metal binding location that could potentially be occupied by copper.

2.4 Kinetic Characterization

After establishing that supplemental zinc must be provided in order to adequately saturate the zinc-binding domain of YqhD, kinetics analysis was done using butyraldehyde and NADPH as substrates. Rate of butanol production was measured indirectly by monitoring the conversion of NADPH, which absorbs at 340 nm, to NADP⁺, which has no absorbance at 340 nm. Analysis was done on an Applied Photophysics SX20 Stopped Flow apparatus equipped with an absorbance detector in the hopes of being able to analyze burst phase kinetics. However, complications in the experiment lead to difficulties in progress curve analysis, so only the steady

state rates measured were used for kinetic parameterization in a Michaelis-Menten fashion. An initial study was done with YqhD that had been treated with EDTA in order to demonstrate the necessity of zinc for catalysis. No significant activity was observed, confirming zinc must be present for catalysis. The K_m and V_{max} for YqhD, in the presence of 5 eq. of zinc, was then measured for the same reaction.

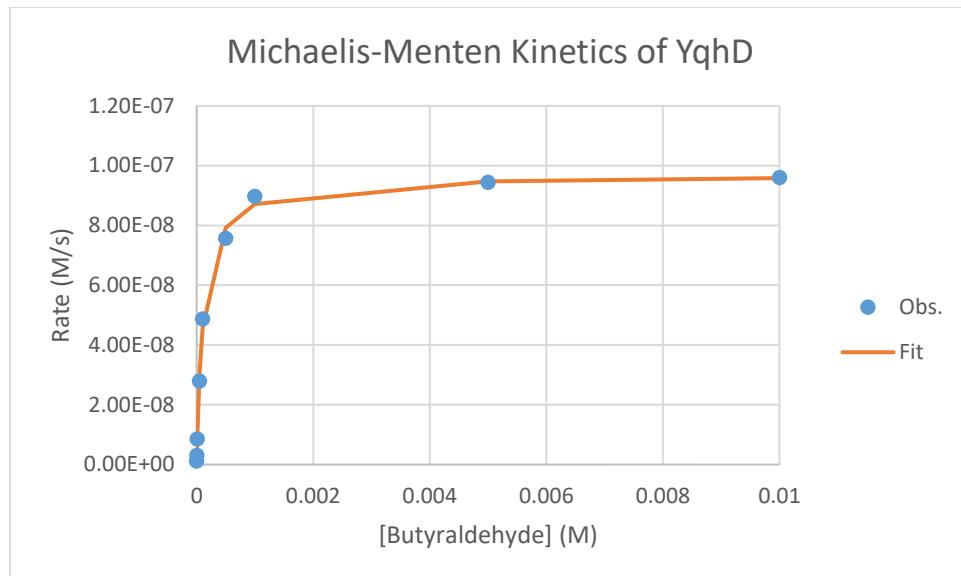


Figure 5: Michaelis-Menten kinetic analysis of YqhD converting butyraldehyde to butanol by measuring consumption of NADPH. Points in blue indicate measured initial velocity rates, while the orange trace indicates the fit of the data to a standard Michaelis-Menten curve.

It was observed that the K_m is $(1.12 \pm 0.088) * 10^{-4} \text{ M}$ with a V_{max} of $(9.69 \pm 0.15) * 10^{-6} \text{ M/s}$, with a corresponding k_{cat} of $0.39 \pm 0.006 \text{ s}^{-1}$. This yields a catalytic efficiency (k_{cat}/K_m) of $3470 \pm 328 \text{ M}^{-1} \cdot \text{s}^{-1}$. This provided a base value for measuring the effects of various perturbations. For example, the metal binding pocket, as observed by observing a previously solved X-ray crystal structure (PDB: 1OJ7), bears residues that would support binding of other various metals. In order to determine if the identity of the metal bound in the catalytic pocket plays a role in catalytic efficiency, a “quick and dirty” preliminary experiment was done on a Cary UV-vis absorbance spectrometer by doping in various divalent metals, along with Fe^{+3} . A

cuvette was initially loaded with enzymatic reaction buffer (described above) containing NADPH so that the end concentration after the final addition would be 100 μ M. A time-course measurement of the absorbance at 340 nm was started to show that no changes in absorbance were occurring due to mixing or spontaneous degradation of NADPH. A volume of butyraldehyde was then added so that the end concentration after the final addition would be 10 mM (spiking in the spectra is due to the presence of the pipette tip in the path of the light beam.) An arbitrary quantity of either metal or enzyme was then added, indicated by the label of the trace (EY indicates addition of YqhD, otherwise the metal indicated was added), in order to show that the reaction is not catalyzed without the presence of both the metal and the enzyme. The final addition was the missing component of the prior mixture, either enzyme or metal.

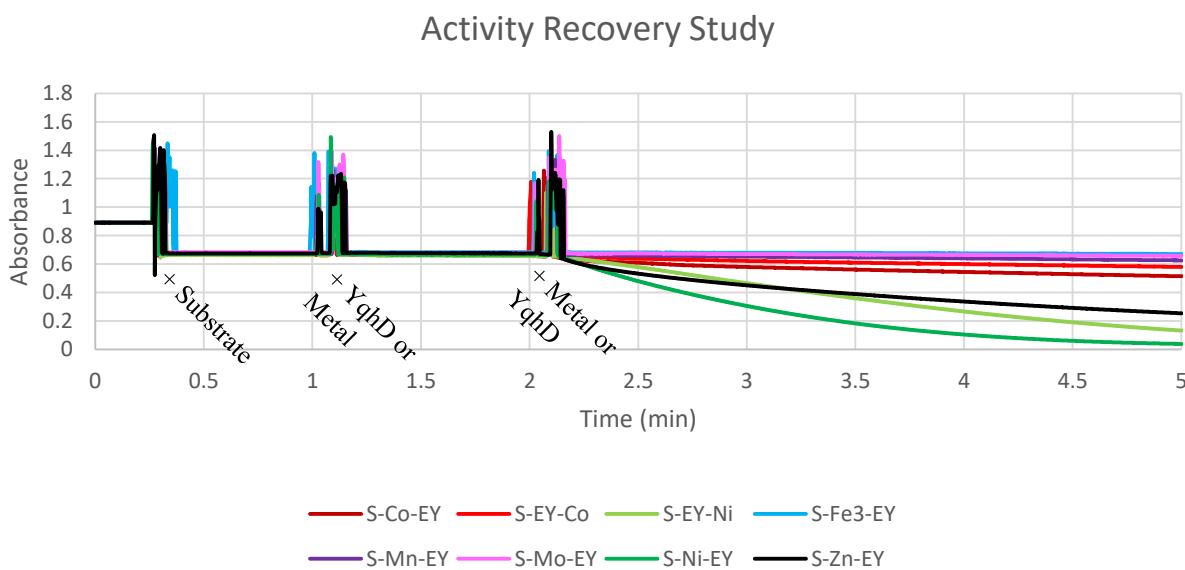


Figure 6: Preliminary evidence of catalysis with non-zinc metals bound to YqhD. Large disturbances in the absorption are due to the presence of the adding/mixing pipette tip. Each addition is labeled with what is being added at that point in time. Line labels indicate the order of addition (S: Butyraldehyde, EY: YqhD, Co: Cobalt (II), Mn: Manganese (II), Mo: Molybdenum (II), Ni: Nickel (II), Zn: Zinc (II), Fe3: Iron (III))

While most metals did not lead to significant catalysis, there were indications of significant catalysis in the presence of metals other than zinc. This experiment also confirms that

the presence of a divalent metal is necessary for catalysis, as no reaction progress was observed until both enzyme and metal were present in solution. With this preliminary experiment in mind, three metals (Co^{2+} , Ni^{2+} , and Cu^{2+}) were chosen for kinetic characterization in a similar fashion as the zinc-containing characterization. The large margin of error is due to only a small amount of data points being taken due to lack of available enzyme at the time.

Table 3: Michaelis-Menten kinetic parameterization of YqhD with butyraldehyde as a substrate

Metal	$K_M (10^{-4} \text{ M})$	$V_{max} (10^{-8} \text{ M/s})$	$k_{cat} (\text{s}^{-1})$	$k_{cat}/K_M (\text{M}^{-1}\text{s}^{-1})$
Zn^{2+}	1.12 ± 0.09	9.69 ± 0.15	0.39 ± 0.01	3470 ± 328
Co^{2+}	0.94 ± 0.29	4.61 ± 0.27	0.18 ± 0.01	1964 ± 723
Ni^{2+}	4.35 ± 2.1	9.27 ± 0.99	0.37 ± 0.04	852 ± 502
Cu^{2+}	27 ± 49	3.98 ± 2.27	0.16 ± 0.09	59 ± 140

As expected, metals other than zinc have a lower catalytic efficiency than the native holo state of YqhD. However, it is interesting that other metals do function, potentially indicating that zinc provides a more structural role to the catalytic function of YqhD, as opposed to directly aiding in the catalytic cycle.

2.5 FRET-Based Cofactor Binding Kinetics

The binding affinity for NADPH was characterized by leveraging the fact that YqhD has a tryptophan in the cofactor binding pocket, and that the fluorescence emission spectrum of tryptophan after excitation at 295 nm overlaps well with the absorbance spectrum of NADPH, which in turn has a fluorescence emission peak at 470 nm. This provides an excellent way to monitor the binding affinity of NADPH to YqhD, and has precedent with prior studies of Carboxy-Terminal Binding Protein (CtBP) binding to NADH²⁰. Again, stopped flow analysis was chosen in the hopes of performing progress curve analysis to extract the k_{on} and k_{off} of

NADPH binding, but complications in spectra acquired lead to difficulties in data processing. As such, only fluorescence intensities measured at equilibria were used for K_d calculation purposes. The fraction of NADPH bound was determined by observing when fluorescence was no longer increasing at an appreciable rate, indicating saturation of the enzyme; setting this value to a fraction bound of 1, and calculating the relative fractions bound by comparing the relative fluorescence intensities of more dilute NADPH titrations.

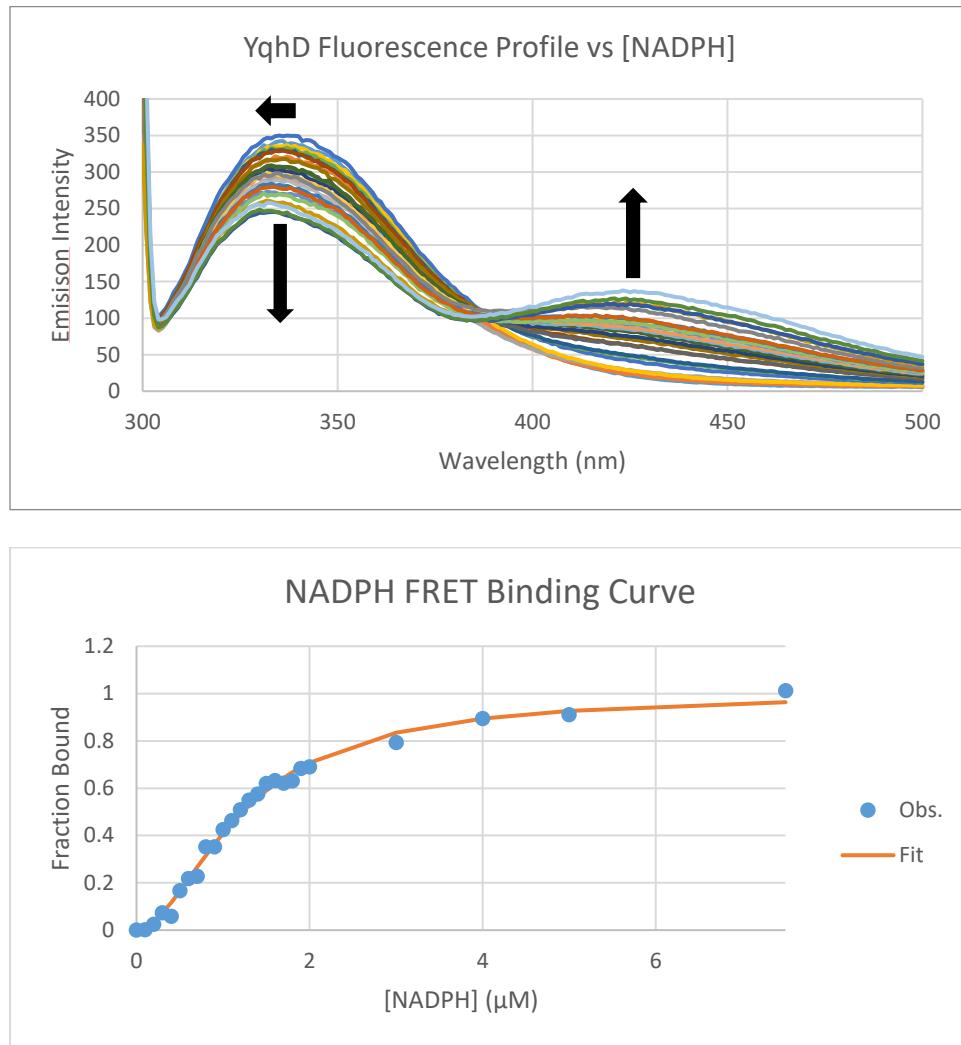


Figure 7: (Top) Fluorescence emission spectra (Ex. 295 nm) of YqhD as a function of increasing concentration of NADPH. Arrows indicate the progression of the spectral change as the concentration of NADPH increases. (Bottom) Fit of the apparent fraction bound at various concentrations of NADPH to the Hill equation. Fraction bound was determined by comparing the emission intensity at 420 nm of sample to that of that of a fully saturated sample.

Interestingly, YqhD appears to have a high degree of cooperativity with respect to NADPH binding, as indicated by the sigmoidal shape of the binding curve. Additionally, the fluorescence profile of the tryptophan emission blue-shifts gradually as the concentration of NADPH increase, perhaps indicating the residue becomes more buried as NADPH binds. Fitting the data using classic Hill equation methods yields a Hill coefficient (n) of 1.81 ± 0.10 , with a K_H of $1.23 \pm 0.05 \mu\text{M}$. Solving for K_d gives a value of $1.45 \pm 0.08 \mu\text{M}$, indicating a moderately tight binding interaction. The binding affinity of NADP^+ was measured indirectly by using a competitive binding assay, as described by Fjeld, C. et al²⁰. Unfortunately, the tight binding of NADPH coupled with the high degree of cooperativity serves to complicate the normally straight-forward non-cooperative binding model used for analysis. However, by assuming that NADP^+ also exhibits cooperative binding and observing that saturation has not occurred yet in the experiment, a decent fit of the assay can be achieved by using the same Hill equation methods as before.

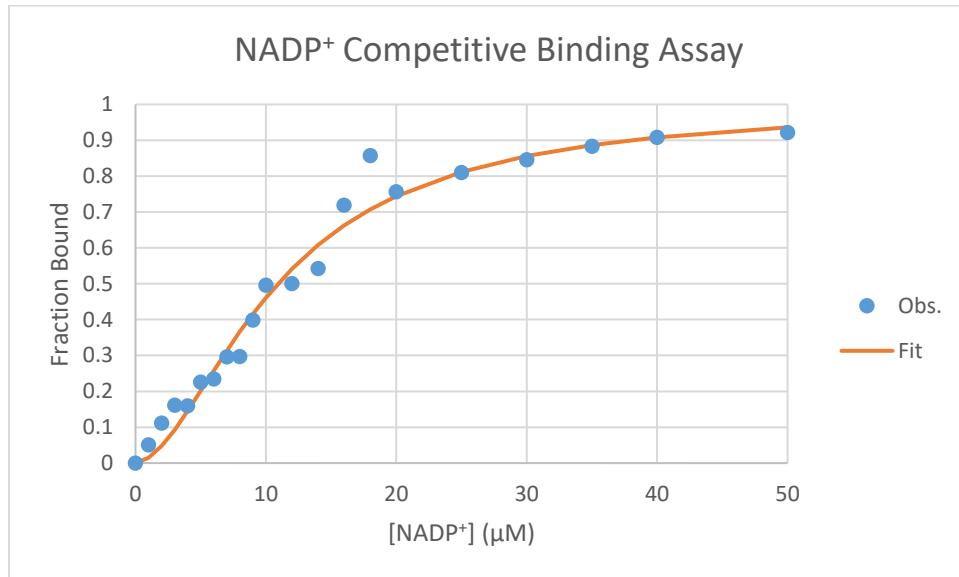


Figure 8: Competitive binding assay of NADP^+ vs NADPH. Fraction bound was determined by comparing the recovery of emission intensity at 340 nm of sample to that of a non-NADPH treated sample.

Fitting this data yields a Hill coefficient of 1.76 ± 0.22 with an apparent K_H of $10.92 \pm 1.06 \mu\text{M}$. Solving again for the apparent K_d of NADP^+ produces a value of $67.90 \pm 28.28 \mu\text{M}$, much higher than the K_d of NADPH. However, in order for a more accurate comparison, the apparent K_d of NADP^+ was converted to the true K_d by using the following equation:

$$K_{d(\text{NADP}^+, app)} = K_{d(\text{NADP}^+)} \left(1 + \frac{[\text{NADPH}]}{K_{d(\text{NADPH})}}\right) \quad (2.1)$$

This produces a value of $15.25 \pm 7.02 \mu\text{M}$ which, while smaller than the apparent K_d , is still greater than the K_d of NADPH, indicating that NADP^+ binds much less tightly than NADPH. A potential explanation for this phenomenon has also been observed in molecular dynamics studies conducted previously by the Mitchell-Koch group. Binding cleft opening/closing dynamics were studied by monitoring the distance between Lys₁₆₀ and His₂₇₁, which reside on either side of the entrance to the cofactor binding cleft. The distribution of distances over the course of a $50 \mu\text{s}$ simulation were analyzed to determine the relative conformational flexibility of the YqhD monomeric subunits in the presence of either NADP^+ or NADPH in both binding pockets.

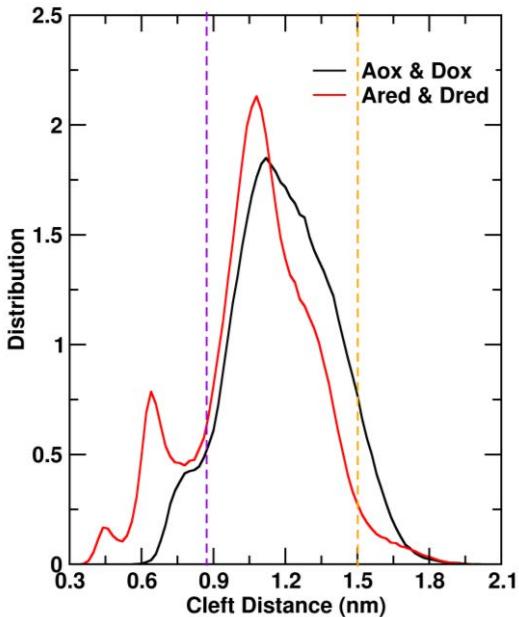


Figure 9: Cofactor-dependent distribution analysis of binding cleft distances by MD. A and D indicate chain identity, ox and red indicate whether the chain has oxidized or reduced cofactor
 (Credit Dr. Rajni Verma, Wichita State University)

Simulations reveal that the presence of NADPH affords an enrichment in the closed conformational state of YqhD, as if the cleft “clamps shut” upon binding of NADPH. This may explain why NADPH has a much higher binding affinity than NAPD⁺, and may lend credence to the observational theory that YqhD seems to preferentially work as an aldehyde reductase, as seen in substrate scope analyses¹. This would also explain why the tryptophan emission fluorescence becomes blue shifted, as a more closed conformation would likely lead to an increase in the hydrophobic environment of the active site tryptophan. More sophisticated protein dynamics studies are needed to confirm this observation, but these computational studies provide promising insights that could be used for future protein engineering studies.

2.6 Conclusions

The degree of binding promiscuity for both substrates and metals makes YqhD an interesting enzyme for potential protein engineering efforts, and the high yield of protein expression could be a definite boon towards efforts to use YqhD in an industrial setting. It is still

ambiguous if the active site metal plays a purely structural role, or if it is involved more deeply in the actual catalytic mechanism. However, the fact that YqhD is able to perform catalysis only in the presence of a divalent metal coupled with there being very little deviation between the holo and apo crystal structure seems to lend credence to a more catalytically oriented role, such as participation in bond polarization enhancement.

YqhD exhibits kinetic qualities of a moderate to low efficiency enzyme with a below average k_{cat} and above average K_M , relative to other enzymes²¹, for the reaction of butyraldehyde and NADPH to form butanol and $NADP^+$. Previously reported catalytic efficiencies for butyraldehyde show a k_{cat}/K_M of ~87,000¹, but these reactions were carried out at vastly different conditions (in particular, at 37 °C), which indicates that the conditions used in the initial parameterization studies conducted within this report were likely not optimal for catalysis. Regardless, comparisons can still be drawn between studies that were carried out at equivalent conditions. Comparison of previously reported catalytic efficiencies for both oxidative and reductive reactions shows almost uniformly higher catalytic efficiency for the conversion of aldehydes to alcohols, implying that this is likely the native function of the enzyme. This hypothesis is strengthened by the observation that the binding affinity for NADPH is greater than that of $NADP^+$, which serves to “prime” the enzyme for catalytic conversion of aldehydes to alcohols by maintaining a higher concentration of NADPH-bound species, followed by preferential release of the spent cofactor after catalysis. This has also been observed computationally via distribution analysis of the cleft opening distance. These preliminary results and method developments lay a foundation for potential future protein engineering efforts, and provide a reference for detecting perturbations to protein dynamics/function by incorporation of fluorinated amino acids for the purpose of ¹⁹F NMR analytical methods.

CHAPTER 3

5F-TRP LABELING OF YQHD

3.1 Experimental

Generation of pASKT-YqhD(W264am)-yPheRS(T415G):

All restriction enzymes and phosphatase were acquired from NEB. A plasmid containing a site specific amber mutant of tryptophan 264 in YqhD was generated using a QuikChange II Site-Directed Mutagenesis Kit (Agilent), using the previously created TOPO-YqhD plasmid. Successful mutagenesis was verified via DNA sequencing. This YqhD amber mutant gene was then subcloned into pASKT (hereby referred to as pASKT-YqhD(W264am)), transformed into Tuner pLysS cells, and tested for protein expression using the previously described protocol. Once it was confirmed that no YqhD expressed, as expected, a yPheRS(T415G) gene was transferred to the pASKT-YqhD(W264am) plasmid, downstream of the YqhD gene. The source of this gene was a pQE16_mDHFR_yPheRS(T415G) generously donated by Prof. David Tirrell (CalTech), encoding for expression of a mutant dihydrofolate reductase and the desired mutant yeast phenylalanine-tRNA synthetase. The yPheRS(T415G) gene was extracted as a blunt-end fragment by digestion of the pQE16am_mDHFR_yPheRS(T415G) plasmid with the PvuII and BglII restriction enzymes followed by agarose gel purification and extraction. pASKT-YqhD(W264am) plasmid was then treated with EcoRV and subsequent dephosphorylation with Antarctic Phosphatase to prevent recyclization. Both the purified yPheRS(T415G) and the dephosphorylated vector were mixed in a 5:1 concentration ratio and mixed with ligation buffer containing PEG 4000. This mixture was directly used to transform competent DNA SmartCells. yPheRS(T415G) insertion was confirmed by digestion of mini-prepped plasmid with XbaI and observation of two fragments.

Expression of 5-fluoro-L-tryptophan Labeled YqhD:

For uniform fluorotryptophan labeling of YqhD, the pASKT-YqhD vector was transformed into the *E. coli* DL41 tryptophan auxotroph strain. For site-specific labeling of YqhD, the pASKT-YqhD(W264am)-yPheRS(T415G) vector was transformed into a Phe/Trp double auxotrophic strain (AFW) containing the pREP4_ytRNA^{Phe}_UG plasmid, which encodes for constitutive expression of an amber-suppressor tRNA that is recognized and chargeable by the yPheRS(T415G) mutant tRNA synthetase. A similar protocol to the wild-type expression and purification protocol was followed, with the exception of using defined minimal media containing measured quantities of all amino acids except tryptophan. Overnight cultures were prepared in the same fashion as before, and these cultures were then used to inoculate defined minimal containing 0.245 mM supplemental tryptophan. These cultures were allowed to grow to an OD₆₀₀ of 0.75-1.0 before harvesting by centrifugation. The harvested pellets were washed twice with normal saline chilled to 4°C by resuspension and centrifugation, then resuspended in minimal media containing 3 mM 5-fluoro-DL-Tryptophan. In the case of site-specific labeling using AFW cells, 15 µM phenylalanine and 10 µM tryptophan were also present in the labeling media. These cultures were repartitioned into 1L growths, and allowed to shake in an incubator at 37°C for 10 minutes before being induced with 100 uL of 2 mg/mL anhydrotetracycline in DMF. Isolation and purification procedure matches that of wild-type YqhD.

¹⁹F-NMR of Labeled YqhD:

NMR samples were prepared at a concentration of approximately 0.3 to 0.5 mM. A coaxial insert containing 0.001% TFA in D₂O was used for internal referencing, and was set to 0 ppm before each experiment. Spectra were acquired on an Oxford 400 MHz Spectrometer

equipped with a Varian 400 ID Triax probe. FID scans were acquired until sufficient spectral signatures were detected, typically after ~6 hours.

Analysis of YqhD by Circular Dichroism Spectroscopy:

Samples were prepared by buffer exchange into 20 mM potassium phosphate buffer, pH 7.5 to an end concentration of 15 μ M YqhD. CD experiments were carried out using a 0.1 mm path length quartz cell which can be attached to a water source. Spectra were acquired on a Jasco J-810 Spectropolarimeter. For thermal melt curves, a computer controlled waterbath was used to automate the melt process.

3.2 Uniform Labeling of YqhD

For the uniformly labeled samples, isolated YqhD was taken through a series of treatments to determine the influence of the zinc metal, cofactors, and an archetypical substrate butyraldehyde. A stock of labeled YqhD was first treated with 5 mM EDTA to facilitate removal of any spare zinc acquired during the expression process, and a small quantity was taken for NMR analysis. This treatment was determined to sufficiently remove traces of zinc in a prior experiment with unlabeled YqhD by ICP-OES analysis. 5 equivalents (relative to enzyme concentration) β -mercaptoethanol was then added, followed by 5 eq. ZnCl₂ to repopulate the zinc binding sites while also accounting for potential binding to the polyhistidine purification tag. A small sample was taken for NMR analysis, and the rest was divided into 3 stocks for cofactor and substrate binding. To one stock was added 5 eq. NADP, to another was added 5 eq. NADPH, and to the final stock was added 50 eq. butyraldehyde. 50 eq. was chosen for butyraldehyde to avoid potential issues with low binding affinity typically associated with non-native substrates, as well as to account for potential loss of substrate due to volatility. Each of these stocks was then made into NMR samples and spectra were acquired. During acquisitions, a high intensity

and sharp resonance was observed uniformly across samples at -53.17 ppm that did not phase properly with the rest of the spectrum. It was determined that this was a fold over peak from the internal TFA standard by acquiring a quick spectrum with only water and the coaxial insert in the NMR tube, and therefore can be safely ignored as an artifact of the acquisition process. Unfortunately, the presence of the fold over peak makes accurate phasing difficult, so peak intensity/quantitation can only be done from a relatively qualitative standpoint. As a future direction, acquiring spectra on more concentrated samples and optimizing sweep width parameters, coupled with more advanced analysis such as Bayesian analysis, may facilitate more accurate discrimination of peaks, especially if peaks begin overlapping into one another. Each monomer of YqhD contains five tryptophans, and so five resonances were predicted to be seen in the ^{19}F -NMR spectra.



Figure 10: Structure of YqhD showing 5-fluorotryptophan positions (orange). Individual monomers are shown in magenta

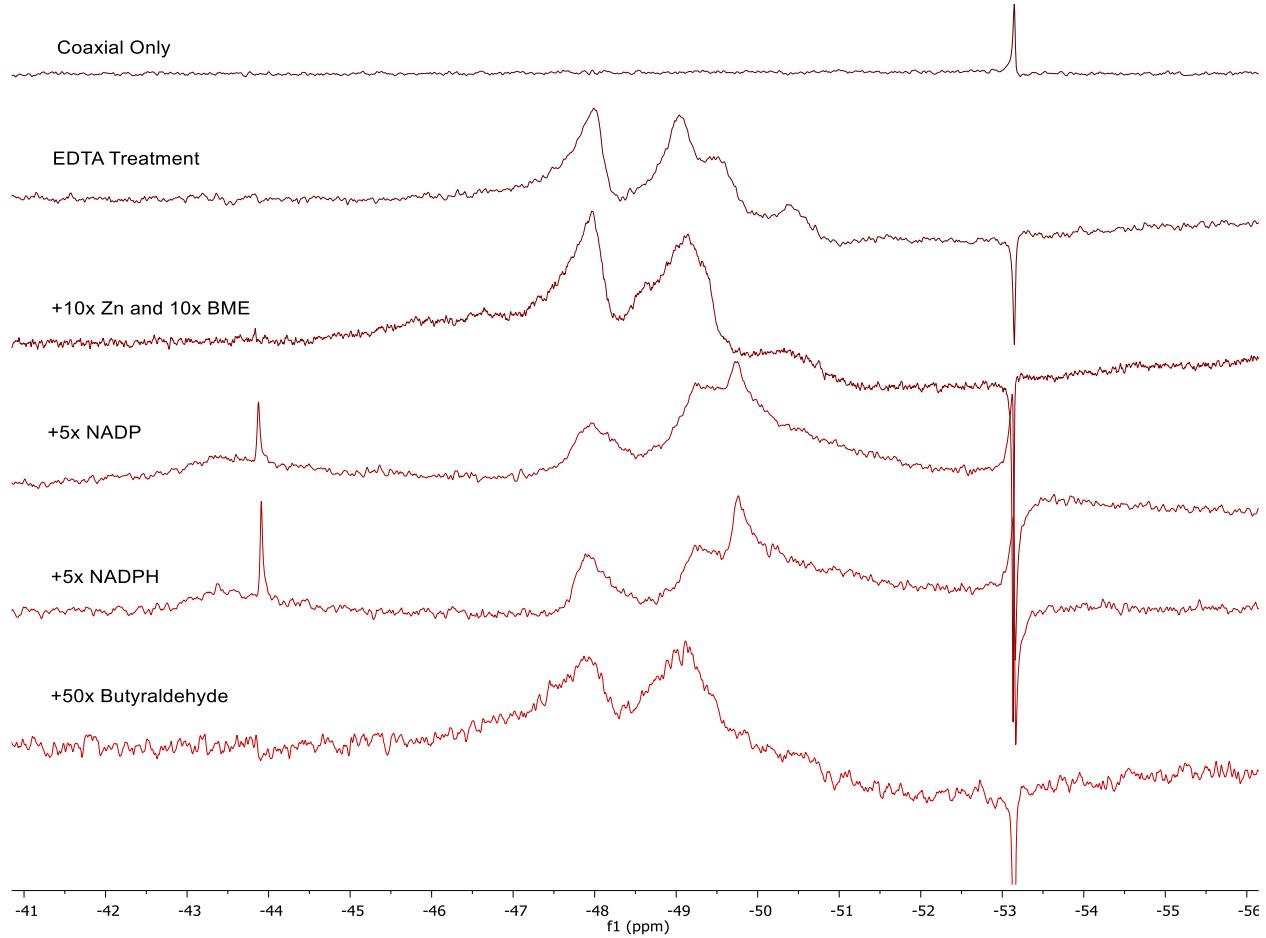


Figure 11: ^{19}F -NMR comparison of 5-fluorotryptophan labeled YqhD in various conditions. Each spectra is referenced to an internal coaxial insert containing 0.001% TFA in D_2O (0.00 ppm). A fold-over peak of this reference is observed at -53.05 ppm. Each fold concentration is relative to the YqhD concentration (0.5 mM)

For the EDTA treated YqhD spectra, three clear resonances were observed at -47.99, -49.04, and -49.46 ppm. A weaker peak appears to potentially exist at -50.36 ppm as well. Upon treatment with β -mercaptoethanol and ZnCl_2 , the weak resonance at -50.36 ppm decreases in intensity, and the peak at -49.46 ppm appears to shift as well. The latter may have shifted closer to the resonance at -49.04 ppm, as the overall peak seems to show more intensity downfield. Addition of both NADP and NADPH appear to result in similar spectral changes, showing a new sharp resonance at -43.91 ppm, with broader resonances at -47.89, -49.23, and -49.76 ppm. There also appears to be a potential broad resonance at -43.50 ppm. As there is a

tryptophan residing very close to the cofactor/metal binding pocket, the sharp resonance could potentially be attributed to tryptophan 264. The other resonance shift at -49.76 ppm may also stem from the tryptophan positioned behind a flexible loop region which, through simulations, has been shown to have variable dynamics depending on the presence and identity of a bound cofactor.

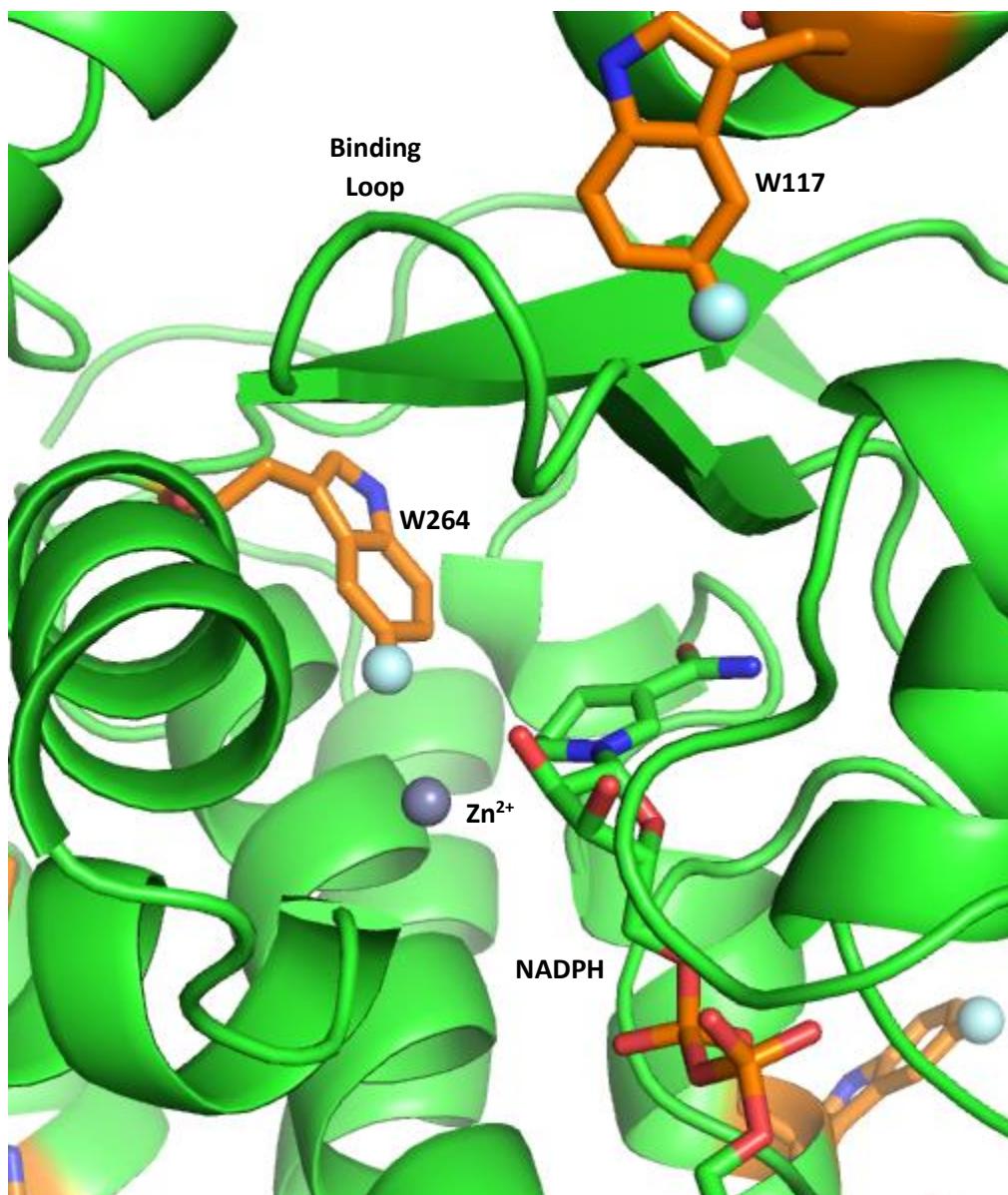


Figure 12: Structure of cofactor/metal binding pocket depicting the close cofactor contact W264 and cofactor binding loop W117

Addition of butyraldehyde leads to a spectrum very similar to the zinc treated enzyme. In addition, acquiring spectra over overnight with the butyraldehyde containing sample lead to a large broadening with an appearance of sharp resonance at -43.87 ppm. While the sharp resonance may be caused by binding of butyraldehyde to the enzyme active site, the broadening might be due to destabilizing and unfolding of the protein. The NMR sample also had visible precipitation after removal from the NMR. The long term change to the spectra may also be due to slow binding of the substrate leading to a long timeframe for equilibria to be achieved. It is difficult to fully discern what has occurred without more sophisticated NMR experiments, such as NOE based studies, but a potential shortcut would be to filter the sample, reacquire spectra, and see if the spectral profile persists. Unfolding studies using chemical denaturants, such as urea or guanidinium chloride, could also be done in order to see where the unfolded protein resonances converge to, in order to determine if any of the resonances observed can be attributed to YqhD in the unfolded/aggregated state.

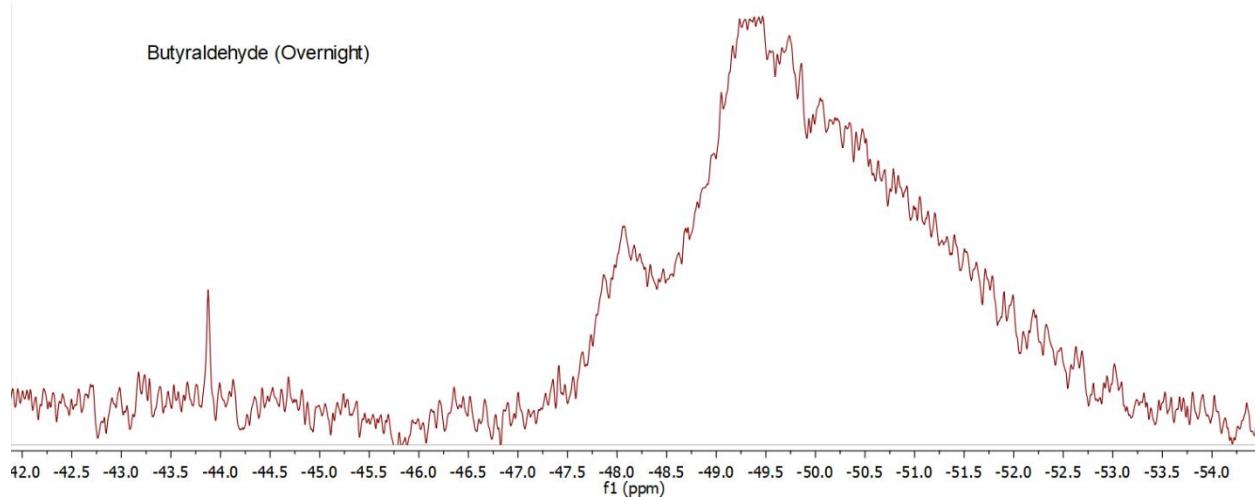


Figure 13: ^{19}F -NMR of fully labeled YqhD incubated with 50 eq. butyraldehyde overnight

It is important to determine if the presence of fluorinated amino acids causes perturbations in the overall structure and dynamics of YqhD. The thermal stability of labeled

and unlabeled YqhD was measured and compared by way of circular dichroism spectroscopy. This was performed to see if the presence of 5-fluorotryptophan significantly impacted the stability or folding ability of YqhD, as there has been previously reported incidence of fluorine labeling leading to instability in other proteins⁷.

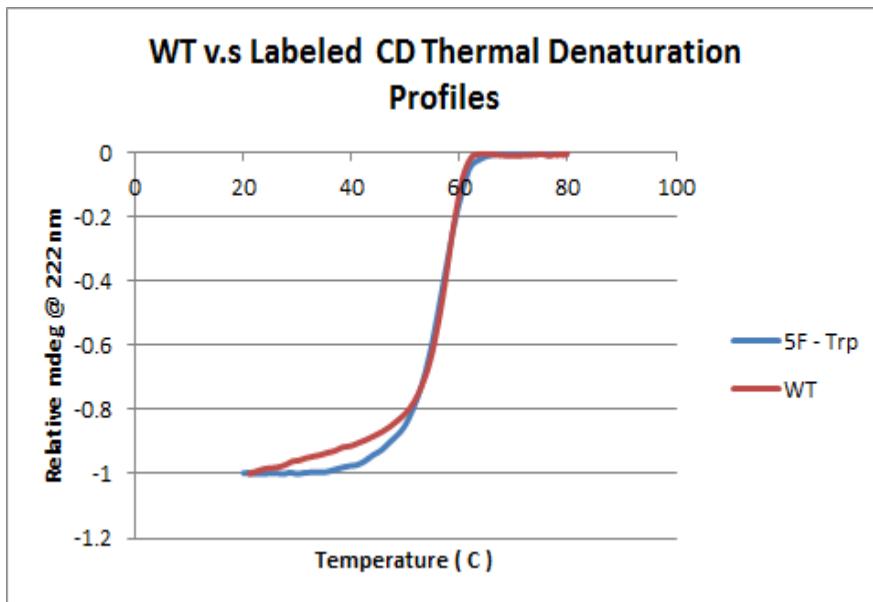


Figure 14: Thermal denaturation curves of labeled and unlabeled YqhD, in 20 mM potassium phosphate buffer pH 7.0, by circular dichroism spectroscopy. Samples were placed in a 0.1 mm quartz CD cuvette. Temperature was varied by a computer controlled Julabo waterbath.

While there does appear to be a small deviation of the thermal denaturation curves, there is no evidence that the presence of 5-fluorotryptophan leads to global destabilization. This provides promising evidence that ¹⁹F NMR can be used as a method for studying the conformational dynamics of YqhD in a focused manner without the complexities associated with more traditional NMR based methods.

The Michaelis-Menten kinetic parameters for the 5-fluorotryptophan labeled species were also measured in the same fashion as wild-type, zinc bearing YqhD. It was predicted that there would likely be a potential change in catalytic efficiency, as one of the five labeled tryptophans

residues in the cofactor binding pocket with the predicted fluorine location being quite close to the Zn²⁺.

Table 4: Comparison of Michaelis-Menten kinetics for unlabeled (WT) and labeled (5FW) YqhD

Protein	K _M (10 ⁻⁴ M)	V _{max} (10 ⁻⁸ M/s)	k _{cat} (s ⁻¹)	k _{cat} /K _M (M ⁻¹ *s ⁻¹)
WT	1.12 ± 0.09	9.69 ± 0.15	0.39 ± 0.01	3470 ± 328
5FW	4.89 ± 0.49	11.85 ± 0.29	0.47 ± 0.01	961 ± 116

Indeed, as expected, the catalytic efficiency of YqhD towards butyraldehyde was modulated by the presence of 5-fluorotryptophan. Studies conducted to determine the binding affinity of NADPH to the labeled protein indicated a much lower binding affinity, with no indications of saturation at a 3-fold concentration of NADPH. This may be due to the fact that fluorine is much more electronegative than the native structure proton, which may impact the ability for the zinc metal to coordinate with the cofactor. However, this is purely speculation, and more sophisticated experiments such as NMR time-course or NOE experiments and substrate soak crystallography would help elucidate what is actually happening.

3.3 Site-Specific Labeling of YqhD

Site-specific techniques were developed in order to add a new, potentially less perturbative assignment method for fluorotryptophan labeling NMR experiments. It had been demonstrated in prior literature that site-specific incorporation of various halogenated tryptophan analogues could be done using a combination of an amber suppressor tRNA with a mutated yeast phenylalanine aminotransferase, but no data was reported on the potential incorporation efficiency of fluorinated analogues⁸. For preliminary experimentation, the tryptophan residing near the cofactor binding site in YqhD (W264) was chosen as it was reasoned to likely undergo the most drastic spectral changes upon binding of cofactor, and would be a potential interesting

probe of protein dynamics for engineering studies if this proved to be true. Once purified protein had been produced, no additional treatments were conducted with EDTA, β -mercaptoethanol, or $ZnCl_2$, as these treatments had proven in the past to potentially lead to significant protein aggregation if done even slightly incorrectly. This decision was made due to the fact that cell growth and protein expression was significantly reduced under the minimal media growth conditions, and only a sufficient quantity of protein for one or two NMR samples was acquired from a 2 liter growth of bacteria. Optimization of growth parameters or utilization of a fermenter apparatus may potentially afford more protein for study, but there was insufficient time to test these ideas.

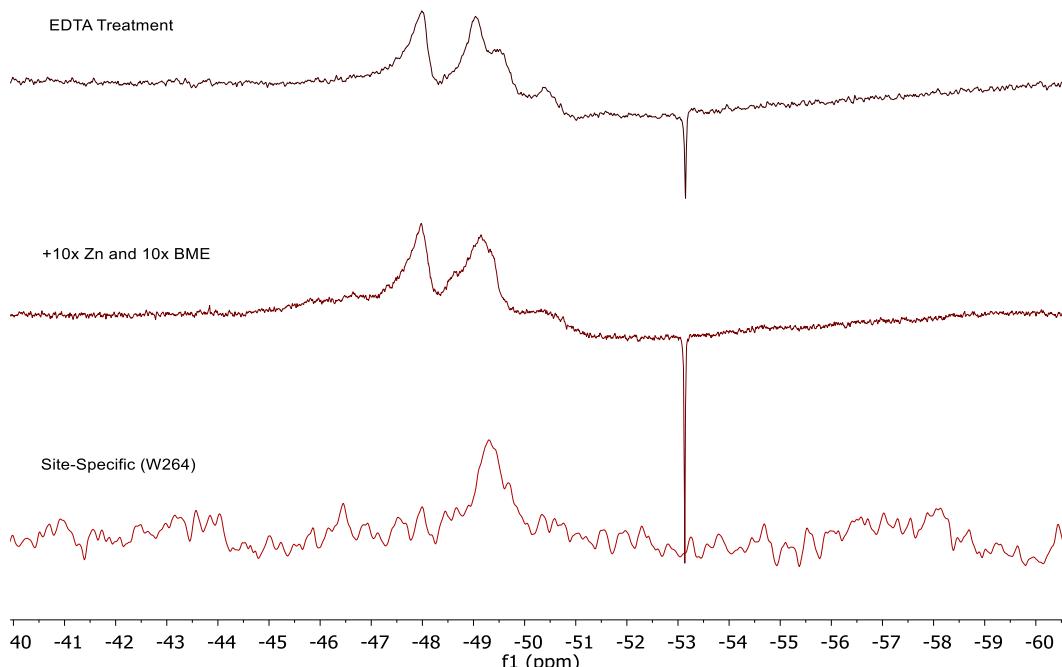


Figure 15: ^{19}F -NMR comparison of site-specifically labeled YqhD (0.3 mM) to fully labeled YqhD. The site-specific was internally referenced to a 0.001% TFA in D_2O standard (0.00 ppm)

Shown is a comparison between fully labeled YqhD (both zinc-free and bound to zinc) and the potential site-specifically labeled variant. Line broadening for the site-specific variant

was increased (40 Hz vs 10 Hz) due to weak signal intensity. A single distinct resonance is observed at -49.30 ppm, which lies reasonably close to peaks seen in the EDTA treated and zinc bound fully labeled spectra (-49.46 and -49.14 ppm, respectively.) This is promising evidence that the site-specific incorporation was successful, and that the aforementioned incorporation technique developed by the Tirrell group is also applicable to fluorinated tryptophan analogues. Spectra were also acquired with the site-specifically labeled protein with NADP added, in order to determine if the protein could be used to determine spectral changes upon binding of cofactor in a more unambiguous manner.

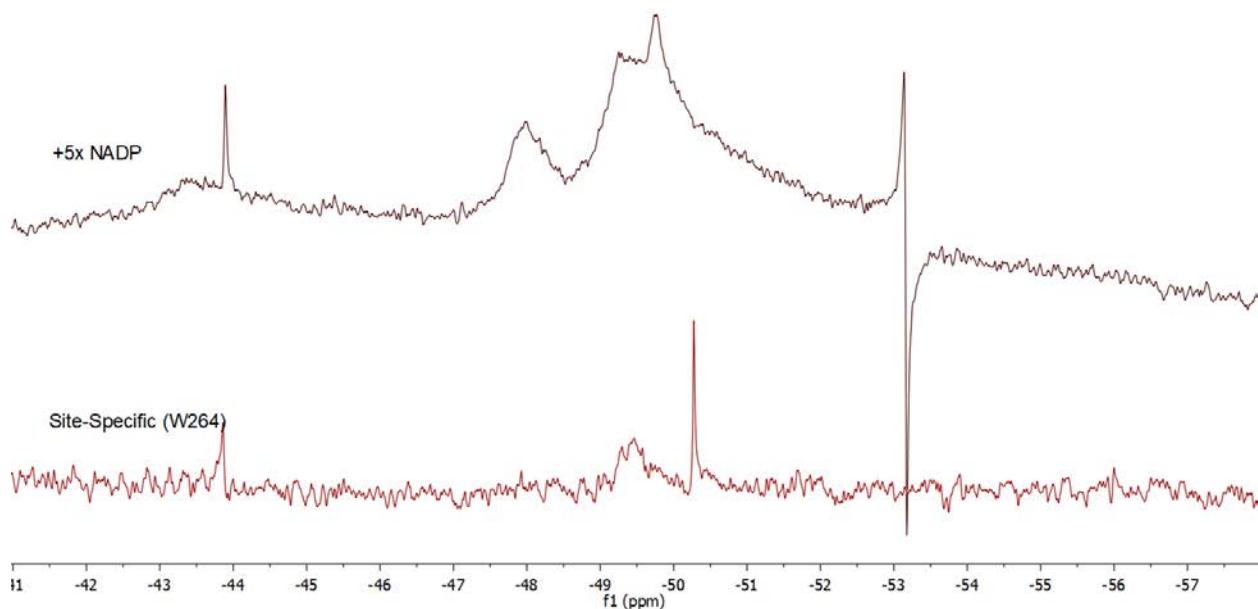


Figure 16: ^{19}F -NMR comparison of site-specifically labeled YqhD (0.3 mM) and fully labeled YqhD in the presence of 5 eq. NADP^+

Upon incubation with NADP, the original resonance observed in the site-specific sample appears to have lost a significant degree of intensity, and two new resonances appear at -43.86 and -50.28 ppm. While this leads to a promising potential assignment of the observed peak at -43.86 ppm to W264, as previously hypothesized, the appearance of a strong resonance at -50.28 ppm is concerning. No prior experiments conducted seemed to show a resonance appear at this

position, and it is unclear if this is due to the protein occupying two distinct states at equilibrium, or if some other phenomenon is the cause. The potential broad resonance at -43.50 ppm also appears to be absent in the site-specific spectra. The narrowing may be due to a shift in the cooperativity between the monomers as a consequence of labeling. Utilization of F-F NOE could afford insight into the changes in dynamics both within an individual monomer and across the dimer interface. The original broad resonance may be due to exchange between the open and closed conformations observed via MD simulations, and the sharp resonance could be potentially result from the increased population of closed conformation upon NADPH binding. Performing a substrate soak experiment with NADPH, or utilizing variable temperature NMR to perturb the exchange rate, would aid in determining if the broad resonance disappearance and new sharp resonance appearance and actual phenomena, or if they are a quirk of the labeling process. Mass spectrometry analysis would also provide unambiguous evidence that the site-specific incorporation occurred as desired, but there was insufficient time to perform such an analysis.

3.4 Conclusions

While the utilization of non-canonical amino acids is not without its potential pitfalls, ¹⁹F-NMR provides a powerful tool for studying changes in protein structure and dynamics under various conditions relevant to the protein's native function. It provides a method of labeling and reporting on local environment changes without the complexities and data density of more traditional protein NMR methods, as was observed in the studies on YqhD. This information, coupled with simulation data and more advanced NMR techniques, may provide insight on what role motional dynamics play during the catalytic process. Of course, whole-protein labeling with fluorinated amino acids is still accompanied with the issue of peak assignment. Traditional methods such as phenylalanine replacement mutagenesis could be used, but confirmation that

fluorotryptophan can be incorporated in a site-specific fashion will hopefully provide a new and useful tool to the ^{19}F -NMR toolbox. These methods, in theory, should be less perturbative in nature while providing a much more focused assignment and protein dynamics analysis method. While other methods exist for site-specific post-translational introduction of diverse non-canonical amino acids, such as BONCAT²², allowing *E. coli* to handle the full labeling process provides a streamlined method for fluorine labeling. A complete labeling system could be envisioned by placing a high-efficiency engineered TrpB^{23,24} encoding gene upstream of the yPheRS (for site-specific labeling) and protein-of-interest genes, allowing one to label and study region of interest for a fraction of the current cost of 5-fluorotryptophan by simply including 5-fluoroindole and supplemental serine in the expression media. This may afford a robust and easy-to-implement method for fluorine NMR studies of proteins by simply swapping out the protein-of interest gene.

CHAPTER 4

COMPUTATIONAL STUDIES OF SMALL MOLECULE NMR PROPERTIES

4.1 Methodology

All QM studies were carried out using the Gaussian 09 software suite²⁵ at either the Wichita State High Performance Computing Center (HiPeCC), or a mini-cluster constructed in the lab using personal funds. Scripts were written to automatically generate, harvest, and process some data sets, and are available upon request. Any NMR values that were not reported in previous literature were acquired in house on an Oxford 400 MHz NMR Spectrometer equipped with a Varian 400 ID Triax probe, using a 0.01% trifluoroacetic acid in D₂O coaxial insert for internal referencing. The observed ¹⁹F resonance for trifluoroacetic acid was set to -76.55 ppm, using CFCl₃ as internal reference (δ CFCl₃ = 0.00 ppm). All NMR simulations were conducted by first performing a ground state optimization of each molecule with a frequency calculation to ensure the structure generated was indeed an energy minimum by validating the absence of imaginary frequencies. These optimized geometries were then automatically fed into a NMR calculation method. For determination of optimal ¹⁹F-NMR simulation methods, a large array of wavefunction solution methods, DFT methods, basis sets, solvation methods, and NMR calculation methods were tested. The following compounds were simulated to determine the accuracy of each computational method: CFCl₃, Trifluoroacetic acid (Protonated), 4-fluorotryptophan, 5-fluorotryptophan, 6-fluorotryptophan, 4-fluoroindole, 5-fluoroindole, 6-fluoroindole, 4-fluorophenylalanine, 4-fluorotoluene, and fluorobenzene. The following calculation methods were tested: HF^{26,27}, MP2²⁸⁻³³ (Both SCF and MP2 nuclear shieldings), B3LYP³⁴, BHandHLYP, BLYP³⁵⁻³⁷, M06³⁸, M062X, PBE^{39,40}, PBE1, PW91^{41,42}, and ω B97X⁴³. The following basis sets were tested: 6-31+G(d) and 6-311++G(3df,2p), and aug-cc-pVTZ. The

following solvation methods were tested: SMD⁴⁴ and CPCM^{45,46}. GIAO^{47–50} was chosen for the NMR calculation method.

For the following methods, solvent dependent chemical shift perturbations were also calculated and compared: B3LYP/6-31+G(d), B3LYP/6-311++G(3df,2p), BHandHLYP/6-31+G(d), BHandHLYP/6-311++G(3df,2p), and MP2/6-31+G(d). The following compounds were used for this dataset: 5-fluorotryptophan, 5-fluoroindole, 4-fluorotryptophan, 4-fluoroindole, 5-fluoroindole, 4-fluorophenylalanine, 4-fluorotoluene, and fluorobenzene. The following solvents were used for this dataset, though not all solvents were used for each compound tested due to either solubility issues or lack of materials: acetonitrile, butyraldehyde, N,N-dimethylformamide, dimethylsulfoxide, ethanol, hexane, isopropanol, methanol, propanol, tetrahydrofuran, and water.

Determination of % error for absolute chemical shifts were performed by comparing the experimentally observed ¹⁹F-NMR chemical shift values (referenced to CFCl₃ at 0.00 ppm) to the simulated chemical shift value by subtracting the compound NMR shift from the simulated CFCl₃ shift of the same method combination.

4.2 Analysis of Computational Methodology

For all methods, utilization of the aug-cc-pVTZ basis set either led to unreasonably long calculation times or non-convergence of structure optimization procedure. The following errors were calculated for the various computational methods for the GIAO NMR calculation method, which was the only one used for the initial screening set:

Table 5: Error in ppm of various ^{19}F -NMR calculation methods (big indicates 6-311++G(3df,2p), while everything else utilized 6-31+G(d).) In bold is the best method observed.

<u>Error (ppm):</u>	Solvation Method:	
Method:	SMD	CPCM
b3	16.33	14.55
b3big	15.93	13.79
bh	3.18	2.6
bhbig	1.4	2.18
BLYP	34.8	32.32
BLYPbig	30.86	28.4
HF	12.59	13.87
HFbig	14.3	15.45
M06	4.24	3.24
M06big	8.28	5.97
M062X	3.26	1.9
M062Xbig	3.53	1.91
mp_SCF	9.89	10.92
mp_MP2	6.69	5.96
PBE1PBE	8.81	6.94
PBE1PBEbig	5.42	3.09
PBEPBE	25.93	24.02
PBEPBEbig	21.46	19.79
PW91PW91	27.29	25
PW91PW91big	23.09	20.93
ω B97X	1.33	0.76
ω B97Xbig	2.68	0.92

The best observed method appears to be ω B97X/6-31+G(d) with CPCM water. This came as somewhat of a relief, as this is a relatively low cost computational method compared to many of the other methods tested. It was decided that simulating ^{19}F -NMR chemical shifts in various solvents would be a good approximation for determining the accuracy of various calculation methods in the various diverse environments found inside of proteins. For example, proteins do not have a uniform dielectric constant in their interior, so a high degree of accuracy across multiple solvents may indicate robustness for simulations in various proteins. 5-fluoroindole, 5-fluorotryptophan, and 4-fluorophenylalanine were chosen, as these were the most

relevant for a preliminary study. For the simulation of solvent dependent chemical shift perturbations, the following errors were determined:

Table 6: Average error, in ppm, of NMR calculation methods with various solvents. In bold is the best method observed.

	b3	b3big	bh	bhbig	mp-SCF	mp-MP2	wb97x
SMD	1.63	1.64	1.48	1.44	1.45	1.77	1.34
CPCM	1.86	1.86	1.56	1.56	1.88	1.83	1.59

ω B97X /6-31+G(d) with SMD solvation proved to be the most robust method among those tested, indicating it's potential future utility in protein NMR simulations. The simulations conducted here will serve as a roadmap for more advanced predictions of ^{19}F -NMR shifts in protein interiors. While it would be unreasonable to apply these methods to a full protein simulation, utilization of methods such as "carving" out regions of interest⁵¹, combined QM/MM methods such as ONIOM¹², and other approximations may afford a useful tool for performing NMR peak assignments without the necessity of performing amino acid replacement studies.

4.3 Simulated Deprotonation of Fluorinated Hydrocarbons

One of the model systems of interest in the Mitchell-Koch lab is that of 2-fluorohistidine and 4-fluorohistidine. This is because 4-fluorohistidine exhibits what is known as "reverse chemical shift behavior", which is to say that deprotonation, which intuitively would lead to shielding due to enhanced electron density in the ring system, results in a downfield shift of the fluorine resonance, indicating deshielding⁹. In order to attempt to explain why this is occurring, simulations have been done to observe how the molecular orbitals change between the protonated and deprotonated states⁵². However, the deconvolution of these changes is difficult from two reference states alone. In an effort to help understand how the electronic structures of 2-fluorohistidine and 4-fluorohistidine evolve over a deprotonation event, simulated deprotonation simulations were conducted. These were facilitated by performing simulations of

the sidechain analogues, 2-fluoro-5-methyl-imidazole (2F-MetImid), 4-fluoro-5-methyl-imidazole (4F-MetImid), and 2,4-difluoro-5-methylimidazole (2,4F-MetImid). 2,4F-MetImid was chosen to see if both reverse and normal chemical shift behavior were observed simultaneously. These molecules are characterized by two potential protonation positions in their neutral state, labeled as the π and τ -tautomers.

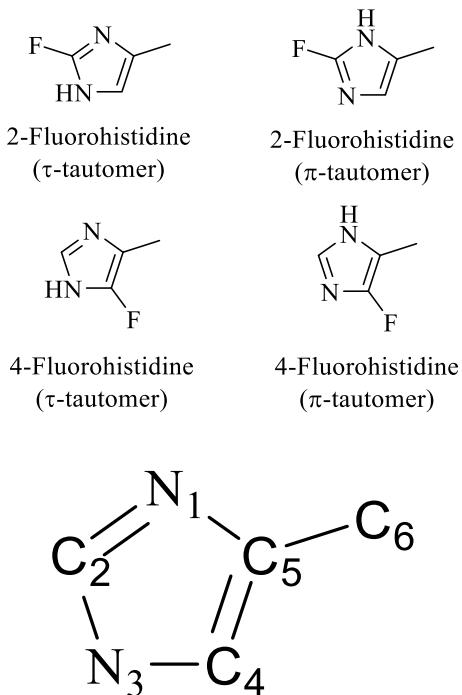


Figure 17: Depiction of the different tautomeric states of fluorohistidines, as well as naming conventions for each heavy atom

Structural alignment for animation generation was facilitated by restricting the geometry of ring system in such a way as to prevent free reorientation in the simulations Cartesian frame. Multiple simulations were then conducted where the distance between a nitrogen of interest and its corresponding proton were restricted to increasing lengths. A closed-shell ω B97x/6-31+G(d) method was employed to ensure homolytic bond cleavage did not occur by mistake. This intrinsically introduces errors in the energies calculated, as prototypical basis set orbital components become stretched which manifests as a steadily increasing molecular energy beyond

where the plateau “should” occur, but these errors are minor in the grand scheme of the simulation effort and were considered acceptable. Molecular orbital visualizations were generated using a combination of bash scripts to create relevant cube files, and Python-based PyMOL scripts to produce individual snapshots of each orbital, which were then compiled into animations. This affords a straightforward method of observing how each orbital changes over the course of a deprotonation event. For characterization of how the molecular properties evolve over the course of the simulated deprotonation, simple graphical methods were employed using a Jupyter notebook function set to aid in data harvesting. The jump in energy observed between cationic and neutral simulations is likely due to the lack of the originally stretched proton, which appears to uniformly contribute an energy of ~0.166 Hartree across all simulations. An offset of 2.5 Å was applied for the neutral-to-cation simulations to visualize how a subsequent deprotonation event continues to affect molecular parameters, which is similar to what would be observed in a simple base titration experiment. Extracted NMR values were shifted so that the ground state cationic species ^{19}F -NMR resonance is at 0.00 ppm to facilitate easier deconvolution of how the chemical shift evolves.

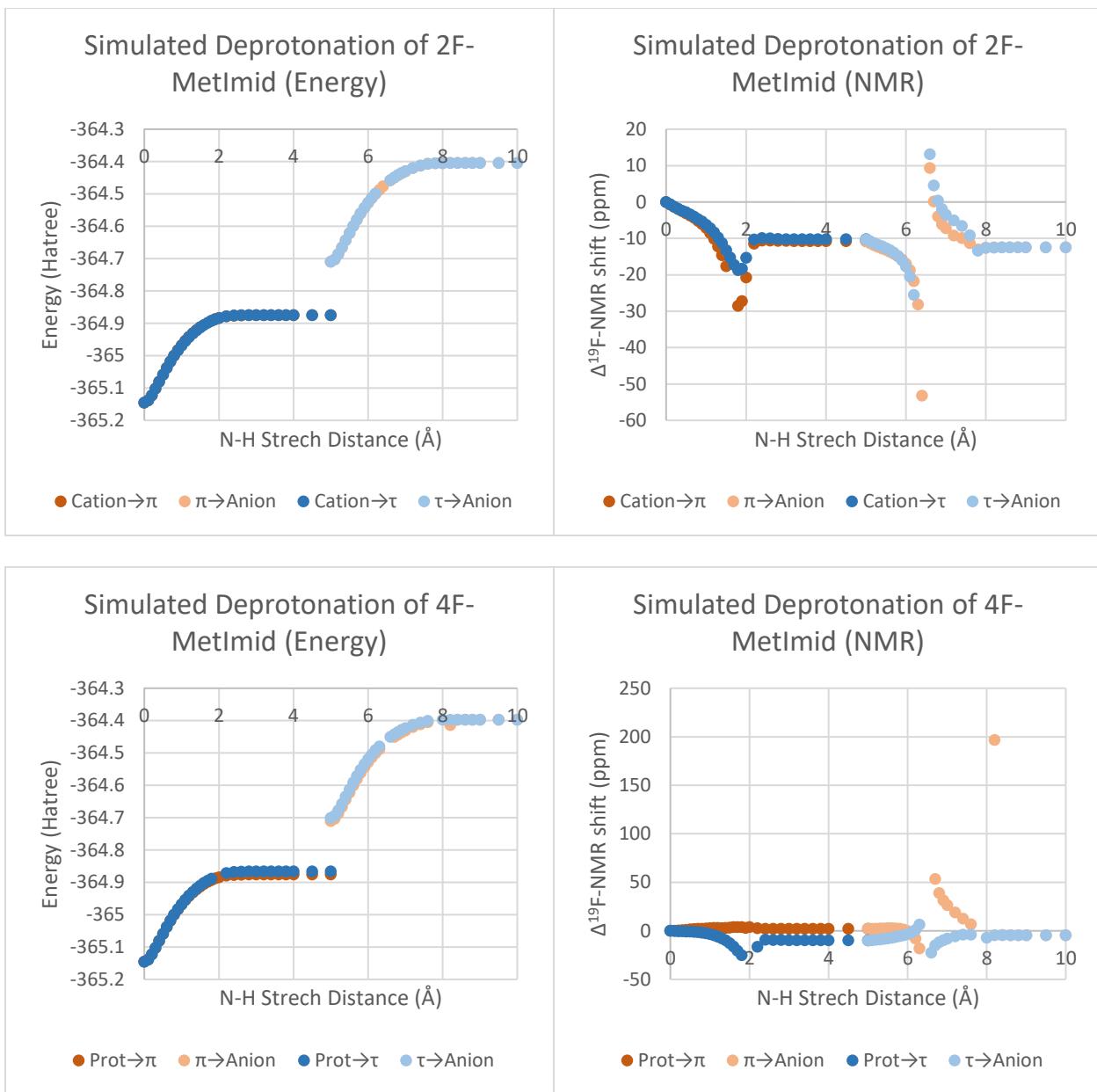


Figure 18: Energies and $^{19}\text{F-NMR}$ values of 2F-MetImid, 4F-MetImid, and 2,4F-MetImid during simulated deprotonation

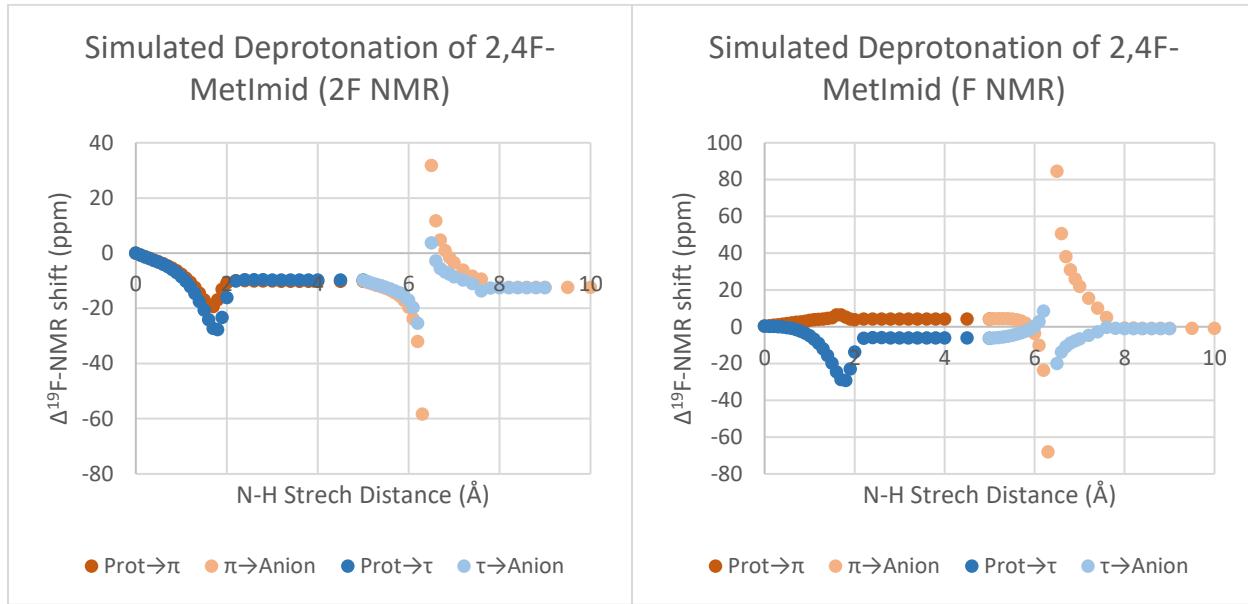
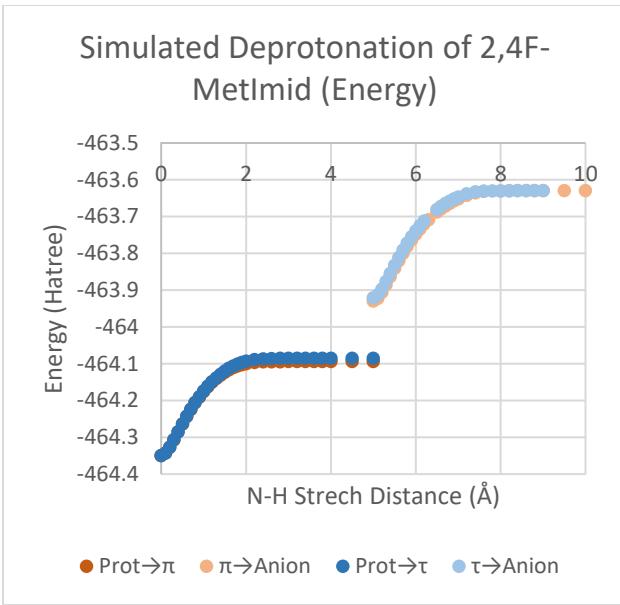


Figure 18 (continued)

The presence of hyperbolas and aberrations appear to be uniformly at the transition point, where the stretched proton is likely undergoing its bonding breaking event. This state is likely poorly modeled by the method chosen, as this sort of event would require multi-reference transition state calculations to properly describe for the purposes refined calculations such as NMR. Interestingly, however, the directionality of the hyperbolic behavior for the 4-position

fluorine depends on which nitrogen is being subjected to deprotonation, which may be a result of whatever underlying phenomena responsible for the reverse NMR chemical shift behavior. The order of deprotonation also seems to matter far more for the 4-position fluorine, whereas the 2-position fluorine behaves similarly regardless of which of the order of deprotonation.

Table 7: Change in ^{19}F -NMR chemical shifts (in ppm) through various deprotonation events.
Deshielding events indicated in orange

	Cation→ π	π →Anion	Cation→ τ	τ →Anion	Cation→Anion
2F-MetImid	-10.76	-1.70	-10.25	-2.22	-12.46
4F-MetImid	2.20	-6.66	-9.81	5.25	-4.51
2,4F-MetImid (F2)	-10.17	-2.25	-9.77	-2.79	-12.49
2,4F-MetImid (F4)	4.15	-5.04	-6.19	5.17	-0.95

As seen above, only the 4-position fluorine underwent deshielding events during the deprotonation process, either when transitioning from the cationic species to π -tautomer, or when transitioning from the τ -tautomer to the anionic species. Analysis of the difference in energy between the π -tautomer and τ -tautomer pathways indicates that the presence of a fluorine at the 4-position causes strong favorability of the π -tautomer pathway, which is consistent with prior observations⁵³.



Figure 19: Boltzmann factor analysis of 2F-MetImid, 4F-MetImid, and 2,4F-MetImid during simulated deprotonation

Electron contribution analyses via Natural Bond Order^{54–60} studies also reveal some interesting features unique to the 4-position fluorine. For example, the methyl carbon (C₆) contributes more electron density to the 2-position fluorine, but changes very little in its contribution to the 4-position fluorine over both pathways, and in the case of 2,4F-MetImid actually ends up withdrawing electron density from the 4-position fluorine relative to the cationic state. The contributions of N-1 also deviate significantly between the 2 and 4-position fluorines.

N-1 starts by contributing electron density to the 2-position fluorine, but starts by withdrawing electron density from the 4-position fluorine. As the deprotonation process progresses, 1-position nitrogen progressively contributes more electron density to the 2-position fluorine as well. However, for the 4-position fluorine, the 1-position nitrogen initial contributes more electron density to the fluorine (still remaining an overall withdrawing species), but then rebounds during the second deprotonation event and withdraws electron density relative to the neutral state. These trends are observed in tandem with the difluorinated species, and are representative to similar effects seen in the monofluorinated species.

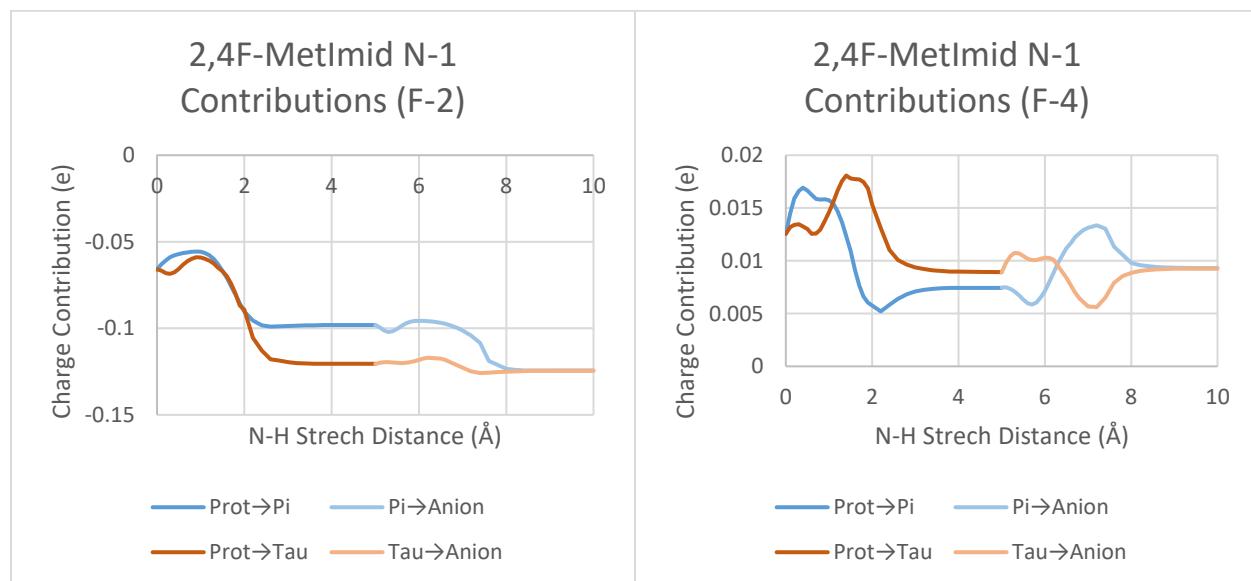


Figure 20: Contributions of N_1 to the electron density of fluorines in 24F-MetImid
However, the true power of molecular orbital visualization provides an intuitive method of observing these changes in a pictorial manner without the need for sifting through large quantities of calculated information. A series of bash and Python scripts (made to adapt with PyMOL) serve to create animations that allow any chemist to intuitively observe how molecular orbitals change over a proposed reaction coordinate.

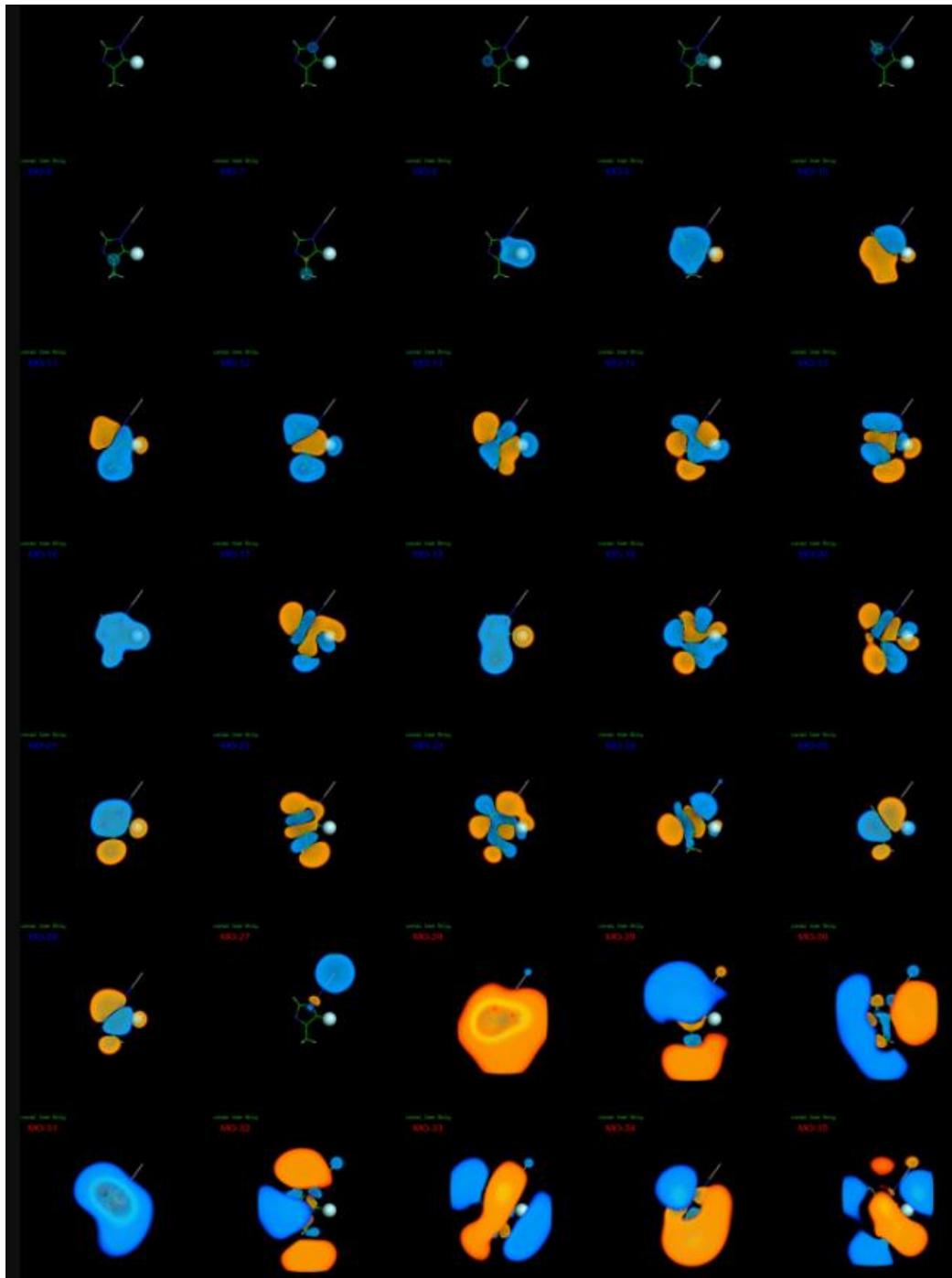


Figure 21: Sample frame of a molecular orbital evolution animation for the deprotonation of 4F-MetImid. Each sub-pane is labeled with its corresponding molecular orbital number. Red labels denote virtual orbitals.

Visualization of the cation to π -tautomer orbital evolution of 4F-MetImid reveals multiple orbitals that lose localization on the fluorine nuclei over the course of the deprotonation event.

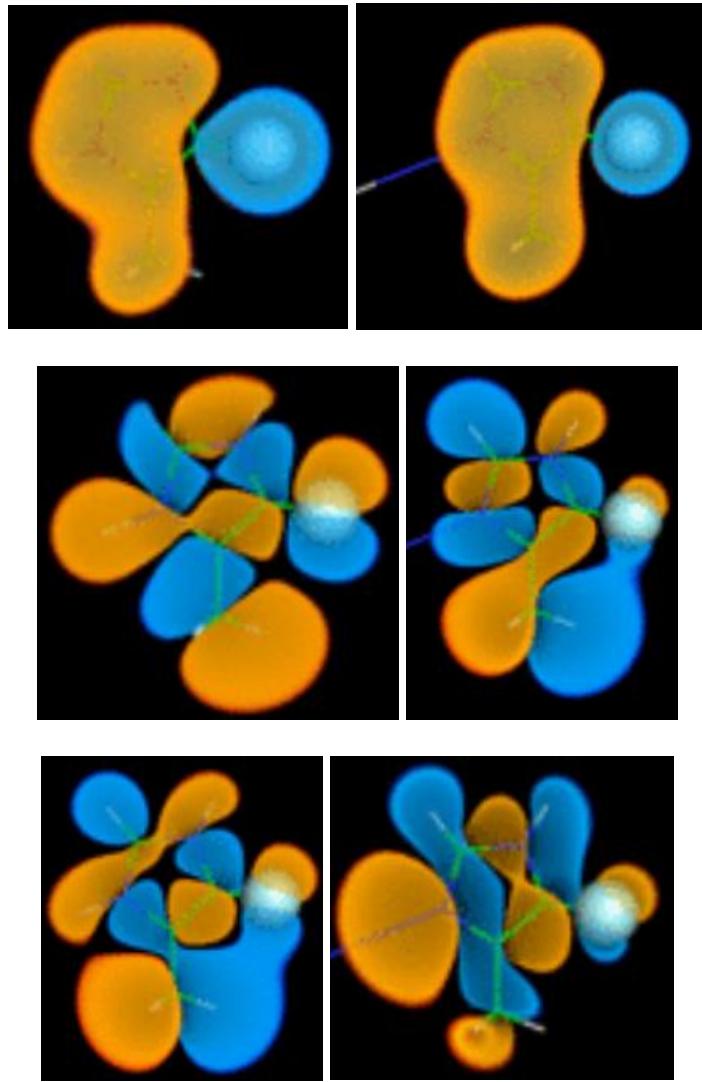


Figure 22: Evolution of MO-20 (Top), MO-23 (Mid), and MO-24 (Bottom) of 4F-MetImid from the cation (Left) to π -tautomer (Right)

These orbitals show a significant change in how much orbital density resides on the fluorine. While there are a few orbitals which show a slight increase in density, these likely do not appear to “equal out” the changes observed in the decreased localization orbitals. In summary, it appears that the unique electronic character and orbital interactions attributed to the

presence of a fluorine in the 4-position of the imidazole ring seems to be one of the core reasons for the reverse chemical shift behavior. Instead of a global increase in orbital density, multiple rearrangements and withdraw events indicate that the fluorine is predicted to experience overall deshielding relative to the protonated state, which is what is observed experimentally.

4.4 Electric Field Effect Results

Fluorine is exquisitely sensitive to its environment, as it has a relatively large electron density packed in a very small locale, which contributes to its wide chemical shift range of almost 350 ppm for C-F bonds. As such, a hypothesis was constructed that a local electric field of modest strength could perturb a fluorinated amino acid ^{19}F -NMR chemical shift in an appreciable fashion. Such an electric field could be potentially generated by partially charged residues of a labeled protein of interest. While it is known that the electronic and stereoscopic features of molecules can be affected by applied electric fields, as is the case in Vibrational Stark Effect (VSE) spectroscopy where IR absorbance features are broadened by the presence of an electric field in isotropic samples of molecules frozen in glasses or in protein interiors⁶¹, no reported data could be found correlating VSE data to NMR chemical shift perturbations. This is likely due to the fact that most VSE data is collected in the solid state with specially designed cells which would be unsuitable for tandem NMR analysis. However, prior research shows that the C-F bond stretch is an excellent VSE reporter on local field strengths⁶², which indicates that fluorinated amino acids may act as dual probes for electric field effects through both NMR and VSE spectroscopy. In order to discern if electric fields are capable of augmenting the native chemical shift value of a fluorinated amino acid, simulations were conducted using Gaussian's electric field function (denoted as Field=X+#, where # denotes the electric field strength applied in 10^{-5} a.u, which equates to approximately 5.14×10^6 V/m) to apply a fixed electric field along

the x-axis of a simulation. Geometric restrictions were applied to prevent possible reorientation due to torque applied from the electric field on the electric dipole of the target molecule.

Fluorobenzene was used as a model system for the initial study. Studies were originally conducted by generating multiple simulation files with rotation of each molecule around the x, y, and z-axes independently of one another to explore how each orientation affected the NMR chemical shift. It was predicted that no dependence on the x-axis rotation should be observed, as the molecule should experience the same effective field orientation regardless of the angle. It was also predicted that shielding should occur when the carbon-fluorine bond vector is aligned with the electric field, as electron density would be “pushed” from the ring system onto the fluorine. Conversely, if the carbon-fluorine bond vector is anti-parallel, deshielding should occur, as electron density from the fluorine would be displaced back onto the ring system.

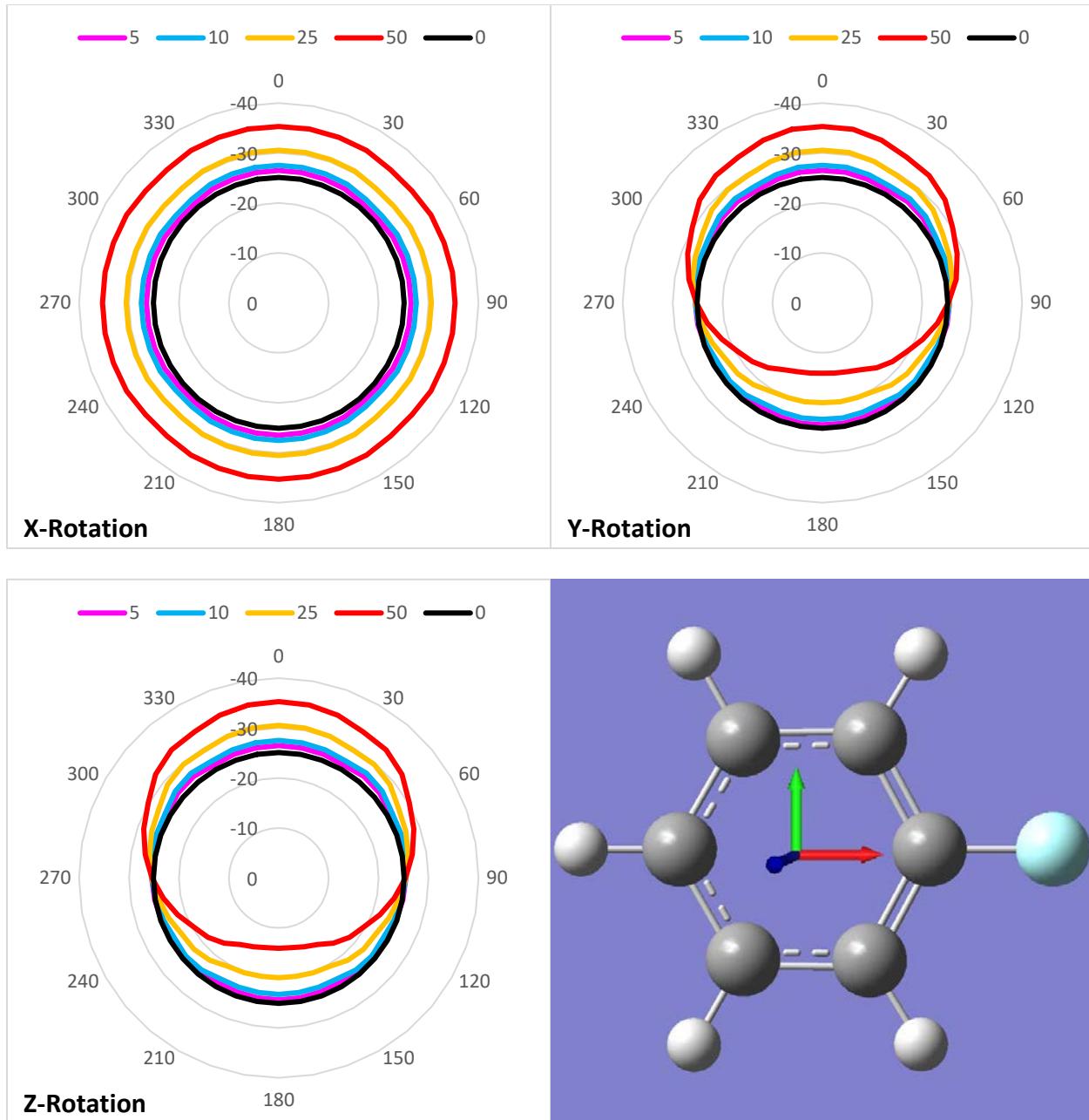


Figure 23: Predicted influence of applied electric field on ^{19}F -NMR chemical shift of fluorobenzene, with a depiction of the x-axis aligned structure. Each series numbers denotes the “Field” value specified in the simulation files. Outer angles denote relative rotation of the carbon-fluorine bond vector from the x-axis vector. Radii labels denote calculated NMR chemical shift value relative to trifluoroacetic acid. Moving inward relative to the zero-field circle (black) implies deshielding, while moving outward implies shielding.

As shown above, many of the predicted features described by classical mechanics principles appear to apply. In all cases, increasing the strength of the applied field increases the

degree of shielding or deshielding observed. Rotation around the x-axis has no effect, while increasing the field strength causes a greater degree of shielding, alignment against the field in the y-axis and z-axis rotations causes a significant degree of deshielding which is dependent on field strength, and there is an overall symmetric and cyclic nature to the chemical shift perturbations. Perturbations ranged from a change of approximately 10 ppm, both shielding and deshielding, depending on orientation and applied field strength.

In order to determine if the vibrational spectra of the molecule are also predicted to change, as seen in VSE spectroscopy, a composite spectrum was generated. A dataset consisting of 500 randomly rotated structures was generated, each with an applied field of 0.0005 a.u along the x-axis. A calculated IR spectra was generated by treating each absorbance as a Lorentzian contribution with a half-height linewidth of 8.0 cm⁻¹. This value was optimized to match the predicted spectra generated by GaussView 5.0 as closely as possible. Each IR absorption was modeled as a Lorentzian using the following equation⁶³:

$$\varepsilon(\tilde{v}) = \frac{200 * \kappa}{\pi * \ln(10)} * \frac{w}{4(\tilde{v} - \tilde{v}_0)^2 + w^2} \quad (4.1)$$

Where κ is the calculated intensity of the absorption, \tilde{v} is the frequency of the absorption, and w is the half-height linewidth (8.0 cm⁻¹). Each absorption spectra was then summed and averaged to form an overall VSE IR spectra. This was then subtracted from a zero-field spectra to generate a VSE difference spectrum.

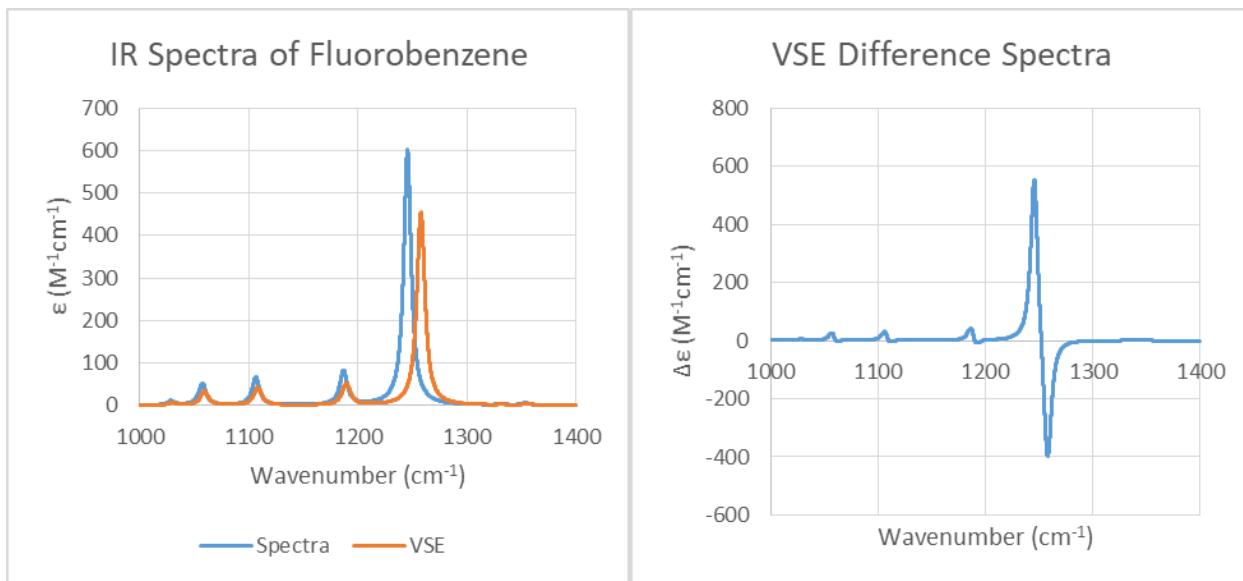


Figure 24: Left: C-F region of predicted IR spectra of fluorobenzene. Right: Simulated Vibrational Stark Effect difference spectra of fluorobenzene

While the line shape does not exactly model observed experimental spectra, the overall features observed are similar in nature. While the dataset was generated in such a way to try and mimic an isotropic mixture as closely as possible, only 1.5% of the possible orientations were sampled (Assuming integer values of y and z rotations ranging from 0 to 180.) With these promising results in hand, a large dataset was generated in order to see if there is a “simple” equation to model how the NMR chemical shift of fluorine changes for a given fluorinated amino acid sidechain analogue. 4-fluorotoluene was chosen for a model of 4-fluorophenylalanine in order to speed up the simulation process by ignoring potential complications from the backbone, but to retain any small delocalization effects than may be caused by the presence of the C β carbon (as observed in the fluorohistidine research above.) Simulation sets were generated for the following SCRF solvents, with the assumption that including such a parameter would act as a proxy for the highly variable dielectric environment in protein interiors: acetone, benzene, N,N-dimethylformamide, dimethylsulfoxide, ethanol, hexane, methanol, pyridine, and water. For each of these, two sets of 500 simulation files were

generated. Each simulation file starts with the 4-fluorotoluene oriented so that the C-F bond vector is placed along the x-axis, with the plane of the ring system contained within the xy-plane. A random rotation (from 0 to 360 by integer value) was then applied, sequentially, about the x-axis, y-axis, and z-axis. The simulation was then given a random Field=X+# tag with # ranging from 0 to 100. Each file was given a unique name which encodes the parameters for the simulation in the following fashion, “x_y_z_field.com,” for easy processing and parameter retrieval. After the full dataset was calculated, an automated Python script was used to filter failed simulations and extract the parameters and relative fluorine NMR chemical shift values (relative to the isotropic tensor value for 4-fluorotoluene without an applied field.)

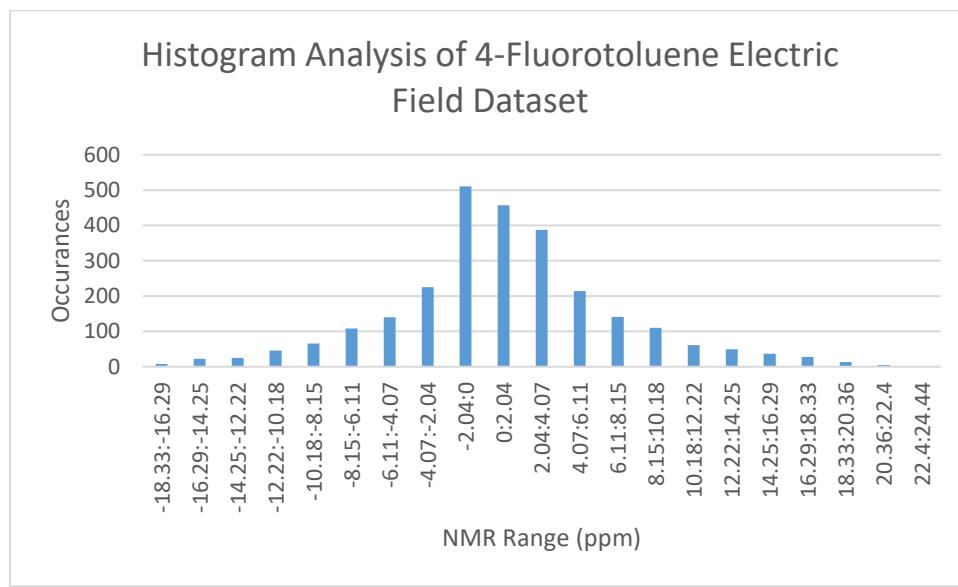


Figure 25: Histogram analysis of the 4-fluorotoluene relative shift perturbation dataset

The dataset appears to show biasing the 0.0 to -2.04 ppm range. This may be due to the limited nature of the dataset, or potential non-uniform sampling. Inheriting biasing may also occur due to the lack of dependence of the NMR chemical shift change on the x-axis rotation. For example, if there is an enrichment in significant rotations around the x, but only a small quantity of rotations around y and/or z, then an increase of shielding occurrences would be expected. This is accounted for in future utilization of the dataset (Described in section 5.4.)

4.5 Conclusions

Prediction of accurate fluorine NMR chemical shifts presents an interesting challenge due to the complex nature of the electronic characteristics of fluorinated small molecules. Through method screening, we have determined computational methods that are capable of predicting reasonably accurate fluorine NMR chemical shifts for relatively low computational cost. This method analysis serves as a foundation for expansion of predictive methods for the fluorine NMR spectra of proteins, such as utilization of ONIOM or other simplification methods.

Development of tools to visualize the evolution of chemical shift influencing events, such as deprotonation and local electric field effects, aid in describing the underlying characteristics of difficult to model systems, such as 4-fluorohistidine. These analyses reveal potential origins for the reverse chemical shift behavior observed in 4-fluorohistidine, demonstrating counterintuitive events of molecular orbital density withdraw upon deprotonation. Application of these techniques to a more diverse array of molecules may afford molecular design principles for steering electron density effects, and could potentially be used for directed synthesis applications (though a much larger dataset would be required for such a grand application.) Probing the effect of local electric fields may also afford an easy to use simplification of a local protein environment, serving to drastically reduce the computational cost associated with protein-based QM/MM simulations.

CHAPTER 5

PRODUCTION OF THE PFNMR SOFTWARE PACKAGE

5.1 Background

As with any NMR labeling project, there is a need to assign measured NMR resonances to their corresponding labeled residues in order to better elucidate protein dynamics changes due to various factors, such as temperature, solvation environment, and binding of cofactors, substrates, and/or partner proteins. The gold standard to tackle this problem for the purposes of fluorotryptophan based labeling has been site-directed mutagenesis swapping of the tryptophan for phenylalanine. Phenylalanine is the closest in size and hydrophobicity to tryptophan, so it presents as the ideal swap partner. By monitoring which resonance vanishes, a labeling assignment can be made. However, the process for doing this can be cumbersome and somewhat expensive, and there is no guarantee that these mutations will leave the protein structure and dynamics unperturbed. As the number of labeling sites increase, so does the amount of time and money required. A more convenient avenue, if protein fluorine NMR were better understood, would be to simply simulate the protein using a molecular dynamics software such as GROMACS⁶⁴, use cluster analysis to generate representative structures of the various protein conformers, and use a quantum mechanical simulation software to simply simulate the NMR spectra. However, proteins present a much larger QM problem than what is traditionally attempted due to the high quantity of atoms, and for especially large proteins, even clever methodology such as ONIOM¹² based calculations would likely pose a significant computational cost. To this end, work has been conducted in order to better understand the origins of fluorine NMR chemical shifts in proteins to refine approximations of these resonances. An avenue of current research has been the potential effect of a local electric field, generated by charged

residues and partially charged atoms, which permeates through the protein environment and may potentially significantly perturb electron densities around fluorine nuclei located in the protein. It is plausible that local electric fields may contribute considerably to the nucleus's chemical shift. These fields will not be averaged out via tumbling, as the fluorine is still contained within the same frame of reference as the tumbling protein, and there is evidence via Vibrational Stark Effect Spectroscopy that local electric fields may be appreciable in strength and may even contribute to some of the catalytic effects of proteins. To this end, work has been conducted in tandem with Aerospace Engineer Bryan Gantt (Masters Graduate, Wichita State University) to develop a simple to use software package to calculate these electric field effects in the hopes of one day easily simulating and generating protein fluorine NMR resonance assignments.

5.2 Methodology

The Protein Fluorine NMR simulation software (PFNMR) has been written in Visual Studio using the C++ programming language in tandem with the CUDA programming language¹³ for GPU based calculations. All open-source code is available under the GNU General Public License at <https://github.com/protein-fnmr/PFNMR>, and is free for use and modification under the provisions of the license. The typical calculations needed for determining properties such as dielectric maps and electric field influences are ideal for CUDA due to their highly parallel nature, resulting in an up to 50-fold decrease in calculation time for a 3D dielectric map calculation (1 month to roughly 15 hours.) The first step in the calculation pathway is to generate dielectric values along the paths from each protein atom to the fluorine nucleus in question (excluding the atoms included in the labeled amino acid; each fluorine nucleus is treated in a separate calculation set.) A method for calculating a smooth dielectric in proteins has been previously described by Lin Li et. al.¹⁴ A portion of code was written to

generate a uniform map of the dielectric environment inside of the protein in order to validate the code written is generating maps similar to what were seen in the original method paper. Slices are taken through the protein at regularly spaced intervals and treated as frames for an animation, similar to a “CT scan” of the protein, where the dielectric values are visualized instead. The methodology described below is taken directly from Lin Li et. al., with small variances made to provide tuning factors. Ongoing efforts are being made to determine if these tuning factors provide a meaningful impact on calculation accuracy, and so the core equations are only described here. Each slice is populated with a regularly spaced grid of points where a dielectric is to be calculated, and the following equations are used to calculate the dielectric at each point:

$$\rho_i(r) = \exp\left(-\frac{r_i^2}{\sigma^2 * R_i^2}\right) \quad (5.1)$$

Where $\rho_i(r)$ is the density at position r , r_i is the distance between atom i and position r , R_i is the van der Waal’s radius of atom i , and σ is the relative variance (A simple tuning factor described in the original publication, in most cases this has a value 0.93 as determined by pKa calculations in the original method literature.)

$$\rho(r) = 1 - \prod_i [1 - \rho_i(r)] \quad (5.2)$$

Where $\rho(r)$ is the density from all atoms in the system (excluding atoms contained within the fluorine labeled residue) and $\rho_i(r)$ is the result of equation 1 from above for each atom.

$$\varepsilon = \rho * \varepsilon_{in} + (1 - \rho) * \varepsilon_{out} \quad (5.3)$$

The dielectric distribution function, ε , is calculated using the density (ρ , the result of equation 2 from above), ε_{in} , which is the reference dielectric value when the density is 1 (usually 4.0), and ε_{out} , which is the reference dielectric value for the water phase (usually 80.4.)

Unfortunately, protein environments result in highly non-uniform dielectric gradients, which make it difficult to calculate accurate electric fields. To this end, two different methods are used to calculate the dielectric influence of each atom to the fluorine nucleus in question. Both of these methods rely on the so-called “Effective Length” method, and differ in how they approach the integration process. Given a distance D between two charged particles, we assume that the dielectric \mathcal{E} at distance x along D is given by some function $f(x)$. For the purposes of calculation simplification, the square root of the dielectric is calculated at each point x so that squaring the integrated result gives the equivalent of $\mathcal{E}D^2$. To calculate the effective length, L , the following equation is utilized:

$$L = \int_0^D f(x)dx \quad (5.4)$$

In the first integration method, roughly 1000 equally spaced points are taken along the path from an influencing atom to the fluorine, each point has a dielectric value generated for it, and trapezoid based integration is used to generate the integral of the dielectric along the path. In the second method, Gaussian quadrature integration is used to find the integral of the dielectric along the path. Currently, a 10-point Gaussian quadrature calculation is utilized, but work is being conducted to implement an adaptive method that should dynamically provide refined results. The Gaussian quadrature method has the advantage that fewer points are needed to calculate the integral value as opposed to the trapezoid method, leading to an appreciable decrease in calculation time. The points taken along the path are based on pre-calculated position values known as “abscissa”, where the distance coordinates are converted to be in a -1 to 1 regime. The following image is provided to better explain the implications of this:

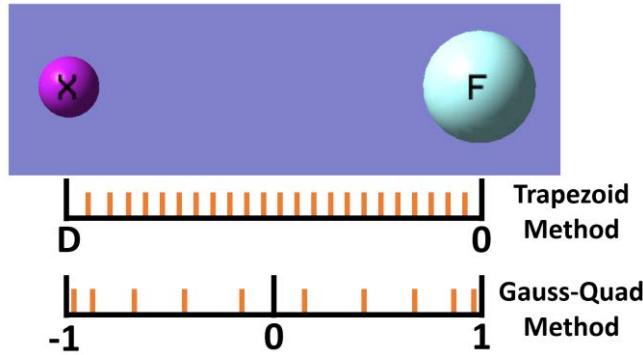


Figure 26: Illustration comparing integration method distance fittings

Each abscissa point's dielectric value is treated with a weight value, and then applied to the following equation:

$$\int_{-1}^1 f(x)dx = \sum_{i=1}^n w_i f(x_i) \quad (5.5)$$

Where w_i is the abscissa weighting factor, $f(x)$ is the function describing the dielectric behavior, and x_i is the abscissa in question. After the effective length is calculated, the electric field contribution of each atom is calculated via:

$$F = \frac{kQ}{L^2} \quad (5.6)$$

Where F is the resulting field vector, k is Coulomb's constant, Q is the charge of the influencing atom, and L is the effective length. Partial charge parameters for each residue are provided as an additional input for the charge value⁶⁵. These field vectors are decomposed into their individual x, y, and z components, the components along each axis are summed together, and a net field vector is calculated through the fluorine nuclei. The geometry of the fluorinated residue is then extracted, and the field vector is aligned to be along the x-axis. This is done because Gaussian 09 only permits linear electric fields along the cardinal axes. The rotation matrix for rotating the C-F bond vector to the x-axis is then also solved in order to extract the x, y, and z-axis rotation angles. In addition to the electric field surrounding the fluorine nucleus,

the average dielectric of the space of the fluorine nucleus is also calculated. These parameters are then used in a fitted NMR prediction equation (described below) to calculate the final predicted NMR value. This process is then repeated for each fluorinated amino acid, and the corresponding geometry, dielectric, electric field, and NMR values are recorded in an output file.

5.3 Dielectric Environment Results

For the purposes of determining if the software created is properly generating dielectric values, animations were generated to visualize the dielectric map of YqhD. Shown below are a few frames of a sample animation, with the frames being taken along the x-axis of a modified version of the 1OJ7 PDB structure from the RCSB:

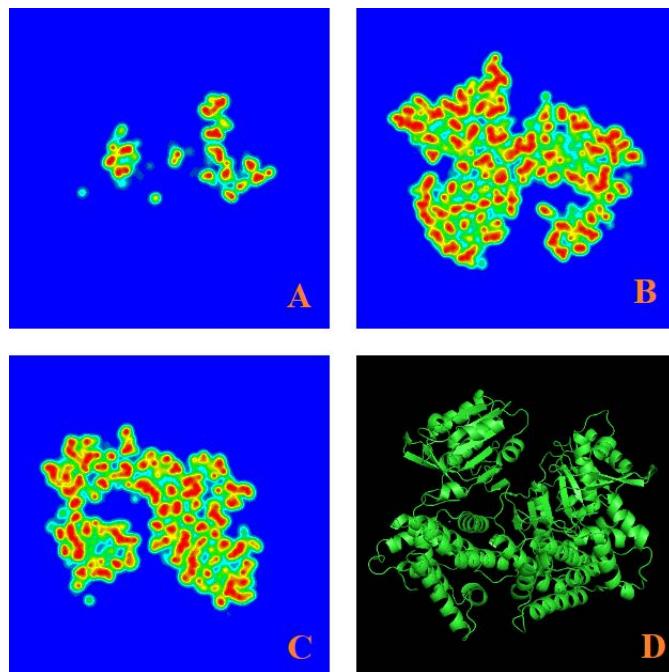


Figure 27: Dielectric environment visualization of YqhD: (PDB: 1OJ7) at various planes through the protein. Blue indicates a water-like dielectric (80.0) while red indicates a hexane-like dielectric (8.0)

Frame A is just as the calculation begins entering the protein, B is roughly 25% through the protein and shows one of the NADPH binding pockets in the center right, C shows the NADPH binding pocket of the other subunit in the center left, and D is a visualization of the

cartoon structure of the input PDB in PyMOL. These maps look very similar to those of other proteins in the published methodology paper from Lin Le, and provided good evidence that the program is accurately performing the calculations desired. In order to test the robustness of the software, a calculation was done on the GroEL/ES complex using the 1AON PDB file from the RCSB. The animation took approximately 3 hours to generate, and small optimizations have since been done to improve performance:

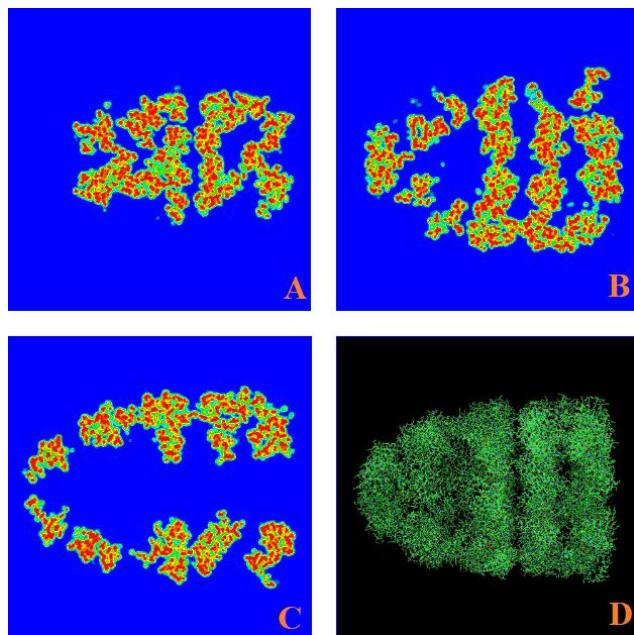


Figure 28: Dielectric environment visualization of GroES (PDB: 1AON) at various planes through the protein. Blue indicates a water-like dielectric (80.0) while red indicates a hexane-like dielectric (8.0)

Again, Frame A is just as the calculation begins to enter the protein, B shows the beginning of the GroES region cap space, C is roughly half way through the protein and shows the intra-chaperon space, and D shows an image of the PDB visualization in PyMOL for reference. This calculation serves to show that large scale calculations are readily accessible to the PFNMR software developed thus far. Many major improvements can still be made to optimize the calculation process, such as multi-GPU utilization or distributed load amongst

multiple computers, and should the preliminary results of the software look promising, these optimizations will be approached.

5.4 Preliminary Electric Field and NMR Prediction Results

To attempt to simplify the calculation process of relative NMR chemical shifts, an equation set was generated to determine the relationship between various geometric and environmental parameters and predicted NMR chemical shift perturbations. The 4-fluorotoluene dataset (described in section 4.4) was chosen in order to determine if a simple equation accurately models the perturbations observed. This dataset was processed through the Eureqa software package to determine equations that model the relationship between 5 parameters (rotations around the x, y, and z-axes, dielectric, and field strength) and the effect on the NMR chemical shift. From this set, two equations were chosen to test and optimize for the PFNMR program. The first equation chosen was based on its position on the complexity vs. error curve, as it appeared to have the largest increase in accuracy for the lowest amount of complexity, while the second equation was chosen based on the lowest amount of error with a “reasonable” complexity.

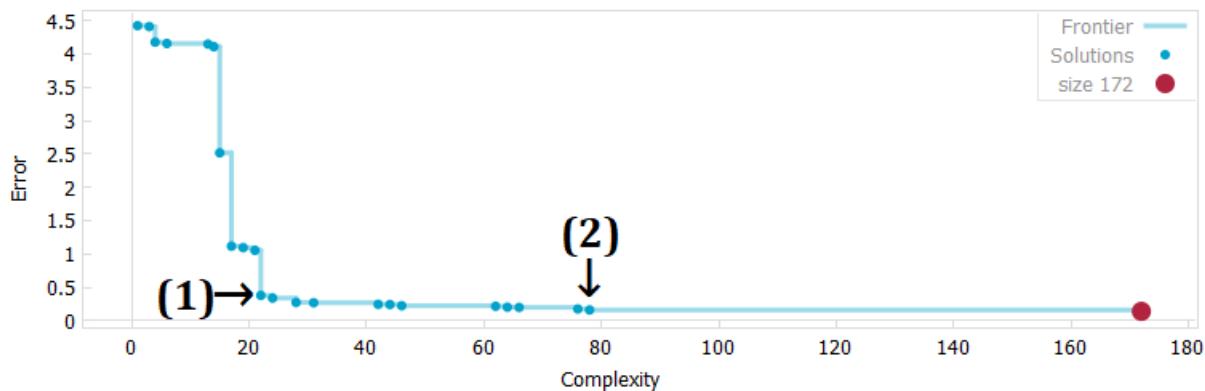


Figure 29: Complexity v.s error curve from Eureqa fitting data. Simple and complex equation positions on the curve indicated as (1) and (2), respectively

$$n_{simple} = \frac{5.579}{d} - (0.2069 * w * \cos(0.01745 * y) * \cos(0.01745 * z)) \quad (5.7)$$

$$\begin{aligned} n_{complex} &= (0.005932 * w) \\ &+ (0.01745 * w * \cos(0.01745 * y)^2 * \cos(0.01745 * z)^2) \\ &+ \frac{6.319 + 0.04333 * w * \cos(0.01745 * y) * \cos(0.01745 * z)}{d} - 0.6561 \\ &- 0.2137 * w * \cos(0.01745 * y) * \cos(0.01745 * z) \end{aligned} \quad (5.8)$$

These were used in the PFNMR program to convert the original output parameters to a predicted NMR value for each resonance. Other parameters and scaling factors were introduced in order to tweak various points of the calculation process. These were introduced to attempt to compensate for inadequacies in the approximations made for the dielectric environment, as well as to try and compensate for potential mid-range effects. Each calculation in the optimization process takes approximately 2 seconds, which is representative of what speed would be expected in the final version of the optimized calculation. These were optimized by using an adaptive steepest descent optimization algorithm, akin to what is used in standard machine learning processes, to reduce the relative error in chemical shift value calculated (in ppm). For initial optimization, rat Intestinal Fatty Acid Binding Protein (IFABP) was chosen, as using this structure would allow for simultaneous optimization of eight 4-fluorophenylalanine residues, and the individual residues have been assigned using traditional methods⁴. The solved crystal structure (PDB: 1IFB) was taken from the RCSB and modified so that each phenylalanine was replaced with 4-fluorophenylalanine. This structure was then used for parameter optimization over the course of 2 days. Simulated were generated by treating each resonance as a Lorentzian for the purpose of visual comparison.

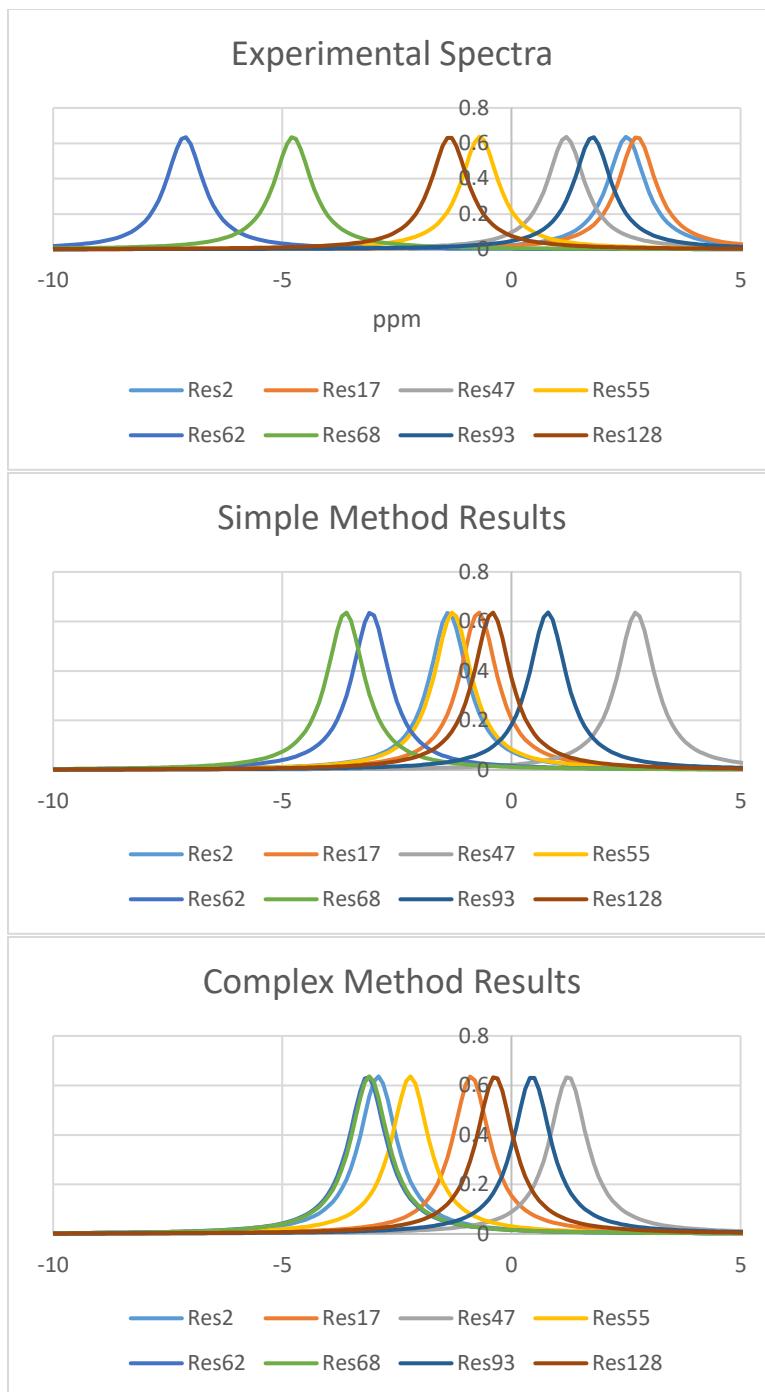


Figure 30: Comparison of IFABP reconstructed experimental ^{19}F -NMR spectrum with PFNMR predicted spectrum using optimized simple and complex calculation methods

Simple and complex denote which equation was used during the end stage of the calculation to convert geometric and electric field parameters to NMR chemical shift values. Utilization of the simple method results in an average error of 2.07 ppm, while the complex

method yields an average error of 2.32 ppm. The overall spectrum from using the simple equation seems to mimic the assignment order more closely than the complex equation. This may be an indication of the more overfit nature of the complex equation, which would presumably lead to inadequacies in diverse protein environments. In order to help validate the robustness of the optimized parameters and simple equation, the same optimized parameters from the IFABP calculation were used to predict the ^{19}F -NMR spectra of PapD⁶⁶. There may be a small degree of offset error, as the resonance assignment values were not fully tabulated.

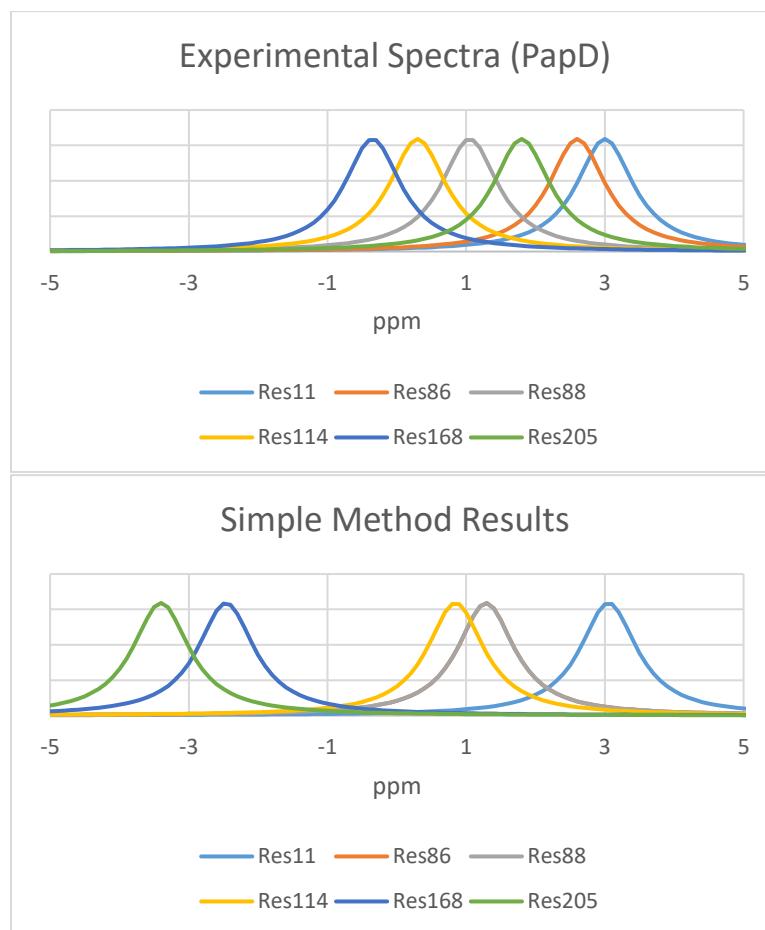


Figure 31: Comparison of reconstructed PapD experimental ^{19}F -NMR spectrum to a PFNMR predicted spectrum with parameters optimized from IFABP (Res86 and Res88 overlap)

The predict spectrum has an average error of 1.58 ppm, and removal of residue 205 from the error calculation as an outlier yields an error of 0.85 ppm. Interestingly, except for residue

205, the predicted resonances appear in the same order as the experimental spectrum. While the PFNMR software, as it stands, is not accurate enough to perform routine resonance assignments, these PapD results are quite promising. Further refinement of the fitting datasets, program architecture, and computational methodology will, hopefully, improve the assignment accuracy even further.

5.5 Conclusions

Advancements in computational power, in particular the advent of easy to use GPU based calculation methods such as CUDA, provide rapid calculation methods for highly complex problems. Leveraging this, the PFNMR software was developed with the intention of providing a rapid method for performing protein ^{19}F -NMR assignments without the need of experimental methods such as site-directed mutagenesis. While the current rendition of the software is not accurate enough to perform assignments of spectra, it is a good step forward in a simple and rapid fluorine assignment method. Using the NMR translation equations in a random fashion provides typical errors in the range of 30 ppm, while using the parameters generated in the software narrow this error range to ~2.2 ppm, and the software is able to complete a full calculation of eight residues in a 131 residue protein in roughly 2 seconds, a massive improvement to more traditional QM/MM approaches. Through expansion of fitting datasets, incorporation of new data such as direct electric potential measurements⁶⁷ and Vibrational Stark Effect measurements⁶⁸, and refinement of computational methodology, this software may afford a new, powerful tool in the “NMR Toolbox” in the relatively near future.

CHAPTER 6

FUTURE DIRECTION

6.1 Exploration of the Fundamental Origins of ^{19}F -NMR Chemical Shifts

Working towards a versatile and uniform method of simulating and predicting ^{19}F -NMR chemical shifts will provide a useful tool for biophysical analysis of proteins, interpretation of fluorinated drug binding, and a myriad of other problems faced by researchers. Understanding the origins of behavior contradictory to chemical intuition may also aid those in the realm of synthesis and molecular characterization. There is a need for studying these fluorinated molecular systems in a more in-depth fashion, both from an experimental and computational perspective. As more labeled protein spectra are published, and more interesting molecules are discovered, our techniques for understanding the origins of chemical shift behavior will become enhanced. These may afford a deeper understanding on the nature of the electronic systems of molecules, which play a major role in governing chemical shift behavior. This could potentially enable the development of more focused methods for protein engineering (by better understanding the environment in which these fluorine labels are placed), as well as synthetic design principles.

6.2 Generation of Diverse Datasets for Electric Field Analysis

Generation and testing of larger and more diverse datasets demonstrating the effects of electric fields on molecules can provide a wealth of knowledge. There are still many questions relating to the origins of ^{19}F -NMR chemical shifts that have a potential wider impact on our understanding of molecular stereoelectronics. For example, is there a “single equation” that describes ^{19}F -NMR chemical shift changes in proteins with a wide array of fluorinated amino acids, and not just 4-fluorophenylalanine alone? Can we accurately predict the Vibrational Stark

Effect spectra of a molecule, and thus predict how a given hypothetical molecule will behave in an electric field? Is there an underlying cause for reverse chemical shift behavior which could be used as a design principle for new non-canonical amino acids? These ideas and principles could potentially be incorporated into new and interesting design paradigms, such as designing small molecule and biochemical catalysts that leverage electric field interactions for the purpose of selective bond polarization. While the research contained here is still leagues away from this sort of goal, these preliminary results and methods provide a simple and relatively “hands-free” approach to creating and parsing large datasets in a “chemical intuitive” fashion. As the quantity and quality of these datasets improves, new trends could potentially be observed which could afford answers to the questions described above.

6.3 Expansion of the Biochemical NMR “Toolbox”

Developing new techniques and methods for studying biomolecules allows chemists to understand, in a more fundamental fashion, important issues facing the realms of biocatalysis development, drug design, and more. While ^{19}F -NMR has its limitations, it still provides a powerful method for studying drug interactions and protein dynamics in a targeted fashion. These studies can be used synergistically with computational studies to afford a deeper understanding of what is occurring at the molecular level when, say, a drug molecule binds to a protein. These new techniques afford researchers a wider array of tools to study difficult to probe proteins, such as the utilization of site-specific labeling techniques to limit stability perturbations in a protein system where global labelling causes misfolding or abolishes catalytic activity. It is our hope that the research contained within this document will enable a deeper understanding of the molecular world around us through the lens of NMR and computational studies.

REFERENCES

REFERENCES

1. Jarboe, L. R. YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol.* **89**, 249–257 (2011).
2. Li, H., Chen, J. & Li, Y. Enhanced activity of yqhD oxidoreductase in synthesis of 1,3-propanediol by error-prone PCR. *Prog. Nat. Sci.* **18**, 1519–1524 (2008).
3. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The Preparation of 19F-Labeled Proteins for NMR Studies. *Methods Enzymol.* **380**, 400–415 (2004).
4. Li, H. & Frieden*, C. NMR Studies of 4-19F-Phenylalanine-Labeled Intestinal Fatty Acid Binding Protein: Evidence for Conformational Heterogeneity in the Native State†. (2005). doi:10.1021/BI047600L
5. Ropson, I. J., Frieden, C., Hultgren, S. J. & Frieden, C. Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by 19F NMR. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7222–6 (1992).
6. Shu, Q. & Frieden*, C. Urea-Dependent Unfolding of Murine Adenosine Deaminase: Sequential Destabilization As Measured by 19F NMR†. (2004). doi:10.1021/BI035651X
7. Sun, X., Dyson, H. J. & Wright, P. E. Fluorotryptophan Incorporation Modulates the Structure and Stability of Transthyretin in a Site-Specific Manner. *Biochemistry* **56**, 5570–5581 (2017).
8. Kwon, I. & Tirrell, D. A. Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells. *J. Am. Chem. Soc.* **129**, 10431–10437 (2007).
9. Yeh, H. J. C., Kirk, K. L., Cohen, L. A. & Cohen, J. S. 19F and 1H nuclear magnetic resonance studies of ring-fluorinated imidazoles and histidines. *J. Chem. Soc. Perkin Trans. 2* **0**, 928 (1975).
10. Giam, C. S. & Lyle, J. L. Medium effects on the fluorine-19 magnetic resonance spectra of fluoropyridines. (2002). doi:10.1021/JA00791A028
11. Adcock, W. & Abeywickrema, A. N. Concerning the origin of substituent-induced fluorine-19 chemical shifts in aliphatic fluorides: carbon-14 and fluorine-19 nuclear magnetic resonance study of 1-fluoro-4-phenylbicyclo[2.2.2]octanes substituted in the arene ring. *J. Org. Chem.* **47**, 2945–2951 (1982).
12. Dapprich, S., Komáromi, I., Byun, K. S., Morokuma, K. & Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct. THEOCHEM* **461–462**, 1–21 (1999).
13. Nickolls, J., Buck, I., Garland, M. & Skadron, K. Scalable Parallel Programming with CUDA. *Queue* **6**, 40–53 (2008).

14. Li, L., Li, C., Zhang, Z. & Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **9**, 2126–2136 (2013).
15. Evans, P. R. & IUCr. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 282–292 (2011).
16. Evans, P. & IUCr. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
17. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135 (2001).
18. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–3 (2012).
19. Evans, P. Resolving Some Old Problems in Protein Crystallography. *Science (80-.)* **336**, 986–987 (2012).
20. Fjeld, C. C., Birdsong, W. T. & Goodman, R. H. Differential binding of NAD⁺ and NADH allows the transcriptional corepressor carboxyl-terminal binding protein to serve as a metabolic sensor. *Proc. Natl. Acad. Sci.* **100**, 9202–9207 (2003).
21. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).
22. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9482–7 (2006).
23. Buller, A. R. *et al.* Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14599–604 (2015).
24. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R. & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie Int. Ed.* **55**, 11577–11581 (2016).
25. Frisch, M. J. *et al.* Gaussian~09 Revision E.01.
26. Pople, J. A. & Nesbet, R. K. Self-Consistent Orbitals for Radicals. *J. Chem. Phys.* **22**, 571–572 (1954).
27. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **23**, 69–89 (1951).
28. Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618–622 (1934).

29. Head-Gordon, M., Pople, J. A. & Frisch, M. J. {MP}2 energy evaluation by direct methods. *Chem. Phys. Lett.* **153**, 503–506 (1988).
30. Head-Gordon, M. & Head-Gordon, T. Analytic {MP}2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.* **220**, 122–128 (1994).
31. Almlöf, S. S. J. Avoiding the integral storage bottleneck in {LCAO} calculations of electron correlation. *Chem. Phys. Lett.* **154**, 83–89 (1989).
32. Frisch, M. J., Head-Gordon, M. & Pople, J. A. A direct {MP}2 gradient method. *Chem. Phys. Lett.* **166**, 275–280 (1990).
33. Frisch, M. J., Head-Gordon, M. & Pople, J. A. Semi-direct algorithms for the {MP}2 energy and gradient. *Chem. Phys. Lett.* **166**, 281–289 (1990).
34. Becke, A. D. Density-functional thermochemistry. {III}. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
35. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
36. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
37. Miehlich, B., Savin, A., Stoll, H. & Preuss, H. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **157**, 200–206 (1989).
38. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **120**, 215–241 (2007).
39. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **78**, 1396 (1997).
40. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
41. Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992).
42. Perdew, J. P., Burke, K. & Wang, Y. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B* **54**, 16533–16539 (1996).
43. Chai, J.-D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **128**, 84106 (2008).

44. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
45. Barone, V. & Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **102**, 1995–2001 (1998).
46. Cossi, M., Rega, N., Scalmani, G. & Barone, V. Energies, structures, and electronic properties of molecules in solution with the C- $\{\text{PCM}\}$ solvation model. *J. Comput. Chem.* **24**, 669–681 (2003).
47. McWeeny, R. Perturbation Theory for the Fock-Dirac Density Matrix. *Phys. Rev.* **126**, 1028–1034 (1962).
48. Ditchfield, R. Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* **27**, 789–807 (1974).
49. Wolinski, K., Hinton, J. F. & Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for $\{\text{NMR}\}$ chemical shift calculations. *J. Am. Chem. Soc.* **112**, 8251–8260 (1990).
50. Cheeseman, J. R., Trucks, G. W., Keith, T. A. & Frisch, M. J. A comparison of models for calculating nuclear magnetic resonance shielding tensors. *J. Chem. Phys.* **104**, 5497–5509 (1996).
51. Caulkins, B. G. *et al.* NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **138**, 15214–15226 (2016).
52. Kasireddy, C., Ellis, J. M., Bann, J. G. & Mitchell-Koch, K. R. The Biophysical Probes 2-fluorohistidine and 4-fluorohistidine: Spectroscopic Signatures and Molecular Properties. *Sci. Rep.* **7**, 42651 (2017).
53. Kasireddy, C., Ellis, J. M., Bann, J. G. & Mitchell-Koch, K. R. Tautomeric stabilities of 4-fluorohistidine shed new light on mechanistic experiments with labeled ribonuclease A. *Chem. Phys. Lett.* **666**, 58–61 (2016).
54. Foster, J. P. & Weinhold, F. Natural hybrid orbitals. *J. Am. Chem. Soc.* **102**, 7211–7218 (1980).
55. Reed, A. E. & Weinhold, F. Natural bond orbital analysis of near-Hartree–Fock water dimer. *J. Chem. Phys.* **78**, 4066–4073 (1983).
56. Reed, A. E., Weinstock, R. B. & Weinhold, F. Natural population analysis. *J. Chem. Phys.* **83**, 735–746 (1985).
57. Reed, A. E. & Weinhold, F. Natural localized molecular orbitals. *J. Chem. Phys.* **83**, 1736–1740 (1985).

58. Carpenter, J. E. & Weinhold, F. Analysis of the geometry of the hydroxymethyl radical by the “different hybrids for different spins” natural bond orbital procedure. *J. Mol. Struct. THEOCHEM* **169**, 41–62 (1988).
59. Reed, A. E., Curtiss, L. A. & Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88**, 899–926 (1988).
60. *The Structure of Small Molecules and Ions*. (Springer US, 1988). doi:10.1007/978-1-4684-7424-4
61. Chattopadhyay, A. & Boxer, S. G. Vibrational Stark Effect Spectroscopy. *J. Am. Chem. Soc.* **117**, 1449–1450 (1995).
62. Suydam, I.T. & Boxer*, S. G. Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins†. (2003). doi:10.1021/BI0352926
63. Spanget-Larsen, J. *Infrared Intensity and Lorentz Epsilon Curve from ‘Gaussian’ FREQ Output*. (2015). doi:10.13140/RG.2.1.4181.6160
64. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
65. Thomas, A., Milon, A. & Brasseur, R. Partial atomic charges of amino acids in proteins. *Proteins Struct. Funct. Bioinforma.* **56**, 102–109 (2004).
66. James G. Bann*, ‡ and & Frieden*, C. Folding and Domain–Domain Interactions of the Chaperone PapD Measured by ¹⁹F NMR†. (2004). doi:10.1021/BI048614U
67. Lu, J., Kobertz, W. R. & Deutsch, C. Mapping the Electrostatic Potential within the Ribosomal Exit Tunnel. *J. Mol. Biol.* **371**, 1378–1391 (2007).
68. Fried, S. D. & Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **48**, 998–1006 (2015).