CHARACTERIZATION AND DESIGN OF HYDROGEN BONDING INTERACTIONS IN OXYGEN REDUCTION BY ENGINEERED MYOGLOBINS

BY

IGOR D. PETRIK

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry in the Graduate College of the University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Yi Lu, Chair, Director of Research
Professor Robert B. Gennis
Professor Thomas B. Rauchfuss
Professor Emad Tajkhorshid
Abstract

Proteins participate in nearly every function of living cells, and an estimated 50% are associated with a metal ion or cofactor. Most, if not all, of the fundamental chemical processes necessary for life, including photosynthesis and respiration, are facilitated by one or more metalloenzymes or metalloproteins. Metal-nucleic acid interactions are also clinically important in cancer and infection treatment, and are useful in industrial purposes. The relationship between macromolecule and metal can almost be viewed as symbiotic – metals often function at extreme limits when free in solution, with minimal ligands; on the other had, amino and nucleic acid polymers have limited functional capabilities. By capturing and precisely holding metal ions, biomolecules tame their reactivities by tuning their electronic and geometric – and therefore functional – properties, allowing the metal and protein together to perform precise chemical functions. However, despite decades of research, the intricate details of how this precision is imparted, in each case and more generally, are not entirely understood. In particular, the roles of long-range (beyond 1-2 bonds), non-covalent interactions are understood to be important; however, exactly how they impart function, is difficult to disentangle, due to the delicate balance resulting from evolution. Moreover, many of these systems, such as the oxidases and photosystems responsible for respiration and photosynthesis, respectively, are large, difficult to purify, and contain multiple cofactors that complicate investigations of the active site chemistry.

Therefore, in the Lu lab, we have been pioneers of the “biosynthetic” modeling approach, in which the active sites of complex enzymes are structurally reconstructed in small, experimentally tractable proteins. In this vein, the active site of heme-copper oxidases (HCOs) – a bimetallic heme-copper center known as CuB – was previously structurally modeled in sperm whale myoglobin, by introducing two additional histidine residues, giving a protein called CuB Mb. Further structural modeling of this protein to include additional structural features, such as a tyrosine residue, brought about not only partial HCO-like oxygen reduction activity, but an unexpected observation – the Cu ion was found to not be necessary for imparting this function, nor did it seem to improve it, in the designed protein. On the other hand the presence and positioning of the tyrosine had a dramatic effect on this activity.

This thesis describes efforts to first understand the reaction of these CuB Mbs with oxygen. It is found that the extended hydrogen bonding network, stabilized by the introduced residues and consisting of water molecules, is critical for imparting oxygen reduction activity to an oxygen binding protein. Based on this observation, further designs are pursued to improve this hydrogen-bonding
network, in an effort to improve activity. It is found that incorporation of a glutamate within hydrogen bonding distance of one of the water molecules composing the extended hydrogen bonding network improves the function of the enzyme by eliminating the ROS release. Moreover, spectroscopic and crystallographic evidence support that this improvement is due to enhanced hydrogen bonding and protonation to the oxygen intermediate. This design demonstrates the critical importance of long-range interactions in enzymes, and should inspire enzyme engineers to pursue incorporation of these types of features.

In addition to these studies, this thesis also reports the computational design of a novel heterobimetallic heme-Fe₄S₄ site, and progress in obtaining the first crystal structure of a metal mediated catalytic DNA construct.
To my parents, Lesya and Oleg, who have supported my curiosity and ambitions. To my sister, Julianna, so she never forgets to pursue her dreams. And to Allan Rozarto – who was and is like a brother to me – for his constant encouragement and philosophical stimulation.

I hope I can return the favor.
ACKNOWLEDGEMENTS

None of our accomplishments would be achievable without the help of many others. Here, I would like to acknowledge everyone who helped me on my journey from Bachelor to Doctor.

First and foremost, I want to thank my research advisor, Professor Yi Lu, for his intellectual and financial support, patience, encouragement, and helpful advice. Thank you for looking out for me and helping me through this process. Additionally, I would like to thank my committee members, Drs. Robert Gennis, Thomas Rauchfuss, and Emad Tajkhorshid, for their insightful guidance and helpful discussion.

I would also like to thank my undergraduate research advisor, Professor Guillermo Moyna, as well as the other professors in the Department of Chemistry at the University of the Sciences in Philadelphia, particularly Preston Moore, Edward Birnbaum, Vojislava Pophristic, Elisabetta Fasella, and Frederick Schaefer, for giving me a foundation without which I would not have gotten into the University of Illinois in the first place.

Most of the work presented here would not have been possible without the help of a large number of collaborators, multi-user facility scientists, and experts, and I would like to thank them in turn. Drs. Roman Davydov and Brian Hoffman from Northwestern University for helping me with EPR investigations. Drs. Vukica Srajer and Robert Henning of BioCARS at the APS for going above and beyond to help me carry out MX+Spec studies. Drs. K. Rajashankar, Frank Murphy, Jonathan Schuermann, and Surajit Banerjee of NECAT at the APS for assistance with DNAzyme data collection and structure solving. Drs. Spencer Anderson, Joseph Brunzelle, and Zdzislaw Wawrzak, for support with data collection at LS-CAT at the APS. Drs. Howard Robinson, Allen Orville, and Babak Andi, for data collection at the NSLS. Dr. Vivian Stojanoff for assistance with data collection at SSRL. The organizers and participants of the RapiData 2012 crystallography workshop – particularly, BNL scientists Drs. Allen Orville, Robert Sweet, Howard Robinson, and Alexei Soares, SSRL scientists, Drs. Graeme Card and Clyde Smith, and participants, Drs. Tom Terwillger, Nat Echols, and Elspeth Garman. The organizers and participants of the CCP4 School 2013 – particularly, APS scientists Drs. Ruslan “Nukri” Sanishvili and Craig Ogata, and participants, Drs. Ronan Keegan, Paul Adams, Zbigniew Dauter, Andrey Lebedev, and Tim Grune. The organizers and participants of the 2013 Neutron X-ray School. Dr. Charles Schultz of Knox College for Mössbauer data collection. Dr. Julio Soares of MRL for laser facility training and assistance. Dr. Raven Hwang for helpful discussion about DNAzyme crystal soaking and cryoprotection.
I am especially grateful to the chemical biology, IMP, and SCS administrative staff, for all of their assistance over the last 6.5 years; my life would have been a mess without your help. Stacy Olson, Susan Lighty, and Becky Duffield for helping me get started when I was new here. Connie Knight, Stacy Dudzinski, Theresa Strauss, and Beth Myler for helping me later in my career. Kara Metcalf, for helping me with the craziness of my final defense. Lori Sage-Karlson, Mike Rhoton, and Jeff Craig for their logistical assistance. Mike Hallock, for helping me expand the lab’s computing resources. And finally, Mike “Hodge” Harland and Kyle Webb, for helping me fix instrument and hardware problems, expand capabilities, and the wisdom they imparted on me through numerous deep conversations.

I would like to thank my mentors in the Lu lab – Drs. Kyle Miner, Nick Marshall, Tiffany Wilson (née Hopper), Nathan Sieracki, Arnab Mukherjee, and Saumen Chakraborty – for all your guidance, support, and friendship in my early years in the lab. Particularly Kyle, for all the relaxing lunches, stressful meetings and deadlines, and much needed drinks. And Nick, for being like a second advisor and a wise guru I could always turn to for help. I am also grateful for the help and friendship of current and past members of the Lu Lab. In particular: Parisa – your energy was infective and you were always a good friend and colleague; and I hope that I was able to be there for you at least 10% of the amount you were there for me. Evan – thank you for the intellectual and social engagement; it was much needed. Jules – your spontaneous and much appreciated deep discussions and hilarious tangents never failed to raise my spirits. Also, thanks for being a great climbing partner. Finally, I am particularly indebted to Mr. Matthew Ross, for his help, friendship, and great, always active sense of humor.

While these people have directly or indirectly contributed to the research performed here, this thesis would not have been possible without the support of friends. During the first year the incoming students stuck together, and I am very grateful for the supportive friendship of Drs. Spencer Peck, Rachel Botham, Kyle Dunbar, Joel Melby, and Maxim Prigozhin. As the years went on and connections changed, I met many wonderful people through my wife in the vet school: game nights with Stephanie and Steven Goff, and Matthew Baxley and Allyn Hanks were always a great break from the stress of work; social gatherings with Kathrine and Larry Brannick were a great way to unwind; thank you Kat, for your delicious goodies; and particularly, thank you to Larry, for teaching me how to be more self-confident – it has been a huge help. And I would not be able to have gotten through these last couple of years without the climbing club. In particular, I want to thank Sean Hible and Fiona Weingartner, for helping me get started and for their constant good vibes and encouragement. Chris Kuntz – you are the man! Thank you so much for helping me learn to lead when I didn’t know anything! Alex Bragg – thanks for your constant encouragement and positive attitude. Alex James –
thank you for helping me get started and get better with trad climbing, and getting me started with multipitch climbing; you are a great teacher and it has opened up whole new worlds. Also, thanks for getting me to the top of my first 14er and back down in one piece. A special thank you to Susan McIntyre, for your constant encouragement, curiosity, ambitiousness, and friendship full of awesome wacky times and deep conversation; I’m very glad to have met you, and hope we don’t grow apart. And finally, I am extremely grateful to Soon-to-be-Dr. Jessica Bruhn, for her scientifically insightful discussions, encouragement, emotional support, and love over the last two and a half years; for showing me a life outside of my career and allowing me to discover a life with purpose; and for helping me become the better person I want to be. Thank you for being an amazing and tolerant partner. It was worth the ride.

Acknowledgment to my family is so obvious it goes without saying. But it should be said. My parents have supported me through my entire life and I do not know where I would be without them. I know I don’t call enough or visit enough, but I appreciate everything that they have done for me and I would not have gotten through without their love, encouragement, and support. Likewise, I am so happy to have a wonderful sister that I can talk to about anything in my life, and know that I will get her honesty, but also her love and support. I am so happy that I have you, and I look forward to us growing old together. My parents didn’t give me a brother, unfortunately, so I decided to adopt one – Allan Rozarto, I can’t even put into words how much you have impacted my life. Thank you for your constant intellectual stimulation and out of the box thinking. You are a rare breed. I know you can accomplish something greater than I, someday. I hope to be by your side when you do.

Lastly, but most importantly, I want and need to thank my amazing wife, soon-to-be-Dr. Tanya Coty, for standing by my side for the last six years of graduate school and longer, for bringing me dinner on late nights at work, for taking care of the laundry and dishes when I was too busy, for planning fun dates, outings, and trips, for baking me amazing food, for her silliness that always brings a smile to my face, for putting up with my ridiculousness, and for supporting me in my neurotic moments; for being rational and level-headed, yet also a dreamer filled with whimsy; but more than anything else, for her unwavering love and commitment. Thank you for being such a wonderful, caring, and giving partner.
# TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION ........................................................................................................1
  1.1. Bioinorganic chemistry ........................................................................................................1
  1.2. Oxygen activation by terminal oxidases – structure and activity .........................................1
  1.3. "Biosynthetic" modeling approach .......................................................................................5
  1.4. Design of HCO models in myoglobin ...................................................................................6
  1.5. Thesis overview ....................................................................................................................11
  1.6. References ..........................................................................................................................12

CHAPTER 2 - CHARACTERIZATION OF TYROSINE CONTAINING MUTANTS OF Cu₈Mb ....... 25
  2.1. Introduction and motivation ................................................................................................25
  2.2. Metal binding properties ....................................................................................................26
  2.3. The extended hydrogen bonding network in oxy-Cu₈Mb proteins ....................................29
  2.4. Materials and methods ......................................................................................................42
  2.5. References ........................................................................................................................50

CHAPTER 3 - EXTENDING HYDROGEN BONDING AND IMPROVING PROTONATION IN Cu₈Mb ..............................................................................................................................57
  3.1. Proton delivery – hypothesis and design ..............................................................................57
  3.2. Proton channel – F33Y-D²-Cu₈Mb ....................................................................................58
  3.3. Proton channel – 2Ser-F33Y-Cu₈Mb ................................................................................60
  3.4. I107E-Cu₈Mb ....................................................................................................................62
  3.5. Materials and methods .......................................................................................................75
  3.6. References ........................................................................................................................81

CHAPTER 4 - DESIGN OF A HEME-Fe₄S₄ SITE IN CcP TO MIMIC SiR AND NiR ............... 86
  4.1. Background – SiR and NiR .................................................................................................86
  4.2. Design ...............................................................................................................................86
  4.3. Further experimental refinement and ERP characterization ..............................................89
  4.4. Materials and methods .....................................................................................................90
  4.5. References .......................................................................................................................101

CHAPTER 5 - TOWARDS CRYSTAL STRUCTURE DETERMINATION OF AN 8–17 DNAZYME ..106
  5.1. Background .......................................................................................................................106
  5.2. Previous attempts and challenges of DNAzyme crystallization .......................................107
1.1. Bioinorganic chemistry

It is taken as well established common knowledge that nucleic acid molecules and protein machines define and produce cellular function and life. On the other hand, the role that metals play in enabling the functions of life often seem glossed over in common understanding. This is despite the fact that more than half of proteins present in living organisms contain metal and most of the complex fundamental processes of life are driven by metal-based chemistry. Given this observation, it is not surprising that the study of metal-mediated chemistry in biology is an established discipline known as bioinorganic chemistry.

The understanding of how metal ions are incorporated in and utilized by natural as well as non-natural proteins and nucleic acids is a major area of bioinorganic chemistry. Interactions of natural and artificial RNA and DNA sequences with metal ions that inhibit their function or impart new function has enabled discoveries that treat cancers, prevent and treat viral infections, and allow metal ion sensing. Similarly, in natural proteins, the functions of metals include structural support and sensing, electron transport and transfer, and direct interaction with and modification of substrates, among others. Given the utility of proteins in mediating metal function, the design of artificial metalloproteins and metallopeptides has seen substantial activity and success.

1.2. Oxygen activation by terminal oxidases – structure and activity

One of the fundamental chemistries of life on earth is the activation of oxygen for reduction in a controlled manner. This is important because oxygen, while highly abundant and a strong oxidant – making it an attractive electron sink for extracting energy from foods and fuels – is kinetically inert. This property of oxygen is generally a good thing and is what allowed oxygen to build up in the atmosphere without the spontaneous combustion of all organic molecules. Chemically, clean oxygen activation is often achieved by allowing oxygen to bind to a metal atom or ion, altering its electronic properties. A prominent example of oxygen activation in biology is in aerobic respiration, where the oxidizing chemical potential of oxygen is utilized to drive synthesis of adenosine triphosphate (ATP), the ubiquitous energy source for biological processes.
extraction of this energy from oxygen is efficiently catalyzed by large membrane-spanning enzymes called terminal oxidases, which activate oxygen for complete reduction and use the energy to generate a proton gradient across a lipid bilayer membrane. Re-equilibration of this proton gradient operates the ATP synthase enzyme, which produces ATP in the cell.

The majority of terminal oxidases comprise a family of enzymes known as heme-copper oxidases (HCOs), so named because they contain heme cofactors and copper ions. The molecular structure of one member of this family, known as cytochrome c oxidase (CcO), is shown in Figure 1.1. The structure of the active site of CcO at which oxygen binds is shown in Figure 1.2. This site lies in the core of CcO and consists of two adjacent metal ions: an iron ion contained in a heme cofactor, and a copper ion ligated by three His residues, named CuB. One of these His sidechains also forms a covalent link to a nearby Tyr – a feature unique to HCOs. Altogether, this active site is known as the CuB site. A small fraction of organisms use an alternative terminal oxidase enzyme, known as cytochrome-bd oxygen reductase, due to its use of two different types of heme, heme-b and heme-d (Figure 1.3). No atomic resolution structure of this enzyme has been published, however it is believed that oxygen binding and reduction occurs at a b-,d- diheme site.

The general mechanism of oxygen reduction at a heme center is shown in Figure 1.4. It can be seen from this mechanism that complete cleavage of the O-O bond after oxygen binding requires efficient delivery of one additional electron and two additional protons to the heme-oxygen moiety. Subsequent recovery of the oxygen-binding state requires delivery of three additional electrons and
two additional protons. Given the size of terminal oxidases (~10^5 Da) and their location within the hydrophobic membrane, the availability of these basic components at the active site is not a trivial concern. Delivery of electrons in HCOs, as in many large metalloenzymes, is enabled by the use of a series of metal ion cofactors, along which electrons hop to reach the active site (Figure 1.5). In order to enable efficient proton delivery, terminal oxidases have evolved water-mediated hydrogen bonding networks called proton channels.

Figure 1.2. Structure of the oxygen binding site of CcO (1OCC). The iron is shown as an orange sphere and copper is shown as a teal sphere. Rendered in VMD.

To elucidate structural features responsible for efficient oxygen activation in these and related enzymes, biochemical and biophysical studies of the enzymes have identified potential amino acid residues that may play key roles in the reaction, and their importance has been confirmed by loss of activity in site-directed mutagenesis (SDM) studies. However, these mutations often cause structural perturbations in addition to those directly being investigated, which may explain the loss of the function. For instance, it is known that removal of CuB from the active site by mutagenesis or chemical methods eliminates oxidase activity, whether the role of the Cu is to cleave the O-O bond or simply to donate electrons upon oxygen binding remains unclear. Moreover, the SDM studies reveal only necessary, but not necessarily minimally sufficient structural features for function. In addition, direct studies of these enzymes are complicated by their large size (~10^5 Da), membrane spanning localization, and presence of additional metal sites and cofactors which interfere with spectroscopic studies of their active sites.
Structures of hemes discussed in this thesis.

To overcome these limitation, structural and functional synthetic models of native enzymes have been prepared and their studies have offered deeper insights into the structural features responsible for oxidase activity. These model systems typically consist of Fe-porphyrin molecules with either covalently linked His analogue moieties or independent tridentate pyridine complexes that bind Cu and form an oxygen mediated complex with the Fe-porphyrin upon oxygen addition. In these systems, a hydrophobic environment is provided by organic...
solvent. Several examples of models incorporating an analogue of the His-Tyr crosslink have been reported, however models incorporating un-crosslinked His and Tyr analogues have not.

**Figure 1.5.** Schematic representation of proton and electron delivery into the active site of *R. sphaeroides*.51

Despite these successes, small molecule mimics have important limitations. These molecules are incompatible with physiological aqueous conditions, despite the fact that water is a key component of HCOs and their activity. In addition, building long, well-defined hydrogen bonding networks, that might mimic the proton channels in HCOs, is a substantial synthetic challenge. Therefore, while studies of the native enzymes and their model systems have yielded numerous insights, an understanding of general features responsible for oxygen activation is still being sought.

1.3. “Biosynthetic” modeling approach

As a complementary approach to the studies of native enzymes and their synthetic models, that overcomes their limitations, use of small designed metalloproteins and metallopeptides has shown remarkable success. In particular, the Lu lab employs a “biosynthetic modeling” approach, based on a small, natural protein scaffold, to gain a fundamental understanding of sufficient features conferring oxidase activity, i.e., 4e⁻ reduction of O₂ to H₂O. In contrast to native proteins, such as HCOs, which are membrane proteins that are not easy to prepare to
homogeneity with high yields, and contain accessory metal-binding sites that complicate investigations, biosynthetic models are much simpler to synthesize and contain metal cofactors only in the desired site. Thanks to recent advances in biology, biosynthetic models of complex metal-binding sites such as in HCOs can be prepared more easily and with higher yields than using organic molecule ligands. Furthermore, the biosynthetic approach, using a small, stable, and well-characterized protein, such as sperm whale myoglobin (Mb), as the scaffold makes the model amenable to study under physiologically relevant conditions. Moreover, the structurally rigid scaffold protein allows defining long-range non-covalent interactions and probing their roles more precisely.

Several uses of Mb as a scaffold for studying alternative activities were reported previously. Watanabe and coworkers enhanced the peroxidase activity of Mb by replacing His64 with Leu and introducing a His at position 29 (L29H) or 43 (F43H), based on structural similarity to cytochrome c peroxidase (CcP).\textsuperscript{86-89} His64 was removed to increase the lifetime of the highly reactive ferryl-porphyrin radical state (\textit{por}^••-Fe(IV)=O; Cpd(I)).\textsuperscript{86-88} The F43H-H64L-Mb did produce enhanced peroxidase activity.\textsuperscript{86,89} Similarly, Watanabe and coworkers gradually rationally engineered Mb to carry out peroxygenase chemistry,\textsuperscript{88,90-94} similar to the peroxide shunted mechanism of cytochromes P450.\textsuperscript{95} The final variant, H64D-V68I-F43W-Mb carried out quantitative peroxygination of the introduced Trp43 to 2,6-dihydro-2,6-dioxoindole.\textsuperscript{94}

1.4. Design of HCO models in myoglobin

1.4.1. Design and characterization of Cu\textsubscript{18}Mb

Using the biosynthetic modeling approach, Sigman et al. previously reported the design of a structural HCO mimic in sperm whale myoglobin (Mb).\textsuperscript{96} Several favorable properties of Mb drove this as the chosen of scaffold for this project: (1) it is a small (17.4 kDa), soluble protein that can be easily produced in high-yield by recombinant expression in \textit{E. coli}; (2) it possesses a heme and is known to bind oxygen – indeed it’s native function is oxygen storage;\textsuperscript{97} (3) it is stable and known to be amenable to mutagenesis;\textsuperscript{86,89,92,94,98-100} (4) it is relatively easy to crystallize and was in fact among the first protein whose crystallographic structure was solved,\textsuperscript{101} facilitating structural characterization (Figure 1.6); (5) it is spectroscopically well characterized, facilitating interpretation of spectroscopic studies.\textsuperscript{97,102-106}

The design of a structural mimic of the Cu\textsubscript{18} active site of HCOs in Mb initially focused on introducing the Cu binding site in the heme pocket of Mb. This was carried out by computational overlay of the heme groups of both proteins and visual inspection to determine which residues to
consider mutating to His, that would allow formation of a tridentate-His Cu binding site,\textsuperscript{96} further confirmed by molecular modeling. The introduced residues – L29H-F43H – in combination with the native His64, were predicted to bind a Cu ion within 5.4 Å of the heme-Fe,\textsuperscript{96} This protein, called Cu\textsubscript{B}Mb, was shown to bind Cu(II) upon titration, with a modest K\textsubscript{d} \sim 9 μM.\textsuperscript{96} It should be noted that the purification of Cu\textsubscript{B}Mb and its variants without Cu in the Cu\textsubscript{B} site allows the metal to be easily substituted, a feat that is impractical or impossible to accomplish in native HCOs.

Figure 1.6. Structures of wtMb. (A) Overall structure of wtMb with helices labeled. (B) and (C) Structures of the active site of wtMb in the resting Fe(III) (B) and oxygen bound Fe(II)-O\textsubscript{2} (C) forms, showing all residues within 7 Å of the distal oxygen atom.\textsuperscript{103}
Subsequently, Cu$_{5}$Mb was shown to catalyze heme-oxygenase (HO)-like chemistry in the presence of Cu(II) and reductant at pH 8.0.$^{107}$ HO-like heme degradation proceeds by reduction of heme-bound oxygen to peroxide, followed by subsequent attack of the heme by the peroxide ligand.$^{10,108-111}$ Chemical reduction of oxy-heme to peroxo-heme is a catalytic step that HO shares in common with terminal oxidases, but not myoglobin, and suggested that Cu may indeed play a role in activating heme-bound oxygen for reduction to peroxide. Given that HO and HCO mechanisms deviate in that the peroxo intermediate of HCOs is rapidly protonated, while in HOs it is not and attacks the heme, it was hypothesized that inadequate protonation during turnover of Cu(II)-Cu$_{5}$Mb may explain its inability to perform oxygen reductase chemistry.$^{10}$ Therefore, incorporating other structural features that may facilitate protonation became a priority.

### 1.4.2. Achieving NO reduction in myoglobin – Fe$_{5}$Mb

In addition to catalyzing oxygen reduction, a subclass of enzymes in the HCO family specialize in reduction of nitric oxide to nitrous oxide.$^{112-114}$ These nitric oxide reductases (NORs) share many similar structural features with the oxygen reductases,$^{112,115-117}$ including a heme-non-heme metal site at which NO reduction occurs.$^{112-114}$ Several key structural differences distinguish nitric oxide reductases from oxygen reductases: (1) the active site – called the Fe$_{5}$ site – consists of heme, as in HCOs, however the non-heme metal site contains Fe rather than Cu; (2) in addition to the three His residues used by HCOs to bind Cu, the non-heme iron is also ligated by a Glu residue; (3) NORs lack the His-Tyr crosslink group and possess a second glutamate in their secondary coordination sphere.$^{112,115}$ Based on a structural overlay of a sequence homology model of an NOR, whose structure was unknown, Cu$_{5}$Mb was modified to contain a Glu residue – V68E – that could ligate the non-heme metal.$^{119}$ This protein, called Fe$_{5}$Mb, was shown to bind Fe in the non-heme site and displayed the activity of native NORs.$^{119}$ Impressively, once the structure of native NOR from *P. aeruginosa* was solved,$^{112}$ it was shown that Fe$_{5}$Mb shared substantial structural similarity to native Fe$_{5}$ sites. The success of this design was a galvanizing force in metalloenzyme engineering, showing that the chemistry of large transmembrane proteins with multiple cofactors can be reproduced in smaller, more experimentally tractable scaffold proteins.

This system has been used as the foundation for elucidating the fundamental chemistry of NO reduction at a heme-non-heme iron site.$^{80,84,120,121}$ Additionally, its success inspired the extension of the model to include more distant structural features, such as the second, non-ligand Glu residue.$^{122}$ This was achieved through the I107E mutation in Fe$_{5}$Mb, which positioned the carboxylate sidechain near, but not ligating, the non-heme Fe (Figure 1.7).$^{122}$ The importance of this
feature was confirmed by the observation of a ~100% increase in NO reduction activity of this mutant compared with the original Fe₈Mb. The success of this new design clearly demonstrates the viability of rational secondary coordination sphere design in improving the activity of designed metalloenzymes, and was a significant motivator for pursuing a similar strategy to improve Cu₈Mb.

Figure 1.7. Structure of active site of I107E Fe₈Mb with (B) and without (A) Fe bound in the non-heme site.

1.4.3. Attempts at improving oxygen reductase activity in Cu₈Mb

In order to improve protonation in Cu₈Mb during reduction of oxygen and enable complete O-O bond cleavage, incorporation of two related structural features into Cu₈ was previously pursued: (1) the incorporation of a hydroxyl group on the heme periphery, mimicking the hydroxyfarnesyl moiety present on o- and a-type hemes; (2) incorporation of a tyrosine residue into the active site. Computational calculations of native HCOs suggest that both of these groups are directly involved the hydrogen bonding network that protonates catalytic intermediates. While the hydroxyfarnesyl group is too large to fit into Mb, synthetic heme derivatives incorporating a hydroxyethyl group are readily available. Moreover, replacement of the heme in myoglobin is well established, allowing the role of the heme structure in oxygen reductase activity to be uncoupled from changes in the overall protein. Wang et al. reported the replacement of the natural heme-b in Cu₈Mb with the synthetic Fe-2(4)-hydroxyethyl-4(2)-vinyldeutroporphyrin (Fe-HVD; heme-o mimic). Assays of this protein showed 19-fold lower heme-degradation rate under Cu-dependent reduction conditions, which was interpreted to suggest that the hydrogen bonding network of the active site was altered and regulated the activity of Cu₈Mb.

Independently, Miner et al. investigated the role of the tyrosine in oxygen reductase activity by Cu₈Mb. In native HCOs, this tyrosine can exist on either the same helix as the histidine it is crosslinked with, or on an adjacent helix. Two Tyr containing Cu₈Mb variants were designed and engineered based on structural overlay with two different known HCO structures (Figure 1.8):
F33Y-Cu₈Mb was engineered based on structure and sequence comparison with bovine CcO, placing Tyr 4 residues downstream from His29; G65Y-Cu₈Mb was designed and engineered based on structural overlay with cbb₃ cytochrom oxidase from P. stutzeri. Oxygen consumption of these proteins under reductive conditions was measured directly using an oxygen electrode. Additionally, based on the hypothesis that proton delivery during turnover needed to be improved, assays were carried out at pH 6.0 rather than 8.0. Under these conditions, it was observed that the tyrosine residues did improve the oxygen reductase activity of Cu₈Mb substantially (Figure 1.9A), with a majority – though notably not all – of the oxygen reduced completely to water. The production of water was confirmed by NMR experiments in which production of H₂¹⁷O from ¹⁷O₂ by the proteins was observed under reductive conditions (Figure 1.9B). Moreover, F33Y Cu₈Mb and G65Y Cu₈Mb were able to achieve over 500 and 1000 turnovers of oxygen reduction, respectively. It is important to note that the oxygen reductase chemistry of these Cu₈Mb proteins was independent of presence and identity of metal ion (see Section 1.5 for further discussion).

![Figure 1.8. Structural overlays of F33Y-Cu₈Mb (crystal structure) and G65Y-Cu₈Mb (molecular model), with their inspirational native HCO structures.](image)

The role of tyrosine was further investigated by Yu, et al., by incorporating a series of unnatural tyrosine analogues with varying reduction potentials and pKₐs into position 33.¹³ It was found that the rate of activity as well as the fraction of water production increased as pKₐ decreased, suggesting a role for proton transfer and the role of tyrosine in defining the oxygen reductase activity of Cu₈Mb. Similarly, the role of the electronic properties of the heme were investigated by Bhagi-Damodaran et al., by incorporating a series of unnatural heme cofactors with varying electron
withdrawing substituents. It was found that the apparent rate of O$_2$ reduction increased as the reduction potential increased, approaching the potential of $\alpha$-type HCOs.

Figure 1.9. (A) Oxygen consumption rates of indicated proteins leading to formation of water or ROS, as indicated. (B) NMR detection of H$_2^{17}$O produced by reaction of proteins with $^{17}$O$_2$ under reductive conditions.

1.5. Thesis overview

The previous results showed myoglobin could be successfully imparted with oxygen reductase function by incorporation of three mutations in the distal site pocket; however, this activity is not entirely complete, as some of the oxygen consumed is released as reactive oxygen species (ROS). Moreover, the structural and electronic features that impart this activity in the absence of metal ion are still being investigated. Chapter 2 describes further characterization of this generation of Cu$_b$Mb proteins, with particular focus on the electronic and structural properties of oxygen bound to the heme. Crystal structures are reported of the G65Y-Cu$_b$Mb, Cu(II)-F33Y-Cu$_b$Mb, and oxy-F33Y-Cu$_b$Mb. Chapter 3 describes progress in improving the oxygen reductase activity of Cu$_b$Mb, based on the conclusions and hypotheses drawn from the results of Chapter 2. First, attempts to improve oxygen reductase activity by introducing a proton channel into F33Y-Cu$_b$Mb are reported. Finally, improvement of Cu$_b$Ms to abolish release of ROS by incorporation of the I107E mutation is reported, and the design hypothesis of this mutation is confirmed using EPR spectroscopy and crystallography of oxy-I107E-Cu$_b$Ms.

The final two chapters contain substantial work that is not related to the understanding or engineering of HCOs, NORs, or their models, but is more generally involved in the design and structural understanding of artificial bioinorganic catalysts. Chapter 4 describes the rational computational design of a unique heme-[Fe$_5$S$_4$] cofactor into CcP, which is expected to facilitate
multielectron reduction of substrates such as sulfite and nitrite. Chapter 5 describes recent progress in obtaining the crystal structure of an active metalloDNAzyme in its active form. Only one atomic resolution structure of an active DNAzyme is known, and such information about metalloDNAzymes would be invaluable in guiding rational design of these versatile artificial catalysts.

Additionally, the thesis contains several appendices which describe improved methods, apparatuses, programs, and protein constructs, as well as a few additional, unrelated crystal structures.

1.6. References


(3) Farrell, N. In Transition Metal Complexes as Drugs and Chemotherapeutic Agents; Springer Netherlands: 1989; Vol. 11, p 222.


(9) Kaila, V. R. I.; Verkhovsky, M. I.; Wikström, M. Proton-coupled electron transfer in cytochrome oxidase Chemical reviews 2010, 110, 7062.


(15) Petrik, I. D.; Liu, J.; Lu, Y. Metalloenzyme design and engineering through strategic modifications of native protein scaffolds *Current Opinion in Chemical Biology* 2014, 19, 67.


(34) Gennis, R. B. Multiple proton-conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine Biochimica et Biophysica Acta (BBA) - Bioenergetics 1998, 1365, 241.


(38) Näsvik Öjemyr, L.; Maréchal, A.; Vestin, H.; Meunier, B.; Rich, P. R.; Brzezinski, P. Reaction of wild-type and Glu243Asp variant yeast cytochrome c oxidase with O2 Biochimica et Biophysica Acta (BBA) - Bioenergetics 2014, 1837, 1012.


(49) Sigman, J. A.; Kim, H. K.; Zhao, X.; Carey, J. R.; Lu, Y. The role of copper and protons in heme-copper oxidases: Kinetic study of an engineered heme-copper center in myoglobin Proceedings of the National Academy of Sciences 2003, 100, 3629.


(52) Lu, C.; Zhao, X.; Lu, Y.; Rousseau, D. L.; Yeh, S.-R. Role of copper ion in regulating ligand binding in a myoglobin-based cytochrome c oxidase model Journal of the American Chemical Society 2010, 132, 1598.

(53) Ganesan, K.; Gennis, R. B. Blocking the K-pathway still allows rapid one-electron reduction of the binuclear center during the anaerobic reduction of the aa3-type cytochrome c oxidase from Rhodobacter sphaeroides Biochimica et Biophysica Acta (BBA) - Bioenergetics 2010, 1797, 619.

(54) Hematian, S.; Garcia-Bosch, I.; Karlin, K. D. Synthetic Heme/Copper Assemblies: Toward an Understanding of Cytochrome c Oxidase Interactions with Dioxygen and Nitrogen Oxides Accounts of chemical research 2015, 48, 2462.


(58) Baeg, J.-O.; Holm, R. H. Covalently supported porphyrins as ligands for the preparation of heme a3/CuB binuclear active site analogues of heme-copper terminal oxidases and metallation under mild conditions Chemical Communications 1998, 571.


(62) Collman, J. P.; Decréau, R. A. Functional biomimetic models for the active site in the respiratory enzyme cytochrome c oxidase Chemical Communications (Cambridge, United Kingdom) 2008, 5065.


(66) Collman, J. P.; Yang, Y.; Decréau, R. A. Synthesis of Nitric Oxide Reductase Active Site Models Bearing Key Components at Both Distal and Proximal Sites Organic Letters 2007, 9, 2855.


(69) Liu, J.-G.; Naruta, Y.; Tani, F. Synthetic models of the active site of cytochrome c oxidase: influence of tridentate or tetradsengate copper chelates bearing a His-Tyr linkage mimic on dioxygen adduct formation by heme/Cu complexes Chemistry--A European Journal 2007, 13, 6365.


(76) Ginovska-Pangovska, B.; Dutta, A.; Reback, M. L.; Linehan, J. C.; Shaw, W. J. Beyond the Active Site: The Impact of the Outer Coordination Sphere on Electro catalysts for Hydrogen Production and Oxidation Accounts of chemical research 2014, 47, 2621.


(95) Hrycay, E. G.; Bandiera, S. M. The monooxygenase, peroxidase, and peroxygenase properties of cytochrome P450 Archives of Biochemistry and Biophysics 2012, 522, 71.


Butland, G.; Spiro, S.; Watmough, N. J.; Richardson, D. J. Two conserved glutamates in the bacterial nitric oxide reductase are essential for activity but not assembly of the enzyme *Journal of Bacteriology* **2001**, *183*, 189.


Thorndycroft, F. H.; Matorin, A. D.; Richardson, D. J.; Watmough, N. J.; Ardelroth, P. Defining the proton entry point in the bacterial respiratory nitric-oxide reductase *Journal of Biological Chemistry* **2008**, *283*, 3839.


(122) Lin, Y.-W.; Yeung, N.; Gao, Y.-G.; Miner, K. D.; Tian, S.; Robinson, H.; Lu, Y. Roles of glutamates and metal ions in a rationally designed nitric oxide reductase based on myoglobin *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107, 8581.


(126) Cukier, R. I. A molecular dynamics study of water chain formation in the proton-conducting K channel of cytochrome c oxidase *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 2005, 1706, 134.

(127) Teale, F. W. Cleavage of the haem-protein link by acid methylethylketone *Biochimica et biophysica acta* 1959, 35, 543.


(129) Hemp, J.; Christian, C.; Barquera, B.; Gennis, R. B.; Martinez, T. J. Helix switching of a key active-site residue in the cytochrome cbb3 oxidases *Biochemistry* 2005, 44, 10766.
CHAPTER 2
CHARACTERIZATION OF TYROSINE CONTAINING MUTANTS OF CuB Mb‡

2.1. Introduction and motivation

While the observed oxygen reducing activity of the designed tyrosine-containing CuB Mb proteins reported by Miner et al. is remarkable, it also raised interesting questions. First – why did the presence of Cu(II) or any other metal in the CuB site of these proteins not significantly affect their oxygen reductase activity? In native HCOs, which Cu(II)-CuB Mb is a structural model of, oxygen reductase activity is seen to be abolished when the Cu in the CuB site is removed, chemically or by mutagenesis.35,36,43-46 However, the CuB site in HCOs is the result of very extensive evolution, and whether it is necessary for oxygen reduction to water in general, or has simply evolved to take on several necessary roles within the context of HCOs is unclear.1-9 Rapid delivery of electrons to oxygen upon binding in the CuB site is important for efficient oxygen reduction, and it is hypothesized that the role of CuB may simply be to allow storage of the necessary electron near the heme.4,6 Indeed, it is shown that electronic coupling between the heme and CuB allows CuB to be reduced before the heme iron and therefore prevents oxygen from binding before CuB is reduced.1 Cytochrome-bd oxygen reductases catalyze the same chemistry and are not known to possess a CuB site, although the two hemes in this protein are proposed to be in close proximity, with one Fe potentially serving the role of CuB.10-13 Similarly, cytochromes P450, which reduce oxygen and promote heterolytic O-O bond cleavage similar to HCOs do not have an additional metal in the active site;14 however, in these enzymes, the heme is ligated to the protein by a Cys residue rather than His used in HCOs and CuB Mb, so they are not directly comparable.

A related question is – if presence of Cu in the CuB site of CuB Mb is not necessary for oxygen reduction to water, what are the structural features of this site in CuB Mb that are responsible for imparting oxidase activity? The known reactivity of wtMb with oxygen is binding and release of oxygen, or autooxidation of oxy-Mb releasing superoxide.15 It is also known that reaction of hydrogen peroxide with Mb leads to O-O bond cleavage.16,17 Therefore, it is hypothesized that the F33Y-CuB Mb and G65Y-CuB Mb promote reduction of the oxy-ferrous state to the peroxo-ferric

‡ Portions of this chapter have been taken from the following manuscripts:

state, in which the O-O bond can be subsequently cleaved by known myoglobin reaction pathways (Figure 2.1); however, the detailed structural features and interactions that promote this reaction are still unclear.

Figure 2.1. Proposed mechanisms of oxygen reduction by F33Y CuB Mb in solution compared with wtMb.

Herein, I first briefly describe some interactions of Cu(II) with F33Y-CuB Mb and G65Y-CuB Mb, including affinities and attempts to achieve electronic coupling between the heme-Fe and CuB. Subsequently, I describe studies of the Cu-free form of the protein attempting to gain insight into how the structural features of CuB Mb interact with bound oxygen and impart oxygen reductase activity.

2.2. Metal binding properties

The Cu(II) binding titration curves and \( K_d \) values for F33Y-CuB Mb and G65Y-CuB Mb are given in Figure 2.2. The ability of Cu(II) to bind in the CuB site of F33Y-CuB Mb was further confirmed by crystallography (Figure 2.3). These affinities, in the μM range, are modest; however, at least for F33Y-CuB Mb (\( K_d = 1.8 \) μM), it is below the concentration used for activity assays. Therefore, it is surprising that Cu(II) has minimal effect on the oxygen reductase activity of F33Y-CuB Mb.

It was hypothesized that reduction of heme and binding of oxygen before Cu(II) is reduced may prevent Cu from improving oxygen reductase activity. In HCOs, Cu reduction before oxygen binding is ensured by electronic coupling between the Fe and Cu.\(^{18-20}\) To determine whether such coupling could be observed in F33Y-CuB Mb or G65Y-CuB Mb under any conditions, EPR spectroscopy was pursued. Based on the observation of Glu mediated coupling in Fe(III)-FeB Mb, acetate was also used to increase the likelihood of coupling. The EPR spectra of F33Y-CuB Mb and
G65Y-CuβMb are shown in Figure 2.4. In phosphate buffer, addition of Cu(II) to both proteins did result in heme signal decrease as observed by EPR. However, this decrease was also observed with wtMb (Figure 2.4), and could be attributed to protein precipitation in the presence of Cu(II) and phosphate at these concentrations. To avoid this, the EPR Cu(II) titrations were again carried out in Bit-Tris buffer (Figure 2.5). Under these conditions, no substantial decrease of heme signal was observed for CuβMb or F33Y-CuβMb. While a decrease of the heme signal was observed in G65Y-CuβMb, given the minimal coupling even at mM concentrations used in EPR experiments, it is unlikely to play a role under turnover conditions at 18 μM concentrations. Based on these results, it seems possible that Cu(II) bound to these proteins may remain Cu(II) throughout their catalytic cycle, because it cannot be reduced before the heme and before oxygen binds. Investigations of oxygen reduction with a variety of metals in FeβMb are currently being actively pursued and producing exciting results.

Figure 2.2. Cu(II) titration curves of F33Y-CuβMb and G65Y-CuβMb, including affinity.

Figure 2.3. Crystal structure of Cu(II)-F33Y-CuβMb (4FWY; stereo). The blue mesh is the anomalous signal from data collected at above the Cu K-edge, while brown mesh is anomalous signal from data collected at below the Cu K-edge energy.
Figure 2.4. EPR spectra and normalized heme signal intensities of wtMb (A), F33Y-Cu₈Mb (B), and G65Y-Cu₈Mb(C), with Cu(II), and with and without acetate, in phosphate buffer.
Figure 2.5. EPR signals of Cu₉Mb (A), F33Y-Cu₉Mb (B), and G65Y-Cu₉Mb (C) with Cu(II), and with and without acetate, in Bis-Tris buffer. No decrease in heme signal was observed in Cu₉Mb or F33Y-Cu₉Mb. A ~2-fold decrease of the high-spin heme signal was observed in G65Y-Cu₉Mb upon addition of 5 eq of Cu(II); however, given the minimal coupling even at mM concentrations of Cu(II), it is not expected to play a substantial role during turnover carried out at 18 μM.

2.3. The extended hydrogen bonding network in oxy-Cu₉Mb proteins

The observation of oxygen reduction completely to water in Cu₉Mb proteins even in the absence of Cu raised the question of how exactly 3 active site mutations converted Mb, an oxygen storage protein, into oxygen reducing enzymes capable of >500 turnovers. Crystal structures of Cu₉Mb and F33Y-Cu₉Mb obtained previously (Figure 2.6) reveal that the His and Tyr mutations introduced an intricate hydrogen bonding network, including additional water molecules in the active site.

Recently, several crystal structures of G65Y-Cu₉Mb in its resting state have been obtained. Interestingly, two low resolution crystal structures revealed two distinct configurations of the active site (Figure 2.7). In one, the sidechain of Tyr65 points into the active site and engages in a hydrogen bond with the waters in the active site. In this configuration, His29 is displaced from the active site.
In the alternative configuration, His29 is involved in hydrogen bonding in the active site, as seen in CuMb, and Tyr33 is displaced. In addition, a high resolution structure (Figure 2.6C,D) of G65Y-CuMb was obtained similarly revealing two overlapping structural configurations of the active site. In addition to the features observed in the first two structures, an additional water is observed here, which likely only exists in the conformation in which Tyr is in the active site. Also, in this configuration, while the His29 is displaced from the active site, it maintains a hydrogen bond with Tyr65, suggesting that it may still be involved in the extended hydrogen bonding network of G65Y-CuMb.

**Figure 2.6.** Crystal structures of CuMb (A), F33Y-CuMb (B), and G65Y-CuMb (C, D).

Computational calculations of HCOs have suggested that a hydrogen bonding network within the active site, linking the tyrosine with the bound oxygen, is necessary for oxygen reduction.\textsuperscript{21-24} Similarly, in cytochrome P450 monooxygenases, which also catalyze reductive O-O bond cleavage of oxygen, a hydrogen bonding network involving water molecules and protic residues is known to be engaged in interactions with the bound dioxygen and necessary for oxygen activation.\textsuperscript{14,25-29} Therefore, the observation of such extensive hydrogen bonding networks mediated by water molecules in our CuMb proteins raised the question of what role, if any, they play in
activating oxygen and converting an oxygen storage protein into a an oxygen reductase. In order to investigate this question, I endeavored to probe the electronic properties of oxygen bound to these proteins.

![Figure 2.7. Low resolution crystal structures of G65Y-CuB Mb showing the two distinct conformations of the active site.](image)

**2.3.1. Characterization by EPR of cryoreduced oxy-CuB Mbs**

EPR has been widely used to probe the environment of $O_2$ activating iron enzymes.\(^{25,30-45}\) Although oxy-heme is an electronically closed shell system not exhibiting any signal in EPR spectra, EPR spectroscopy has been used to offer insight into the properties of oxy-heme, such as its interactions with residues around it and its intermediates upon reduction, by a method called cryoreduction.\(^{46}\) Briefly, the oxy-heme protein, trapped at 77K, is exposed to gamma radiation, releasing “free” electrons into the sample matrix, which efficiently reduce EPR-silent oxy-heme into EPR-active peroxo-heme. Because cryogenic trapping suppresses thermal relaxation and equilibration of the protein and its active site, EPR spectra of the resulting sample, maintained at 77K, provide valuable structural information on the oxy-heme precursor (Figure 2.8).\(^{30,35-42,46}\) Furthermore, as the radiolytically generated peroxo-heme products are intermediates along the pathway of oxygen reduction, EPR spectra of the species that arise in their subsequent decay provide further mechanistic insight into the reaction (Figure 2.1).\(^{30,46-51}\) This decay is achieved by stepwise warming of the samples at increasing temperatures, followed by rapid re-cooling to 77K after each step to trap intermediates – a procedure called annealing. As a result, EPR spectroscopy
in combination with cryoreduction and annealing has been successfully used over the last two decades to probe the structure and interactions of dioxygen in many heme proteins, including myoglobins and monooxygenases. \cite{25,30,35-51} Because wtMb, the scaffold for F33Y-CuB Mb, was among the first and most studied proteins by the above method, the abundance of results facilitates understanding of the latter by direct comparison.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Formation and trapping of (hydro)peroxo products reflective of oxy protein states by cryoreduction.}
\end{figure}

\subsection{EPR and ENDOR characterization of cryoreduced oxy-F33Y CuB Mb}

Figure 2.9 shows the EPR spectra of oxy-wtMb and oxy-F33Y-CuB Mb cryoreduced at 77K, obtained under the same conditions. The wtMb spectrum shows the signal from a single species with \( g = [2.22, 2.13, 1.96] \) (see Figure 2.10 for spectra at other temperatures) whose EPR parameters (Table 2.1) identify it as a ferric-peroxo state, in which the peroxo ligand is H-bonded to the distal His-64. \cite{35,38,46} In contrast, the cryoreduced oxy-F33Y-CuB Mb exhibits two rhombic EPR signals (Figures 2.9, 2.10, and 2.11). The first, with \( g = [2.24, 2.13, 1.95] \), is similar to that of the ferric-peroxo species in wtMb. The second, more rhombic EPR spectrum, has \( g \)-values \( g = [2.29, 2.17, 1.96] \) that are in the range characteristic of ferric-hydroperoxo intermediates. \cite{25,30,37,40,41} This assignment is supported by \( ^1 \)H ENDOR measurements presented in Figure 2.12, in which the \( g_{\text{max}} = 2.24 \) species has been removed by annealing, which show a maximum proton coupling of \( \sim 9.5 \) MHz. Ferric-peroxo intermediates show strongly coupled exchangeable proton signals with \( A_{\text{max}} \sim 14-18 \) MHz, \cite{30,38,43} while hydroperoxo- exhibit a smaller coupling, \( A_{\text{max}} \leq 13 \) MHz.\cite{39,40,45} The absence of signal with \( A_{\text{max}} > 13 \) MHz confirms that the more rhombic species observed in F33Y CuB Mb is hydroperoxo-, not peroxo-.
Figure 2.9. EPR spectra of oxy-wtMb and oxy-F33Y-Cu₈Mb after radiolytic reduction with a ~3Mrad dose of γ radiation from ⁶⁰Co, and after subsequent stepwise annealing for one minute at indicated temperatures (sharp signal marked by asterisk is due to radiolytically generated hydrogen atoms at 77K in quartz EPR tube).

Table 2.1. Signal parameters for EPR signals observed in cryoreduced oxy-wtMb and F33Y-Cu₈Mb.

<table>
<thead>
<tr>
<th>Protein</th>
<th>g₁</th>
<th>g₂</th>
<th>g₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>wtMb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peroxo</td>
<td>2.22</td>
<td>2.12</td>
<td>1.965</td>
</tr>
<tr>
<td>hydroperoxo²</td>
<td>2.32</td>
<td>2.18</td>
<td>1.943</td>
</tr>
<tr>
<td>F33Y-Cu₈Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peroxo</td>
<td>2.24</td>
<td>2.13</td>
<td>1.964</td>
</tr>
<tr>
<td>hydroperoxo¹</td>
<td>2.29</td>
<td>2.16</td>
<td>1.950</td>
</tr>
<tr>
<td>hydroperoxo³</td>
<td>2.34</td>
<td>2.19</td>
<td>1.937</td>
</tr>
</tbody>
</table>

¹ Primary hydroperoxo product formed at 77K. ² Hydroperoxo product formed upon annealing.
**Figure 2.10.** Full collection of EPR spectra of oxy-wtMb and oxy-F33Y-CuMb after cryoreduction and annealing at various temperatures.
Figure 2.11. Time and temperature dependent annealing EPR studies of cryoreduced oxy-F$_{33}$Y-Cu$_{8}$Mb.
To gain further insight into oxygen reduction by F33Y-CuB Mb, we monitored the decay of the above cryoreduced species by annealing. As previously reported, annealing the cryoreduced oxy-wtMb at 190K converts the peroxo signal with $g_{\text{max}} = 2.22$ to the more rhombic EPR signal of a ferric-hydroperoxo intermediate, with $g = [2.32, 2.18, 1.93]$ (Figures 2.9 and 2.10), through protonation of the peroxo moiety. Upon annealing of cryoreduced oxy-F33Y-CuB Mb at 173-175K for at least 1 minute, a signal with $g = [2.34, 2.19, 1.94]$ begins to appear and continues to grow in at 190K (Figures 2.9, 2.10, and 2.11). These EPR parameters are very similar to those of ferric-hydroperoxo-wtMb. Appearance of this signal during annealing is interpreted to arise both from protonation of the ferric-peroxo centers with $g_{\text{max}} = 2.24$ to form ferric-hydroperoxo, and by relaxation/repositioning of the hydroperoxo ligand of the $g_{\text{max}} = 2.29$ species of the heme iron (Figure 2.9).
Based on the known reactivity of the (hydro)peroxo-hemes, these intermediates are expected to thermally decompose to the compound II (CpdII), a ferryl species that is EPR silent. Consistent with this prediction, upon further annealing both samples at 210K and 220K (Figures 2.9 and 2.10), we observed obvious decay of the (hydro)peroxo signals to an EPR silent state. It has been shown previously that further cryoreduction of the EPR silent CpdII (Fe(IV)\(=\)O\(^{2-}\)) at 77 K generates a low spin rhombic EPR signal from the Fe(III)-O\(^{2-}\) or Fe(III)-OH\(^{-}\) species, depending on the protonation state of ferryl precursor. Subsequent cryoreduction of the above EPR-silent species in wtMb resulted in a signal with \(g_{\text{max}} = 2.43\) and \(g_{\text{min}} = 1.93\) (Figure 2.13), characteristic of Fe(III)-O\(^{-}\) (ferric oxo) species observed in cryoreduced CpdII (\(g = [2.43, 2.12, 1.93]\)), thereby confirming that the EPR silent state arising after annealing of (hydro)peroxo-wtMb as the Fe(IV)=O\(^{2-}\) CpdII. A minor signal with \(g_{\text{max}} = 2.52\) and \(g_{\text{min}} = 1.90\) (Figure 2.13) was also observed, previously assigned as protonated ferric-oxo, i.e. ferric-hydroxo (Fe(III)-OH\(^{-}\)).

**Figure 2.13.** EPR spectra of oxy-ferrous proteins after cryoreduction-annealing-cryoreduction.

In the case of F33Y-Cu\(_{8}\)Mb, further cryoreduction of the EPR annealed silent species yielded a dominant EPR signal with \(g = [2.52, 2.15, 1.90]\) (Figure 2.13), likewise assigned as ferric-hydroxo. This result supports the hypothesis that the hydroperoxo-F33Y-Cu\(_{8}\)Mb converts to CpdII, as in wtMb. The assignment of these species is confirmed by comparison with the same signals observed in cryoreduced samples of CpdII in both proteins generated by reaction with H\(_2\)O\(_2\) at ambient temperature (Figure 2.14).

The observation of two distinct signals in cryoreduced oxy-F33Y Cu\(_{8}\)Mb at 77K suggests the presence of two structural states of the oxy-ferrous precursor. The peroxo signal with \(g = [2.24, 2.13,\)
1.95] is nearly identical to the signal observed in wtMb, and well within the range of other oxygen binding proteins such as hemoglobin, suggesting that the oxy-ferrous population leading to this state has similar interactions and behaves in a similar manner. On the other hand, the observation of a hydroperoxo species with \( g = [2.29, 2.17, 1.96] \) at 77K requires a configuration of the oxygen in which it can be easily protonated at 77K (Figure 2.9). The observation of protonated Fe(III)-OH in F33Y-CuMb upon subsequent cryoreduction of annealed species provides further support for the facile protonation of intermediates.

![EPR spectra of hydrogen peroxide formed Cpd(II) in wtMb and F33Y CuMb after cryoreduction.](image)

**Figure 2.14.** EPR spectra of hydrogen peroxide formed Cpd(II) in wtMb and F33Y CuMb after cryoreduction.

### 2.3.2. Crystal structure of oxy-F33Y CuMb

Previous cryoreduction studies of a variety of oxy-hemeproteins showed that proton transfer to a cryoreduced oxy-heme at 77K requires the presence of an extended H-bonded proton delivery network in the oxy-ferrous protein that includes protic residues and at least one ordered water molecule H-bonded to the terminal oxygen atom of the dioxygen ligand. Based on these observations, we hypothesize that the F33Y-CuMb may contain a similar H-bonding network for faster delivery of protons to the oxy-heme than in wtMb.

To test this hypothesis, we obtained the crystal structure of oxy-F33Y-CuMb. The oxy-F33Y-CuMb crystal was prepared using an established method of soaking ferric-Mb crystals in dithionite, followed by exposure to oxygen. A UV-Vis spectrum of this single crystal prior to diffraction confirmed spectral features indicative of oxy-protein (Figure 2.15). This oxy-F33Y-CuMb structure refined at 1.27 Å resolution is shown in Figure 2.16, compared with the structure of oxy-wtMb. An omit-map showing the electron density of the oxygen ligand is shown in Figure 2.17. While the
refinement of this structure yielded an incomplete occupancy for the oxygen ligand, likely due to the reduced affinity of F33Y-CuB Mb for oxygen, this structure nevertheless provides a foundation supporting the above interpretation of the EPR data. As suggested by the EPR results, an extended H-bonding network involving two water molecules is present in oxy-F33Y-CuB Mb, with W1 stabilized by H-bonds with the engineered His43 and Tyr33 residues, and W2 stabilized by H-bonds to W1 and the engineered His29 (Figure 2.16). Together, this H-bonding network creates H-bonding interactions between W2 and the distal oxygen atom of bound oxygen, as postulated above. Furthermore, as a result of these changes in the active site pocket, the oxygen atom in F33Y-CuB Mb is rotated from that in wtMb by ~13° about the Fe-O bond, apparently to maximize its ability to accept hydrogen bonds from His64 and W2. Such an interaction of the oxygen with the W2 would be expected to enable the facile protonation, by W2, of the peroxo intermediate formed on cryoreduction at 77K, yielding the hydroperoxo observed by the cryogenic EPR results.

Figure 2.15. Single crystal microfocus electronic absorption spectrum of the oxy-ferrous F33Y CuB Mb crystal.

Figure 2.16. Crystal structure of oxy-F33Y-CuB Mb determined at 1.27 Å resolution, in comparison with that of oxy-wtMb (1A6M). Alternative W2 position shown translucent.
2.3.2.1. EPR characterization of cryoreduced oxy-Cu₈Mb and oxy-G65Y-Cu₈Mb.

In order to obtain a more complete understanding of oxygen reduction by the family of Cu₈Mb proteins, the EPR of cryoreduced oxy-ferrous protein was extended to Cu₈Mb and G65Y-Cu₈Mb as well. Crystal structures of both of these proteins revealed similar hydrogen bonding networks to F33Y-Cu₈Mb, and these would be expected to have similar but distinct EPR spectra.

The EPR spectra of oxy-Cu₈Mb and oxy-G65Y-Cu₈Mb cryoreduced at 77K are shown in Figure 2.18, and their g-values are given in Table 2.2. Comparison with F33Y-Cu₈Mb and wtMb clearly reveals that cryoreduced oxy-Cu₈Mb and oxy-F33Y-Cu₈Mb possess similar spectroscopic features. Minor shifts in the ERP signals are observed, most noticeably, the change in g-value of the hydroperoxo peak at 2.29 shifting to 2.28 in Cu₈Mb upon annealing. Such a shift could be attributed to a minor rearrangement of the hydrogen bonding network, resulting in the proton on the hydroperoxo ligand engaging in a stronger hydrogen bond. While it cannot be stated with certainty why this occurs in Cu₈Mb and not F33Y-Cu₈Mb, it may be that the presence of Tyr33 and the additional water observed in the hydrogen bonding network of F33Y prevents locks the orientation of one of the water molecules and prevents reorganization of its hydrogen bonding interaction.

G65Y-Cu₈Mb reveals qualitatively similar results to the other Cu₈Mb variants, with distinct differences. First, a fairly significant increase is observed in the g_{max}-values of the species formed immediately upon annealing. This is indicative of an environment that more strongly polarizes the iron-oxygen orbitals and consistent with strong hydrogen bond donation to the bound ligand. Secondly, the intensity and therefore population of the hydroperoxo signals is significantly lower.
than in the other Cu₈Mb mutants. This is unexpected, given that G65Y-Cu₈Mb achieves faster rates of oxygen reduction with similar product ratios. However, it is also clear from crystal structures that G65Y-Cu₈Mb engages in significantly different hydrogen bonding interactions than Cu₈Mb and F33Y-Cu₈Mb (Figures 2.6 and 2.7). Given that the configuration in which Tyr65 is displaced from the active site is expected to have a similar active site arrangement to Cu₈Mb, it seems likely that, given the significant spectroscopic differences between the proteins, that G65Y-Cu₈Mb does not adopt this configuration upon oxygen binding. Therefore, it seems likely that in oxy-G65Y-Cu₈Mb the His29 is displaced from the active site and can no longer engage in the hydrogen bonding network. It is also likely that substantial protonation of the peroxo intermediate at cryogenic temperatures requires the presence of His29 to achieve. However, given that G65Y-Cu₈Mb still achieves oxygen reduction with a similar product ratio to the other Cu₈Mb proteins, it seems likely that the slight polarization observed is sufficient to inhibit ROS release and drive oxygen reduction to water (See Section 3.4.8 for further discussion of trends).

Figure 2.18. EPR spectra of oxy-wtMb, oxy-Cu₈Mb, oxy-F33Y-Cu₈Mb, and oxy-G65Y-Cu₈Mb after radiolytic reduction with a ~3Mrad dose of γ radiation from ⁶⁰Co and subsequent stepwise annealing at indicated temperatures (sharp signal marked by asterisk is due to radiolytically generated hydrogen atoms at 77K in quartz EPR tube).
Table 2.2. ERP parameters of cryoreduced oxy-Cu\textsubscript{\textit{b}}Mb and oxy-G65Y-Cu\textsubscript{\textit{b}}Mb

<table>
<thead>
<tr>
<th>Protein</th>
<th>(g_1)</th>
<th>(g_2)</th>
<th>(g_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu\textsubscript{\textit{b}}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peroxo</td>
<td>2.24</td>
<td>2.13</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo\textsuperscript{1}</td>
<td>2.28-2.29</td>
<td>2.16</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo\textsuperscript{2}</td>
<td>2.33</td>
<td>2.18</td>
<td>1.94</td>
</tr>
<tr>
<td>G65Y-Cu\textsubscript{\textit{b}}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peroxo</td>
<td>2.25</td>
<td>2.13</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo\textsuperscript{1}</td>
<td>2.28-2.30</td>
<td>2.16</td>
<td>nd</td>
</tr>
<tr>
<td>hydroperoxo\textsuperscript{2}</td>
<td>2.33</td>
<td>2.19</td>
<td>1.94</td>
</tr>
</tbody>
</table>

\textsuperscript{1} Primary hydroperoxo product formed at 77K. \textsuperscript{2} Hydroperoxo product formed upon annealing.

2.3.3. Conclusions

Given the spectroscopic and crystallographic evidence for an H-bonding network that interacts with the O\textsubscript{2} in oxy-F33Y-Cu\textsubscript{\textit{b}}Mb, and its absence in oxy-wtMb (Figure 2.16), we infer that this H-bonding network is the important structural feature that has transformed wtMb, a protein that can only bind O\textsubscript{2} reversibly, into F33Y-Cu\textsubscript{\textit{b}}Mb, a protein that can activate O\textsubscript{2} and reduce it to water. We propose that this activation is achieved by polarization of the O-O bond by the H-bonded water, allowing facile reduction and protonation of the oxygen and intermediates. While H-bonding networks involving water have been observed in native heme enzymes such as cytochromes P\textsubscript{450\textsuperscript{25,27,41,54-56}} and a similar network has been proposed to occur in HCOs based computational and isotope studies,\textsuperscript{21-24} design of metalloenzymes to incorporate this important structural feature has been difficult to perform and to confirm. The combination of cryoreduction EPR spectroscopy and high-resolution crystallography presented in this work provides a clear example of the importance of H-bonding networks involving water in conferring activity to designed metalloenzymes. We believe this structural feature will be critical to enhancing the future success of high activity metalloenzyme design.

2.4. Materials and methods

2.4.1. Determination of \(K_d\) for F33Y-Cu\textsubscript{\textit{b}}Mb and G65Y-Cu\textsubscript{\textit{b}}Mb for copper

F33Y-Cu\textsubscript{\textit{b}}Mb and G65Y-Cu\textsubscript{\textit{b}}Mb were titrated with increasing amounts of copper and monitored using UV-Visible spectroscopy using a previously described method\textsuperscript{57} to determine \(K_d\). The spectra were analyzed using the method described by Bidwai, A. et al.\textsuperscript{58}

2.4.2. EPR spectroscopy of Cu(II) titrations

EPR samples were prepared by exchanging protein into fresh 100 mM KPi pH 7.0 and concentrating to \(\sim 400-500\) \textmu M. Sodium acetate or water were added from a stock to bring the acetate concentration to desired quantities. Copper sulfate was added from a stock, slowly in 4
additions over 4 minutes, with stirring, on ice. Glycerol was mixed in to 30% final concentration. The solution was loaded into an EPR tube and flash frozen in liquid nitrogen.

X-band EPR spectra were collected on a Varian E-122 spectrometer at the Illinois EPR Research Center (IERC). EPR spectra were recorded at 20K, a modulation frequency of 100 kHz, a modulation amplitude of 5 G, microwave power of 2 mW and with a sweep rate of 8.3 G/s. Spectra were recorded at microwave frequencies of approximately 9.364 GHz, Magnetic fields were calibrated with a Varian NMR gaussmeter, and the frequencies were measured with an EIP frequency counter.

2.4.3. Cryoreduced oxy-ferrous EPR sample preparation

Each protein was concentrated to ~1-3 mM concentration and reduced with excess dithionite anaerobically under nitrogen atmosphere. Dithionite was removed under anaerobic conditions by a small GE Sephadex G-25 desalting column. The protein was reconcentrated anaerobically. The protein solution was removed from the anaerobic chamber and immediately, slowly, and gently sparged with pure oxygen gas from a syringe. The protein was briefly centrifuged to remove bubbles and mixed with pure glycerol to yield 15% final glycerol concentration. This solution was quickly transferred to an EPR tube and slowly frozen with liquid nitrogen.

Samples were transferred to a glass dewar filled with liquid nitrogen and irradiated with ~3 Mrad of gamma radiation over 440 minutes using a Gammacell 220 irradiator containing a $^{60}$Co source. Irradiated samples were stored and transported in liquid nitrogen at all times.

2.4.4. EPR/ENDOR spectroscopy – cryoreduced samples

X-band CW EPR spectroscopy was performed using a Bruker ESP-300 spectrometer equipped with an Oxford Instruments ESR-900 helium flow cryostat. EPR spectra were recorded at 20K, a modulation frequency of 100 kHz, a modulation amplitude of 5 G, microwave power of 2 mW and with a sweep rate of 10 G/s. Spectra were recorded at microwave frequencies of approximately 9.364 GHz, with the precise microwave frequencies recorded for individual spectra using a Hewlett-Packard Microwave Frequency Counter (HP5352B). CW Q-band EPR/ENDOR spectra were recorded on a modified Varian E-109 spectrometer described previously.$^{30,38,40}$ All CW Q-band EPR/ENDOR spectra were recorded at 2 K in dispersion mode, under “rapid passage” conditions, which gives absorption line shape. Derivative spectra were obtained numerically using program LabCalc. Asymmetry in the $^1$H ENDOR spectra commonly seen in the ENDOR spectra of cryoreduced oxyferroheme proteins is due to the effects of spin relaxation.$^{30}$
2.4.5. Protein crystallography

2.4.5.1. F33Y-CuMb crystallization

Resting state (aquo-met-ferric) crystals of F33Y CuMb were grown under conditions previously reported (for Cu(II) structure), or with slight optimizations detailed here (for subsequent crystals; see Appendix). Sitting drop vapor diffusion was utilized to increase drop size and allow for larger crystals. Protein was concentrated to 1 mM in 20 mM tris sulfate pH 8.0 buffer. Crystallization solution was composed of 100 mM sodium 2-(N-morpholino)ethansulfonate (MES) pH 6.75, 200 mM sodium acetate, and 30% w/v PEG 10,000. Solutions were mixed 1:1 to yield a final drop volume of 200 μL. This was equilibrated by vapor diffusion against 30 mL of crystallization solution, at 4°C, in dark. Crystals formed within 1 week.
2.4.5.2. G65Y-Cu₉Mb crystallization – low resolution

Met-G65Y-Cu₉Mb was exchanged into fresh 20 mM tris sulfate pH 8.0 and concentrated to 1.5 mM. The protein was mixed 1:1 with crystallization buffer consisting of 30 % PEG 10K, 100 mM tris sulfate pH 8.4, and 200 mM potassium acetate and crystallized by hanging drop vapor diffusion against this buffer. The crystals were cryoprotected with 50% PEG 400.

Table 2.3. Data collection and refinement statistics for low resolution G65Y-Cu₉Mb structures.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>YL108-3</th>
<th>YL108-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>NSLS x29a</td>
<td>NSLS x29a</td>
</tr>
<tr>
<td>Date</td>
<td>ADSC Q315</td>
<td>ADSC Q315</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.075</td>
<td>1.075</td>
</tr>
<tr>
<td>Spacegroup</td>
<td>P₂₁</td>
<td>P₂₁</td>
</tr>
<tr>
<td>Unit cell (a x b x c; Å / β; °)</td>
<td>34.41 x 31.86 x 70.62 / 101.38</td>
<td>34.56 x 31.60 x 70.94 / 101.36</td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>50-2.00 (2.24-2.20)</td>
<td>50-2.30 (2.34-2.30)</td>
</tr>
<tr>
<td>Overall B-factor, Wilson (Å²)</td>
<td>25.24</td>
<td>28.52</td>
</tr>
<tr>
<td>Unique reflection</td>
<td>7791 (366)</td>
<td>6740 (293)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (95.8)</td>
<td>98.0 (86.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.1 (8.7)</td>
<td>6.5 (4.7)</td>
</tr>
<tr>
<td>&lt;1/σ(I)&gt;</td>
<td>11.5</td>
<td>6.3</td>
</tr>
<tr>
<td>CC½</td>
<td>(0.968)</td>
<td>(0.657)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.115 (0.224)</td>
<td>0.140 (0.753)</td>
</tr>
<tr>
<td>Rρim</td>
<td>0.038 (0.111)</td>
<td>0.058 (0.372)</td>
</tr>
<tr>
<td>Rmeas/rim</td>
<td>0.121 (0.282)</td>
<td>0.152 (0.844)</td>
</tr>
</tbody>
</table>

2.4.5.3. G65Y-Cu₉Mb crystallization – high resolution

G65Y-Cu₉Mb was exchanged into 100 mM KPi pH8.0 and concentrated to 1.7 mM concentration. To this was added 5eq of potassium cyanide and after 10 min, the protein exchanged

45
into fresh 20 mM tris sulfate pH 8.0. This was concentrated back to 1.7 mM and used to set a hanging drop crystal screen tray. The protein crystal was found after ~6 months in dark at 4°C in a drop mixed 1:1 with 1.6 M sodium citrate tribasic dihydrate pH 6.5. No additional cryoprotectant was used.

**Table 2.4.** Data collection and refinement statistics for high resolution G65Y-Cu₈Mb structure.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>NSLS x26c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>NSLS x26c</td>
</tr>
<tr>
<td>Date</td>
<td>ADSC Q315</td>
</tr>
<tr>
<td>Detector</td>
<td>1.000</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td></td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>50-1.55 (1.58-1.55)</td>
</tr>
<tr>
<td>Overall B-factor, Wilson (Å²)</td>
<td>98.5 (82.0)</td>
</tr>
<tr>
<td>Unique reflection</td>
<td>22041 (899)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.5 (82.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.8 (4.5)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt;</td>
<td>28.0</td>
</tr>
<tr>
<td>R_merge</td>
<td>0.040 (0.097)</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th>Software</th>
<th>Phenix v1.8.2_1309 (Linux)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (highest shell)</td>
<td>41.087-1.550 (1.570-1.550)</td>
</tr>
<tr>
<td>Reflections</td>
<td>41107 (1208)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.22 (79)</td>
</tr>
<tr>
<td>Free set (%; random)</td>
<td>9.12</td>
</tr>
<tr>
<td>R_work</td>
<td>0.1724 (0.1860)</td>
</tr>
<tr>
<td>R_free</td>
<td>0.2109 (0.2299)</td>
</tr>
<tr>
<td>Molecules in ASU</td>
<td>1</td>
</tr>
<tr>
<td># of Atoms</td>
<td>1309</td>
</tr>
<tr>
<td>Protein</td>
<td>168</td>
</tr>
<tr>
<td>Solvent</td>
<td>44</td>
</tr>
<tr>
<td>Ramachandran favored/disallowed (%)</td>
<td>97.6/0.00</td>
</tr>
</tbody>
</table>

**2.4.5.4. Cu(II)-F33Y-Cu₈Mb crystal preparation**

Small crystals of met-F33Y-Cu₈Mb obtained from hanging drop crystallization were used soaked with 10 eq. CuSO₄ in a solution of F33Y-Cu₈Mb and well buffer consistent with the original drop conditions to avoid dissolving the crystals and subsequently with 15 eq. potassium cyanide, in a mixture of F33Y-Cu₈Mb and well buffer using pH 7.0 sodium cacodylate.
2.4.5.5. Cu(II)-F33Y-CuB*Mb data collection and structure solving

The crystals were first soaked briefly in cryoprotectant (30% polyethylene glycol 400) and were flash cooled in liquid nitrogen. The diffraction data sets summarized in Table 2.5 were collected at the National Synchrotron Light Source beamline X29A (Upton, NY) and were processed with HKL2000 software.60

The crystal structure was solved by the molecular replacement method using MOLREP in the CCP4 Package.61 Initial refinement was performed using X-plor62 and SHELX’9763. The position of Tyr33 was rebuilt using the program O.64 Final refinement of Cu-F33Y-CuB*Mb were carried out with the aid of Refmac565 and Coot.66

Table 2.5. Data collection and refinement statistics for Cu(II)-F33Y-CuB*Mb structure.

<table>
<thead>
<tr>
<th>Source</th>
<th>Data Collection</th>
<th>BNL x29a</th>
<th>BNL x29a</th>
<th>BNL x29a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>BNL x29a</td>
<td>1.075</td>
<td>1.380</td>
<td>1.372</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>HKL2000 (Linux)</td>
<td>HKL2000 (Linux)</td>
<td>HKL2000 (Linux)</td>
<td></td>
</tr>
<tr>
<td>Spacegroup</td>
<td>P2,2,2,</td>
<td>P2,2,2,</td>
<td>P2,2,2,</td>
<td></td>
</tr>
<tr>
<td>Unit cell (a x b x c; Å)</td>
<td>39.73 x 48.29 x 77.69</td>
<td>39.68 x 48.18 x 77.48</td>
<td>39.69 x 48.19 x 77.51</td>
<td></td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>50-1.80 (1.86-1.80)</td>
<td>50-2.12 (2.20-2.12)</td>
<td>50-2.12 (2.20-2.12)</td>
<td></td>
</tr>
<tr>
<td>Unique reflection</td>
<td>14457</td>
<td>8795</td>
<td>8976</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0 (100.0)</td>
<td>100.0 (99.9)</td>
<td>100.0 (100.0)</td>
<td></td>
</tr>
<tr>
<td>&lt;1/σ(I)&gt;</td>
<td>23.3 (4.6)</td>
<td>35.9 (4.7)</td>
<td>23.3 (4.6)</td>
<td></td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.111 (0.637)</td>
<td>0.094 (0.584)</td>
<td>0.098 (0.628)</td>
<td></td>
</tr>
<tr>
<td>Rwork</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rfree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Refinement

| Software | Phenix, REFMAC (Linux) |
| Resolution range (highest shell) | 41.01-1.80 (1.84-1.80) |
| Reflections | 13729 |
| Rwork | 0.199 (0.221) |
| Rfree | 0.247 (0.331) |
| Molecules in ASU | 1 |
| # of Atoms | Protein: 1218, Solvent: 48, Other: 42HEM/1FE/1CU/4CU |
| Average Isotropic B-factor | Protein: 34.0, Solvent: 36.5, Other: 34.5/27.1/38.4/61.6 |
| ESD (ESU) | 0.097 |
| Bond RMSD (Å) | 0.019 |
| Angle RMSD (°) | 3.182 |
2.4.5.6. Oxy-F33Y-Cu8Mb crystal preparation

Large crystals of met-F33Y-Cu8Mb obtained from sitting drop crystallization were used. Reduction was carried out anaerobically. Several crystals were harvested into a small container. This was exchanged into an anaerobic atmosphere chamber (Coy) by 27 headspace purges, down to 8 inHg of vacuum each, in the antechamber. Subsequently, crystals were transferred to a small finely-fritted filter reservoir filled with degassed crystallization solution containing 20 mM dithionite. These crystals were allowed to soak for 10 minutes to reduce protein and eliminate residual oxygen. Soaking solution was gently removed from the reservoir through the frit using a syringe, driven by a syringe pump and connected with plastic tubing to the bottom of the filter reservoir. Solution was replaced by adding fresh degassed solution into the filter. This washing was repeated four times total. (See Appendix for detailed descriptions of apparatuses used). Simultaneously, an aliquot of crystallization solution was supersaturated with oxygen, by exposing to ~1500 psi of oxygen pressure in a steel bomb for ~30 min. Subsequently, deoxy crystals were harvested into small (~2 μL) drops on slides. Upon removing the slide from the anaerobic chamber, ~30 μL of oxygen supersaturated crystallization solution was immediately added and the drop was gently mixed. Crystals were harvested, cryoprotected (50% PEG 400 supersaturated with O2), and cryocooled in LN2 1-3 minutes after exposure to oxygen.

2.4.5.7. Oxy-F33Y-Cu8Mb data collection and structure solving

Single crystal spectra were collected under cryogenic conditions at APS Sector 14 using an OceanOptics DH-2000-S light source, OceanOptics USB4000 diode array detector, 4DX microfocus optics, 10 μm source fiberoptic, and 200 μm collector fiberoptic. X-ray data collection was carried out on beamline 14-BM-C. Data reduction and scaling were performed in HKL2000. The structure was solved by isomorphous replacement with the backbone and heme of the resting state model, 4FWX. R-free flags were copied from the 4FWX dataset and extended to the resolution of this dataset with random selection. Refinement was performed in Phenix. Model building was carried out in Coot. Sidechains and water molecules were rebuilt manually and iteratively with refinement. The O-O bond was restrained at 1.22Å. Graphics were generate with CCP4MG and VMD with Tachyon.
Table 2.6. Data collection and refinement statistics for oxy-F33Y CuMb structure.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>APS 14-BM-C</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2014-11-14</td>
</tr>
<tr>
<td><strong>Detector</strong></td>
<td>ADSC Q315</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9787</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Reduction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software</strong></td>
<td>HKL2000 v704u (Linux)</td>
</tr>
<tr>
<td><strong>Spacegroup</strong></td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td><strong>Unit cell (a x b x c; Å)</strong></td>
<td>39.63 x 47.61 x 76.53</td>
</tr>
<tr>
<td><strong>Resolution range (highest shell)</strong></td>
<td>50-1.27 (1.29-1.27)</td>
</tr>
<tr>
<td><strong>Overall B-factor, ML (Å²)</strong></td>
<td>12.45</td>
</tr>
<tr>
<td><strong>Overall B-factor, Wilson (Å²)</strong></td>
<td>12.98</td>
</tr>
<tr>
<td><strong>Unique reflection</strong></td>
<td>39055 (1853)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.6 (94.8)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>4.4 (2.6)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt;</td>
<td>17.9</td>
</tr>
<tr>
<td>CC½</td>
<td>(0.856)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.039 (0.329)</td>
</tr>
<tr>
<td>R&lt;sub&gt;pim&lt;/sub&gt;</td>
<td>0.020 (0.229)</td>
</tr>
<tr>
<td>R&lt;sub&gt;meas/rim&lt;/sub&gt;</td>
<td>0.044 (0.404)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software</strong></td>
<td>Phenix v1.9-1692 (Linux)</td>
</tr>
<tr>
<td><strong>Resolution range (highest shell)</strong></td>
<td>23.84-1.268 (1.286-1.268)</td>
</tr>
<tr>
<td><strong>Reflections</strong></td>
<td>72950 (2356)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>98.35 (82)</td>
</tr>
<tr>
<td><strong>Free set (%; random)</strong></td>
<td>4.76</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;work&lt;/sub&gt;</strong></td>
<td>0.146 (0.2218)</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;free&lt;/sub&gt;</strong></td>
<td>0.178 (0.2288)</td>
</tr>
<tr>
<td><strong>Molecules in ASU</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong># of Atoms</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>1277</td>
</tr>
<tr>
<td><strong>Solvent</strong></td>
<td>243</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>45</td>
</tr>
<tr>
<td><strong>Average Isotrop B-factor</strong></td>
<td>18.362</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>16.69</td>
</tr>
<tr>
<td><strong>Solvent</strong></td>
<td>27.91</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>12.94</td>
</tr>
<tr>
<td><strong>Standard Uncertainty (ML)</strong></td>
<td>0.1100</td>
</tr>
<tr>
<td><strong>Bond RMSD (Å)</strong></td>
<td>0.017</td>
</tr>
<tr>
<td><strong>Angle RMSD (Å)</strong></td>
<td>2.097</td>
</tr>
<tr>
<td><strong>Ramachandran favored/disallowed (%)</strong></td>
<td>97.5/0.0</td>
</tr>
</tbody>
</table>
2.5. References


(10) Borisov, V. B.; Gennis, R. B.; Hemp, J.; Verkhovsky, M. I. The cytochrome bd respiratory oxygen reductases Biochimica et Biophysica Acta 2011, 1807, 1398.


(13) Borisov, V. B.; Verkhovsky, M. I. Accommodation of CO in the di-heme active site of cytochrome bd terminal oxidase from Escherichia coli *Journal of inorganic biochemistry* 2013, 118, 65.


(19) Van Gelder, B. F.; Beinert, H. Studies of the heme components of cytochrome c oxidase by EPR spectroscopy *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1969, 189, 1.


(24) Cukier, R. I. A molecular dynamics study of water chain formation in the proton-conducting K channel of cytochrome c oxidase Biochimica et Biophysica Acta (BBA) - Bioenergetics 2005, 1706, 134.


(29) Makris, T. M.; von, K. K.; Schlichting, I.; Sligar, S. G. Alteration of P450 distal pocket solvent leads to impaired proton delivery and changes in heme geometry Biochemistry 2007, 46, 14129.


(34) Loewen, P. C.; Villanueva, J.; Switala, J.; Donald, L. J.; Ivancich, A. Unprecedented access of phenolic substrates to the heme active site of a catalase: Substrate binding and peroxidase-
like reactivity of Bacillus pumilus catalase monitored by X-ray crystallography and EPR spectroscopy Proteins: Structure, Function, and Bioinformatics 2015, 83, 853.


(54) Denisov, I. G.; Sligar, S. G. In *Cytochrome P450*; Ortiz de Montellano, P. R., Ed.; Springer International Publishing: 2015, p 69.


(71) Stone, J., University of Missouri-Rolla, 1998.
CHAPTER 3
EXTENDING HYDROGEN BONDING AND IMPROVING PROTONATION IN CuMb

3.1. Proton delivery – hypothesis and design

The secondary coordination spheres of metalloenzymes provide important interactions to bound substrates and intermediates during enzymatic turnover. In particular, hydrogen bonding interactions are involved in tuning side-chain pK\textsubscript{a} values, stabilizing water networks, and facilitating proton shuttling. For instance, heme-copper oxidases (HCOs), whose active site resides in a hydrophobic membrane protein, have evolved networks of hydrogen bonded water molecules for efficient delivery of protons into the active site, and across the membrane.\textsuperscript{1-5} Within the active sites of most known HCOs are protic residues and cofactor moieties, such as tyrosine and the hydroxylfarnesyl group of \textit{a} and \textit{o}-type hemes.\textsuperscript{1,3-6,9} Computational calculations suggest that these protic groups are responsible for proton shuttling between the end of the K-channel and the oxygen substrate during turnover, and that this function must be mediated by a network of water molecules \textit{within} the active site.\textsuperscript{10-13} It is also known that the first two protons for oxygen reduction are delivered to the oxygen by the D-channel, which terminates in a glutamate near the CuB site.\textsuperscript{14} In cytochromes P450, which also activate oxygen for heterolytic O-O bond cleavage, hydrogen bond interactions between protic residues and the oxygen, mediated by water molecules, have been shown to be critical in proper oxygen activation.\textsuperscript{14,25-29}

The work in Chapter 2 has already discussed the important role that extensive hydrogen-bonding networks likely play in imparting oxidase activity to CuMb.\textsuperscript{15,46} Briefly, it was hypothesized that the presence of an extensive hydrogen binding network consisting of the introduced protic residues and – most importantly – mediated by water, is the critical feature that imparted oxygen reductase activity to myoglobin. The extension of the hydrogen bonding network by positioning Tyr at one of two different positions demonstrated a dramatic effect.\textsuperscript{15} While rational protein design has been extremely successful, it has largely focused on design of interactions directly to the substrate or cofactor.\textsuperscript{17-19} On the other hand, extended hydrogen bonding networks, particularly those mediated by water molecules are well known in natural enzymes; however, rational design of such water-mediated interactions in enzymes has been limited.\textsuperscript{20-22} Here we focus on improving the oxygen reduction activity of CuMb and F33Y-CuMb by rational extension of the water mediated

\textsuperscript{8} Portions of this chapter have been taken from the following manuscript:
hydrogen bonding network in their active sites to promote oxygen activation and proton transfer. Based on the nominal reaction catalyzed by our designed oxidases, complete reduction of oxygen to water requires delivery of two external protons and electrons to promote the O-O bond cleavage step. Two structural features were pursued to improve rapid delivery of protons during catalytic turnover, based on known features of oxidases and monooxygenases, as well as based on chemical intuition: 1) extending the hydrogen bonding network out of the protein for facile delivery of protons during turnover, mimicking the K-channel of HCOs; 2) extending the hydrogen bonding network by introducing a carboxylate sidechain into the active site, inspired by the Glu residue at the end of the D-channel that is responsible for delivering the first two catalytic protons.

3.2. Proton channel – F33Y-D²-Cu₈Mb

3.2.1. Design hypothesis

In CcOs, two protons for complete catalytic turnover are delivered to the Cu₈ site via a water-based hydrogen bonding proton channel called the K-channel (see Chapter 1 for more detail). This channel terminates with the hydroxyl groups of the tyrosine and the hydroxyfarnesyl moiety on the catalytic heme. In F33Y-Cu₈Mb, a hydrogen bonding network of waters is observed between the heme Fe and Tyr33/His43. It was hypothesized that proton delivery could be facilitated by extending this channel through the protein to bulk solvent, to mimic the structure and function of the K-channel. Structural analysis revealed that the shortest route from Tyr33 to bulk solvent is achieved by going between the G-helix and BC-loop region of the protein.

Analysis of this region of the protein reveals that the closest residue to the hydrogen bonding network in F33Y-Cu₈Mb is Thr39. Since this residue is already capable of forming hydrogen bonds, it was left, and attention was turned to designing the remainder of the hydrogen bonding network leading from bulk solvent to Thr39. Structural analysis revealed that access to Thr39 by bulk solvent is blocked by residues His36 and Pheo6. Initial designs, inspired by the D-channel of CcOs, incorporated two aspartate residues at these positions. Modeling suggested that these groups could stabilize a channel of water molecules between helix G and loop BC.
3.2.2. Engineering and activity assays

The F106D and H36D mutations were introduced into F33Y-Cu₈Mb by site directed mutagenesis and the protein was purified from inclusion bodies (see Appendix A). The mutagenesis was confirmed by ESI-MS. The yield of this protein was remarkably low. The results of activity assays are shown in Figure 3.2. Unfortunately, this construct yielded lower activity than the parent F33Y-Cu₈Mb with comparable effect of catalase (indicating similar amounts of H₂O₂ release; see Appendix). It was hypothesized that introduction of two negatively charged residues in close proximity destabilized the scaffold.

Figure 3.2. Oxygen reduction catalyzed by F33Y_D²-Cu₈Mb compared with F33Y-Cu₈Mb.
3.3. Proton channel – 2Ser-F33Y-CuB Mb

3.3.1. Design rationale and hypothesis

In order to improve on the design of the proton channel, residues were chosen at F106 and H36 that were predicted to cause less interference with each other. Serine was chosen due to its small protic nature. Molecular dynamics simulations were carried out and supported the possible formation of a hydrogen bonding network between bulk solvent and the active site pocket, through Thr39 (Figure 3.3). This protein, called 2S-F33Y-CuB Mb was engineered, purified, and characterized.

![Figure 3.3. Molecular model of proposed design of 2S-F33Y-CuBMb.](image)

3.3.2. Engineering and basic characterization

The F106S and H36S (and T39V, discussed later) mutations were introduced into F33Y-CuB Mb by site directed mutagenesis and the protein was purified from inclusion bodies (see Appendix A). The mutagenesis was confirmed by ESI-MS. Yield of this protein was lower than of F33Y-CuB Mb, but reasonable. The absorption spectrum of this protein is shown in Figure 3.4. The features are typical of high-spin ferric protein,23 with a soret at 408 nm with an extinction coefficient $\varepsilon_{408} = 136 \text{ mM}^{-1}$. 

$\varepsilon_{408} = 136 \text{ mM}^{-1}$.
3.3.3. Activity assays

The effect of the mutations was characterized by oxygen reduction activity assays, as previously reported (Figure 3.5). These assays revealed a ~2-fold increase in total oxygen consumption rate, with roughly the same product ratio as F33Y-CuMb. In order to attempt to determine whether the mutations were functioning by forming an extended proton channel out of the active site, a knock-out study was performed. Since Thr39 is a native residue of Mb proposed to form part of the designed channel, this residue was mutated to a valine, which is effectively a single atom substitution of oxygen to carbon on the amino-acid side-chain. Such a substitution, while structurally minimal, would be expected to completely break the proposed 2S hydrogen bonding network. Incorporation of this mutation resulted in no change in rate, and in fact a slight increase in the ratio of water production, within error. A similar relative effect was observed in the T39V mutant of F33Y-CuMb lacking the two Ser residues, suggesting that Thr39 and the 2S mutations are independent of each other (Figure 3.5).
3.3.4. Discussion

It is not entirely clear why the 2S mutations result in doubling of the oxygen consumption rate of F33Y Cu₈Mb. In wtMb crystal structures, Phe106 and His36 form a pi-stacking interaction that may impart inter-helix stability. It is proposed that the effect of 2S mutations may be due to a general destabilization of the inter-helix interactions between the G helix and the BC loop. The three residues introduced in F33Y-Cu₈Mb with respect to wtMb – His29, Tyr33, His43 – all reside on the B and C loops of the protein. An increase in the dynamics of these residues may be responsible for improving oxygen turnover rates.

3.4. I107E-Cu₈Mb

3.4.1. Design hypothesis

Alternative rational extension of the water mediated hydrogen bonding network of Cu₈Mb an F33Y Cu₈Mb was based on analysis of their crystal structures and inspiration from native terminal oxidases. While the K-channel, ending in the hydroxyfarnesyl group and tyrosine, is known to be important for delivering some catalytic protons, in fact, the first two catalytic protons, responsible for the O-O bond cleavage step, are delivered by the D-channel. This channel terminates in a Glu residue, which is critical for function. In Chapter 2, it was proposed that the presence of W1 in the active site and it’s interactions with active site residues is key for imparting complete oxidase activity. In order to improve this activity, I chose to incorporate an acidic residue into this hydrogen bonding network, based on the hypothesis that a negatively charged carboxylate group would stabilize hydronium or protonated histidine in the active site during turnover, which would be expected to improve oxygen activation – by donating stronger hydrogen bonds to the bound oxygen – facilitate intermediate protonation, and minimize release of ROS. Given the extensive mutagenesis already performed, the available locations to introduce glutamate were limited (Figure 1.6). Location and choice of mutation were based on visual analysis and basic modeling in PyMol. The I107E mutation showed the best interaction with W1, with minimal perturbation of the active site (Figure 3.6). As discussed in Chapter 1, I107E was also found to improve nitrous oxide production in Fe₈Mb, proposed to be due to improved protonation, further inspiring my choice of mutation.
3.4.2. Basic characterization

Proper incorporation of the I107E mutation into CuB Mb and F33Y-CuB Mb was confirmed by sequencing and mass spectrometry. As purified, the proteins displayed a soret peak at 408nm and a visible spectrum indicative of a high spin ferric state, and consistent with the aquo-ferric (met) resting state of Mb (Figure 3.7). The extinction coefficients of the soret peak in the resting states of I107E-CuB Mb and I107E-F33Y-CuB Mb are 181 mM$^{-1}$ and 175 mM$^{-1}$, respectively, consistent with high-spin ferric myoglobin.

3.4.3. Crystallography

The successful incorporation of a carboxylate group into the water-containing hydrogen bonding network of CuB Mb in I107E mutants was confirmed by X-ray crystallography. The crystal structures of the resting states of I107E-CuB Mb and I107E-F33Y-CuB Mb were determined at 1.67 Å
and 1.47 Å resolution, respectively (Figure 3.8). In both structures, as in previous structures of CuMb variants, the active site contains a water molecule coordinating the heme (W0), which engages in a hydrogen bond with His64, and a second water molecule near His29 (W1), which engages in a hydrogen bond with W0 and His29. As desired in our design of the I107E variant, W1 near His29 is also engages in a hydrogen bond with Glu107. Having shown that, at least in the resting state, the engineered Glu107 engages the hydrogen bonded water network in the desired manner, we proceeded to investigate its effect on the functional properties of the engineered oxygen reductase enzymes.

**Figure 3.8.** Crystal structures of (A) I107E CuMb and (B) I107E F33Y CuMb, in their resting aquoferric (met) states.

### 3.4.4. Oxygen reduction assays

The oxygen consumption activities of both I107E-CuMb and I107E-F33Y-CuMb are shown in Figure 3.9, as compared with their Ile107 counterparts. As intended by the design, in both cases,
the I107E mutation has resulted in effective elimination of ROS release. The rate of water formation of CuB Mb was roughly doubled by the I107E mutation. The I107E mutation did not increase the rate of water production of F33Y-CuB Mb.

Interestingly, the raw oxygen consumption traces as a function of oxygen concentration for I107E mutants showed rate beginning to saturate at lower oxygen concentrations than in their parent variants (Figure 3.10), suggesting that the I107E mutation imparts enhanced oxygen affinity. This was confirmed by measuring the oxygen affinities of these variants (Table 3.1 and Figure 3.11), which indeed showed that the I107E mutation resulted in substantial increases in oxygen affinity for both proteins. Previous studies at tuning the oxygen binding affinity of myoglobin have shown that mutations to Leu29 and Ile107 are critical for tuning oxygen affinity,26-28 although, typically increased oxygen affinity has been associated with larger hydrophobic sidechains and water exclusion in this region. The observation of increase in affinity with the introduction of a charged sidechain is noteworthy. Furthermore, it is known that increased oxygen affinity slows auto-oxidation and release of ROS from myoglobin,29 which may contribute to the cleanliness of oxygen reduction by our I107E variants.

![Figure 3.9](image-url) Rates of O2 consumption yielding water (blue) or reactive oxygen species (ROS) (red), measured and calculated as reported previously.15,30,31 Rates of CuB Mb, F33Y CuB Mb, have been reported previously15 and are included here for visual comparison.
Table 3.1. Oxygen affinities of myoglobin mutants

<table>
<thead>
<tr>
<th>Protein</th>
<th>$k_{on}$ (μM⁻¹·s⁻¹)</th>
<th>$k_{off}$ (s⁻¹)</th>
<th>$K_d$ ($k_{off}/k_{on}$; μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wtMb26</td>
<td>8.8</td>
<td>8.3</td>
<td>0.94</td>
</tr>
<tr>
<td>CuB Mb</td>
<td>0.04</td>
<td>9.4</td>
<td>235</td>
</tr>
<tr>
<td>I107E-CuB Mb</td>
<td>0.19</td>
<td>3.9</td>
<td>20.5</td>
</tr>
<tr>
<td>F33Y-CuB Mb</td>
<td>0.026</td>
<td>9.2</td>
<td>350</td>
</tr>
<tr>
<td>I107E-F33Y-CuB Mb</td>
<td>0.19</td>
<td>25.9</td>
<td>136</td>
</tr>
</tbody>
</table>

Figure 3.10. Oxygen consumption rate curves as a function of oxygen concentration.

3.4.5. EPR characterization of cryoreduced oxy-ferrous state

In order to gain further understanding of how the introduced carboxylate residue tunes the reactivity of CuB Mb and F33Y-CuB Mb, I sought to probe the interactions of the oxygen bound to the active site, as performed on the original CuB Mb variants in Chapter 2, using ERP spectroscopy of cryoreduced oxy-ferrous proteins. The EPR spectra of 77K cryoreduced oxy-CuB Mb and oxy-F33Y-CuB Mb, and their I107E variants are shown in Figure 3.12 and their $g$-values are reported in Table 3.2. As discussed in Chapter 2, Both CuB Mb and F33Y-CuB Mb spectra at 77K reveal two species. The EPR parameters of the first species, with $g_{max} = 2.24$ in both proteins, suggest that it is a ferric peroxo state in which the peroxo ligand engages in a hydrogen bond with the distal His64. The second
species with $g_{\text{max}} = 2.29$ in both proteins is in the range characteristic of ferric-hydroperoxo intermediates. Annealing of these two proteins at 199K leads to formation of a state with $g_{\text{max}} = 2.34$. This is proposed to be due to protonation of the $g_{\text{max}} = 2.24$ peroxo state and rearrangement of the $g_{\text{max}} = 2.29$ hydroperoxo state into a more rhombic hydroperoxo state.

Results of EPR of cryoreduced oxy-I107E-CuB\textsubscript{Mb} variants appear qualitatively similar, but reveal distinct differences. In both cases, the major peroxo signal produced at 77K is significantly more rhombic, with $g_{\text{max}} = 2.27$ in both proteins. Furthermore, the highly rhombic hydroperoxo signal with $g_{\text{max}} = 2.32$, which was not observed in Cu\textsubscript{B}Mb and F33Y Cu\textsubscript{B}Mb until they were annealed at 175K, is immediately observed at 77K in I107E variants. The lower rhombicity hydroperoxo ferric signal, $g_{\text{max}} = 2.28-2.29$, is observed as a shoulder of the major peroxo peak in both variants. The observation of these signals of much higher rhombicity than observed in the 77K spectrum of the parent proteins is a strong indicator of stronger hydrogen bond donation to and polarization of the bound O-O species. In the case of F33Y-Cu\textsubscript{B}Mb and I107E-F33Y-Cu\textsubscript{B}Mb, a minor signal is observed in I107E-F33Y-Cu\textsubscript{B}Mb that is consistent with the signals observed in F33Y-Cu\textsubscript{B}Mb. The presence of this mixture of signals in I107E-F33Y-Cu\textsubscript{B}Mb not observed in I107E-Cu\textsubscript{B}Mb may indicate a hysteresis of the water mediated hydrogen-bonding network due to an over-abundance of potential hydrogen bonding partners to W1. Nevertheless, the effect of the Glu107 still provides substantial benefit, driving the oxygen reduction reaction to lower ROS release (Figure 3.9).

Table 3.2. EPR spectroscopic parameters of cryoreduced oxy- proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu\textsubscript{B}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peroxo</td>
<td>2.24</td>
<td>2.13</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo$^1$</td>
<td>2.29</td>
<td>2.16</td>
<td>1.93</td>
</tr>
<tr>
<td>hydroperoxo$^2$</td>
<td>2.34</td>
<td>2.19</td>
<td>1.94</td>
</tr>
<tr>
<td>F33Y Cu\textsubscript{B}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peroxo</td>
<td>2.24</td>
<td>2.13</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo$^1$</td>
<td>2.29</td>
<td>2.17</td>
<td>1.95</td>
</tr>
<tr>
<td>hydroperoxo$^2$</td>
<td>2.34</td>
<td>2.19</td>
<td>1.94</td>
</tr>
<tr>
<td>I107E Cu\textsubscript{B}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peroxo</td>
<td>2.27</td>
<td>2.16</td>
<td>1.96</td>
</tr>
<tr>
<td>hydroperoxo A$^1$</td>
<td>2.29</td>
<td>2.20</td>
<td>nd</td>
</tr>
<tr>
<td>hydroperoxo B$^1$</td>
<td>2.32</td>
<td>2.17</td>
<td>1.95</td>
</tr>
<tr>
<td>I107E F33Y Cu\textsubscript{B}Mb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peroxo A</td>
<td>2.24</td>
<td>2.12</td>
<td>1.96</td>
</tr>
<tr>
<td>peroxo B</td>
<td>2.27</td>
<td>2.16</td>
<td>nd</td>
</tr>
<tr>
<td>hydroperoxo A$^1$</td>
<td>2.28</td>
<td>2.17</td>
<td>nd</td>
</tr>
<tr>
<td>hydroperoxo B$^1$</td>
<td>2.32</td>
<td>2.17</td>
<td>1.947</td>
</tr>
<tr>
<td>hydroperoxo C$^2$</td>
<td>2.34</td>
<td>2.20</td>
<td>1.91</td>
</tr>
</tbody>
</table>

$^1$ Primary hydroperoxo product formed at 77K. $^2$ Hydroperoxo product formed upon annealing.
To further understand the reactivities of I107E variants with oxygen, annealing of the samples was carried out. The full spectral series are shown in Figure 3.13 and a summary of the $g_{\text{max}}$ signal intensities as a function of annealing is given in Figure 3.14. The first interesting feature observed was the immediate, nearly complete decay of the major peroxo signals in the I107E mutants after annealing at 145K for one minute. The decay of the peroxo signal is concomitant with the increase of the $g_{\text{max}} = 2.32$–2.34 hydroperoxo signal, indicative of conversion of the peroxo species into a hydroperoxo state, requiring protonation. This observation of rapid protonation of the major peroxo species at 145K to yield hydroperoxo is in stark contrast with the results of annealing of the parent proteins, and strongly supports the presence of a very labile proton in the active site.
Further careful annealing of the original Cu\textsubscript{B}Mb and F\textsubscript{33}Y-Cu\textsubscript{B}Mb proteins reveals that their peroxo and hydroperoxo states persist to about ~200K. Upon annealing at 200K for up to 200 min, the (hydro)peroxo species decay to an EPR silent state. In contrast, the (hydro)peroxo species of the I\textsubscript{107}E proteins begin to decay rapidly at 180K, and decay to an EPR silent state in less than 200 min at this temperature (Figure 3.13 and 3.14). Consistent with previous results on myoglobin, subsequent cryoreduction of this EPR-silent state produces $g_{\text{max}} = 2.54$ signals consistent with ferric hydroxo (Fe(III)-OH). This confirms that (1) the EPR-silent state is the ferryl compound II (Fe(IV)=O; Figure 3.15), and furthermore that (2) the active site environment of I\textsubscript{107}E variants of Cu\textsubscript{B}Mb generally facilitates protonation of the Fe ligand.
Figure 3.14. Temperature dependence of the $g_{\text{max}}$ peak intensity of the cryoreduced states, showing faster decay at lower temperatures in I107E mutants.

The contrast between the decay kinetics of the (hydro)peroxo species of I107E and parent enzymes at cryogenic temperatures further supports the design hypothesis, that Glu107 stabilizes an additional proton in the active site, enabling the hydrogen bonding network to efficiently activate oxygen and quickly deliver two protons and promote heterolytic O-O bond cleavage during turnover.
3.4.6. Crystal structure of oxy-I107E-CuBMB

![Crystal structure diagram](image)

Figure 3.15. EPR spectrum of oxy-I107E-F33Y-Cu8Mb after cryoreduction-annealing-cryoreduction, showing ferric hydroxo signal.

In order to gain further insight into the structural basis for oxygen activation of these efficient enzymes, we sought to obtain the crystal structure of the oxy- intermediate. Oxy-I107E-Cu8Mb crystals were obtained as previously reported, by reducing ferric crystals, followed by rapid exposure to oxygen. The active site structure of oxy-I107E-Cu8Mb, solved at 1.18 Å resolution, is shown in Figure 3.16. While the occupancy of the dioxygen atom is incomplete, likely due to large crystals and short soak times, the structure still provides a structural basis for understanding the role of Glu107. As expected from design, a water molecule is present within hydrogen bonding distance of the bound dioxygen, and furthermore, this water molecule engages in hydrogen bonds with Glu107 and His29. The geometry of the hydrogen bonding interactions with oxygen is close to...
ideal, suggesting that they do likely polarize the oxygen bond and could facilitate reduction and O-O bond cleavage.

### 3.4.7. Characterization of I107E-G65Y-Cu\textsubscript{b}Mb.

Inspired by the success of the I107E mutation in Cu\textsubscript{b}Mb and F33Y-Cu\textsubscript{b}Mb, I also investigated if the same effect could be achieved in G65Y-Cu\textsubscript{b}Mb. I107E-G65Y-Cu\textsubscript{b}Mb was expressed and purified as previously reported. Oxygen consumption assays (Figure 3.17) show an elimination of ROS release, consistent with the results obtained for I107E-Cu\textsubscript{b}Mb and I107E-F33Y-Cu\textsubscript{b}Mb, although an overall decrease in the rate of oxygen consumption is observed. Oxygen affinity measurements (Figure 3.18 and Table 3.3) reveal that I107E increased the oxygen affinity of G65Y-Cu\textsubscript{b}Mb from $K_d = 81$ to $50.6 \ \mu$M.

EPR spectra (Figures 3.19 and 3.20), parameters (Table 3.3), and annealing decay traces (Figure 3.21) of G65Y-Cu\textsubscript{b}Mb and I107E-G65Y-CuBMb are shown below. The effect of the I107E mutation on G65Y-Cu\textsubscript{b}Mb is similar to the effect observed in Cu\textsubscript{b}Mb. First, the parent peroxo species is more rhombic, with $g_{\text{max}} = 2.29$. Again, the high-rhombicity hydroperoxo species is observed immediately at 77K, indicating facile protonation, as observed in I107E-Cu\textsubscript{b}Mb and I107E-F33Y-Cu\textsubscript{b}Mb. Furthermore, the complete decay of (hydro)peroxo signals at 180K is also observed. These results all suggest that the effect of I107E on G65Y-Cu\textsubscript{b}Mb is the same as observed in I107E-Cu\textsubscript{b}Mb and I107E-F33Y-Cu\textsubscript{b}Mb.

![Figure 3.17. Oxygen consumption of G65Y-Cu\textsubscript{b}Mb and I107E-G65Y-Cu\textsubscript{b}Mb.](image)
Figure 3.18. Oxygen binding plot for I107E-G65Y-CuB Mb.

Table 3.3. Oxygen binding properties of I107E-G65Y-CuB Mb and G65Y-CuB Mb.

<table>
<thead>
<tr>
<th>Protein</th>
<th>$k_{on}$ (μM⁻¹s⁻¹)</th>
<th>$k_{off}$ (s⁻¹)</th>
<th>$K_d$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G65Y-CuB Mb</td>
<td>0.46</td>
<td>37.5</td>
<td>81</td>
</tr>
<tr>
<td>I107E-G65Y-CuB Mb</td>
<td>0.35</td>
<td>17.7</td>
<td>50.6</td>
</tr>
</tbody>
</table>

Figure 3.19. EPR spectra of G65Y-CuB Mb and I107E-G65Y-CuB Mb

Figure 3.20. EPR annealing kinetics spectra.
Figure 3.21. Temperature dependence of the \( g_{\text{max}} \) peak intensity of the cryoreduced states of G65Y-Cu\( \beta \)Mb and I107E-G65Y-Cu\( \beta \)Mb, showing faster decay at lower temperatures in I107E mutant.

3.4.8. Trends of oxygen polarization and oxygen reduction by Cu\( \beta \)Mbs.

Given the large quantity of cryo-reduction EPR and oxygen reduction activity data obtained through these investigations, we sought to identify trends that may help guide general principles of improving oxygen reduction. The strongest trend that was observed was between the rhombicity of the primary peroxo species and the ratio of water released vs. ROS (Figure 3.22). It is seen that as the rhombicity of this signal increases, the amount of ROS released decreases, down to 0%. Therefore, it is postulated that polarization of the bound dioxygen is a key determinant in whether oxygen reduction proceeds to completeness or results in release of ROS.

Figure 3.22. The apparent co-variance of water/ROS ratio of oxygen reduction products with the rhombicity of the primary peroxo EPR signal.
3.4.9. Conclusion

Taken together, the results above demonstrate that the extension of the hydrogen bonding network within the active site of our designed oxidase myoglobin by introduction of Glu107 facilitates protonation of intermediates and improves oxygen binding, resulting in cleaner oxygen reductase chemistry, by eliminating ROS release. It is postulated that this is a result of Glu107 engaging in strong hydrogen bonds with the active site water network, stabilizing protons in the active site (Figure 3.23), which are then available for efficient oxygen polarization, activation, and protonation during turnover. These results support the role of the D-channel-terminating Glu residue in providing protons to the bound dioxygen during the first steps of oxygen reduction. Moreover, the results presented here and in Chapter 2, extensively characterizing the hydrogen bonding networks of 7 different variants of Mb, clearly demonstrate that designing long-range, water-mediated hydrogen-bonding networks – a feature that has seen some but, minimal pursuit in rational enzyme design strategies – can play a key role in determining enzymatic function, and should be pursued more actively in contemporary enzyme design strategies.

![Figure 3.23. Schematic representation of the proposed differences in interactions between I107E and parent CuMb s that results in cleaner oxygen reductase chemistry in I107E mutants.](image)

3.5. Materials and methods

3.5.1. Molecular modeling

Modeling of 2S-F33Y-CuMb was carried out with the help of molecular dynamics simulations carried out in NAMD. The crystal structure of F33Y-CuMb was used as the starting model, and the crystallographic water molecules in the heme were kept. Mutations were made using PSFGEN. Standard CHARMM topology and parameter files were used. The water box size was at least 3 times the size of the protein in each direction. Simulation was carried out with 5000 steps of minimization, followed by 1 ns of dynamics.

Modeling of I107E-CuMb was done simply by using the mutate feature of pymol and seeing that a conformer of Glu at position 107 would engaging in a hydrogen bond with the water in the active site.
3.5.2. **Oxygen affinity assays**

Oxygen affinity was determined by flash methods in which CO was flashed off in the presence of O\(_2\) to measure \(k_{on}\), and \(k_{off}\) was measured directly.

3.5.3. **Crystallization, I107E-Cu\(_B\)Mb**

Resting state (aquo-met-ferric) crystals of I107E-Cu\(_B\)Mb were grown under modified conditions previously reported.\(^{15}\) Sitting drop vapor diffusion was utilized to increase drop size and allow for larger crystals. Protein was concentrated to 1.5-2.0 mM in 20 mM tris sulfate pH 8.0 buffer. Crystallization solution was composed of 100 mM tris sulfate pH 8.6, 200 mM sodium acetate, and 30% w/v PEG 10,000. Solutions were mixed 1:1 to yield a final drop volume of 150-200 μL. This was equilibrated by vapor diffusion against 30 mL of crystallization solution, at 4°C, in dark. Crystals formed within 1 week. Two crystal forms were observed – long needles which diverged into fan shapes (P2\(_1\)) and thicker rectangular crystals (P2\(_{2,1,2,1}\)). The orthorombic crystal form was propagated by seeding new crystallization drops.

3.5.4. **Crystallization, I107E-F33Y-CuBMb**

Resting state (aquo-met-ferric) crystals of I107E-Cu\(_B\)Mb were grown under modified conditions previously reported.\(^{15}\) Sitting drop vapor diffusion was utilized to increase drop size and allow for larger crystals. Protein was concentrated to 1.5-2.0 mM in 20 mM tris sulfate pH 8.0 buffer. Crystallization solution was composed of 100 mM tris sulfate pH 8.6, 200 mM sodium acetate, and 30% w/v PEG 10,000. Solutions were mixed 1:1 to yield a final drop volume of 150-200 μL. This was equilibrated by vapor diffusion against 30 mL of crystallization solution, at 4°C, in dark. Crystals formed within 1 week.

3.5.5. **Crystal preparation, oxy-I107E-Cu\(_B\)Mb**

Oxy-I107E Cu\(_B\)Mb crystals were prepared by anaerobic reduction and exposure to oxygen, based on previously reported methods.\(^{35,36}\) Several orthorhombic crystals were harvested into a small container. This was brought into an anaerobic atmosphere chamber (Coy) by 27 headspace purges, down to 8 inHg of vacuum each, in the antechamber. Subsequently, crystals were transferred to a small finely-fritted filter reservoir, filled with degassed crystallization solution containing 20 mM dithionite. These crystals were allowed to soak for 10 minutes to reduce protein and eliminate residual oxygen. Soaking solution was gently removed from the reservoir through the frit using a syringe, driven by a syringe pump and connected with plastic tubing to the bottom of the filter reservoir. Solution was replaced by adding fresh degassed solution into the filter reservoir. This washing was repeated four times total. Deoxy crystals were harvested into small (~2 μL) drops on
slides and sealed in a high pressure bomb. The bomb was removed from the chamber and immediately pressurized to ~1500 psi O$_2$ for 1-10 minutes. The bomb was carefully vented over 30s and the crystal harvested after brief soak in O$_2$ supersaturated cryoprotectant (50% PEG 400 supersaturated with O$_2$).

3.5.6. Crystallography, general method

Unless otherwise noted, X-ray diffraction data were collected under cryogenic conditions. Data reduction and scaling were carried out with HKL2000.$^{37}$ Refinement was carried out in Phenix.$^{38}$ Model building was carried out using Coot.$^{39}$ Graphics were generated using VMD with Tachyon.$^{70,71}$

3.5.7. Crystallography, met-I107E-Cu$_B$Mb

X-ray diffraction data were collected on beamline x29 at NSLS. The structure was solved using molecular replacement using the structure of Cu$_B$Mb (4FWZ). Sidechains and water molecules were rebuilt manually and iteratively with refinement.

3.5.8. Crystallography, met-I107E-F33Y-Cu$_B$Mb

X-ray diffraction data were collected on beamline 21-ID-G at APS. The structure was solved by isomorphous replacement using the backbone and heme of a related model, 4FWX. R-free flags were copied from the 4FWX dataset and extended to the resolution of this dataset with random selection. Sidechains and water molecules were rebuilt manually and iteratively with refinement.
Table 3.4. Data collection and refinement statistics for oxy-F33Y CuMb structure.

<table>
<thead>
<tr>
<th>Sample</th>
<th>1107E-CuMb</th>
<th>1107E-F33Y-CuMb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>NSLS x29</td>
<td>APS 21-ID-G</td>
</tr>
<tr>
<td>Date</td>
<td>2014-08-23</td>
<td>2013-06-03</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Q315-906</td>
<td>MAR 300 CCD</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.075</td>
<td>0.9786</td>
</tr>
<tr>
<td><strong>Data Reduction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>HKL2000 v2.3.8 (Linux)</td>
<td>HKL2000 v2.3.1 (Linux)</td>
</tr>
<tr>
<td>Spacegroup</td>
<td>P2₁</td>
<td>P2₁₂,₂,</td>
</tr>
<tr>
<td>Unit cell (a x b x c; Å)</td>
<td>34.02 x 31.72 x 71.28</td>
<td>39.67 x 47.87 x 77.88</td>
</tr>
<tr>
<td>(β; °)</td>
<td>102.4</td>
<td>N/A</td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>50-1.70 (1.76-1.70)</td>
<td>50-1.48 (1.51-1.48)</td>
</tr>
<tr>
<td>Overall B-factor, ML (Å²)</td>
<td>23.28</td>
<td>18.84</td>
</tr>
<tr>
<td>Overall B-factor, Wilson (Å²)</td>
<td>25.02</td>
<td>20.18</td>
</tr>
<tr>
<td>Unique reflection</td>
<td>16343 (1600)</td>
<td>25784 (1268)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (98.5)</td>
<td>99.8 (99.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.6 (5.8)</td>
<td>3.8 (3.7)</td>
</tr>
<tr>
<td>&lt;1/σ(I)&gt;</td>
<td>11.5</td>
<td>12.0</td>
</tr>
<tr>
<td>CC½</td>
<td>(0.843)</td>
<td>--</td>
</tr>
<tr>
<td>Rmerge/symmm</td>
<td>0.073 (0.635)</td>
<td>0.047 (0.534)</td>
</tr>
<tr>
<td>Rsym</td>
<td>0.025 (0.285)</td>
<td>--</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.077 (0.698)</td>
<td>--</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>Phenix v1.10.1_2155 (Win)</td>
<td>Phenix v1.8.2_1309 (Linux)</td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>34.82-1.708 (1.735-1.708)</td>
<td>20.48-1.467 (1.499-1.467)</td>
</tr>
<tr>
<td>Reflections</td>
<td>31358 (1171)</td>
<td>48282 (2811)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.1 (83)</td>
<td>99.25 (93)</td>
</tr>
<tr>
<td>Free set (%; random)</td>
<td>10</td>
<td>4.74</td>
</tr>
<tr>
<td>Rwork</td>
<td>0.159 (0.262)</td>
<td>0.212 (0.313)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.208 (0.305)</td>
<td>0.239 (0.352)</td>
</tr>
<tr>
<td>Molecules in ASU</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># of Atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1237</td>
<td>1274</td>
</tr>
<tr>
<td>Heme</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Solvent</td>
<td>114</td>
<td>121</td>
</tr>
<tr>
<td>Other</td>
<td>113</td>
<td>--</td>
</tr>
<tr>
<td>Average Isotropic B-factor</td>
<td>31.32</td>
<td>22.65</td>
</tr>
<tr>
<td>Protein</td>
<td>28.57</td>
<td>22.34</td>
</tr>
<tr>
<td>Heme</td>
<td>21.60</td>
<td>18.54</td>
</tr>
<tr>
<td>Solvent</td>
<td>41.30</td>
<td>27.36</td>
</tr>
<tr>
<td>Other</td>
<td>55.07</td>
<td>--</td>
</tr>
<tr>
<td>Standard Uncertainty (ML)</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Bond RMSD (Å)</td>
<td>0.017</td>
<td>0.004</td>
</tr>
<tr>
<td>Angle RMSD (Å)</td>
<td>1.561</td>
<td>1.489</td>
</tr>
<tr>
<td>Ramachandran favored/disallowed (%)</td>
<td>96.75/0.65</td>
<td>96.82/0.00</td>
</tr>
</tbody>
</table>
3.5.9. Crystallography, oxy-I107E-CuBMB

Single crystal spectroscopic data and X-ray diffraction data were collected at SSRL beamline 11-1. Data reduction and scaling was carried out with XDS, pointless, aimless, and truncate, automated using the script ‘autoxds’. The structure was solved by isomorphous replacement using the backbone and heme of a related model, 4FWX. R-free flags were copied from the 4FWX dataset and extended to the resolution of this dataset with random selection. Refinement was performed with Phenix and REFMAC. Sidechains and water molecules were rebuilt manually and iteratively with refinement. The O-O group was modeled as peroxide (PER), due to reduction by the X-ray beam.
**Table 3.5.** Data collection and refinement statistics for oxy-F33Y Cu₈Mb structure.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Oxy-I107E-Cu₈Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>SSRL 11-1</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2015-07-20</td>
</tr>
<tr>
<td><strong>Detector</strong></td>
<td>DECTRIS Pilatus 6M</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data Reduction</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Software</strong></td>
<td>XDS v2015-03-01</td>
</tr>
<tr>
<td></td>
<td>Pointless v1.9.33</td>
</tr>
<tr>
<td></td>
<td>Aimless v0.5.12</td>
</tr>
<tr>
<td></td>
<td>Truncate v6.5.013 (Linux)</td>
</tr>
<tr>
<td><strong>Spacegroup</strong></td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td><strong>Unit cell (a x b x c; Å)</strong></td>
<td>39.84 x 46.98 x 78.03</td>
</tr>
<tr>
<td><strong>Resolution range (highest shell)</strong></td>
<td>39-1.18 (1.20-1.18)</td>
</tr>
<tr>
<td><strong>Overall B-factor, ML (Å²)</strong></td>
<td>7.19</td>
</tr>
<tr>
<td><strong>Overall B-factor, Wilson (Å²)</strong></td>
<td>11.18</td>
</tr>
<tr>
<td><strong>Unique reflection</strong></td>
<td>47352 (1510)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>97.0 (63.4)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>3.9 (2.1)</td>
</tr>
<tr>
<td><strong>&lt;I/σ(I)&gt;</strong></td>
<td>31.7 (4.1)</td>
</tr>
<tr>
<td><strong>CC½</strong></td>
<td>(0.902)</td>
</tr>
<tr>
<td><strong>Rmerge/symm</strong></td>
<td>0.024 (0.216)</td>
</tr>
<tr>
<td><strong>Rpin</strong></td>
<td>0.013 (0.169)</td>
</tr>
<tr>
<td><strong>Rmeas/rim</strong></td>
<td>0.028 (0.275)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Software</strong></td>
<td>Phenix v1.10.1_2155 (Linux)</td>
</tr>
<tr>
<td><strong>Resolution range (highest shell)</strong></td>
<td>39.02-1.18 (1.19-1.18)</td>
</tr>
<tr>
<td><strong>Reflections</strong></td>
<td>86347 (1303)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>92.81 (41)</td>
</tr>
<tr>
<td><strong>Free set (%; random)</strong></td>
<td>4.77</td>
</tr>
<tr>
<td><strong>Rwork</strong></td>
<td>0.1555 (0.2329)</td>
</tr>
<tr>
<td><strong>Rfree</strong></td>
<td>0.1874 (0.3038)</td>
</tr>
<tr>
<td><strong>Molecules in ASU</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong># of Atoms</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>1238</td>
</tr>
<tr>
<td><strong>Heme + Oxygen</strong></td>
<td>45</td>
</tr>
<tr>
<td><strong>Solvent</strong></td>
<td>307</td>
</tr>
<tr>
<td><strong>Average Isotropic B-factor</strong></td>
<td>17.48</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>15.33</td>
</tr>
<tr>
<td><strong>Heme + Oxygen</strong></td>
<td>10.46</td>
</tr>
<tr>
<td><strong>Solvent</strong></td>
<td>27.19</td>
</tr>
<tr>
<td><strong>Standard Uncertainty (ML)</strong></td>
<td>0.10</td>
</tr>
<tr>
<td><strong>Bond RMSD (Å)</strong></td>
<td>0.015</td>
</tr>
<tr>
<td><strong>Angle RMSD (Å)</strong></td>
<td>1.722</td>
</tr>
<tr>
<td><strong>Ramachandran favored/disallowed (%)</strong></td>
<td>97.42/0.00</td>
</tr>
</tbody>
</table>
3.5.10. Cryoreduced oxy-ferrous EPR sample preparation

Each protein was concentrated to ~1-3 mM concentration and reduced with excess dithionite anaerobically under nitrogen atmosphere. Dithionite was removed under anaerobic conditions by a small GE Sephadex G-25 desalting column. The protein was reconcentrated anaerobically. The protein solution was removed from the anaerobic chamber and immediately, slowly, and gently sparged with pure oxygen gas from a syringe. The protein was briefly centrifuged to remove bubbles and mixed with pure glycerol to yield 15% final glycerol concentration. This solution was quickly transferred to an EPR tube and slowly frozen with liquid nitrogen.

Samples were transferred to a glass dewar filled with liquid nitrogen and irradiated with ∼3 Mrad of gamma radiation over 440 minutes using a Gammacell 220 irradiator containing a ^{60}Co source. Irradiated samples were stored and transported in liquid nitrogen at all times.

3.5.11. EPR spectroscopy

X-band CW EPR spectroscopy was performed using a Bruker ESP-300 spectrometer equipped with an Oxford Instruments ESR-900 helium flow cryostat. EPR spectra were recorded at 20K, a modulation frequency of 100 kHz, a modulation amplitude of 5 G, microwave power of 2 mW and with a sweep rate of 10 G/s. Spectra were recorded at microwave frequencies of approximately 9.364 GHz, with the precise microwave frequencies recorded for individual spectra using a Hewlett-Packard Microwave Frequency Counter (HP5352B).

3.6. References

(1) Kaila, V. R. I.; Verkhovsky, M. I.; Wikström, M. Proton-coupled electron transfer in cytochrome oxidase Chemical reviews 2010, 110, 7062.


(3) Gennis, R. B. Multiple proton-conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine Biochimica et Biophysica Acta (BBA) - Bioenergetics 1998, 1365, 241.


(13) Cukier, R. I. A molecular dynamics study of water chain formation in the proton-conducting K channel of cytochrome c oxidase Biochimica et Biophysica Acta (BBA) - Bioenergetics 2005, 1706, 134.

(14) Wikström, M.; Verkhovsky, M. I. The D-channel of cytochrome oxidase: An alternative view Biochimica et Biophysica Acta (BBA) - Bioenergetics 2011, 1807, 1273.


(16) Petrik, I. D.; Davydov, R.; Ross, M.; Zhao, X.; Hoffman, B.; Lu, Y. Spectroscopic and crystallographic evidence for the role of a water-containing hydrogen bonding network in
oxidase activity of an engineered myoglobin Journal of the American Chemical Society 2016, accepted.


(34) Schroedinger LLC 2015.


(40) Gonzalez, A.; Tsai, Y. 2010.

CHAPTER 4
DESIGN OF A HEME-Fe₄S₄ SITE IN CcP TO MIMIC SiR AND NiR

4.1. Background – SiR and NiR

Sulfite and nitrite reductases (SiR and NiR) are two related classes of enzymes that perform the intensive 6 e⁻ reductions of sulfite and nitrite directly to hydrogen sulfide and ammonia, respectively.¹⁻⁵ In bacteria, SiRs play a crucial role in the Cys synthesis pathway. These reactions are catalyzed at a unique active site, consisting of a novel Cys-ligated, five-coordinate Fe tetrapyrole cofactor, bridged via the Cys-ligand to a Fe₄S₄ cluster.¹⁻⁹ Using this arrangement, these enzymes catalyze sulfite and nitrite reduction efficiently, without release of partially reduced intermediates.¹⁻⁵ It is believed that this unique coupled cofactor acts as a “molecular battery,” capable of rapid and complete 6e⁻ reduction of the substrate, in clean 2e⁻ or 3e⁻ transfers.⁴⁻¹⁰⁻¹⁹ DFT calculations suggest that catalytic intermediates of SiRs/NiRs may be analogous to those of NO and O₂ reducing enzymes.²⁰ Therefore, given the success of our lab in engineering simplified enzymes with similar function to native enzymes to expand our understanding of that function, we sought to apply the biosynthetic to the design of the unique heterobimetallic heme-Fe₄S₄ site of SiR/NiR into the CcP scaffold.

4.2. Design

4.2.1. Scaffold selection

The first criterion for scaffold selection was that the protein be crystallized and known to have a tetrapyrole binding site. The MARKUS server²¹ identified the peroxidase fold as a promising fold due to the amount of space potentially available on the proximal side of the heme, if native sidechains are made smaller. Fortunately, the Lu lab has experience working with cytochrome c peroxidase (CcP)²²⁻²⁶ and has previously engineered the Cu₅ center into CcP.²⁶ Furthermore, CcP has been previously successfully mutated to have Cys ligation to the heme.²³ Therefore, this scaffold was used for subsequent design. Prior to searching for how to incorporate the Fe₄S₄-cluster in the proximal pocket of the heme, molecular modeling was carried out to change the heme ligand to Cys, and this structure was used for subsequent work.

4.2.2. RosettaMatch – “Theozyme” design

The Rosetta suite, developed by David Baker and coworkers at University of Washington,²⁷⁻³⁰ has proven extremely powerful in designing new proteins and enzymes.³⁻³⁵ In particular, the RosettaMatch algorithm²⁹ is an extremely powerful, thorough, and automated search protocol to identify sites, and corresponding mutations in a protein that can accommodate specific interactions
with a ligand. This was used to identify sites that could bind the transition states of the retro-aldol reaction\textsuperscript{32} and the Kemp-elimination,\textsuperscript{33} and to re-engineer natural proteins to carry out these unnatural reactions. Here, we use this powerful tool to identify mutations that can allow the Fe\textsubscript{4}S\textsubscript{4} cluster to bind in the proximal pocket of the heme.

In order to use RosettaMatch, the desired interactions of the cofactor with the protein must be defined. This definition is called a “theozyme” – short for theoretical enzyme, due to the fact that it typically describes the key minimal interactions that are expected to impart the desired enzymatic activity. In this case, the theozyme describes the allowable binding modes of the Fe\textsubscript{4}S\textsubscript{4} cluster with its ligands. In order to define the theozyme, the structure of native SiR was inspected.\textsuperscript{16,36,37} The structure of the Fe\textsubscript{4}S\textsubscript{4} cluster itself was built by measuring the average Fe-Fe and S-S distances. The average Fe-Fe distance was measured as 2.78 Å and the average S-S distance as 3.70 Å. The cluster was obtained by building two tetrahedra with edge lengths corresponding to these distances, with Fe and S atoms at the vertices, and intersecting these to form a cuboid structure. The resulting Fe-S distances were 2.336 Å, which is roughly half way between the distances known for oxidized and reduced [Fe\textsubscript{4}S\textsubscript{4}] clusters.

In SiR, the cluster is held in place by Cys coordination of each of the Fe vertices.\textsuperscript{16,36,37} The possible placements of the cluster were determined relative to the heme-ligating Cys. The constraints used to define these placements are shown graphically in Figure 4.1. To narrow down these placements, necessary interaction constraints with additional Cys residues were defined, including the Fe-S bond length and the C\textsubscript{\beta}-S\textsubscript{Cys}-Fe and S\textsubscript{Cys}-Fe-S angles. These are shown graphically in Figure 4.1. The constraints definition file is given in Materials and methods.

![Figure 4.1](image-url)

**Figure 4.1.** Graphical depiction of constraints used to search for [Fe\textsubscript{4}S\textsubscript{4}] cluster binding site.
4.2.3. **RosettaMatch – Results**

RosettaMatch was carried out using these constraints, searching within the proximal pocket of the heme scaffold. The first attempt, finding 4-Cys residues to bind the FeS cluster within the proximal pocket, was unsuccessful. Only 3 ligand constraints could be satisfied, but the program could not find a fourth. In order to determine whether any of the intermediate hits might give viable starting points for manually designing the fourth ligand, RosettaMatch was rerun to look for only 3 ligands. Three hits were identified, however only one of these did not cause clashes with the heme – H175C/W191C/L232C (Figure 4.2).

![Figure 4.2. Best result of RosettaMatch attempt 2.](image)

4.2.4. **MD simulations**

In order to identify the best place, if any existed, to build the fourth Cys ligand, the result of RosettaMatch was first be subject to molecular dynamics simulation to allow the structure to relax around the cluster. In order to carry out this simulation, the force field parameters for the Fe₄S₄ cluster and its interactions with the Cys residues and heme needed to be defined. These were derived from Chang and Kim, and the relevant parameters are given in Materials and methods. In order to make room for the cluster, residues 230 and 235 were additionally mutated to Val or Ala. The results of the best simulation are shown in Figure 4.3.
4.2.5. Manual design of final ligand

After this simulation, it was identified that residue 180 could act as a Cys ligand to the $\text{[Fe}_4\text{S}_4\text{]}$-cluster if the backbone loop is allowed to move slightly. This mutation was made in silico and the resulting cluster site was simulated. The overlay of this designed site with the heme-$\text{[Fe}_4\text{S}_4\text{]}$ site of native SiR shows relatively close structural agreement between the two (Figure 4.4).

4.3. Further experimental refinement and ERP characterization

This construct has been engineered by mutagenesis and subject to basic characterization. Incorporation of the $\text{Fe}_4\text{S}_4$ cluster was carried out according to published techniques, and been confirmed by EPR spectroscopy (Figure 4.5). Activity assays and further characterization are actively underway.
Figure 4.5. EPR spectra of H175C/T180C/W191C/M230A/L232C/D235V-CcP, reduced with dithionite.

4.4. Materials and methods

4.4.1. RosettaMatch constraint files


```plaintext
CST::BEGIN
  TEMPLATE::  ATOM_MAP: 1 atom_name: FE1 S4 FE2
  TEMPLATE::  ATOM_MAP: 1 residue3: F40
  TEMPLATE::  ATOM_MAP: 2 atom_name: SG CB CA
  TEMPLATE::  ATOM_MAP: 2 residue1: C
  CONSTRAINT:: distanceAB:  2.40 0.30 100.00 1 2
  CONSTRAINT::  angle_A: 106.00 6.00 100.00 360.00 1
  CONSTRAINT::  angle_B: 170.00 10.00 50.00 360.00 2
  CONSTRAINT::  torsion_A: 000.00 10.00 50.00 10.00 0
  CONSTRAINT::  torsion_B: 000.00 60.00 25.00 360.00 6
  CONSTRAINT::  torsion_AB: 000.00 10.00 0.00 10.00 0

  ALGORITHM_INFO:: match
    CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
    IGNORE_UPSTREAM_PROTON_CHI

CST::END

CST::BEGIN
  TEMPLATE::  ATOM_MAP: 1 atom_name: FE2 S3 FE1
  TEMPLATE::  ATOM_MAP: 1 residue3: F40
  TEMPLATE::  ATOM_MAP: 2 atom_name: SG CB CA
  TEMPLATE::  ATOM_MAP: 2 residue1: C
  CONSTRAINT:: distanceAB:  2.40 0.30 100.00 1 2

CST::END
```
Code 4.1 (cont.)

```
CONSTRrAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRrAINT:: angle_B: 170.00 10.00 50.00 360.00 2

ALGORITHM_INFO:: match
  CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
  IGNORE_UPSTREAM_PROTON_CHI
  SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END

CST::BEGIN

TEMPLATE:: ATOM_MAP: 1 atom_name: FE3 S2 FE1
TEMPLATE:: ATOM_MAP: 1 residue3: F4O

TEMPLATE:: ATOM_MAP: 2 atom_name: SG CB CA
TEMPLATE:: ATOM_MAP: 2 residue1: C

CONSTRrAINT:: distanceAB: 2.40 0.30 100.00 1 2
CONSTRrAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRrAINT:: angle_B: 170.00 10.00 50.00 360.00 2

ALGORITHM_INFO:: match
  CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
  IGNORE_UPSTREAM_PROTON_CHI
  SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END

CST::BEGIN

TEMPLATE:: ATOM_MAP: 1 atom_name: FE4 S1 FE1
TEMPLATE:: ATOM_MAP: 1 residue3: F4O

TEMPLATE:: ATOM_MAP: 2 atom_name: SG CB CA
TEMPLATE:: ATOM_MAP: 2 residue1: C

CONSTRrAINT:: distanceAB: 2.40 0.30 100.00 1 2
CONSTRrAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRrAINT:: angle_B: 170.00 10.00 50.00 360.00 2

ALGORITHM_INFO:: match
  CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
  IGNORE_UPSTREAM_PROTON_CHI
  SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END
```

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: FE1 S4 FE2
TEMPLATE:: ATOM_MAP: 1 residue3: F40

TEMPLATE:: ATOM_MAP: 2 atom_name: SG CB CA
TEMPLATE:: ATOM_MAP: 2 residue1: C

CONSTRAINT:: distanceAB: 2.40 0.30 100.00 1 2
CONSTRAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRAINT:: angle_B: 170.00 10.00 50.00 360.00 2
CONSTRAINT:: torsion_A: 000.00 10.00 50.00 10.00 0
CONSTRAINT:: torsion_B: 000.00 60.00 25.00 360.00 6
CONSTRAINT:: torsion_AB: 000.00 10.00 0.00 10.00 0

ALGORITHM_INFO:: match
   CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
       IGNORE_UPSTREAM_PROTON_CHI
   SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: FE2 S3 FE1
TEMPLATE:: ATOM_MAP: 1 residue3: F40

TEMPLATE:: ATOM_MAP: 2 atom_name: SG CB CA
TEMPLATE:: ATOM_MAP: 2 residue1: C

CONSTRAINT:: distanceAB: 2.40 0.30 100.00 1 2
CONSTRAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRAINT:: angle_B: 170.00 10.00 50.00 360.00 2

ALGORITHM_INFO:: match
   CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
       IGNORE_UPSTREAM_PROTON_CHI
   SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: FE3 S2 FE1
TEMPLATE:: ATOM_MAP: 1 residue3: F40

TEMPLATE:: ATOM_MAP: 2 atom_name: SG CB CA
TEMPLATE:: ATOM_MAP: 2 residue1: C

CONSTRAINT:: distanceAB: 2.40 0.30 100.00 1 2
CONSTRAINT:: angle_A: 106.00 6.00 100.00 360.00 1
CONSTRAINT:: angle_B: 170.00 10.00 50.00 360.00 2

ALGORITHM_INFO:: match
   CHI_STRATEGY:: CHI 1 EX_FOUR_HALF_STEP_STDDEVS
       IGNORE_UPSTREAM_PROTON_CHI
   SECONDARY_MATCH: DOWNSTREAM
ALGORITHM_INFO::END

CST::END
**Code 4.3.** Rosetta ligand parameter file for [Fe₄S₄] cluster.

```plaintext
NAME F40
IO_STRING F40 Z
TYPE LIGAND
AA UNK
ATOM FE1  Fe3p  X   1.58
ATOM  S2  S     X  -1.58
ATOM FE4  Fe3p  X   1.58
ATOM FE2  Fe3p  X   1.58
ATOM  S4  S     X  -1.58
ATOM FE3  Fe3p  X   1.58
ATOM  S3  S     X  -1.58
ATOM S1  S     X  -1.58
BOND S2  FE4
BOND S2  FE1
BOND FE4  FE1
BOND FE4  FE2
BOND FE4  S3
BOND FE4  S4
BOND FE4  FE3
BOND FE1  FE2
BOND FE1  S3
BOND FE1  FE3
BOND FE1  S1
BOND FE2  S4
BOND FE2  FE3
BOND FE2  S1
BOND S3  FE3
BOND S4  FE3
BOND FE3  S1
NBR_ATOM FE1
NBR_RADIUS 3.967548
ICOOR_INTERNAL FE1  0.000000  0.000000  0.000000  FE1  S2  FE4
ICOOR_INTERNAL S2  0.000000  100.000000  2.335732  FE1  S2  FE4
ICOOR_INTERNAL FE4  0.000000  107.421730  2.343594  S2  FE1  FE4
ICOOR_INTERNAL FE2  -76.847206  126.420234  2.769515  FE4  S2  FE1
ICOOR_INTERNAL S4  -159.917952  126.420234  2.343594  FE2  FE4  S2
ICOOR_INTERNAL FE3  -76.847206  107.421730  2.335732  S4  FE2  FE4
ICOOR_INTERNAL S3  93.812278  75.246352  2.335732  FE3  S4  FE2
ICOOR_INTERNAL S1  -110.438908  74.935730  2.343594  FE3  S4  S3
```

**Code 4.4.** Flags file for running attempt 1

```plaintext
-s ../Cys-heme_CcP_frame1003_res1Met.pdb
-match:lig_name F4O
-match:scaffold_active_site_residues_for_geomcsts ../Cys-heme_CcP_attempt1.pos
-match:geometric_constraint_file ../Cys-heme_CcP_attempt1.cst
-extra_res_fa ../4FE4S.params
-use_input_sc
-match:output_format CloudPDB
-match_grouper SameSequenceAndDSPositionGrouper
#-output_matchres_only true
-enumerate_ligand_rotamers
-only enumerate_non_match_redundant_ligand_rotamers
-packing:ex1
-packing:ex2
```
4.4.2. Molecular dynamics simulations

Molecular dynamics simulations were carried out in NAMD. Mutations were made using PSFGEN. Standard CHARMM topology and parameter files were used, with the added parameters derived from Chang and Kim in the next section. The water box size was at least 3 times the size of the protein in each direction. Simulation was carried out with 5000 steps of minimization, followed by several nanoseconds of dynamics.

4.4.3. NAMD topology and parameter files

Code 4.5. Positions file for attempt 1 (for attempt to change N_CST to 3 and remove line 4)

N_CST 4
1: 175
2: 172 180 191 202 230 232 235 238
3: 172 180 191 202 230 232 235 238
4: 172 180 191 202 230 232 235 238

Code 4.6. Topology file for Fe4S4 cluster

! [4Fe-4S]2+ resname F40
MASS 5 FEJO 55.84700 !
MASS 6 SJO 32.06600 !

!!! Oxidized [4Fe4S] cluster segments HYDA, HYDB, HYDC
RESI F40 0.54050 !

GROUP
ATOM FE1 FEJO 1.10417
ATOM FE2 FEJO 1.10417
ATOM FE3 FEJO 1.10417
ATOM FE4 FEJO 1.10417
ATOM S1 SJO -0.96905
ATOM S2 SJO -0.96905
ATOM S3 SJO -0.96905
ATOM S4 SJO -0.96905
BOND FE1 S1 FE1 S2 FE1 S3 FE2 S1
BOND FE2 S2 FE2 S4 FE3 S2 FE3 S3
BOND FE3 S4 FE4 S1 FE4 S3 FE4 S4
BOND FE1 FE2 FE1 FE3 FE1 FE4 FE2 FE3
BOND FE2 FE4 FE3 FE4
ANGL FE1 S3 FE4 FE1 S3 FE3 S4 FE2 FE1
ANGL FE1 S2 FE3 FE1 S2 FE2 FE1 S1 FE4
ANGL FE1 S1 FE2 FE1 S1 FE4 FE1 FE4 FE2
ANGL FE1 FE3 FE4 FE1 FE1 FE2 FE2 FE4
ANGL FE1 FE2 FE3 FE2 S4 FE4 FE2 S4 FE3
ANGL FE2 S2 FE3 FE2 S1 FE4 FE2 FE4 FE3
ANGL FE2 FE3 FE4 FE2 FE1 FE4 FE2 FE1
ANGL FE3 S4 FE4 FE3 S3 FE4 FE3 FE2 FE4
ANGL FE3 FE1 FE4 S1 FE4 FE3 S1 FE4 FE2
ANGL S1 FE4 FE1 S1 FE4 S4 S1 FE4 S3
ANGL S1 FE2 FE4 S1 FE2 FE3 S1 FE2 FE1
ANGL S1 FE2 S4 S1 FE2 S2 S1 FE1 FE4
ANGL S1 FE1 FE3 S1 FE1 FE2 S1 FE1 S3
ANGL S1 FE1 S2 S2 FE3 FE4 S2 S2 FE3 S3
ANGL S2 FE3 FE1 S2 FE3 S4 S2 FE3 S3
ANGL S2 FE2 FE4 S2 FE2 FE3 S2 FE2 FE1
ANGL S2 FE2 S4 S2 FE1 FE4 S2 FE1 FE3
Code 4.6. (cont.)

ANGL S2 FE1 FE2 S2 FE1 S3 S3 FE4 FE3
ANGL S3 FE4 FE2 S3 FE4 FE1 S3 FE4 S4
ANGL S3 FE3 FE4 S3 FE3 FE2 S3 FE3 FE1
ANGL S3 FE3 S4 S3 FE3 FE1 S3 FE3 FE1
ANGL S3 FE1 FE2 S4 FE4 FE3 S4 FE4 FE2
ANGL S4 FE4 FE1 S4 FE4 S4 FE3 FE2 FE3
ANGL S4 FE3 FE1 S4 FE2 FE4 S4 FE2 FE3
DIHE FE1 S3 FE3 FE2 FE1 S3 FE3 FE4
DIHE FE1 S3 FE4 FE2 FE1 S3 FE4 FE3
DIHE FE1 S2 FE2 FE3 FE1 S2 FE2 FE4
DIHE FE1 S2 FE3 FE2 FE1 S2 FE3 FE4
DIHE FE1 S1 FE2 FE3 FE1 S1 FE2 FE4
DIHE FE1 S1 FE4 FE2 FE1 FE2 FE4 FE3
DIHE FE1 FE4 S4 FE3 FE1 FE4 S4 FE2
DIHE FE1 FE4 FE2 S2 FE1 FE4 FE2 S4
DIHE FE1 FE4 FE2 FE3 FE1 FE4 FE2 S1
DIHE FE1 FE4 S3 FE3 FE1 FE4 S3 FE2
DIHE FE1 FE4 FE3 FE2 FE1 FE3 S2 FE2
DIHE FE1 FE4 FE3 S3 FE1 FE3 S3 FE3
DIHE FE1 FE4 FE3 S4 FE1 FE3 S4 FE4
DIHE FE1 FE3 FE2 S2 FE1 FE3 S2 FE2
DIHE FE1 FE3 S4 FE4 FE1 FE3 S4 FE2
DIHE FE1 FE3 FE2 S4 FE1 FE3 S4 FE4
DIHE FE1 FE3 S1 FE1 FE3 S1 FE1 FE4
DIHE FE1 FE3 S1 FE4 FE1 FE3 S1 FE4
DIHE FE1 FE3 S1 FE4 FE2 FE1 FE4 S1
DIHE FE1 FE3 FE4 S4 FE2 FE1 S1 FE4
DIHE FE1 FE3 FE4 FE2 FE1 FE3 S2
DIHE FE1 FE3 FE4 FE3 S3 FE1 FE3 S3 FE3
DIHE FE1 FE3 FE4 S3 FE2 FE4 S3 FE3 S4
DIHE FE1 FE3 FE4 S4 FE2 FE4 S4 FE3 S1
DIHE FE1 FE3 S2 FE1 FE3 S2 FE1 FE4
DIHE FE1 FE4 FE2 S2 FE1 S2 FE2 S2 FE2
DIHE FE1 FE4 FE2 S4 FE1 FE3 S4 FE2
DIHE FE1 FE4 FE4 S4 FE1 FE3 S4 FE4
DIHE FE1 FE4 FE4 FE2 FE1 FE4 S4 FE3
DIHE FE1 FE4 S4 FE2 FE1 S4 FE4 S1
DIHE FE1 FE4 S4 FE2 FE1 FE3 S2
DIHE FE1 FE4 S4 FE2 FE1 FE3 S4 FE3
DIHE FE1 FE4 S4 FE2 FE1 FE3 S4 FE4
DIHE FE1 FE4 S4 FE2 FE1 FE3 S1
DIHE FE1 FE4 S4 FE2 FE1 FE3 S1
DIHE FE1 FE4 S4 FE2 FE1 FE3 S1
DIHE FE1 FE4 S4 FE2 FE1 FE3 S1
Code 4.6 (cont.)

DIHE FE3 FE2 FE4 S3 FE3 FE2 FE4 S4
DIHE FE3 FE1 S1 FE4 FE3 FE1 S3 FE4
DIHE FE3 FE1 FE2 S1 FE3 FE1 FE2 S2
DIHE FE3 FE1 FE2 S4 FE3 FE1 FE2 FE4
DIHE FE3 FE1 FE4 S1 FE3 FE1 FE4 S3
DIHE FE4 FE3 FE1 S2 FE4 FE3 FE1 S3
DIHE FE4 FE3 FE2 S1 FE4 FE3 FE2 S2
DIHE FE4 FE3 FE2 S4 FE4 FE2 FE1 S1
DIHE FE4 FE2 FE1 S2 FE4 FE2 FE1 S3
DIHE FE4 FE2 FE3 S2 FE4 FE3 FE2 S1
DIHE FE4 FE2 FE3 S4 FE4 FE1 FE2 S1
DIHE FE4 FE2 FE4 S1 FE4 FE2 S1
DIHE FE4 FE3 FE1 S4 FE4 FE3 FE1 S3
DIHE FE4 FE3 FE2 S1 FE4 FE2 S1
DIHE FE4 FE3 FE2 S4 FE4 FE1 FE2 S1
DIHE FE4 FE2 FE1 S2 FE4 FE2 FE1 S3
DIHE FE4 FE2 FE3 S2 FE4 FE3 FE2 S1
DIHE FE4 FE2 FE3 S4 FE4 FE1 FE2 S1
DIHE FE4 FE1 FE2 S2 FE4 FE1 FE2 S4
DIHE FE4 FE1 FE3 S2 FE4 FE1 FE3 S3
DIHE FE4 FE1 FE3 S4 FE4 FE1 FE3 S4
DIHE S1 FE4 FE1 S3 S1 FE4 FE1 S2
DIHE S1 FE4 FE2 S4 S1 FE4 FE2 S2
DIHE S1 FE4 FE3 S3 S1 FE4 FE3 S2
DIHE S1 FE2 FE1 S2 S1 FE2 FE1 S3
DIHE S1 FE2 FE3 S2 S1 FE2 FE3 S3
DIHE S1 FE2 FE3 S4 S1 FE2 FE3 S3
DIHE S1 FE2 FE4 S4 S1 FE2 FE3 S3
DIHE S1 FE1 FE2 S4 S1 FE1 FE3 S2
DIHE S1 FE1 FE3 S4 S1 FE1 FE3 S2
DIHE S1 FE1 FE4 S3 S1 FE1 FE3 S4
DIHE S2 FE3 FE1 S3 S2 FE3 FE2 S4
DIHE S2 FE3 FE4 S3 S2 FE3 FE4 S4
DIHE S2 FE2 FE1 S3 S2 FE2 FE3 S3
DIHE S2 FE2 FE3 S4 S2 FE2 FE4 S3
DIHE S2 FE2 FE4 S4 S2 FE2 FE4 S3
DIHE S2 FE1 FE3 S3 S2 FE1 FE3 S4
DIHE S2 FE1 FE4 S3 S2 FE1 FE3 S4
DIHE S3 FE4 FE2 S4 S3 FE4 FE3 S4
DIHE S3 FE4 FE2 S4 S3 FE4 FE3 S4
DIHE S3 FE1 FE2 S4 S3 FE1 FE3 S4
DIHE S3 FE1 FE4 S4 S3 FE1 FE3 S4

PRES F4-10 -0.78829 ! patch for CY5 to FE1 of oxidized [4Fe4S] (1-F4O 2-CYS).
ATOM 2N NH1 -0.47
ATOM 2HN H 0.31
ATOM 2CA CT1 0.07
ATOM 2HA HB 0.09
ATOM 2CB CT2 -0.898872
ATOM 2HB1 HA 0.342857
ATOM 2HB2 HA 0.342857
ATOM 2SG S -0.575132
DELETE ATOM 2HG1
BOND 1FE1 2SG
ANGL 1FE2 1FE1 2SG
ANGL 1FE3 1FE1 2SG
ANGL 1FE4 1FE1 2SG
ANGL 1S1 1FE1 2SG
ANGL 1S2 1FE1 2SG
ANGL 1S3 1FE1 2SG
ANGL 1FE1 2SG 2CB
DINE 1FE1 2SG 2CB 2HB1
DINE 1FE1 2SG 2CB 2HB2
DINE 1FE2 1FE1 2SG 2CB
Code 4.6. (cont.)

    DIHE 1FE2 1FE3 1FE1 2SG
    DIHE 1FE2 1FE4 1FE1 2SG
    DIHE 1FE2 1S1 1FE1 2SG
    DIHE 1FE2 1S2 1FE1 2SG
    DIHE 1FE2 1S3 1FE1 2SG
    DIHE 1FE3 1FE1 2SG 2CB
    DIHE 1FE3 1FE2 1FE1 2SG
    DIHE 1FE3 1FE2 1FE1 2SG
    DIHE 1FE3 1S1 1FE1 2SG
    DIHE 1FE3 1S2 1FE1 2SG
    DIHE 1FE3 1S3 1FE1 2SG
    DIHE 1FE4 1FE1 2SG 2CB
    DIHE 1FE4 1FE2 1FE1 2SG
    DIHE 1FE4 1FE3 1FE1 2SG
    DIHE 1FE4 1FE4 1FE1 2SG
    DIHE 1FE4 1S1 1FE1 2SG
    DIHE 1FE4 1S2 1FE1 2SG
    DIHE 1FE4 1S3 1FE1 2SG
    DIHE 1S1 1FE2 1FE1 2SG
    DIHE 1S1 1FE4 1FE1 2SG
    DIHE 1S1 1FE1 2SG 2CB
    DIHE 1S2 1FE2 1FE1 2SG
    DIHE 1S2 1FE3 1FE1 2SG
    DIHE 1S2 1FE1 2SG 2CB
    DIHE 1S3 1FE3 1FE1 2SG
    DIHE 1S3 1FE4 1FE1 2SG
    DIHE 1S3 1FE1 2SG 2CB
    DIHE 1S4 1FE2 1FE1 2SG
    DIHE 1S4 1FE3 1FE1 2SG
    DIHE 1S4 1FE4 1FE1 2SG

PRES F4-2O -0.78829 ! patch for CYS to FE2 of oxidized [4Fe4S] (1-F4O 2-CYS).

ATOM 2N    NH1  -0.47
ATOM 2HN   H    0.31
ATOM 2CA   CT1  0.07
ATOM 2HA   HB   0.09
ATOM 2CB   CT2  -0.898872
ATOM 2HB1  HA   0.342857
ATOM 2HB2  HA   0.342857
ATOM 2SG   S    -0.575132
DELETE ATOM 2HG1
BOND 1FE2 2SG
ANGL 1FE1 1FE2 2SG
ANGL 1FE3 1FE2 2SG
ANGL 1FE4 1FE2 2SG
ANGL 1S1 1FE2 2SG
ANGL 1S2 1FE2 2SG
ANGL 1S4 1FE2 2SG
ANGL 1FE2 2SG 2CB
DIHE 1FE1 1FE2 2SG 2CB
DIHE 1FE1 1FE3 1FE2 2SG
DIHE 1FE1 1FE4 1FE2 2SG
DIHE 1FE1 1S1 1FE2 2SG
DIHE 1FE1 1S2 1FE2 2SG
DIHE 1FE2 2SG 2CB 2HB1
DIHE 1FE2 2SG 2CB 2HB2
DIHE 1FE3 1FE1 1FE2 2SG
DIHE 1FE3 1FE2 2SG 2CB
DIHE 1FE3 1FE4 1FE2 2SG
DIHE 1FE3 1S2 1FE2 2SG
DIHE 1FE3 1S4 1FE2 2SG
DIHE 1FE4 1FE1 1FE2 2SG
Code 4.6. (cont.)

DIHE 1FE4 1FE2 2SG 2CB
DIHE 1FE4 1FE3 1FE2 2SG
DIHE 1FE4 1S1 1FE2 2SG
DIHE 1FE4 1S4 1FE2 2SG
DIHE 1S1 1FE4 1FE2 2SG
DIHE 1S1 1FE1 1FE2 2SG
DIHE 1S1 1S4 1FE2 2SG
DIHE 1S1 1S4 1FE3 1FE2 2SG
DIHE 1S1 1S4 1FE4 1FE2 2SG
DIHE 1S2 1FE3 1FE2 2SG
DIHE 1S2 1FE1 1FE2 2SG
DIHE 1S2 1FE4 1FE2 2SG
DIHE 1S3 1FE3 1FE2 2SG
DIHE 1S3 1FE1 1FE2 2SG
DIHE 1S3 1S4 1FE2 2SG
DIHE 1S4 1FE3 1FE2 2SG
DIHE 1S4 1FE2 2SG 2CB
DIHE 1S4 1FE4 1FE2 2SG

PRES F4-30 -0.78829 ! patch for CYS to FE3 of oxidized [4Fe4S] (1-F40 2-CYS).

ATOM 2N NH1 -0.47
ATOM 2HN H 0.31
ATOM 2CA CT1 0.07
ATOM 2HA HB 0.09
ATOM 2CB CT2 -0.898872
ATOM 2HB1 HA 0.342857
ATOM 2HB2 HA 0.342857
ATOM 2SG S -0.575132
DELETE ATOM 2HG1
BOND 1FE3 2SG
ANGL 1FE1 1FE3 2SG
ANGL 1FE2 1FE3 2SG
ANGL 1FE4 1FE3 2SG
ANGL 1S2 1FE3 2SG
ANGL 1S3 1FE3 2SG
ANGL 1S4 1FE3 2SG
ANGL 1FE3 2SG 2CB
DIHE 1FE1 1FE2 1FE3 2SG
DIHE 1FE1 1FE4 1FE3 2SG
DIHE 1FE1 1FE3 2SG 2CB
DIHE 1FE1 1S2 1FE3 2SG
DIHE 1FE1 1S3 1FE3 2SG
DIHE 1FE2 1FE1 1FE3 2SG
DIHE 1FE2 1FE3 2SG 2CB
DIHE 1FE2 1FE4 1FE3 2SG
DIHE 1FE2 1S2 1FE3 2SG
DIHE 1FE2 1S4 1FE3 2SG
DIHE 1FE3 2SG 2CB 2HB1
DIHE 1FE3 2SG 2CB 2HB2
DIHE 1FE4 1FE1 1FE3 2SG
DIHE 1FE4 1FE2 1FE3 2SG
DIHE 1FE4 1FE3 2SG 2CB
DIHE 1FE4 1S3 1FE3 2SG
DIHE 1FE4 1S4 1FE3 2SG
DIHE 1S1 1FE1 1FE3 2SG
DIHE 1S1 1FE2 1FE3 2SG
DIHE 1S1 1FE4 1FE3 2SG
DIHE 1S2 1FE1 1FE3 2SG
DIHE 1S2 1FE2 1FE3 2SG
DIHE 1S2 1FE3 2SG 2CB
Code 4.6. (cont.)

DIHE 1S3 1FE1 1FE3 2SG
DIHE 1S3 1FE3 2SG 2CB
DIHE 1S3 1FE4 1FE3 2SG
DIHE 1S4 1FE2 1FE3 2SG
DIHE 1S4 1FE3 2SG 2CB
DIHE 1S4 1FE4 1FE3 2SG

PRES F4-40 -0.635126 ! patch for CYS to FE4 of oxidized [4Fe4S] (F4O 2-CYS).

ATOM 2N NH1 -0.47
ATOM 2HN H 0.31
ATOM 2CA CT1 0.07
ATOM 2HA HB 0.09
ATOM 2CB CT2 -0.82229
ATOM 2HB1 HA 0.342857
ATOM 2HB2 HA 0.342857
ATOM 2SG S -0.49855
DELETE ATOM 2HG1
BOND 1FE4 2SG
ANGL 1FE1 1FE4 2SG
ANGL 1FE2 1FE4 2SG
ANGL 1FE3 1FE4 2SG
ANGL 1S1 1FE4 2SG
ANGL 1S3 1FE4 2SG
ANGL 1S4 1FE4 2SG
ANGL 1FE4 2SG 2CB
DIHE 1FE1 1FE2 1FE4 2SG
DIHE 1FE1 1FE3 1FE4 2SG
DIHE 1FE1 1FE4 2SG 2CB
DIHE 1FE1 1S1 1FE4 2SG
DIHE 1FE1 1S3 1FE4 2SG
DIHE 1FE2 1FE1 1FE4 2SG
DIHE 1FE2 1FE3 1FE4 2SG
DIHE 1FE2 1FE4 2SG 2CB
DIHE 1FE2 1S1 1FE4 2SG
DIHE 1FE2 1S4 1FE4 2SG
DIHE 1FE3 1FE1 1FE4 2SG
DIHE 1FE3 1FE2 1FE4 2SG
DIHE 1FE3 1FE4 2SG 2CB
DIHE 1FE3 1S3 1FE4 2SG
DIHE 1FE3 1S4 1FE4 2SG
DIHE 1FE4 2SG 2CB 2HB1
DIHE 1FE4 2SG 2CB 2HB2
DIHE 1S1 1FE1 1FE4 2SG
DIHE 1S1 1FE2 1FE4 2SG
DIHE 1S1 1FE4 2SG 2CB
DIHE 1S2 1FE1 1FE4 2SG
DIHE 1S2 1FE2 1FE4 2SG
DIHE 1S2 1FE3 1FE4 2SG
DIHE 1S3 1FE1 1FE4 2SG
DIHE 1S3 1FE3 1FE4 2SG
DIHE 1S3 1FE4 2SG 2CB
DIHE 1S4 1FE2 1FE4 2SG
DIHE 1S4 1FE3 1FE4 2SG
DIHE 1S4 1FE4 2SG 2CB

PRES HCB ! patch to bond bridging CYS to FE1 of H-Cluster (1-HCR 2-CYS). Must apply this after F4-X-type patch!
BOND 1FE1 2SG
Code 4.6. (cont.)

ANGLE  2SG  1FE1  1S1  2SG  1FE1  1S2  1FE2  1FE1  2SG
ANGLE  1FE1  2SG  2CB  2SG  1FE1  1C4  2SG  1FE1  1C3
IMPR  2SG  1SI  1FE1  1C5  1SI  1FE1  2SG  1C3

PRES PHCF         0.00 ! Patch for HEME to thiolate (CYS-(F4O)) link
! Patch residues must be 1-CYS, and 2-HEME.
! do NOT use AUTOgenerate ANGLes DIHedrals after this patch

DELETE ANGLE 2NA 2FE 2NC  2NB 2FE 2ND
BOND  1SG 2FE
ANGLE 1CB 1SG  2FE  1CB 1SG  2FE
ANGLE 1SG  2FE 2NA  1SG  2FE 2NB 1SG  2FE  2NC  1SG  2FE  2ND
IC 1CB 1SG  2FE 2NA  0.0000  0.0000  0.0000  0.0000  0.0000
IC 1CB 1SG  2FE 2NB  0.0000  0.0000  0.0000  0.0000  0.0000
IC 1CB 1SG  2FE 2NC  0.0000  0.0000  0.0000  0.0000  0.0000
IC 1CB 1SG  2FE 2ND  0.0000  0.0000  0.0000  0.0000  0.0000

END

Code 4.7. Additional force-field parameters.

! Heme to Sulfate (PSUL) link
S    FE    250.0       2.3200 ! force constant a guess
! equilibrium bond length optimized to reproduce
! CSD survey values of
! 12.341pm0.01 (mean, standard error)
! adm jr., 7/01
SS    FE    250.0       2.3200 ! force constant a guess
! equilibrium bond length optimized to reproduce
! CSD survey values of
! 12.341pm0.01 (mean, standard error)
! adm jr., 7/01

!!! F4O bonds !!!
FEJO  FEJO    38.18   2.628
FEJO  SJO    115.84   2.175
FEJO  S     178.70   2.251

! Heme to Sulfate (PSUL) link
CT2  S    FE    50.0       100.6 ! force constant a guess
! equilibrium angle optimized to reproduce
! CSD survey values
! 107.5pm0.6 (mean, standard error)
! adm jr., 7/01
S    FE    NPH   100.0      90.0 ! force constant a guess
! adm jr., 7/01

! CS    SS    FE    50.0       100.6 ! force constant a guess
! equilibrium angle optimized to reproduce
! CSD survey values
! 107.5pm0.6 (mean, standard error)
! adm jr., 7/01
SS    FE    NPH   100.0      90.0 ! force constant a guess
! adm jr., 7/01

!
Code 4.7. (cont.)

!!! F4O angles
FEJO  FEJO  FEJO  0.713  60.00
SJO  FEJO  SJO  4.313  103.69
FEJO  SJO  FEJO  2.725  74.35
FEJO  FEJO  SJO  2.290  100.49
S  FEJO  SJO  2.449  114.73
FEJO  FEJO  S  0.850  144.61
FEJO  S  CT2  74.621  106.29
FEJO  S  FEM1  500.000  118.00
FEJO  S  FEL1  500.000  118.00

!Heme to Sulfate (PSUL) link
X  FE  S  X  0.0000  4  0.00 ! guess
!adm jr., 7/01
X  FE  SS  X  0.0000  4  0.00 ! guess
!adm jr., 7/01
CT1  CT2  S  FE  0.2000  3  0.00 ! guess
!from methanethiol, HS S CT3 HA
!adm jr., 7/01
X  CS  SS  X  0.2000  3  0.00 ! guess
!from methanethiol, HS S CT3 HA
!adm jr., 7/01

!!! F4O dihedrals !!!
FEJO  S  CT2  HA  14.03  3  0.0
FEJO  FEJO  S  CT2  1.04  3  180.0
CT2  S  FEJO  SJO  1.17  3  0.0
S  FEJO  FEJO  FEJO  0.48  5  180.0
S  FEJO  FEJO  SJO  1.44  0  180.0
FEJO  FEJO  FEJO  SJO  1.16  10  180.0
SJO  FEJO  FEJO  SJO  2.16  0  0.0
S  FEJO  SJO  FEJO  0.64  5  180.0
FEJO  FEJO  SJO  FEJO  0.72  10  180.0
FEJO  FEJO  FEJO  FEJO  0.36  5  180.0
FEJO  SJO  FEJO  SJO  0  0  0

! SJO  0.000000  -0.380000  1.975
FEJO  0.000000  -0.011500  1.442
!

4.5. References


(17) Oliveira, T. F.; Vonrhein, C.; Matias, P. M.; Venceslau, S. S.; Pereira, I. A. C.; Archer, M. The crystal structure of *Desulfovibrio vulgaris* dissimilatory sulfite reductase bound to DsrC provides novel insights into the mechanism of sulfate respiration *Journal of Biological Chemistry* **2008**, *283*, 34141.


(39) Zhao, J.; Li, N.; Warren, P. V.; Golbeck, J. H.; Bryant, D. A. Site-directed conversion of a cysteine to aspartate leads to the assembly of a N iron-sulfur[3Fe-4S] cluster to PsaC of photosystem I. The photoreduction of FA is independent of FB *Biochemistry* **1992**, *31*, 5093.


CHAPTER 5
TOWARDS CRYSTAL STRUCTURE DETERMINATION OF AN 8-17 DNAZYME

5.1. Background

DNA is well known as the long term storage medium of cellular functions. The genes encoded in DNA store the information needed to synthesize proteins which carry out numerous complex chemical and regulatory functions. Translation of DNA code to protein is mediated by RNA.\(^1\) However, in addition to functioning as messengers between DNA and protein synthesis, RNA has been found to perform a number of important functions of its own.\(^2-3\) Splicing of messenger RNA plays an important role in proper protein synthesis.\(^4,5\) RNAs constitute a substantial fraction of the ribosome\(^3\) and are responsible for catalyzing peptide bond formation during protein synthesis. In plants and bacteria, RNA sequences known as riboswitches which control gene expression by rapidly responding to metabolite binding.\(^5-8\) These natural examples of functional nucleic acids set the stage for discovery of non-natural counterparts.

The \textit{in vitro} isolation of functional nucleic acids began in 1990 with the development of a process known as Systematic Evolution of Ligand by EXponential enrichment (SELEX).\(^9,10\) This resulted in isolation of ligand binding RNA aptamers, and novel ribozymes.\(^9-11\) Concurrently, the strategy was also applied to DNA and resulted in the first DNA aptamers for similar targets.\(^12,13\) Shortly thereafter, a DNA sequence capable of phosphodiester bond cleavage was isolated,\(^14\) and the term DNAzyme was coined to describe the class of catalytically active DNA. Since then, numerous DNAzymes have been isolated that perform myriad functions under various specific conditions.\(^15-31\)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure5.1.png}
\caption{The secondary structure and catalytic function of an 8-17 DNAzyme.}
\end{figure}

The best characterized DNAzyme is known as the 8-17 DNAzyme,\(^32\) whose secondary structure is shown in Figure 5.1. This DNAzyme catalyzes phosphodiester bond cleavage of a ribonucleobase in the presence of divalent metal ions, such as Mg\(^{2+}\), Pb\(^{2+}\), Zn\(^{2+}\), and Mn\(^{2+}\). It has been the subject of countless biochemical and biophysical studies, which have shown light on its mechanism.\(^33-43\) Its structure has been probed by numerous techniques, including crosslinking,\(^38,39,42\) FRET,\(^35,40,44,45\) and circular dichroism (CD).\(^40\) Although these studies have
confirmed the secondary structure and provided insights into conformational changes that occur upon interaction with metal, no atomic resolution structure of 8-17 DNAzyme, in its active configuration has been reported. Such a structure would be invaluable in opening a new frontier in the rational engineering of DNAzymes. Recently, the structure RNA-ligating DNAzyme was reported, and is the first and only active DNAzyme structure known.\textsuperscript{46}

5.2. Previous attempts and challenges of DNAzyme crystallization

The only known crystal structure of a DNAzyme sequence is of the so-called 12-23 DNAzyme sequence.\textsuperscript{47} Obtained in 1999 by Joyce and coworkers, the structure reveals the enzyme and substrate strands forming a 2:2 enzyme-substrate complex and adopting a four-way junction structure, inconsistent with its known active configuration. Given the high value of an atomic resolution structure of a DNAzyme, it is all but certain that work to obtain such structures continues to be pursued, though no further progress has been reported.

Nucleic acids pose two major unique crystallographic challenges: (1) uniform negative charge of the backbone results in weak, non-specific crystal packing, resulting in disordered lattices;\textsuperscript{48} (2) the structures adopted by nucleic acids are highly heterogeneous in solution.\textsuperscript{49} Two methods of overcoming these problems are protein co-crystallization and so called “helix-engineering”. Helix engineering involves incorporation of a known crystallizable sequence into a non-essential part of the construct to facilitate crystallization. Co-crystallization involves incorporation of a protein binding sequence or motif onto the DNA to allow the protein to drive crystallization of the DNAzyme. Prior efforts in the Lu lab utilizing these approaches and others to attempt to obtain diffraction quality crystals of active 8-17 DNAzyme are reported briefly. More detailed information about these attempts can be found in the thesis of Dr. Tian Lan (2012).

Co-crystallization of the 8-17 DNAzyme sequence was attempted with two constructs: (1) biotinylated DNA co-crystallized with avidin and (2) lysozyme co-crystallized with a DNAzyme bearing the lysozyme aptamer sequence. The lysozyme co-crystallization did not yield crystals, likely due to the aptamer interfering with lysozyme packing. Co-crystallization with avidin did in fact yield diffraction quality crystals. However, only biotin could be resolved, with the remainder of the DNAzyme too disordered.

Helix-engineering was carried out by introducing a commonly crystallized double-stranded DNA sequence known as the Drew-Dickerson dodecamer (DDD) sequence\textsuperscript{50-63} in the non-catalytic double-stranded arm of the 8-17 DNAzyme. Crystallization of many different constructs with various overhanging nucleotides were attempted under a variety of crystallization screening
conditions. Finally, crystals of two constructs, E-3'T/S-5'A and E/S-3'T, were obtained under a small range of crystallization conditions (see Materials and methods). However, after several attempts to obtain diffraction data from these crystals, the highest resolution achieved was 5 Å, not sufficient to solve the structure.

5.3. Crystal preparation and diffraction screening

As stated above, despite successes in crystallizing the 8-17 DNAzyme sequence, no substantial diffraction could be obtained. This could be due to two main reasons: (1) the quality of crystals themselves is too low to obtain significant diffraction, or (2) the crystal quality deteriorates during crystal handling in preparation for diffraction. In order to determine which is true in this case, I proposed to screen the crystals for diffraction at room temperature. In this way, the crystal could be maintained in the exact same solution and atmosphere under which it grew, minimizing chances of deterioration.

Crystal trays of 8-17 DNAzyme E/S-3'A were carefully transported to the APS on ice. A single ~50 μm diameter crystal (Figure 5.2) was harvested using the mitogen RT sleeve system, with the sleeve filled with motherliquor from the drop. This crystal was placed on LS-CAT beamline 21-ID-F. Screening revealed very good diffraction to ~3.0 Å resolution (Figure 5.3). Indexing of the images revealed a trigonal or hexagonal unit cell with dimensions 64 x 64 x 66 Å. This result confirmed that the quality of the crystals as grown is sufficient for modest diffraction data collection, and suggested that crystal preparation for cryogenic diffraction was responsible for deterioration. The most likely explanation is that the cryoprotectant conditions were incompatible with the crystals.

Figure 5.2. Example drop of 8-17 DNAzyme E/S-3'A crystals.
5.4. Optimization of cryoprotection conditions

Previously, crystals were cryoprotected by mixing PEG 400 or glycerol, up to 30% w/v final concentration, with the well buffer against which the DNA was crystallized. Both the addition of these additives and the dilution of the precipitant concentration by its addition may be responsible for deterioration of the crystal. In order to stabilize the crystal during soaking, the concentration of precipitant was first increased, by soaking the crystal in 1.1 M lithium sulfate. Subsequently, the crystal was soaked in freshly prepared solutions containing 1.1 M lithium sulfate and increasing concentrations of glycerol, up to 20%. No visible deterioration was observed, and the cryocooled crystal scattered X-rays to 2.7 Å resolution. Later, sodium gluconate was also successfully used as a cryoprotectant, also using increased precipitant concentration.

5.5. Preparation of heavy-atom derivatives

One method of solving novel crystal structures involves preparation of one or more heavy-atom derivatives of the crystal (see Section 5.7.2). The native crystal was grown in the presence of lead, but no lead signal could be detected in the crystal by X-ray fluorescence. This is likely due to precipitation of lead in the presence of high sulfate concentration. Likewise, a crystal grown in Sr" did not yield anomalous absorption signal that could be used for structure solving. Attempts to soak the crystal with Zn" and Mn" did not yield anomalous signal. Attempts to grow crystals containing Br-dU and Se-dA have not produced diffractable crystals. Finally, soaking with 1 mM
TaBr cluster produced measurable anomalous diffraction in the crystallographic data, which might be usable for structure solving.

5.6. Data collection

Data collection was carried out at APS on NECAT beamlines 24-ID-C and 24-ID-E. Three datasets were collected: (1) native DNAzyme crystal grown in presence of Pb but no other heavy atoms added; (2) DNAzyme crystals soaked with TaBr clusters; (3) DNAzyme crystals grown in presence of Sr. The data collection statistics are given below.

Table 5.1. Data collection statistics for DNAzyme crystal datasets

<table>
<thead>
<tr>
<th>Source</th>
<th>DNAzyme (native)</th>
<th>DNAzyme (+TaBr)</th>
<th>DNAzyme (Sr)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>2013-03-17</td>
<td>2015-02-13</td>
<td>2015-02-13</td>
</tr>
<tr>
<td>Detector</td>
<td>DETRIS Pilatus 6M</td>
<td>ADSC HF-4M</td>
<td>DETRIS Pilatus 6M</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9497</td>
<td>0.9791</td>
<td></td>
</tr>
</tbody>
</table>

Data Reduction

<table>
<thead>
<tr>
<th>Software</th>
<th>APS 24-ID-C</th>
<th>APS 24-ID-E</th>
<th>APS 24-ID-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spacegroup</td>
<td>P6_22</td>
<td>P6_22</td>
<td>P6_22</td>
</tr>
<tr>
<td>Unit cell (a/b x c; Å)</td>
<td>63.74 x 66.33</td>
<td>63.64 x 66.49</td>
<td>63.64 x 66.49</td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>42.43-2.69 (2.82-2.69)</td>
<td>66.49-2.91 (3.08-2.91)</td>
<td>66.49-2.91 (3.08-2.91)</td>
</tr>
<tr>
<td>Unique reflection</td>
<td>2489 (317)</td>
<td>1988 (298)</td>
<td>1988 (298)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0 (100.0)</td>
<td>99.9 (99.8)</td>
<td>99.9 (99.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>17.3 (15.1)</td>
<td>17.0 (18.0)</td>
<td>17.0 (18.0)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt;</td>
<td>39.7 (2.1)</td>
<td>22.7 (1.1)</td>
<td>22.7 (1.1)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.045 (1.336)</td>
<td>0.105 (3.006)</td>
<td>0.105 (3.006)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.015 (0.470)</td>
<td>0.035 (0.986)</td>
<td>0.035 (0.986)</td>
</tr>
<tr>
<td>Rmeas/rim</td>
<td>0.046 (1.387)</td>
<td>0.046 (1.387)</td>
<td>0.111 (3.166)</td>
</tr>
<tr>
<td>Anomalous slope</td>
<td>1.195</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This dataset could not be processed successfully; see Section 5.7.4.

5.7. Structure solving and troubleshooting

5.7.1. Apparent pathology

While the collection of several high quality datasets at modest resolution was exciting, it was clear from the beginning that something was amiss with this crystal form. Given the unit cell dimensions and symmetry determined, presence of both strands of the DNAzyme in the asymmetrical unit would take up > 100% of the volume of the unit cell. Typical values are ~50%, and often lower for nucleic acid structures. It was hypothesized that the space group might be apparently too high due to twinning, in which two lower symmetry unit cells co-crystallized in random distribution of orientations around one or more axes, to give data that appear a higher symmetry than they are. Luckily, statistical analyses have been developed that can detect twinning pathologies.66-69 This is
based on the fact that twining will result in overlapping of weak and strong reflections, resulting in a greater number of apparently strong reflections than expected. Running these tests on these data, however, did not indicate that twinning is likely (Figure 5.4).

![L-Test](image)

**Figure 5.4.** Results of the L-test for twinning on the native dataset.

An alternative crystal pathology that can result in higher apparent spacegroup is pseudo-symmetry. If half of the DNAzyme structure as crystalized is similar in structure to the other half – after a rotation around an axis parallel to a crystallographic axis – then the point group of the asymmetric unit is higher than the true crystal symmetry and will cause the data to appear to be of higher symmetry than the true crystal symmetry. Alternatively, it might be possible that the crystal only contains one of the strands of DNA, homo-hybridized. In this case, it would be not be possible that the crystal structure is of the real active form. In subsequent attempts to solve the structure of this crystal form, both possibilities we left open, and solutions were pursued in P6,22 and all possible lower symmetry spacegroups.

### 5.7.2. Experimental phasing

Given that no structures of DNAzymes are known, substantial effort was made to phase the structures using heavy-atom derivative difference phasing. A variety of experimental phasing techniques of increasing complexity and robustness have been established. Single-wavelength anomalous difference (SAD) phasing uses differences in intensities of opposed reflections in a single crystal due to X-ray absorption to locate heavy atom sites in structures. Multiple-wavelength
anomalous difference (MAD) phasing measures such differences at wavelengths below and above the absorption edge energy of the absorbing element to increase the S/N-ratio of the difference, but requires more robust crystals to prevent deterioration during several data collections. Single isomorphous replacement (SIR) phasing is carried out using the differences in intensities between two crystals, one containing a heavy atom and the other not. This technique offers similar signal improvements to MAD, but requires that the derivitization not cause major changes in the structure or unit cell parameters (>-1-2%). Multiple isomorphous replacement (MIR) is the most robust technique in terms of quality of the experimentally determined phases and was the first technique used to solve macromolecular crystal structures. However, as with SIR, it requires preparation of several derivatives with minimal perturbation of the structure, which can be challenging to achieve. Additionally, these techniques can be combined in a variety of ways to improve the quality of experimental phases. For instance, single isomorphous replacement with anomalous scattering (SIRAS) combines differences obtained between native and derivative crystals with anomalous differences within the derivative crystal to obtain more robust phases.

![Figure 5.5. SAD phasing solutions in P3 (A) and P6 (B) obtained from TaBr soaked data.](image)

The only crystal that gave measurable anomalous signal was the crystal soaked with TaBr. The crystal grown in Sr was subject to MAD data collection, but further crystal pathologies (see Section 5.7.4) prevented its use for MAD phasing. First attempts to phase the TaBr data using SAD phasing gave partial solutions in P6 and P3 space groups only. Analysis of the resulting electron density (Figure 5.5) showed distinct electron density spacing in certain regions consistent with B-DNA base stacking distances. It was presumed that the overlapping and noisy density was a result of the fact that the spacegroup was wrong. It was suggested that a partial model built into the noisy
density could be a better model to use for molecular replacement to identify the correct spacegroup; however, even a reasonable partial model could not be built from these solutions.

SIR and SIRAS phasing were also attempted with the TaBr data and the native data. Several partial solutions were obtained in a number of spacegroups (Figure 5.6). Again, several of these showed evidence of B-DNA base stacking. Attempts to build models into this density, however, also yielded no results.

Figure 5.6. Selected maps and map contours of solutions obtained from SIR/SIRAS phasing of TaBr and native datasets in P6₁ (A) and P3₂ (B) spacegroups.
5.7.3. Patterson analysis and molecular replacement

A valuable analysis that gives insights into structural features of a crystal before its structure is solved is the Patterson map. This map is generated by mathematical convolution of the data with itself, and yields a representation of distance vectors between atoms. At high resolution (~1.2 Å or better) and with fewer than ~1000 atoms, this map can be used to directly build the structure of a molecule from scratch. At low resolution, this map is used to identify heavy atom derivative sites in experimental phasing experiments, or to solve the phase problem by fitting the known structure of a similar molecule into the data of the structure being solved. Previously, it was shown that long stretches of B-DNA aligned along a cell axis give rise to distinctive peaks in the Patterson map (Figure 5.7A). Patterson analysis of the DNAzyme crystal data shows peaks consistent with this feature (Figure 5.7B). This observation reveals not only that some B-form DNA is present, but also the alignment of the helical axis within the crystal.

Figure 5.7. (A) Patterson map cross-section of ideal pseudo-infinite B-DNA helix crystals. (B) Patterson map cross section of DNAzyme crystal data. Peaks consistent with B-DNA base stacking vectors are labeled with their vector lengths.

The observation of B-DNA is not unexpected, as the construct specifically incorporated the Drew-Dickerson dodecamer (DDD) sequence, which is expected to exist as B-DNA. Recently, Robertson and Scott reported a novel, fragment-based approach to solving RNA crystal structures. They hypothesized that while the 3D structures of new RNA sequences may not be known, the structures of many secondary structural features are well established, and that these fragments could be used as MR search models. Based on this approach, and the observation that B-DNA is present in the crystal, the structure of DDD was used for molecular replacement of our crystal structure in P6_2_2 and lower spacegroups; however, this did not yield any results. Based on a hypothesis that part of the DDD may be disordered, a truncated octamer of the DDD structure was used as the search model. Again, no results were obtained. The molecular replacement protocol
was broken down into steps to attempt to troubleshoot the failure. The first step of MR is a rotation function (RF) search to orient the model within the crystal. Independent RF searches using the DDD octamer yielded successful results. Comparison of the Patterson map of the oriented model with that of the data confirmed that indeed that base-stacking features arise from B-DNA base stacking, and the hypothesis about the orientation of the helical axis. However subsequent translation function searches failed. Finally, the DDD was further broken down into three individual tetrameric B-DNA fragments. Use of these fragments in MR finally yielded solutions with statistics that were on the cusp that separates real solutions from false positives (Figure 5.8). Subsequent refinement of these solutions yielded statistics that were again on the cusp separating false positives and real solutions. However, such statistics were consistent with the results reported by Roberston and Scott, who successfully used this approach to phase a novel RNA structure. They reported that, while the statistics were not suggestive of a real solution, the resulting electron density maps clearly revealed real unmodeled density and areas that were mis-modeled. Visual inspection of the refinement results obtained here did not reveal clear unmodeled structural features. Attempts to auto-build DNA into the resulting electron density did not provide an improved structure from any of the solutions.

![Figure 5.8. Best obtained solutions from MR using tetrameric fragments of DDD structure.](image)

5.7.4. Identification of multilattice pathology

Upon attempted MAD data collection of the Sr-derivatized crystals, there arose some difficulties processing the data. The typical RAPD beamline pipeline failed to integrate and scale the dataset. Careful manual reanalysis showed apparently split spots along an off-diagonal vector
(Figure 5.9). In HKL2000, one set of spots could be indexed giving unit cell with dimensions 63 x 63 x 65 Å. The remaining spots could then be indexed in 63 x 63 x 63 Å. Previously collected datasets were reanalyzed and also seemed to contain this pathology. It seems that there are a mixture of conformations, potentially due to a mixture of hybridization states between the substrate and enzyme strand. Due to the palindromic nature of the DDD sequence, it is possible for the hybridization to be not only E-S, but E-E or S-S as well, with overhangs on either end. Dissolving, quantitative labeling, and gel electrophoresis of a crystal showed that indeed, the strands seem present in unequal ratios in the crystal, suggesting at least two different hybridization modes.

**Figure 5.9.** Image showing double spots.

**Figure 5.10.** Gel of crystal showing unequal ratio of enzyme and substrate strands
5.8. Alternative approaches and future directions

Given that this crystal form seems to have crystallographic pathology consisting of mis-hybridization of the enzyme and substrate strands, it will be necessary to produce crystals of new constructs. Constructs will be screened with various mutations to the DDD to prevent mis-hybridization, and extensions or truncations of the non-catalytic regions to optimize different packing interactions. Linked E-S constructs (called cis-constructs) will be attempted to encourage proper hybridization. In addition, co-crystallization with known DNA-co-crystallizable proteins will be attempted. Proteins such as polymerases bind DNA rigidly, and are known to co-crystallize with DNA. Engineering polymerase binding sequences into the non-catalytic portion of the DNA could allow the structure of the DNA to be determined. Co-crystallization with other prolifically crystallized proteins, such as MMLV-reverse transcriptase will also be pursued. Given the recent successful crystallization of a DNAzyme, we will also attempt to incorporate the non-catalytic regions of this DNAzyme into 8-17 DNAzyme, to enable crystal growth via the same intermolecular contacts.

5.9. Materials and methods

5.9.1. Materials

All DNA sequences were obtained from IDT, unless otherwise noted. Chemicals were used as purchased from Sigma-Aldrich, unless noted otherwise.

5.9.2. DNAzyme crystallization

DNAzyme solution preparation and crystallization trials were carried out by members of the DNA subgroup of the Lu lab. All crystals were obtained by hanging drop vapor diffusion at 4°C after about 6 months. Molecularly imprinted polymers (MIPs) were employed to help nucleate crystal growth. The crystallization conditions are given below.

For room temperature diffraction and the native dataset, DNA solution consisting of 1.25 mM DNA, 1.25 mM lead acetate, 6.25 mM sodium cacodylate pH 6.5, 2.5 mM magnesium chloride, and 2.5 mM H33342, was mixed with crystallization buffer consisting of 0.5 M ammonium sulfate, 100 mM sodium cacodylate pH 5.5, 250 mM magnesium chloride, 1 mM lead acetate, and 5% isopropanol, and equilibrated by hanging drop vapor diffusion against the crystallization buffer. The drop was rehydrated with solution consisting of 1 M ammonium sulfate, 100 mM sodium cacodylate pH 5.5, 250 mM magnesium chloride, and 3% isopropanol, due to dehydration during harvesting.
For the native dataset, the crystal was soaked overnight in 1.1 M lithium sulfate, 100 mM sodium cacodylate pH 5.5, and 250 mM magnesium chloride. The crystal was subsequently transferred to solutions of the above containing increasing amounts of glycerol – 10% then 20%. The crystal was soaked in each of these for 15 min and cryocooled in liquid nitrogen after the 20% soak.

For the TaBr derivatized dataset, the crystal was grown under the same conditions as above. The crystals was soaked in solution containing 1.1 M lithium sulfate, 100 mM MES pH 5.5, 250 mM magnesium chloride, and 10% w/v sodium gluconate for 20 minutes. The crystal was subsequently transferred to the same solution containing 1 mM TaBr cluster and soaked at 4° C overnight. The crystal obtained a dark green/blue color. This crystal was soaked in 1.1 M lithium sulfate, 100 mM MES pH 5.5, 250 mM magnesium chloride, and 30% w/v glycerol and harvested and cryocooled.

Sr co-crystallization was carried out using crystallization buffer consisting of 0.5 M lithium chloride, 250 mM strontium chloride, 100 mM potassium cacodylate pH 5.5, and 25% ethylene glycol. This crystal was harvested and cryocooled directly due to the high ethylene glycol concentration.

5.10. References


(8) Sudarsan, N.; Barrick, J. E.; Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes *Rna* 2003, 9, 644.


(38) Liu, Y.; Sen, D. A contact photo-cross-linking investigation of the active site of the 8-17 deoxyribozyme *J Mol Biol* 2008, 381, 845.


(41) Peracchi, A. Preferential activation of the 8-17 deoxyribozyme by Ca2+ ions. Evidence for the identity of 8-17 with the catalytic domain of the MG5 deoxyribozyme *J Biol Chem.* 2000, 275, 11693.

(42) Sekhon, G. S.; Sen, D. A stereochemical glimpse of the active site of the 8-17 deoxyribozyme from iodine-mediated cross-links formed with the substrate’s scissile site *Biochemistry* 2010, 49, 9072.

(43) Wang, B.; Cao, L.; Chiuman, W.; Li, Y.; Xi, Z. Probing the function of nucleotides in the catalytic cores of the 8-17 and 10-23 DNAzymes by abasic nucleotide and C3 spacer substitutions *Biochemistry* 2010, 49, 7553.


(52) Malinina, L.; Soler-Lopez, M.; Aymami, J.; Subirana, J. A. Intercalation of an acridine-peptide drug in an AA/TT base step in the crystal structure of [d(CGCGAATTCGCG)]₂ with six duplexes and seven Mg(2+) ions in the asymmetric unit *Biochemistry* 2002, 41, 9341.

(53) Liu, J.; Subirana, J. A. Structure of d(CGCGAATTCGCG) in the presence of Ca(2+) ions *J Biol Chem* 1999, 274, 24749.

(54) Ding, D.; Gryaznov, S. M.; Wilson, W. D. NMR solution structure of the N₃' --> P₅' phosphoramidate duplex d(CGCGAATTCGCG)₂ by the iterative relaxation matrix approach *Biochemistry* 1998, 37, 12082.


(64) In *International Tables of Crystallography*; Vol. Volume C.

(65) In *International Tables of Crystallography*; Vol. Volume D.


APPENDIX A
GENERAL METHODS

A.1. Myoglobin expression and purification

Proteins were expressed and purified as previously reported, with minor changes to improve yield and purity. Expression was carried out with E. coli BL-21 DE3 Star strain (Life Technologies); other strains produced inferior yield and purity. For purification from inclusion bodies, dialysis of protein after refolding was carried out against 10 mM tris sulfate pH 8.0. Dialysis was limited to two exchanges – first for 4 hours, second overnight. Protein was loaded onto the size exclusion column within 12 hours of completing dialysis. Protein batch activity was validated by oxygen reduction activity assays.

A.2. Oxygen activity assays

The oxygen reduction rate assays were carried out as previously reported, using a Clark type oxygen electrode, ascorbate and TMPD as reductant and mediator, respectively, and ROS scavengers.

A.2.1. Calculation of reactive oxygen species release rate

The method of oxygen reduction data collection does not directly measure the production of water or reactive oxygen species. These are inferred indirectly from the apparent rate reduction imposed by ROS scavengers, catalase and superoxide dismutase, which disproportionate peroxide and superoxide, respectively, releasing ½ eq. of oxygen. In order to determine production of ROS, the rates were measured in the absence and presence of these scavengers.

For every equivalent of oxygen that is converted to peroxide or superoxide, half of an equivalent will be converted back to oxygen by each of the scavengers. This means the rate of peroxide and superoxide production can be calculated as twice the difference in rate between the rates with and without scavengers. Likewise, the rate of oxygen conversion to water, is estimated as the total rate, less the calculated ROS production rate. Error propagation is used to calculate the errors. This calculation has been built into a standard Excel spreadsheet deposited with this thesis.
Figure A.1. Apparent oxygen consumption rates of G65Y-CuMb in the presence and absence of ROS scavengers. The difference in rates is equal to half the rate of ROS production.

\[
Rate_{O_2 \rightarrow ROS} = 2 \times (Rate - Rate_{Cat/SOD})
\]

\[
\sigma_{O_2 \rightarrow ROS} = \sqrt{4 \times \sigma^2 - 4 \times \sigma_{Cat/SOD}^2}
\]

\[
Rate_{O_2 \rightarrow H_2O} = 2 \times Rate - Rate_{Cat/SOD}
\]

\[
\sigma_{O_2 \rightarrow H_2O} = \sqrt{4 \times \sigma^2 - \sigma_{Cat/SOD}^2}
\]

A.3. Myoglobin crystallization condition optimization

F33Y-CuMb and FeMb crystallization using PEG as the precipitant was optimized to produce larger, higher quality crystals in several steps. A screen was designed to vary multiple parameters simultaneously in four 24-well plates. This was done by dividing a 96 well box into 4 quadrants and designing a 4 x 6 screen in each quadrant, such that one variable is changes vertically and another horizontally in each quadrant. The designed screen is shown below (Figure A.2). Each parameter was chosen to vary by ± 10% of its original value. The default conditions are 30% PEG 10K, 100 mM sodium cacodylate pH 6.5, 200 mM sodium acetate. This custom screen was ordered from Emerald Bio (now Rigaku).

These conditions were then used to screen crystallization of F33Y-CuMb and FeMb. An example is given below (Figure A.3). For both of these proteins, pH and [PEG] were found to be the most sensitive parameters. No trend was seen with salt concentration. Conditions around the best condition found from these screens were then further subject to fine optimization. In order to be able to vary pH efficiently, the pH of each buffer was not measured directly. Instead, 1M stock solutions were prepared that were assumed to consist entirely of acid (HA) or conjugate base (A⁻).
These were then mixed in appropriate ratios based on the Henderson-Hasselbalch equations. An example is given below in Table A.1 for MES buffer in the range of 5.8 to 6.8. The formula used is given to the right of the table. This method is based on the method used by Emerald Bio.

<table>
<thead>
<tr>
<th>→[NaOAc]</th>
<th>220.0 mM</th>
<th>206.7 mM</th>
<th>193.3 mM</th>
<th>180.0 mM</th>
<th>→[protein]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[PEG] ↓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[PEG] ↓</td>
</tr>
<tr>
<td>27.0 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27.0 %</td>
</tr>
<tr>
<td>28.2 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28.2 %</td>
</tr>
<tr>
<td>29.4 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.4 %</td>
</tr>
<tr>
<td>30.6 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30.6 %</td>
</tr>
<tr>
<td>31.8 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31.8 %</td>
</tr>
<tr>
<td>33.0 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33.0 %</td>
</tr>
<tr>
<td>27.0 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.85</td>
</tr>
<tr>
<td>28.2 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.11</td>
</tr>
<tr>
<td>29.4 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.37</td>
</tr>
<tr>
<td>30.6 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.63</td>
</tr>
<tr>
<td>31.8 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.89</td>
</tr>
<tr>
<td>33.0 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.15</td>
</tr>
</tbody>
</table>

Figure A.2. Crystal optimization screen conditions.

Figure A.3. Results of Fe₃Mb optimization screen.
Table A.1. Sample Excel spreadsheet for calculating mixing ratios for pH screening.

<table>
<thead>
<tr>
<th>pKa</th>
<th>pH</th>
<th>V[total]</th>
<th>V[HA]</th>
<th>V[A-]</th>
<th>HA/A-</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>5.8</td>
<td>55.7</td>
<td>44.3</td>
<td>33.4</td>
<td>2.953</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>55.7</td>
<td>44.3</td>
<td>33.4</td>
<td>1.258</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>44.3</td>
<td>55.7</td>
<td>66.6</td>
<td>0.794</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>33.4</td>
<td>66.6</td>
<td>76.0</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>24.0</td>
<td>76.0</td>
<td>83.4</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>6.8</td>
<td>16.6</td>
<td>83.4</td>
<td>93.4</td>
<td>0.199</td>
</tr>
</tbody>
</table>

A further optimization was to use MES buffer instead of cacodylate in the pH 6 range, for two reasons: (1) cacodylate is an arsenate toxin; (2) cacodylate causes high absorption upon X-ray irradiation, complicating correlated crystallography and spectroscopy. Similar optimization was carried out for I107E-CuBMb and G65Y-CuBMb. The final conditions for each of the proteins are given below (Table A.2).

Table A.2. Final crystal buffer conditions. Protein mixed 1:1 with buffer.

<table>
<thead>
<tr>
<th>Protein</th>
<th>PEG 10K</th>
<th>buffer</th>
<th>salt</th>
<th>[Protein]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe3 Mb</td>
<td>27%</td>
<td>100 mM NaMES pH 6.8</td>
<td>200 mM NaOAc</td>
<td>2 mM</td>
</tr>
<tr>
<td>F33Y-CuB Mb</td>
<td>30%</td>
<td>100 mM NaMES pH 6.7</td>
<td>200 mM NaOAc</td>
<td>1 mM</td>
</tr>
<tr>
<td>I107E-CuB Mb</td>
<td>30%</td>
<td>100 mM tris sulfate pH 8.6-8.7</td>
<td>200 mM NaOAc</td>
<td>1.5 mM</td>
</tr>
<tr>
<td>G65Y-CuB Mb</td>
<td>30%</td>
<td>100 mM tris sulfate pH 8.4</td>
<td>200 mM NaOAc</td>
<td>1.5 mM</td>
</tr>
</tbody>
</table>

Finally, the crystallization was optimized by growing in large (100-200 μM) sitting drop conditions to grow crystals 200-500 μm in with and several mm in length.

A.4. References


APPENDIX B

APPARATUS, METHODS, AND AUTOMATION FOR CRYSTALLOGRAPHY

B.1. Inexpensive microscope display for facilitating anaerobic crystal manipulation

Handling crystals in the glove bag is facilitated by use of a microscope. However, microscope use is hampered by the vinyl or plastic of the glove bag. A camera would alleviate these problems, but cameras for microscopes can be prohibitively expensive for some labs. Luckily, integrated consumer devices with cameras and displays are ubiquitous. I describe how a camera phone was attached to the microscope eyepiece in the glove bag, to act as a display.

First, a rack consisting of metal bars and right angle connectors was fastened to the neck of the microscope. This was used as a bar on to which a clamp could be attached. This clamp is used to hold the camera in place near the eyepiece. The proper distance between the eyepiece and the camera optics was maintained by a rubber gasket, made from a septum by coring it. The camera was held in the correct position using a rubber band.

![Image of setup with camera attached to microscope eyepiece](image)

**Figure B.1.** Photos of setup to use camera phone in place for use as a microscope display.

B.2. Device and method for efficient and careful simultaneous soaking of many crystals

Multicrystal data collection requires efficient preparation of many identical crystals. In case they need to be prepared before by multiple soaking steps, it is more efficient to soak them all simultaneously than individually. Here I describe a jury-rigged device and setup used to carefully
soak many identical crystals simultaneously (Figure B.2). This device has been successfully used numerous times for efficient anaerobic crystal soaking.

![Figure B.2](image1.png) Photos of device and setup used for simultaneous multicrystal, multistep soaking.

The filter device was repurposed from a used, disposable PD-10 Minitrap G25 column (GE Life Sciences). The beads were disposed of and the tube was cut to about 0.5 cm above the bottom frit. A large septum was used as a support; the top of the septum was removed to make a hole straight through. A plastic tube of the appropriate diameter to fit snugly in the outlet of the filter was inserted into the filter. A notch was cut so that the tube could pass under the device while it remains flat against the working surface. The tube was glued in place using super glue. The other end was attached to a syringe of appropriate volume by fitting the tube over a needle of appropriate diameter. This syringe could then be gently pumped with a syringe pump to withdraw solution.
through the filter. Fresh solution is added through the top of the filter. A recommended withdrawal rate is 25 μL / min.

**B.3. MX+Spec collection at BioCARS**

The online microspectrophotometer setup at BioCARS has been set up to collect spectroscopic (Spec) data interleaved with macromolecular X-ray diffraction (MX) data collection, to monitor crystal oxidation state. This is run through a script given below (Code B.2). The spectrophotometer is set to software trigger and started before data collection is initiate. The position at which to collect spectra and X-ray data are stored in a window on the computer and moves to these positions are initiated by the script as needed. The script saves 2 spectra per scan – the second one is more stabilized.

**Code B.1. Data collection steps for collecting data on 14-BM-C with interleaved spectroscopy.**

Procedure

```plaintext
|| Out of hutch ||
- Press 'Mount Sample'
- Close shutter and open door
- Bring in dewar, canes, cryotongs

|| In hutch ||
- ** Cover lens **
- Remove sample
- Place next sample
- ** Uncover lens **

|| Out of hutch ||
- Put cane in shipper
- Turn on light
- Center sample for X-ray
- ** Save X-ray position **
- ** Set zero for microspec here **
- Find best spec angle
- ** Save microspec position **
- Turn of light
- Close lamp shutter - record dark
- Open shutter
- Set up time acqu.
- Set up screening script
- ** Confirm focus (beam size) **

|| In hutch ||
- Put up 32 foils for ~95-99% atten.
- Search station

|| Out of hutch ||
- **! Confirm attenuation and saved positions !**
- Run screening script

|| Index in MosFlm (!Remember to swap phi rotation!) ||
- or integrate in HKL ||

|| Open hutch ||
- Set desired attenuation for data collection
- ** Ensure total range and phi start **
- Run collection script
```
**Code B.2.** Script to collect interleaved single crystal spectra during X-ray data collection.

```python
#!/usr/local/bin/python

# Script to take spectra at a specific position.
# -Drive to microspec position.
# -Trigger microspec
# -Drive back to previous position
from epics import caget,caput,PV
import Numeric
import time,os,sys,urllib

# DATA COLLECTION
#
# Data collection will start at starting_phi value. If phi_range_start != starting_phi, will
# continue collecting from phi_range_start to starting_phi
# Ex.
# phi_range_start = 95
# starting_phi = 120
# phi_range_end = 185
# osc_width = 0.5
# base_filename = Crystal_1
#
# Will result in files:
# Crystal_1_A_001.img [phi 120.0->120.5]
# Crystal_1_A_002.img [phi 120.5->121.0]
# ...
# Crystal_1_A_129.img [phi 184.0->184.5]
# Crystal_1_A_130.img [phi 184.5->185.0]
# Crystal_1_B_131.img [phi 95.0->95.5]
# Crystal_1_B_132.img [phi 95.5->96.0]
# ...
# Crystal_1_B_180.img [phi 119.5->120.0]
#
phi_range_start = -118 # Start of full data collection range
starting_phi = -54 # Phi-start of first frame
phi_range_end = 2 # End of full data collection range
osc_width = 0.5
exposure_time = 3
beamstop_distance = 40
detector_distance = 170

# MICROSPEC ACQ
microspec_bins = [0,10,60,1000] # Frame ranges for microspec frequency determination
microspec_frequency = [2,10,30] # Number of x-ray images between microspec spectrums

# FILE OUTPUT
directory = "/data/lu_1411/YL145-3/H1/" # Have to create manually
base_filename = "YL145-3_H1" # This script will overwrite files!!!!
restartNum = 1 # Flag (0|1) whether to restart numbering
fileNum = 1 # First file number

spec_num = 1

def spectrum():
    caput("14BMC:DAC1_7.VAL",0)
    # Drive to Save Sample position 1
```
Code B.2. (cont.)

caput("14BMC:XYZ_phi_Select", 0)
time.sleep(0.1)
print "Moving to microspec position 1!!!"
while (caget("14BMC:phi.DMOV") == 0 or caget("14BMC:ESP300X.DMOV") == 0 or caget("14BMC:ESP300Y.DMOV") == 0 or caget("14BMC:ESP300Z.DMOV") == 0):
    #print "Moving"
    time.sleep(.5)
time.sleep(.5)

    # Trigger DG535
    print "Acquiring spectrum!!!"
    caput("14BMC:DG535_1genSingleShotTrigBO", 1)
time.sleep(4)
caput("14BMC:DG535_1genSingleShotTrigBO", 1)
##     for i in range(10):
##         caput("14BMC:DG535_1genSingleShotTrigBO", 1)
##         time.sleep(0.1)
time.sleep(5)
f.write("Collected Spec. 
")
f.flush()
caput("14BMC:DAC1_7.VAL",2)

    # Move back to previous position
    caput("14BMC:XYZ_phi_Select", 1)
    print "Moving to x-ray position 2!!!"
    time.sleep(0.1)
while (caget("14BMC:phi.DMOV") == 0 or caget("14BMC:ESP300X.DMOV") == 0 or caget("14BMC:ESP300Y.DMOV") == 0 or caget("14BMC:ESP300Z.DMOV") == 0):
    #print "Moving"
    time.sleep(.5)

def xray_image(phi):
caput("14BMC:phi.VAL", phi, wait=True)
caput("14ADSC1:cam1:Acquire", 1)
caput("14BMC:xiaStart.PROC", 1)
time.sleep(0.1)
IC1=caget("14BMC:sclS1_cts1.D")
IC2=caget("14BMC:sclS1_cts2.A")
RC=caget("5:SRcurrentAI.VAL")
print "acq1"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
    IC2=caget("14BMC:sclS1_cts2.A")
time.sleep(0.1)
time.sleep(0.5)
print "acq2"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
    IC2=caget("14BMC:sclS1_cts2.A")
time.sleep(0.1)
time.sleep(0.5)
print "acq3"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
time.sleep(0.1)
time.sleep(1)
f.write(str(phi)+" "+str(IC1)+" "+str(IC2)+" "+str(RC)+" 
")
val = pvFullFileName_RBV.get(as_string=True)
n=0
while (val == "" and n < 30):
    time.sleep (1)
    val = pvFullFileName_RBV.get(as_string=True)
    n=n+1
    print "File written (" + str(n) + "): " + val
    f.write("File written: " + val + 

def xray_dark():
    #caput("14ADSC1:cam1:AutoSave", 0, wait=True)
    res = caput("14ADSC1:cam1:FileName", "dark_file", wait=True)
    if (res != 1):
        sys.exit()
    time.sleep(0.5)
    res = caput("14ADSC1:cam1:AutoIncrement", 0, wait=True)
    print "AutoIncrement: " + str(res)
    if (res != 1):
        sys.exit()
    time.sleep(0.5)
    res = caput("14ADSC1:cam1:Acquire", 1)
    print "Acquire: " + str(res)
    if (res != 1):
        sys.exit()
    time.sleep(0.1)
    print "acq1"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
    time.sleep(0.5)
    time.sleep(0.5)
    print "acq2"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
    time.sleep(0.5)
    time.sleep(0.5)
    print "acq3"
while (caget("14ADSC1:cam1:DetectorState_RBV") != 0):
    print ('.'),
    time.sleep(0.5)
    time.sleep(0.5)
    #caput("14ADSC1:cam1:AutoSave", 1, wait=True)
    time.sleep(0.5)
    res = caput("14ADSC1:cam1:AutoIncrement", 1, wait=True)
    print "AutoIncrement: " + str(res)
    if (res != 1):
        sys.exit()
    time.sleep(0.5)
    val = pvFullFileName_RBV.get(as_string=True)
    n=0
    while (val == "" and n < 30):
        time.sleep (1)
        val = pvFullFileName_RBV.get(as_string=True)
        print "Wrote Dark (" + str(n) + "): " + val
        f.write("Wrote Dark: " + val + "\n")
        res = caput("14ADSC1:cam1:FileName", base_filename, wait=True)
        if (res != 1):
            sys.exit()
        f.flush
print "Please confirm"
print "Current IC1: " + str(caget("14BMC:sclS1_cts1.D"))
print "Vertical JJ: " + str(caget("14BMC:jjvl.RBV")) + "\nHorizontal focal length " + str(caget("14BMA:SDD.RBV"))
print str(exposure_time) + " s per " + str(osc_width) + " deg oscillation"
answ = raw_input("=== (Y)es?: ")
if answ != "Y":
    sys.exit()
print "Detector distance: " + str(detector_distance)
print "Wedges: " + str(starting_phi) + " -> " + str(phi_range_end) + ", " + str(phi_range_start) + " -> " + str(starting_phi)
if ((starting_phi < phi_range_start) or (starting_phi > phi_range_end)): print "ERROR: starting phi must be within phi range!"; sys.exit()
print "Files will be saved as " + directory+base_filename+"_[A,B]_###.img"
answ = raw_input("=== (Y)es?: ")
if answ != "Y":
    sys.exit()
if restartNum == 1:
    print "Numbering will be restarted at " + str(fileNum)
else:
    print "Numbering will not be restarted!!!
answ = raw_input("=== Continue? (Y): ")
if answ != "Y":
    sys.exit()
print "Spectrum collection will be triggered after frame (spectrum file number):"
bin = 0
for i in range(int((phi_range_end-phi_range_start)/osc_width)):
    if (i >= microspec_bins[bin+1]):
        bin=bin+1
        if ((i % microspec_frequency[bin]) == 0):
            if (i == 0): print "0 (before exposure) (" + str(temp_spec_num) + ")"
            else: print str(i) + "(" + str(temp_spec_num) + ")"
            temp_spec_num = temp_spec_num + 1
answ = raw_input("=== (Y)es?: ")
if answ != "Y":
    sys.exit()
answ = raw_input("=== Did you start spectrum acquisition? (Y) ")
if answ != "Y":
    sys.exit()
caput("14BMC:XYZ_phi1.DO2", 11111)
print "X-ray: " + str(caget("14BMC:XYZ_phi1.DO2")) + " " + str(caget("14BMC:XYZ_phi1.DO3")) + " " + str(caget("14BMC:XYZ_phi1.DO4")) + " " + str(caget("14BMC:XYZ_phi1.DO5"))
print "Spec: " + str(caget("14BMC:XYZ_phi0.DO2")) + " " + str(caget("14BMC:XYZ_phi0.DO3")) + " " + str(caget("14BMC:XYZ_phi0.DO4")) + " " + str(caget("14BMC:XYZ_phi0.DO5"))
answ = raw_input("=== Are these positions correct? (Y) ")
if answ != "Y":
    sys.exit()

f = open(directory+base_filename+".log",'a')
time.strftime('%l:%M%p %Z on %b %d, %Y')+"\nAngle IC1 IC2 RingCurrent\n"
Code B.2. (cont.)

```python
caput("14ADSC1:cam1:FilePath", directory, wait=True)
#caput("14ADSC1:cam1:FileName", base_filename, wait=True)
#caput("14ADSC1:cam1:FileNumber", 1, wait=True)
caput("14ADSC1:cam1:AcquireTime", exposure_time+0.8, wait=True)
caput("14ADSC1:cam1:ADSCBeamX", 156.9, wait=True)
caput("14ADSC1:cam1:ADSCBeamY", 157.0, wait=True)
caput("14ADSC1:cam1:ADSCDistance", detector_distance, wait=True)
caput("14ADSC1:cam1:ADSCWavelen", 0.9787, wait=True)
caput("14ADSC1:cam1:ADSCImWidth", osc_width, wait=True)
caput("14ADSC1:cam1:ADSCPhi", osc_width, wait=True)
caput("14BMC:xiaExposureTime", exposure_time, wait=True)
caput("14BMC:oscWidth", osc_width, wait=True)
caput("14BMC:bsz", beamstop_distance, wait=True)
caput("14BMC:dz", detector_distance, wait=True)

n=1

pvFullFileName_RBV = PV("14ADSC1:cam1:FullFileName_RBV")
#spectrum()
xray_dark()
if (restartNum == 1):
    res = caput("14ADSC1:cam1:FileNumber", fileNum, wait=True)
    if (res != 1):
        exit
time.sleep(0.5)

initial_range = range(int((phi_range_end-starting_phi)/osc_width))
ms_bin = 0

for i in initial_range:
    if (i >= microspec_bins[ms_bin+1]):
        ms_bin=ms_bin+1
    if (((i % microspec_frequency[ms_bin]) == 0):
        spectrum()
        caput("14ADSC1:cam1:ADSCPhi", (starting_phi+i*osc_width), wait=True)
        print "X-ray image: " + str(caget("14ADSC1:cam1:FileNumber_RBV")) + " (phi " + str(starting_phi+i*osc_width) + " to " + str(starting_phi+(i+1)*osc_width) + ")"
xray_image(starting_phi+i*osc_width)

caput("14ADSC1:cam1:FileName", base_filename + "_B", wait=True)
if (len(initial_range) > 0):
    modific = initial_range[-1]
else:
    modific = -1

for i in range(int((starting_phi-phi_range_start)/osc_width)):
    if (((i + 1 + modific) >= microspec_bins[ms_bin+1]):
        ms_bin=ms_bin+1
    if (((i + 1 + modific) % microspec_frequency[ms_bin]) == 0):
```

136
Code B.2. (cont.)

    directory+"Snap2_"+base_filename+str(caget("14ADSC1:cam1:FileNumber_RBV"))+".jpg")
    spectrum()
    caput("14ADSC1:cam1:ADSCPhi", (phi_range_start+i*osc_width), wait=True)
    print "X-ray image: " + str(caget("14ADSC1:cam1:FileNumber_RBV")) + " (phi " +
    str(phi_range_start+i*osc_width) + " to " + str(phi_range_start+(i+1)*osc_width) + ")"
    xray_image(phi_range_start+i*osc_width)
    spectrum()
    f.close()
    print "Data Collection Finished!!!"

B.4. References

(1) Pearson, A. R.; Pahl, R.; Kovaleva, E. G.; Davidson, V. L.; Wilmot, C. M. Tracking X-ray-
derived redox changes in crystals of a methylamine dehydrogenase/amicyanin complex
using single-crystal UV/Vis microspectrophotometry *Journal of Synchrotron Radiation*
2007, 14, 92.
APPENDIX C
SOFTWARE FOR MULTI-CRYSTAL STRATEGY PREDICTION

C.1. Background

X-ray diffraction data collection requires exposure of crystals to intense X-ray beams for prolonged periods of time, and requiring $10^4$-$10^8$ Gy of X-ray dose.\textsuperscript{1-3} At low doses within this range, redox active cofactors, such as metals and sulfur atoms, undergo redox state changes that may perturb the local structure.\textsuperscript{3} At high doses, the X-rays begin destroying macromolecular bonds, and deterioration of crystal diffraction becomes a concern.\textsuperscript{1-3} Due to this, multicrystal data-collection is sometimes necessary to capture minimally perturbed states.\textsuperscript{4-7} In order to efficiently obtain the most complete data from the least amount of crystals, the relative orientations of the crystals with respect to the x-ray beam must be determined and the new crystal data collection must be optimized to collect the most unique data and least redundant data, and two tools have been made to help guide the data collection strategy, given information about past crystals.

C.2. Current available tools

The commonly used pipelines for strategy prediction for multiple crystals are in the Mosflm\textsuperscript{8} and XDS\textsuperscript{9} pipelines. However, these pipelines are limited in their application. Mosflm can roughly predict what data was collected on previous crystals given their orientations, as determined by Mosflm previously, and predict where to best collect data on the current crystals. However, since it doesn’t take into account what data was actually collected and used from prior crystals, but only a prediction, it may miss holes. XDS in the other hand does read in actual collected data and predict data collection strategies based on the useable data of previous crystals. However, like Mosflm, XDS will only read in previously collected data that has been processed and merged using XDS. XDS itself has limited adoption among users due to being run from the command line. XDS is incorporated into specific beamline processing pipelines (BluIce at GM/CA, autoxds at SSRL, RAPD at NECAT) and its multicrystal strategy has been made accessible through these pipelines in some cases. However, these pipelines are not transferable between beamlines. Therefore, I sought to develop a multicrystal strategy program that would be more intuitive than XDS to save time during data collection, and compatible with multiple beamline and data-processing tools.

C.3. Overview and usage

My code is currently written to read in actually collected data processed from HKL2000,\textsuperscript{10} an intuitive graphical data reduction software, and orientations from HKL2000 and Mosflm. It is straightforward to write a reader for other file types, and these will be added as time permits. The
interface is shown below (Figure C.1). Datasets from multiple crystals can be read-in, and then combined in a variety of combinations in which they might possibly be merged, by transferring to a dataset group in the list on the right. The data in the dataset group that has been collected can be plotted as a function of orientation of the currently selected OM crystal within the X-ray beam.

![Overview of program interface.](image)

**Figure C.1.** Overview of program interface.

![Screenshot of the plot of the above configuration.](image)

**Figure C.2.** Screenshot of the plot of the above configuration.

Once plotted, yellow spots indicate data that has been measured in the other crystals in the dataset group with “Data” selected. Red indicates uncollected data. The lower plot indicates the completeness at each angle of rotation of this crystal: the blue line is the total relative amount of unique reflections, yellow line is the amount that have been collected in other crystals, and pink is the fraction of total unique reflections collected.
A wedge of data can be marked by dragging horizontally to visualize what reflections will be collected in a sweep of those angles (Figure C.3).

**Figure C.3.** Plot with scan from Phi 118-130° marked for collection.

When a dataset with collected data is plotted, the data collected from that dataset show up as black (for the phi range that was actually collected) and blue (for symmetry related reflection).

**Figure C.4.** Plot in orientation of a collected dataset.
Using this tool, one can continue to visually add data as it is collected, remove data that may be bad, and plan the data collection on the next crystal, with their choice of data processing software.

**C.4. Future directions**

This code will continue to be developed in collaboration. Bugs need to be worked out and new file type support needs to be added. Then the program will be further extended to help analyze datasets and determine which ones are compatible with each other and mergable. The ability to view correlated spectroscopic data is planned. An eventual portable suite is planned that can be used for multicrystal dataset management and planning on any beamline. The code is written in Java, and so is portable to any common OS.

**C.5. Code**

Once completed, the code will be made available through SourceForge.

**C.6. References**


(9) Kabsch, W. *XDS Acta Crystallographica Section D* 2010, 66, 125.