Gas Sensing in Bacteria:
Understanding Carbon Monoxide-Mediated Allostery in Heme-Dependent Transcriptional Regulators

By
Matthew R. Dent

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Chemistry)

at the
UNIVERSITY OF WISCONSIN-MADISON
2019

Date of final oral examination: 10 May 2019

The dissertation is approved by the following members of the Final Oral Committee:

Judith N. Burstyn, Professor and Irving Shain Chair of Chemistry
Thomas C. Brunold, Professor of Chemistry
Silvia Cavagnero, Professor of Chemistry
John F. Berry, Lester R. McNall Professor of Chemistry
M. Thomas Record, Professor of Chemistry and Biochemistry

Gas Sensing in Bacteria:

Understanding Carbon Monoxide-Mediated Allostery

in Heme-Dependent Transcriptional Regulators

by

Matthew R. Dent

Under the supervision of Professor Judith N. Burstyn

At the University of Wisconsin-Madison

Despite the potent toxicity of carbon monoxide (CO), a diverse array of microorganisms utilize this molecule as a source of energy and/or carbon. These microorganisms occupy drastically different niches, from deep-sea hydrothermal vents to plant root nodules in the soil. The highly complex CO-metabolizing machinery employed by such organisms reflect this ecological diversity: Enzymes associated with anaerobic and aerobic CO metabolism are structurally and phylogenetically distinct. The complex nature of the enzymatic machinery required to carry out CO metabolism necessitates tightly-controlled regulation, and this regulation largely occurs at the transcriptional level. The primary focus of this dissertation is characterization of transcriptional regulators that employ heme to regulate gene expression in a CO-dependent manner.

The first chapter briefly introduces anaerobic and aerobic CO metabolism pathways in microbes and then takes an in-depth look at transcriptional regulation of both processes. It comes as no surprise that the diversity associated with CO metabolism has given rise to diversity amongst

CO-dependent transcriptional regulation mechanisms. This chapter examines two such mechanisms in detail, reviewing CO-mediated allosteric activation in two heme-dependent transcriptional regulators, CooA (CO oxidation activator) and RcoM (regulator of CO metabolism). In reviewing what is known about CO-induced transcriptional activation in CooA and RcoM, it is clear that we have only begun to understand how these the versatile transcription factors regulate a variety of CO-dependent physiological processes. This chapter also highlights exciting new discoveries of putative regulators of CO metabolism, many of which possess metallocofactor binding motifs. Future studies of these uncharacterized proteins may unlock entirely new CO-dependent regulatory pathways and CO sensing mechanisms.

The second chapter of this dissertation presents our work aimed at understanding the role of protein dynamics in heme-mediated, CO-dependent allosteric activation in CooA. A growing number of studies demonstrate that the modulation of protein dynamics, particularly ps-ns timescale conformational dynamics, represents a valid thermodynamic strategy employed by allosterically-activated proteins. This chapter describes how we have employed site-directed spin label electron paramagnetic resonance (SDSL-EPR) spectroscopy to probe changes in protein dynamics in CooA as a function of heme ligation and oxidation state.

Chapters three and four detail spectroscopic and biophysical characterization of the RcoM-1 orthologue from *P. xenovorans*. To fully enumerate the redox-mediated ligand switch at heme, we employed magnetic circular dichroism, electronic absorption, and electron paramagnetic resonance spectroscopies and unequivocally identified Met^{104} as the CO-replaceable ligand at heme. RcoM proteins exhibit a unique domain architecture that has not been structurally characterized. Using a variety of biophysical tools, we discovered that *Px*RcoM-1 exists predominantly as a homodimer in solution and developed an experimentally-validated structural

homology model of this dimeric species. Prompted by the discovery that *Px*RcoM-1 is homodimeric, we re-examined the proposed RcoM binding sites in the *coxM* promoter region using bioinformatics. We identified a direct repeat that exhibited strong consensus among 40 different *coxM* promoter regions; however, this highly conserved repeat does not match a high-affinity cognate motif discovered through *in vitro Px*RcoM-1 binding experiments, This interesting results suggests that *coxM* promoters may have evolved to attenuate binding of the RcoM transcription factor.

The fifth chapter takes an in-depth look at hemoproteins bearing an axial, cysteine-derived thiolate ligand (heme-thiolate proteins). Specifically, this chapter examines the role of the hydrogen bonding (H-bonding) environment surrounding the coordinating Cys(thiolate) in tuning the function and reactivity of heme-thiolate proteins. Understanding second coordination sphere effects in heme-thiolate proteins was a critical aspect of this dissertation, as many homologues of the CO-sensing transcription factors studied herein are heme-thiolate proteins. In an effort to develop EPR spectroscopy as a tool to characterize H-bonding in heme-thiolate proteins, we elucidated the electronic origins of the uniquely narrow rhombic EPR signal associated with six-coordinate, Fe(III) heme-thiolates and explored how H-bonding to thiolate influences this signal. Through synthesis and characterization of a series of six-coordinate, aryl-thiolate-ligated Fe(III) porphyrin complexes bearing a tunable intramolecular H-bond, we developed a new system in which to model the thiolate H-bonding network and its role in tuning the reactivity of heme-thiolate proteins.

Acknowledgements

I have an overwhelming number of people to thank for helping me complete my graduate work and for inspiring me, motivating me, and giving me the tools required to be the scientist that I am today. In this section, I will do my best to acknowledge all of those important people in my life who have, directly or indirectly, made my completion of this dissertation possible.

I'll begin with those who are (more or less) contractually obligated to read this document in its entirety (and may therefore like to skip ahead after reading this paragraph), my thesis committee members. Be it through classes, committee meetings, and/or random visits to your offices, you all have provided critical information, unique perspectives, and excellent questions that have enhanced my knowledge of chemistry and have wholly enriched my graduate experience. To Silvia and Tom: I thank you for teaching this inorganic chemist how to think like a biophysicist. To John: I thank you for showing us the ever-evolving wonders of the periodic table and for being the perfect example of a genuinely curious inorganic chemist. To Thomas: Of all for which I must thank you, I am most grateful for your unwavering support throughout my graduate career. When I committed to studying bioinorganic chemistry at UW-Madison, I had no idea that I would essentially get two advisors (and groups) for the price of one. You and the excellent members of your group have been my mentors, mentees, sounding board, and sources of complimentary expertise since day one. I can honestly say that my graduate experience would have been far less enriching without your guidance and support.

The next person that I must thank is Judith Burstyn, my graduate research advisor. Judith, above all else, it is unequivocal that you have provided me with an entirely unique graduate experience. Not many graduate students can boast that their advisor allowed them to take two trips halfway around the world within a month. It is also pretty rare to find an advisor who is willing to

put her faith in a second year graduate student to help write an NSF grant from scratch. Actually, I think that these two examples succinctly capture some of the qualities that make you an excellent mentor and research advisor. You very clearly appreciate the health and wellbeing of your graduate students, and you understand (and try to get us to understand) that graduate school is so much more than what happens in the lab. You have always encouraged me to take advantage of unique opportunities in science, and in life, and I will never forget that. Your unwavering encouragement, enthusiasm, and faith have inspired and challenged me to be the very best scientist I can be. So, I say that my graduate experience has been a unique one because, above all else, you have shown me how to be both a successful scientist and a decent person, and I am incredibly grateful to have you as a mentor.

In addition to Judith and the members of my committee, there are many other mentors whom I must thank for their support and guidance, particularly in the early days of my graduate career. To Stephanie Dillon: Thank you for being such an awesome student host back in February 2014 when the polar vortex extended my stay in Madison by two days. I am so glad that I came here and that you (along with Nuru, Mickie, Chris, Ivan, and Kristine) were my Brunold group buddies. You all trained me on MCD and taught me the basics of doing electronic structure computations, both of which have been critical to my success in grad school. I have done my best to pay this debt forward and lend my support to the younger members of the Brunold Group, and I have so thoroughly enjoyed my experiences mentoring and learning from Rebeca, Josh, Elizabeth, and all of the rest. To the Burstyn group of old: Thank you Hannah, Judy, Reb, and Eli for being such inviting and supportive group members and for welcoming me to the Heme Team with open arms! I must especially thank Hannah and Judy for the many hours you spent training me to design and carry out good spectroscopic experiments and (of course) express lots of

beautiful, red heme protein. I am very thankful to have worked with you two on a number of exciting and fruitful projects.

I had the incredible good fortune to work with a number of excellent collaborators at UW-Madison and around the country throughout my graduate career. You all have added such an incredible level of depth and impact to my projects that simply would not be possible without your varied expertise and unique perspectives. Here at UW-Madison, I must thank Dr. Kai Cai at NMRFAM for collecting SAXS data on RcoM, as well as Prof. Aseem Ansari and members of his group, Dr. Devesh Bhimsaria and Prof. Jose Rodriguez-Martinez for their excellent work in carrying out and analyzing SELEX experiments on RcoM. At James Madison University (my *alma mater*, Go Dukes!), I must thank Prof. Chris Berndsen and Hamilton Young for helping us develop a better RcoM homology model and for keeping JMU on my mind. At Indiana University, Bloomington, I must thank Prof. Megan Thielges and her outstanding graduate student, Sashary Ramos. In addition to probing fast dynamics at heme in CooA using 2D-IR, Sashary has been a sort of academic pen pal, and I have really enjoyed our correspondence regarding research progress and life in graduate school. Finally, I must thank Prof. Nicolai Lehnert and his ever-patient graduate student, Andrew Hunt for their awesome expertise in assembly of porphyrin-aryl-thiolate complexes. For helping us both conceptually and practically (to a *very* large extent), I cannot thank you both enough.

A very special group of people that I must thank are those undergraduates that I have had the honor to mentor throughout my graduate career. Watching you all find success has probably been the most gratifying experience of my graduate career, and I cannot thank you enough for letting me share in that success. To Richard Dulli: You were the first undergraduate student that I mentored, and the CooA variant that you helped create, D134C, has been an integral part of our

work on heme-mediated allostery. I know that you will make an excellent veterinarian one day soon. To Luc de la Villefromoy: You are such an enthusiastic, talented, and genuinely nice human, and I know that you will find success because *not a single one* of the many synthetic failures we experienced together managed to dampen your spirits. It's that tenacity (along with your overt passion for teeth) that leaves me confident you'll have no trouble breezing your way through dental school. Finally, I have to talk about Michael Milbauer. To thank you as an undergraduate mentee wouldn't do our productive *collaboration* and valuable *friendship* justice. I've had the privilege of watching your interests in chemistry grow from your lion-maned days as my student in Chem 311, through our time in the research lab on Hall-and-Oates-Fridays, and up to now as you begin research as an organometallic chemist in the Sanford Lab at the University of Michigan (we'll be watching your career with great interest). Along the way, you've found success in our research lab (Chapter 5 of this thesis would, at best, outline a theoretical hypothesis without your ligands), and we've grown together as friends. That is truly a rare opportunity, and I am so incredibly grateful to have had such an excellent mentee with which I have shared so many similar interests in and out of lab. Here's to the many insightful scientific discussions and games of foosball we have yet to enjoy in the future!

Speaking of rare opportunities, I had the incredible good fortune to see the number of graduate students in our research group *triple* in size over the past two years. I must say that my colleagues in the Burstyn group are some of the brightest, most inquisitive, and hard-working scientists that I've known. Brian, I am so incredibly grateful that you joined our group when you did. With Michael leaving the lab at the end of the summer in 2018, I was destined to be the lone wolf howling at the heme (see funny door card). Thankfully, you joined in the fall of 2017, bringing your incredible energy and a much-needed knowledge of biology to the group. In a rather short

time, you've helped me uncover new directions for our CO-sensing transcriptional regulators, while also doing incredibly cool work on Co-substituted hemoproteins. In our cozy three-person office, you have provided a sounding board that is equal parts support and skepticism. In this capacity, you have been invaluable, especially as I've worked to see the bigger picture in my last few months. Thank you. I can think of no scientist more deserving of the NSF-GRFP, and I can't wait to see how far your lead the Burstyn group in the next few years. Of course, my expectations for the group have been bolstered by our newest addition, Madeleine. As a fellow inorganic chemist, I'm very impressed at how quickly and enthusiastically you've embraced the biological nature of our group's projects. If anyone can coax the ever-enigmatic RcoM to bind DNA *in vitro*, it's you Madeleine. You didn't hesitate to jump right into a very difficult project, and that enthusiasm gives me confidence in your ability to find success in our group and in your career as a scientist. Again, thanks to the both of you for being such awesome scientists and for making our little group office such a fun place to hang out!

You'll not find a more social group of highly intelligent people than the graduate students in the chemistry department at UW-Madison, and I am incredibly thankful to the many colleagues-turned-friends (and friends-turned-colleagues) that have helped me throughout my Ph.D. First, I must acknowledge the members of my cohort with whom I began this journey in August of 2014. To my roommates, Brandon, Jack, and Jesús, I'll always appreciate our time on Ingersoll and Spring streets, where our living room played host to a wide variety of activities both academic (studying for those silly qualifying exams, TBO practice talks, etc.) and extra-curricular (parties, charades, intense darts matches, the Sunday pit, and Jesús's personal favorite, mouse-wrangling). To Sarah, I thank you for all of your friendship and support as we walked the inorganic path together. Not only do I value the incredible friendship we developed over the past five years, but

I am grateful for your support as an outstanding colleague. I'm very happy that you're coming to Pittsburgh, and I look forward to following your success throughout your career! To Eric, as with Sarah, I owe you tremendous thanks for being such a great friend *and* colleague. In the chemistry building, you've helped pull me through Shannon's organometallics class, provided a "non-biological" perspective for my Ph.D. milestones, and supported Michael and I in our efforts to carry out synthetic organic chemistry (no easy feat, especially when I was involved). Outside of chemistry, we've had many a great evening at Church Key, amongst other places, talking about science and life. The last cohort member that I must thank is Maddy. While it's true that Lindy and Matt have brought us closer together, I really value the friendship that we've developed of our own accord. You've been especially helpful in the waning months of my time here, listening to my practice job talk and thesis defense, providing me with meals and baked goods during the hellacious time preceding my defense, and giving me the chance to hang out with Meowmers, which can be very relaxing when she wants it to be. You all are excellent chemists and even better friends, and I'm excited to see just how far you go.

Sports (both watching and playing) helped keep me sane during my time here, and I am thankful to the many friends that I made through various sports-affiliated activites along the way. I'm really going to miss 8 AM beers and brats with the Badger tailgate crew (shout out to the Zoology moped lot!), chanting and dancing at hockey games, and the various road trips we made in the name of Badger football (thanks to Matt, Maddy, Scott, and Angela for being great passengers and putting up with my road rage/see-saying). I must thank Brandon for introducing me to ultimate Frisbee and the members of team Hammersonly.com for making summer Frisbee such an enjoyable experience! To Captain Diane, "Trigger Happy" Tom, Garrett, Gokul, Matt Stolt, Maddy, Lindy, Laura, Aaron, Aurora, and all the rest, thanks for being such awesome

teammates! When snow covered the fields, we turned to dodgeball to stay active. I had a blast playing with the Buckyballs during the winter, so thanks to Tom, Nick, and Jessi for being such high energy, fun-spirited captains, as well as to the Zanni group for contributing so many excellent players to the team. May you continue to show Mad Dog what Buckyballs are made of for years to come! The last "sport" that I came to enjoy in Madison was fishing, and I have fellow JMU Duke, Aidan McKenzie, to thank for that. Aidan, thanks for pushing me to broaden my horizons and try my hand at fishing in the heat of summer and in the dead of winter. Ice fishing was a wholly unique and fun experience (once you got the hut), and I thank you for getting me out on the lakes. You have always been a great friend and fellow fisherman, so I'm also grateful for your support (and Katie's) throughout the years in Madison.

I must also thank my wild, crazy, but always loving family for helping me to become the person that has been able to tackle graduate school. I have not just two, but three awesome parents, Kathleen, Mike, and Maureen, to thank for helping me get where I am today. Whether it was telling me how proud you were of me after a band concert or being there to talk things through when stuff hit the fan, you have always demonstrated how much you care for me. The same rings true for my extended family. I couldn't have dreamed for more supportive grandparents, aunts, uncles, or cousins. I can always count on your support and on having a great time when we all get together. I also have to thank my younger sisters, Courtney, Brooklyn, and Madelyn for putting up with my desire to teach chemistry *all* of the time. I've really enjoyed our at-home science experiments (sorry about the mess, Mom!), and I hope that, like me, you will look back those experiences and smile.

Finally, I must thank my soon-to-be-wife, Lindy. Your positive attitude and generosity of spirit have not only helped me navigate the trials and triumphs of graduate school, but have

challenged me to be the very best person that I can be. You will always be the first person with whom I hope to share an intriguing result, as well as the first person I seek out after a frustrating day. Either way, I know that you'll be waiting for me with an infectious smile and a Lord of the Rings quote, and I cannot thank you enough for your love and support. With all that we've experienced and accomplished together in just five years, I can only imagine what exciting new adventures await us in the years to come!

Table of Contents

List of Tables

List of Schemes

Chapter 5

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

A

AUC-SE	analytical ultracentrifugation-sedimentation equilibrium
1-MeIm	1-methylimidazole
ACS	acetyl-coenzyme A synthase

B

BLAST	basic local alignment search tool
bp	base pair

C

cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
Ch	*Carboxydothermus hydrogenoformans*
CO	carbon monoxide
CoA	coenzyme A
CODH	carbon monoxide dehydrogenase
COG	cluster of orthologous group of proteins
CooA	CO oxidation activator
CRP	cAMP receptor protein
CSI	cognate sequence identifier
Cyt P450	cytochrome P450

D

Da	Dalton
DFT	density functional theory
Dmax	maximum end-to-end distance
DMSO	dimethlysulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol

E

EcDOS	Escherichia coli direct oxygen sensor
ECH	energy-conserving hydrogenase
EFI	Enzyme Function Initiative
EPR	electron paramagnetic resonance
EST	Enzyme Similarity Tool

F

FAD	flavin adenine dinucleotide
FNR	fumarate and nitrite reductase regulator

FPLC	fast protein liquid chromatography

G

GNT	Genome Neighborhood Tool

H

HBD	heme-binding domain
H-bond	hydrogen bond

I

IPTG	isopropyl β-D-thiogalactopyranoside
IR	infrared
ITC	isothermal titration calorimetry

M

MAL-6	(N-(2,2,6,6-tetramethylpiperidin-4-yl-1-oxyl)maleimide)
MCD	magnetic circular dichroism
MO	molecular orbital
MOPS	3-(N-Morpholino)propanesulfonic acid
MTSL	(2,2,5,5-tetramethyl-1-oxyl-3-methyl methanethiosulfonate spin label)
MWCO	molecular weight cutoff

N

NMR	nuclear magnetic resonance
NO	nitric oxide

P

PAS	Per-Arnt-Sim
PDB	Protein Data Bank
Pr	pairwise distribution
Px	Paraburkholderia xenovorans

R

RcoM	regulator of CO metabolism
Rg	radius of gyration
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RNA	ribonucleic acid
RNAP	RNA polymerase
Rr	Rhodospirillum rubrum

S

SAXS	small angle X-ray scattering

SDSL	site-directed spin label
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SELEX	systematic enhancement of ligands by exponential enrichment
SSN	sequence similarity network

T

TD-DFT	time-dependent density functional theory
TPP	tetraphenylporphyrin

U

UNO	unrestricted natural orbital

V

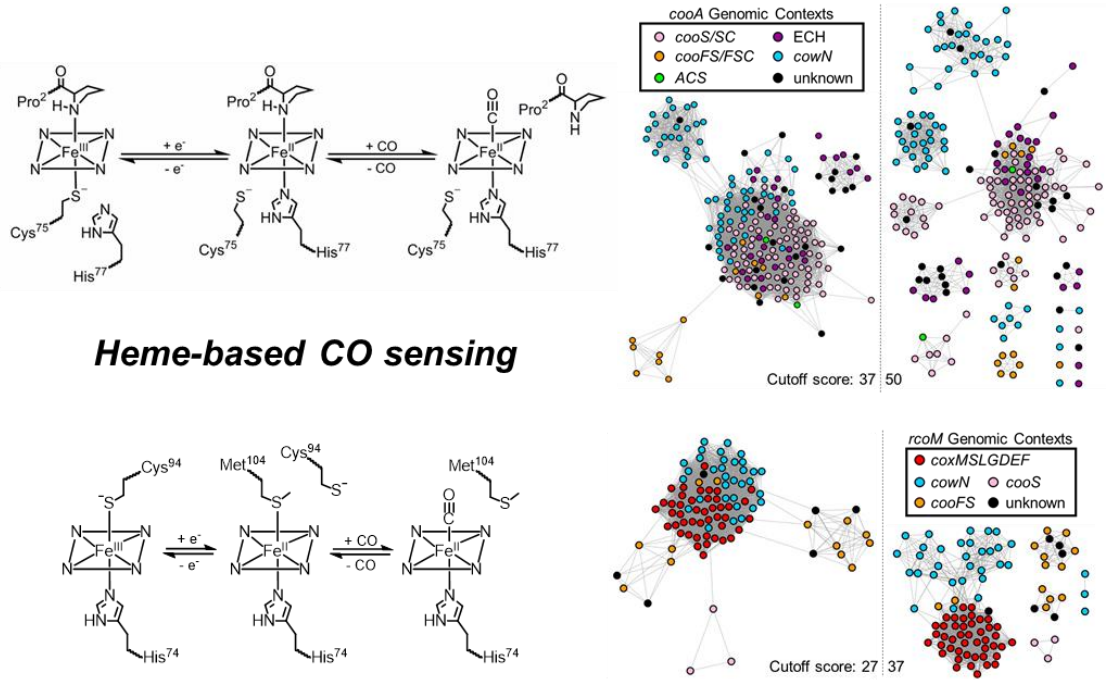
Vc	correlation volume
VTVH	Variable temperature variable field

W

WT	wild type

Chapter One

Transcriptional Regulation of Microbial Carbon Monoxide Metabolism



A version of this chapter is to be submitted as a mini-review:

Dent, M.R.; Weaver, B.R.; Burstyn, J.N. Transcriptional Regulation of Microbial Carbon Monoxide Metabolism.

MRD and BRW wrote the manuscript of this draft, and BRW carried out bioinformatics analyses of CooA and RcoM.

1.1 Introduction

Microbes that utilize carbon monoxide (CO) as an energy and/or carbon source play important roles in a variety of biological contexts and occupy diverse niches. In soils, CO-oxidizing bacteria remove an estimated 145 to 163 Tg of CO from the atmosphere each year, helping to maintain atmospheric CO concentrations below toxic levels.[1-4] Microbes that metabolize dissolved CO, largely derived from photodegraded organic matter, are also found in freshwater and marine environments under both aerobic and anaerobic conditions.[5,6] Important, albeit poorly understood, symbiotic relationships are believed to exist between plants and CO metabolizers in the rhizosphere,[7,8] and CO-metabolizing bacteria likely provide nutrients to a gutless marine worm, *O. algervensis*, in nutrient-poor coastal sediments.[9,10]

To metabolize CO, bacteria utilize CO dehydrogenase (CODH) enzymes, which catalyze the oxidation of CO to CO_2 via Equation 1.

$$CO + H_2O \leftrightarrows CO_2 + 2\ H^+ + 2\ e^- \qquad (\mathbf{1})$$

CODH enzymes belong to one of two classes: O_2-sensitive Ni-Fe CODHs or O_2-tolerant Cu-Mo CODHs. Enzymes in both of these classes utilize highly specialized cofactors that require a myriad of accessory proteins and are costly to the cell. As a result, regulation of the expression of genes that encode CO metabolism is critical.

This review article seeks to summarize recent advances in our understanding of the regulation of microbial CO metabolism in the context of heme-containing transcriptional regulators and introduce a number of putative regulators of CO metabolism that may utilize heretofore uncharacterized mechanisms of CO-mediated transcriptional regulation. To provide biological context and motivate the need to regulate CO metabolism, we briefly summarize known

CO metabolism pathways and describe the highly-specialized enzymes involved in aerobic and anaerobic CO oxidation. We then examine allosteric activation in two heme-dependent, CO-sensing transcriptional regulators, CooA (CO oxidation activator) and RcoM (regulator of CO metabolism). CooA is a well-studied bacterial CO sensor and presents the paradigm in CO-dependent transcriptional regulation of anaerobic CO metabolism. Like CooA, RcoM utilizes heme to sense CO, but the RcoM domain architecture is distinct from that of CooA and unique amongst single-component transcription factors. In addition to being associated with anaerobic and aerobic CO metabolism genes, CooA and RcoM have also been identified in the genomic context of CowN, a small accessory protein that mitigates CO-dependent inactivation of Mo-containing nitrogenase enzymes. We also review a growing body of evidence that points to several strategies which may enable these transcription factors to regulate such a diverse set of physiological processes.

In addition to summarizing what is known (and what remains unknown) about heme-mediated, CO-dependent allosteric regulation in CooA and RcoM, we review the small, yet compelling body of evidence pointing to the existence of other putative regulators of CO metabolism. Many of these putative regulators appear to contain metallocofactor binding motifs, suggesting that these uncharacterized transcription factors may directly sense CO. Future studies of these proteins may uncover heretofore unknown strategies for regulating CO metabolism in microorganisms.

1.2 Microbial CO Metabolism

A wide variety of bacteria and archaea found in both aerobic and anaerobic environments utilize CO as a source of carbon and/or energy. Anaerobic CO metabolism represents one of the most ancient metabolic pathways and has been conserved in microorganisms for more than 3.5 billion

years.[11] A large number of living microorganisms that utilize anaerobic CO metabolism have been identified and characterized. Most of these organisms are carboxydotrophic extremophiles found near undersea hydrothermal vents or volcanic hot springs (where CO concentrations can reach up to 100 nM),[12] although a number of photosynthetic, acetogenic, and methanogenic microorganisms also undergo anaerobic CO metabolism.[13] Aerobic CO metabolizers occupy an equally diverse range of niches and are found in soils, sediments, and aquatic environments.[5] These organisms are able to utilize CO at much lower concentrations than those found in anaerobic environments, suggesting that they must utilize high-affinity CO metabolizing machinery that is also O_2-tolerant. Given that drastically different environments are populated by anaerobic and aerobic CO metabolizers, it is perhaps unsurprising that drastic differences exist between the enzymatic machinery that drives anaerobic and aerobic CO metabolism.

Both anaerobic and aerobic microorganisms capture energy from CO oxidation via reducing equivalents generated in CODH-catalyzed reactions.[6] These reducing equivalents are shuttled into respiratory chains and eventually funneled to a terminal oxidant. In anaerobic CO-oxidizing organisms, the electrons derived from CO-oxidation can drive a variety of respiratory processes including sulfate reduction to sulfide (desulfurication), proton reduction to molecular hydrogen (hydrogenogenesis), and carbon dioxide reduction to acetate (acetogenesis) or methane (methanogenesis).[14] In aerobic CO-oxidizing bacteria, the terminal oxidant is most commonly molecular oxygen, although reducing equivalents may also drive dissimilatory nitrate reduction or denitrification.[5]

Anaerobic and aerobic microorganisms also utilize divergent pathways to fix CO. In anaerobic bacteria and archaea, CO is fixed directly to acetyl-coenzyme A (CoA) via the Wood-Ljungdahl pathway. In this pathway, CODH catalyzes the reduction of CO_2 to produce CO, which

is subsequently shuttled to acetyl-coenzyme A synthase (ACS). ACS condenses CO, CoA and a methyl group delivered by a corrinoid iron-sulfur protein to produce acetyl CoA. The pathway regenerates the methyl group on the corrinoid iron-sulfur protein through the six-electron reduction of CO_2 in via tetrahydrofolate-dependent process.[15] Alternatively, in aerobic bacteria and purple sulfur bacteria, the CO_2 produced by the oxidation of CO is fixed via the Calvin-Benson-Bassham cycle.[16]

Anaerobic CO metabolism: Enzymatic machinery and genomic contexts

Anaerobic oxidation of CO to CO_2 is reversibly catalyzed by Ni,Fe-dependent, O_2-sensitive CODH enzymes at a specific metallocluster known as the C-cluster. These unique clusters have a cubane-like [Ni-4Fe-4S] moiety with an additional iron atom linked by a bridging sulfide (Figure 1.1A,B).[17,18] The kinetics of CO oxidation at the C-cluster have been characterized for CODH II from *Carboxydothermus hydrogenoformans*, which exhibits a k_{cat} of 31,000 s^{-1} and a K_M for CO of 18 μM at physiological temperature and pH (70 °C, pH 8.0).[19] These parameters give rise to a catalytic efficiency ($k_{cat}/K_M = 1.7 \bullet 10^9$ $M^{-1}s^{-1}$) that approaches the diffusion limit, demonstrating that anaerobic CODHs are highly efficient enzymes. Importantly, the oxidation of CO to CO_2 is fully reversible, meaning that under CO_2-replete conditions, CODH reactivity favors formation of CO, which can be assimilated into acetyl-CoA via the Wood-Ljungdahl pathway.[15]

Anaerobic CODHs are *structurally* classified as either CooS-type or Cdh-type CODHs.[20] CooS-type CODHs are characterized as homodimeric proteins in which each monomer contains a catalytic C-cluster along with two [4Fe-4S] clusters, the B- and D-clusters (Figure 1.1A).[21] All three clusters are localized to the solvent-accessible dimer interface, and the D-cluster is positioned to exchange electrons with soluble electron-accepting proteins, such as ferrodoxins.[17] Cdh-type

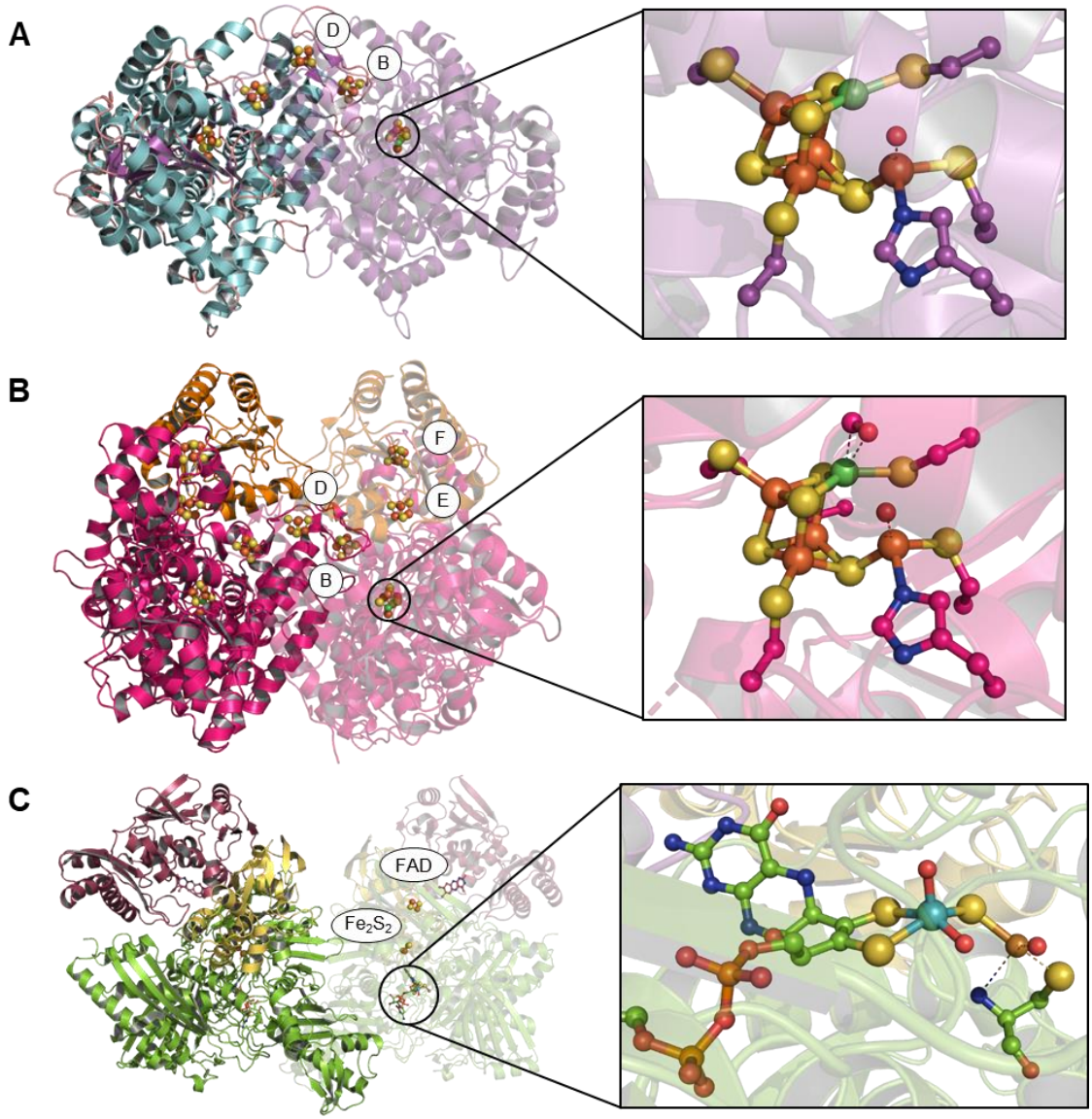


Figure 1.1 Crystal structures of carbon monoxide dehydrogenase (CODH) enzymes. (A) Homodimeric, CooS-type anaerobic CODH from *C. hydrogenoformans*, PDB: 3B51.[21] Atoms in the [4Fe-4S] clusters B and D are shown as spheres, and the catalytic C-cluster is shown to the right with balls and sticks. One monomer is shown in light blue and the other in purple. (B) Cdh-type anaerobic CODH from *M. barkerii*, PDB: 3CF4.[18] Atoms in the [4Fe-4S] clusters B, D, E, and F are shown as spheres, and the catalytic C-cluster is shown to the right with balls and sticks. For the $(\alpha\beta)_2$ tetramer, one heterodimer is highlighted with the small subunit in orange and the large subunit in magenta. (C) Form I aerobic CODH from *O. carboxydovorans*, PDB: 1N5W.[22] Atoms in the [2Fe-2S] cluster are shown as spheres, the FAD molecule is depicted with sticks, and the bimetallic Cu/Mo cofactor is shown to the right with balls and sticks. For the dimer of heterotrimers, one heterodimer is highlighted with the small subunit in yellow, medium subunit in plum, and large subunit in green. Fe atoms are shown in orange, S atoms in yellow, Ni atoms in green, and C atoms in the same color as the protein cartoon. Structures were visualized using PyMOL (Version 1.3 Schrödinger, LLC.).

CODHs differ from CooS-type enzymes in that they exist as $(\alpha\beta)_2$ tetramers with two additional [4Fe-4S] clusters (E & F) that may fulfill the electron transfer role of the D-cluster (Figure 1.1B). These additional clusters allow the D-cluster of Cdh-type enzymes to serve as an electron path to/from FAD.[18] Recent phylogenetic studies suggest that, even amongst members of the same CODH class, there is a large amount of structural diversity associated with the protein scaffold of the active site and Fe-S clusters.[11,20,23] This observation has prompted bioinformatics studies to characterize anaerobic CODHs in terms of genetic context.

Anaerobic Ni,Fe-dependent CODHs are encoded by CO oxidation (*coo*), adenosyl coenzyme A synthase (*acs*), or carbon monoxide dehydrogenase (*cdh*) operons with diverse genomic contexts (Figure 1.2). Each operon encodes a CODH enzyme (*cooS*, *acsA*, or *cdhAB*) and a wide variety of accessory proteins. In *acs* and *cdh* operons, the accessory genes *acsBCDE* and *cdhCDE* encode ACS and other enzymes critical to carbon fixation via the Wood-Ljungdahl pathway.[20] Importantly, expression of genes in *acs* and *cdh* operons is not strongly regulated by CO,[24] and no known CO-dependent transcription factors are associated with these operons. The number of accessory genes adjacent to *coo* operons is highly variable: entire clades of CODHs contain no CODH-associated accessory genes, while others contain up to ten.[23,25] Experimental evidence suggests that the accessory protein *cooC* appears to be involved in maturation of the [Ni-4Fe-4S] cofactor, and putative roles in cofactor maturation have also been proposed for *cooT* and *cooJ*;[26-30] however, accessory protein-mediated cofactor assembly is poorly understood and likely not required for CODHs from *coo* operons with no accessory proteins.[31] Other CODH-linked accessory genes play a role in energy and/or carbon utilization. For example, the *cooMKLXUH* genes upstream of *cooS* in *R. rubrum* encode hydrogenase enzymes needed to carry out hydrogenogenesis.[32] Another accessory protein upstream of *cooS* in *R. rubrum* is *cooF*, a

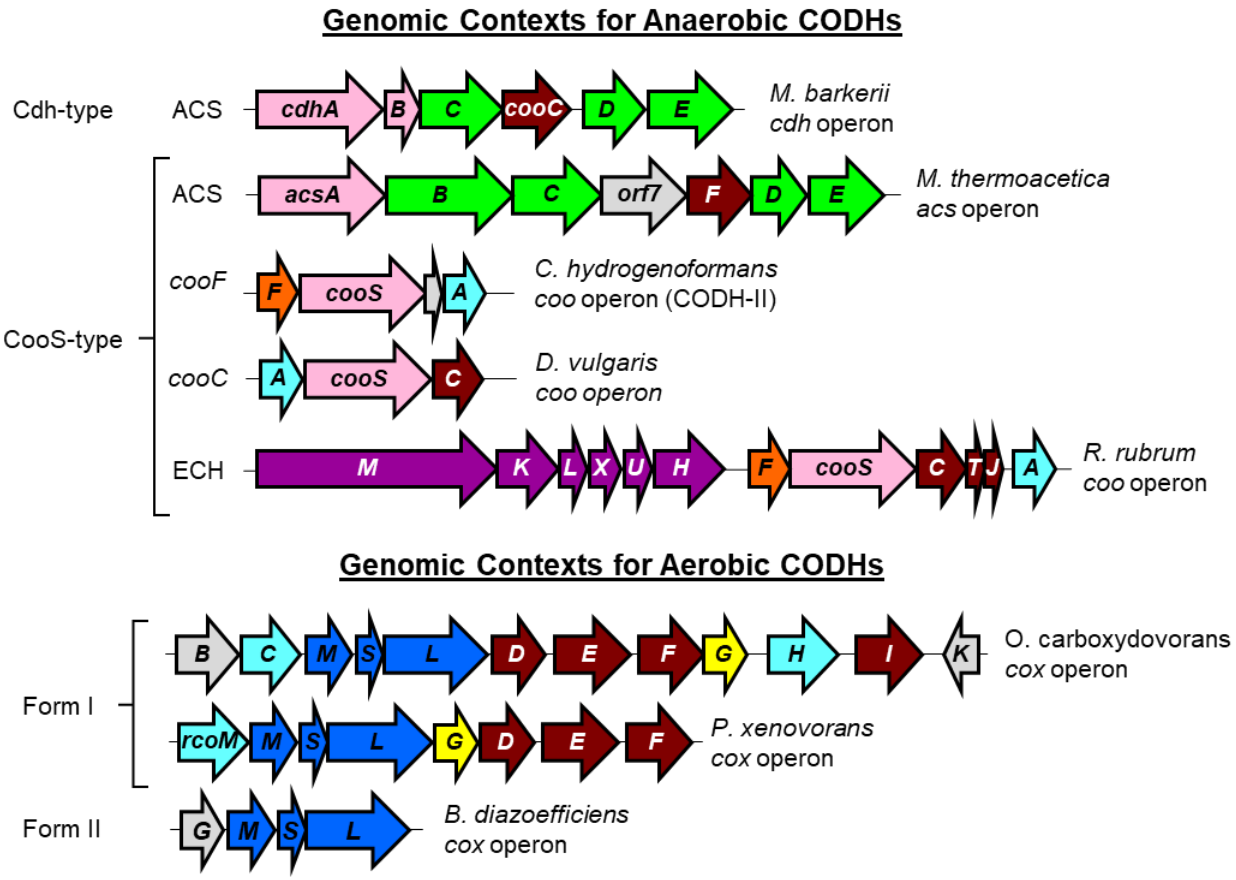


Figure 1.2 Exemplary genomic contexts for anaerobic and aerobic CODH enzymes. Arrows represent genes and open reading frames. Colors denote the function of the gene product: Ni,Fe-dependent CODH enzyme (pink); Wood-Ljundahl pathway-associated enzyme (green); cofactor assembly protein (maroon); electron transfer protein (orange); transcription factor (light blue); membrane anchoring protein (yellow); Cu,Mo-dependent CODH enzyme (dark blue); protein of unknown function (grey).

membrane-associated, Fe-S cluster-containing protein required for optimal CODH activity.[33] The CooF protein likely connects CODH activity to a membrane-associated electron transport chain. Given the diverse environments in which anaerobic CO metabolizers are found, it is unsurprising that such genetic diversity exists in CODH-encoding operons. As a result of this genetic diversity, mapping phylogenetic relationships and enzyme structural and functional properties has been difficult.

Aerobic CO metabolism: Enzymatic machinery and genomic contexts

CO oxidation in aerobic bacteria is catalyzed using O_2-tolerant, Cu,Mo-dependent CODHs, which are members of the molybdenum hydroxylase/xanthine oxidase enzyme family.[34-36] Aerobic CODHs exist as a dimer of heterotrimers, with small (18 kDa), medium (30 kDa), and large (89 kDa) subunits.[37] The small and medium subunits that contain two [2Fe-2S] clusters and a flavin adenine dinucleotide (FAD) cofactor, respectively. The large (89 kDa) subunit harbors a unique bimetallic catalytic site, which consists of a molybdenum-bound molybdopterin cytosine dinucleotide (MCD) cofactor linked to a mononuclear copper ion via a Mo-S-Cu sulfide bridge (Figure 1.1C).[22] Non-reversible CO oxidation occurs at this unique cofactor, and varied results for i*n vitro* steady state kinetics of CuMo-CODH from *Oligotropha carboxidovorans* have been observed. The value of k_{cat} for fully active enzyme is between 90-100 s^{-1}, and K_M values for CO range from 10.7 μM to 0.52 μM.[22,38] Quasi single-turnover kinetics indicate that the initial mechanistic step is rapid CO binding. Taken together, these observations suggest that although aerobic CODHs may be slower enzymes, they likely exhibit a higher affinity for CO than their anaerobic counterparts, giving rise to comparable catalytic efficiencies. This hypothesis cannot be fully validated given the apparent variability in experimental steady state parameters; however, it

is likely that aerobic CODHs bind CO with a high affinity, as the organisms that utilize aerobic CO metabolism are not typically exposed to extremely elevated concentrations of CO.

Aerobic CODH enzymes are classified into two groups based on active site sequence motifs AYXCSFR and AYRGAGR for form I and form II, respectively.[5] The form I active site motif is unique to CODH enzymes, whereas the form II active site resembles that of other molybdenum hydroxylase/xanthine oxidase enzymes. This difference in active site structure gives rise to variable CODH activity, and form I enzymes carry out CODH reactions 100-1000 times faster than form II enzymes *in vitro*.[39,40] In fact, *in vivo* activity of a form II Cu,Mo-CODH has only been demonstrated in a single study.[40] As a result, the physiological role of form II Cu,Mo-CODH is poorly understood, and the structural, mechanistic and physiological distinctions between type I and type II Cu,Mo-CODHs remain unclear.

Aerobic Cu,Mo-CODH genes are found within the carbon monoxide oxidase, or *cox*, operon, and organization of this operon differs between form I and form II CODHs. The operon contains *coxS*, *coxM*, and *coxL*, encoding the small, medium, and large subunits of the CODH, in addition to a variable set of accessory genes (Figure 1.2).[5] From numerous studies of the form I CODH from *O. carboxidovorans*, *coxDEF* and *coxI* accessory genes were determined to be involved in posttranslational assembly of the [$CuSMoO_2$]-MCD cluster, and *coxG* was identified as critical in anchoring the Cu,Mo-CODH complex to the cytoplasmic membrane.[38,41-43] Two distinctions between the organization of the *cox* clusters from form I and form II CODHs exist. First, the subunit organization is *coxMSL* in form I CODHs and is *coxSLM* in form II CODHs. Second, the accessory genes identified in form I CODHs do not necessarily appear in *cox* operons of form II CODHs.[5] At present, it is unclear how these differences in operon organization translate to differences in aerobic CO metabolism.

1.3 Regulation of Anaerobic and Aerobic Microbial CO Metabolism

As demonstrated in the above section, CO metabolism is a complex process that requires highly-specialized enzymes along with a host of accessory proteins. Given the energy input required to generate CO-metabolizing machinery in the cell, microorganisms utilizing this cellular process must tightly regulate expression of genes involved in CO metabolism. In microorganisms, rapid response to environmental stimuli is often accomplished through transcriptional regulation, and CO-metabolizing microorganisms employ CO-sensing transcription factors to regulate expression of CO metabolism. Despite the genetic diversity exhibited by CO-metabolizing machinery in microorganisms, only two CO-sensing transcription factors have been characterized.

In this section, we compare a detailed model of anaerobic transcriptional regulation, enumerated in the well-studied transcriptional activator CooA, to what is known about aerobic transcriptional regulation in the less well-studied transcription factor, RcoM. While CooA and RcoM are likely not the only transcriptional regulators of CO metabolism, they are the only such regulators to be characterized to date. As a result, CO-dependent allosteric activation models in CooA and RcoM represent mechanistic archetypes that lay the foundation for discovery of new regulators of CO metabolism. We caution that the models of regulation in these archetypes are themselves incomplete, as CooA and RcoM have only been studied in one or two of their many genomic contexts (see Section 4). This section also highlights experimental approaches used to study the thermodynamics of CO-dependent allostery and to understand this fundamental biological phenomenon at the atomic level.

CO-mediated allostery in CooA, the archetypical CO-sensing transcription factor

CooA is a member of the cAMP receptor protein/fumarate and nitrate reductase (CRP/FNR) structural superfamily and interacts with its DNA promoter via a helix-turn-helix motif. Each ~200 amino acid-long monomer in the homodimeric CooA protein contains an N-terminal regulatory domain that binds heme and a C-terminal DNA binding domain.[44] CO binding to Fe(II) heme in the regulatory domain allosterically activates CooA to interact with promoter sites within the *coo* regulon and enhance binding of RNA polymerase (RNAP) upstream of genes encoding hydrogenases and CO dehydrogenases.[45,46] Like other CRP/FNR proteins, CooA interacts with DNA via a helix-turn-helix motif, which binds to a set of 5 base pair (bp) inverted repeat sites spaced 6 bp apart.[47-49] Changes in transcriptional activation activity in several gain- and loss-of-function CooA variants led to the identification of three "activating regions" which are proposed to interact directly with RNAP.[45] These activating regions, located in the DNA binding and effector binding domains, and on the flexible 4/5 loop, make specific contacts with RNAP and thereby enhance transcription of downstream genes.

The current allosteric activation model for CooA invokes a structural change upon CO binding to Fe(II) heme. In CooA from *R. Rubrum*, CO cooperatively binds to the Fe(II) heme of each monomer following a reversible redox-mediated ligand switch in which a charged Cys(thiolate) heme ligand is replaced with a neutral His ligand (Figure 1.3, top).[50-55] Unique structural conformations are associated with inactive Fe(II) *"ready-off" Rr* CooA and a constitutively active variant of CooA from *Carboxydothermus hydrogenoformans* (*Ch*) (Figure 1.3, bottom).[44,56] These structural data, in conjunction with small-angle X-ray scattering (SAXS) data for *Rr*CooA, led to the development of a "swinging hinge" model in which the DNA binding domains undergo a propeller-like rotation to expose the DNA-binding F-helices upon CO

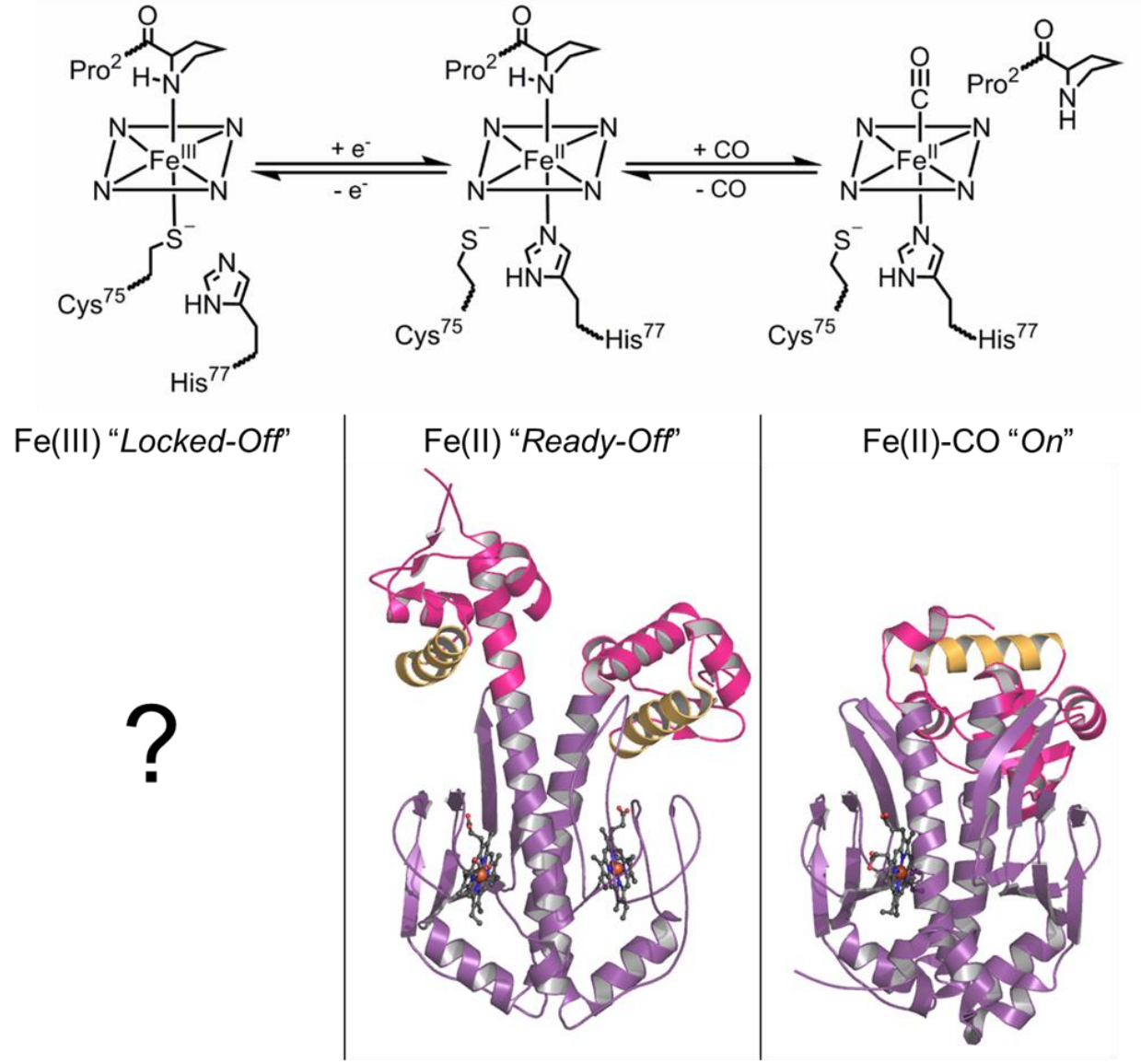


Figure 1.3 The three functional states of CooA and corresponding structural data for the Fe(II) "*ready-off*" state in CooA from *R. rubrum* (PDB 1FT9)[44] and Fe(II)–CO *"on"* state of CooA from *C. hydrogenoformans* (PDB 2HKX).[56] For each structure, the effector binding domain is shown in magenta, DNA binding domain in purple, and heme cofactors as balls and sticks. The DNA-binding F-helices are highlighted in orange. Structures were visualized using PyMOL (Version 1.3 Schrödinger, LLC.). The unique heme ligation/oxidation state associated with each functional state is depicted above the structural data.

binding.[57] This rigid body rotation may be facilitated by CO binding through loss of the Pro^2 heme ligand and subsequent rotation of the C-helices in the dimer interface.[58-62] Citing constitutive activity of a covalently cross-linked variant of *Ch* CooA, Tripathi and Poulos recently suggested that, in *Ch* CooA, a structural transition to the active conformation may be facilitated by stabilizing interactions between the N-terminal tail, heme domain, and DNA-binding domain;[63] however, it has yet to be determined whether the N-terminus of *Ch*CooA, which is several residues shorter than that of *Rr*CooA, coordinates to heme.[64,65]

While a well-enumerated structural model for CO-dependent allosteric activation in CooA exists, CO-induced changes in protein dynamics may represent a heretofore unrecognized factor in the CooA allosteric activation mechanism. New research in a growing number of bacterial transcriptional regulators, including CRP, a structural homologue of CooA, suggests that effector-induced changes in protein dynamics plays a key role in allosteric activation.[66-69] Specifically, effector binding can lead to changes in fast (ns-ps timescale) protein motions, giving rise to changes in conformational entropy. Conformational heterogeneity, which may be modulated in response to effector binding, appears to occur in CooA. CO recombination following flash photolysis exhibited nonexponential behavior when probed by sub-picosecond mid-IR spectroscopy.[70] Additionally, biphasic kinetics for CO association and dissociation were observed using a combination of rapid mixing electronic absorbance, flash photolysis, and time-resolved resonance Raman spectroscopy.[54] These data provide spectroscopic evidence that Fe(II)–CO *"on"* CooA samples multiple conformational states on the fast (ns-ps) timescales. Additionally, recent experiments using site-directed spin label electron paramagnetic resonance spectroscopy demonstrated that Fe(III) *"locked-off" Rr*CooA exhibits conformational heterogeneity, and this

technique is currently being applied to probe changes in protein dynamics as a function of heme redox/ligation state.[71]

The CO binding affinity of CooA is relatively weak; however CO binding to heme is cooperative, and active CO-bound CooA exhibits a very high affinity for its promoter site. Using the Hill equation to fit CO titration data, a P_{50} value of 2.2 μM with n = 1.4 was observed for CooA, indicative of relatively weak CO binding.[54] These data suggest that very high (μM) concentrations of CO are required to activate CooA, but that activation occurs rapidly once the appropriate CO levels are reached. The CooA CO binding affinity appears to be tuned to that of the CODH enzymes found in anaerobic CO metabolizers, which exhibit K_M values of approximately 20 μM.[19,72] Furthermore, Fe(II)–CO *"on"* CooA exhibits nanomolar affinity for its promoter site,[71] implying that upregulation of *coo* expression occurs rapidly upon activation of CooA. Taken together, these observations present a model for physiological response to CO in anaerobic CO metabolizers. When these microorganisms experience a drastic spike in local CO concentration, potentially caused by a sporadic volcanic event or escape of a CO gas bubble from a mat of decaying organic matter, CooA is quickly activated and binds upstream of the organism's *coo* operon. This leads to rapid expression of CODH and other enzymes needed to capitalize on an abundant, but potentially fleeting source of energy and carbon.

RcoM regulates aerobic CO metabolism in a manner distinct from CooA

Genes encoding the heme-dependent transcription factor RcoM were originally identified upstream of *coxMSL* genes in the soil bacterium *Paraburkholderia xenovorans*. RcoM was originally annotated as a transcriptional regulator based on sequence homology, and the protein is predicted to adopt a unique domain architecture comprised of an N-terminal sensory domain that adopts a heme-binding PAS (Per-Arnt-Sim) fold and a C-terminal DNA binding domain that

adopts a LytTR fold (Figure 1.4).[73] RcoM utilizes heme to sense CO in a manner similar to that observed in CooA: reduction of Fe(III) heme results in a redox-mediated ligand switch in which a charged Cys(thiolate) ligand is replaced by a neutral Met ligand (Figure 1.4).[74] In *Px*RcoM-1, one of two RcoM orthologues in *P. xenovorans*, Met[104] is replaced by CO, resulting in activation of the protein to bind to its promoter upstream of *coxM*.[75,76] Unlike CooA, RcoM exhibits a very high affinity for CO (K_d < 100 pM for a heme-binding domain truncate of *Px*RcoM-2).[77]

We are only just beginning to understand the biophysical properties of the structurally-distinct RcoM protein, and no crystallographic data for RcoM exists to date. A combination of size exclusion chromatography and analytical ultracentrifugation sedimentation equilibrium data revealed that *Px*RcoM-1 is primarily homodimeric in solution, that dimerization occurs via the heme-binding PAS domain, and that dimerization does not require heme (*vide infra*). These observation are consistent with the behavior of other prokaryotic PAS domain-containing proteins, which tend to form homo-oligomers in solution.[78] Small angle X-ray scattering (SAXS) data further revealed that the Fe(III) RcoM homodimer is elongated in solution, and a dimeric RcoM homology model (based on individual structures of known PAS and LytTR domains)[79-82] fits within the molecular envelope determined by SAXS. These observations represent an important first step in characterizing the structure of RcoM, which is the only known single-component fusion of a PAS domain and a LytTR domain.

The exact identity of the RcoM promoter site has not been fully enumerated. Unlike CooA, RcoM interacts with DNA via a LytTR domain in which several residues of the protein β-β-β fold interact with DNA nucleobases, resulting in significant DNA bending and (presumably) activation of *cox* transcription.[82] The canonical DNA binding site for LytTR proteins is a pair of direct repeats located upstream of the -35 region of the relevant operon.[83-87] These repeats are typically imperfect,

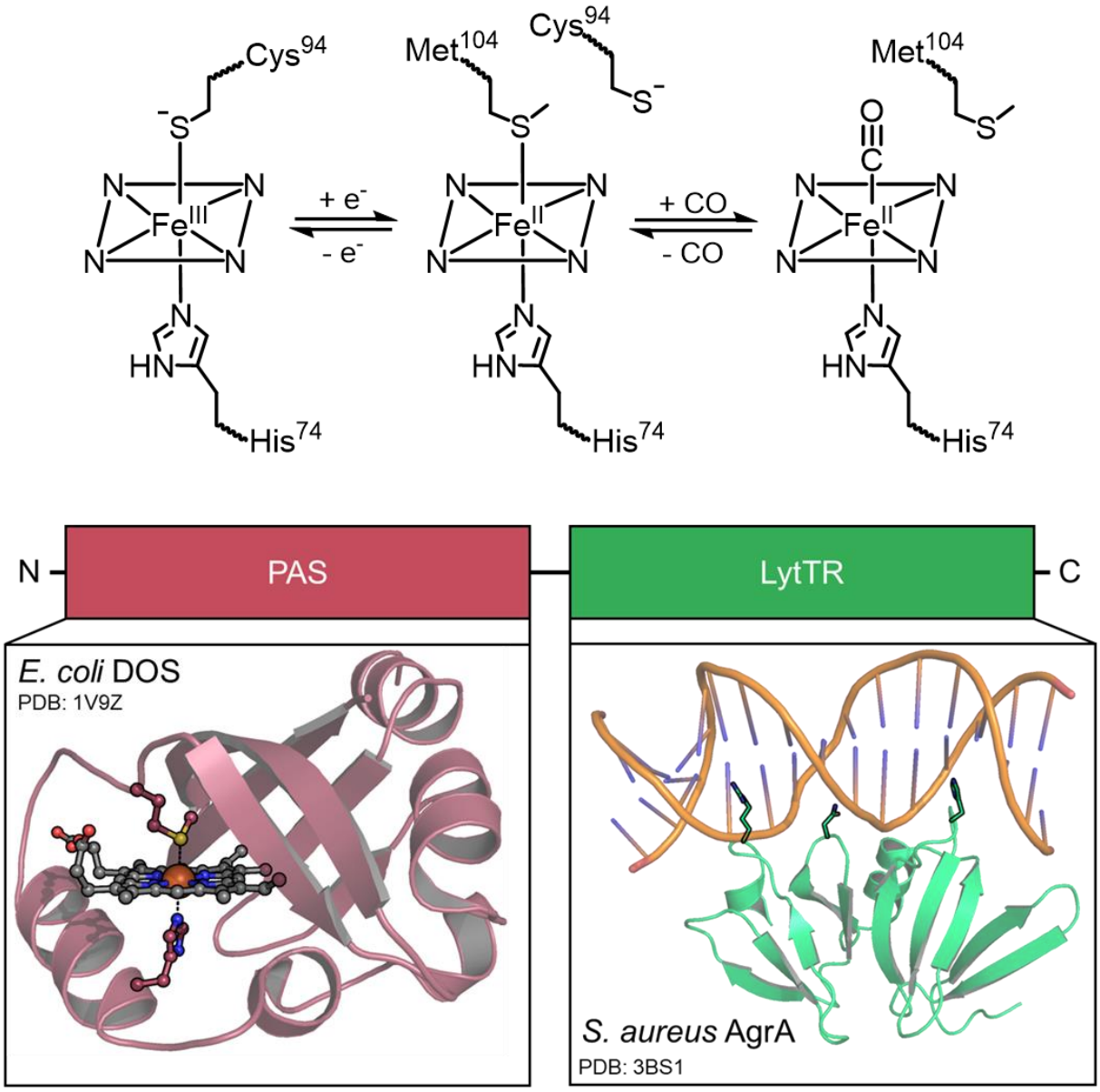


Figure 1.4 Redox-mediated ligand switch and domain architecture of in RcoM-1 from *P. xenovorans*. Crystal structures of individual domains from homologous proteins are displayed. The PAS domain of DOS from *E. coli* is depicted in plum to the left.[81] Heme and protein-derived ligands are shown as balls and sticks. The LytTR domain of AgrA from *S. aureus* bound to DNA is depicted in green to the right.[82] Protein residues that make specific contacts with DNA base pairs are shown as sticks. Structures were visualized using PyMOL (Version 1.3 Schrödinger, LLC.).

9 bp in length, and spaced 10-13 bp apart; importantly, the spacing between repeats is critical in determining transcription factor binding strength.[88,89] The originally-proposed promoter binding site for *Bx*RcoM-1 differs from well-established LytTR DNA-binding motifs. A combination of DNA footprinting, *in vivo* reporter, and *in vitro* fluorescence anisotropy data, led to the identification of two *Bx*RcoM-1 binding regions spaced in the intergenic region between the $rcoM_1$ and $coxM_1$ genes.[75] One of these two binding regions is located directly upstream of the -10/-35 extension preceding the $coxM_1$ gene, while the second binding region is ~100 bp further upstream. Both putative *Bx*RcoM-1 binding regions contain three direct repeats with a 5'-TTnnnG-3' motif that modestly resembles repeat motifs observed in other LytTR-containing transcription factors;[83,84,90,91] however, the spacing between direct repeats is nearly doubled for RcoM (21 bp) compared to other LytTR motifs, and it is difficult to envision a binding model in which dimeric RcoM binds to a triplet binding motif.

A refined RcoM promoter binding site has recently been proposed which more closely resembles the canonical LytTR-binding promoter motif. The new putative RcoM-binding motif was identified through alignment of the upstream promoter regions of 40 different *cox* operons that encode *rcoM* genes (Figure 1.5). A conserved motif was identified just upstream of the -10/-35 region of *coxM* and consists of three 8 bp repeats, each with the sequence TT[CG][GA][TC]G[CT][AG], that are each separated by a 12-13 bp linker. This conserved motif encompasses one of the two originally-proposed triplet repeat sites, but the second upstream triplet repeat was not conserved amongst *cox* promoter regions. The first and second repeats of the triplet motif are much more highly conserved than the third repeat, suggesting RcoM binds a *pair* of direct repeats (Figure 1.5). This pair of direct repeats is highly reminiscent of those found in other LytTR-containing transcription factors, and would be compatible with binding of an RcoM dimer.

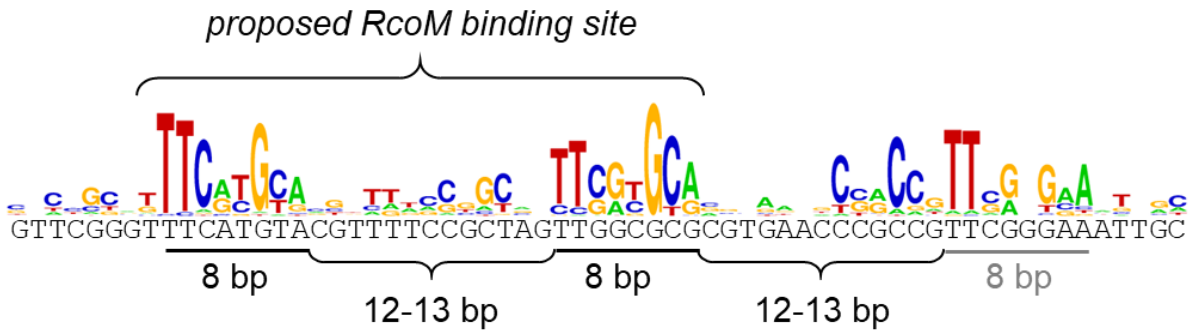


Figure 1.5 Logo plot generated from multiple sequence alignment of putative *cox*-associated RcoM promoter regions. The nucleotide sequence of the *coxM*$_1$ promoter site in *P. xenovorans* is shown below the logo plot. Conserved repeats and a consensus site for *Px*RcoM-1 binding are also shown.

While RcoM exhibits an incredibly high affinity for CO, this transcriptional regulator appears to exhibit a relatively low affinity for its promoter site upstream of *coxM*. Preliminary *in vitro* fluorescence anisotropy data suggest that *Px*RcoM-1 exhibits micromolar binding affinity for individual triplet repeats in the originally-proposed promoter sites as well as for the newly-proposed pair of direct repeats (*vide infra*). These low binding affinities of individual repeats may suggest that, at least in the case of *Px*RcoM-1, cooperative binding of higher order oligomers to both triplet sites is required for maximal promoter binding; however, this model is at odds with sequence alignment data, which shows no conservation of a second upstream triplet motif. Interestingly, the native *cox* promoter region does not elicit maximal DNA binding of RcoM, as evidenced by 1) enhanced *in vivo Bx*RcoM-1 binding for reporter strains with certain base pair substitutions in the promoter region and 2) the identity of an 8 bp motif, which bound to *Px*RcoM-1 with high affinity during *in vitro* systematic enhancement of ligands by exponential enrichment (SELEX) experiments and *did not* match the sequence of any proposed repeat motifs.

Aerobic CO metabolizers may exploit changes in the DNA sequence of the *coxM* promoter region to attenuate CO-dependent gene expression. If RcoM exhibited a high affinity for its promoter site, then RcoM-regulated genes would be constitutively expressed in the presence of picomolar CO, given the RcoM's extremely high CO binding affinity.[77] It is therefore possible that *cox* promoter sites have evolved to attenuate RcoM binding as a means to limit the expression of CO-metabolizing enzymes at low concentrations of CO. By altering the extent of this attenuation, RcoM orthologues would be able to operate in drastically different CO concentration regimes. We hypothesize that this strategy contributes to RcoM's versatility as a regulator of anaerobic CO metabolism, aerobic CO metabolism, and *cowN* expression.

1.4 Diversity in the Genomic Contexts of CooA and RcoM

As the amount of genomic data for CO-oxidizing organisms continues to grow, researchers have developed an appreciation for the diversity of composition and organization of genes encoding CO-metabolizing cellular machinery. Genomic studies of *coo* operons have mostly focused on genomic diversity within CODH-encoding genes, which gives rise to the unique reactivity observed in anaerobic CODH enzymes.[11,14,20,72] Recently, Inoue et al. carried out a broadened bioinformatics analysis of operons associated with anaerobic CO metabolism.[23] The authors identified many new putative anaerobic CODH-associated genes and demonstrated that genomic diversity amongst anaerobic CO metabolizers is even greater than originally anticipated. Genomic diversity is also prevalent in aerobic CO metabolizers; however, the distinct physiological roles associated with different *cox* operon motifs are not well enumerated.[5,38,72]

In this section, we highlight the genomic diversity observed in regulators of CO metabolism by taking an in-depth look at the genomic contexts associated with *cooA* and *rcoM*. Surprisingly, sequence similarity network analysis shows that CooA and RcoM *protein* sequences correlate well with the *genomic context* of the associated *cooA* and *rcoM* genes. *cooA* genes involved in regulation of CO metabolism appear exclusively in the context of *coo* operons; however, *rcoM* genes are associated with both *coo* and *cox* operons. Interestingly, both *cooA* and *rcoM* exhibit an additional genomic context near *cowN*, which encodes a small (~100 amino acid residues) accessory protein. CowN is believed to mitigate CO toxicity towards the Mo-containing nitrogenase encoded by the *nifHDK* operon, as demonstrated in two cases where *cowN* expression sustains *in vivo* nitrogen fixation activity in the presence of CO. [92,93] CooA- and RcoM-induced expression of the *cowN* gene was observed in a CO-dependent manner in one of each of these studies, although it is unknown how CowN acts to ameliorate detrimental reaction between CO

and nitrogenase. The properties of *cowN*-linked CooA and RcoM proteins have also not been explored.

cooA genomic contexts

cooA genes are observed near operons encoding anaerobic CO metabolism machinery, as well as near *cowN* genes. To obtain an up-to-date picture of the genomic contexts associated with *cooA* genes, we developed a protein sequence similarity network (SSN) for CooA proteins using the Enzyme Function Initiative-Enzyme Similarity Tool (see Appendix 1). From the initial 1,000 hits with highest similarity to the query sequence (CooA from *R. rubrum*), we identified 182 *cooA* genes with a discernable genomic context and 24 with unknown genomic contexts using the Enzyme Function Initiative-Genome Neighborhood Tool (Table 1.1). Four genomic contexts associated with anaerobic CO metabolism were identified in which a putative *cooA* gene was found adjacent to: a standalone CODH-encoding *cooS* gene or *cooS* and the putative cofactor assembly gene, *cooC* (*cooS/SC*), *cooS* and the rubredoxin-like *cooF* gene or *cooS*, *cooF*, and *cooC* (*cooFS/FSC*), genes encoding acetyl coenzyme A synthase (ACS), or *cooS* and an energy-conserving hydrogenase (ECH). The fifth *cooA* genomic context was identified as adjacency to *cowN*, described above.

Mapping the genomic context of each *cooA* gene onto its corresponding node in the protein SSN reveals a connection between *cooA* genomic context and CooA protein sequence (Figure 1.6). At a lower alignment cutoff score, the majority of CooA nodes cluster in a central hairball, suggesting that there is a good degree of sequence similarity between most CooA proteins. Three divergent sub-clusters associated with *cowN*, *cooF/FSC*, and ECH genes were also observed at the lower alignment cutoff score, suggesting that some CooA protein sequences diverge in a manner consistent with their associated genomic context. These genome-specific effects are magnified at

Table 1.1 Summary of genomic contexts for CooA- and RcoM-encoding genes.

cooA		***rcoM***	
genomic context	**number of genes**	**genomic context**	**number of genes**
cooS/SC	77	*cooS*	3
cooFS/FSC	19	*cooFS*	13
ACS	2	*coxMSLGDEF*	44
ECH	32	*cowN*	39
cowN	65	unknown	13
unknown	24		
total	**219**		**112**

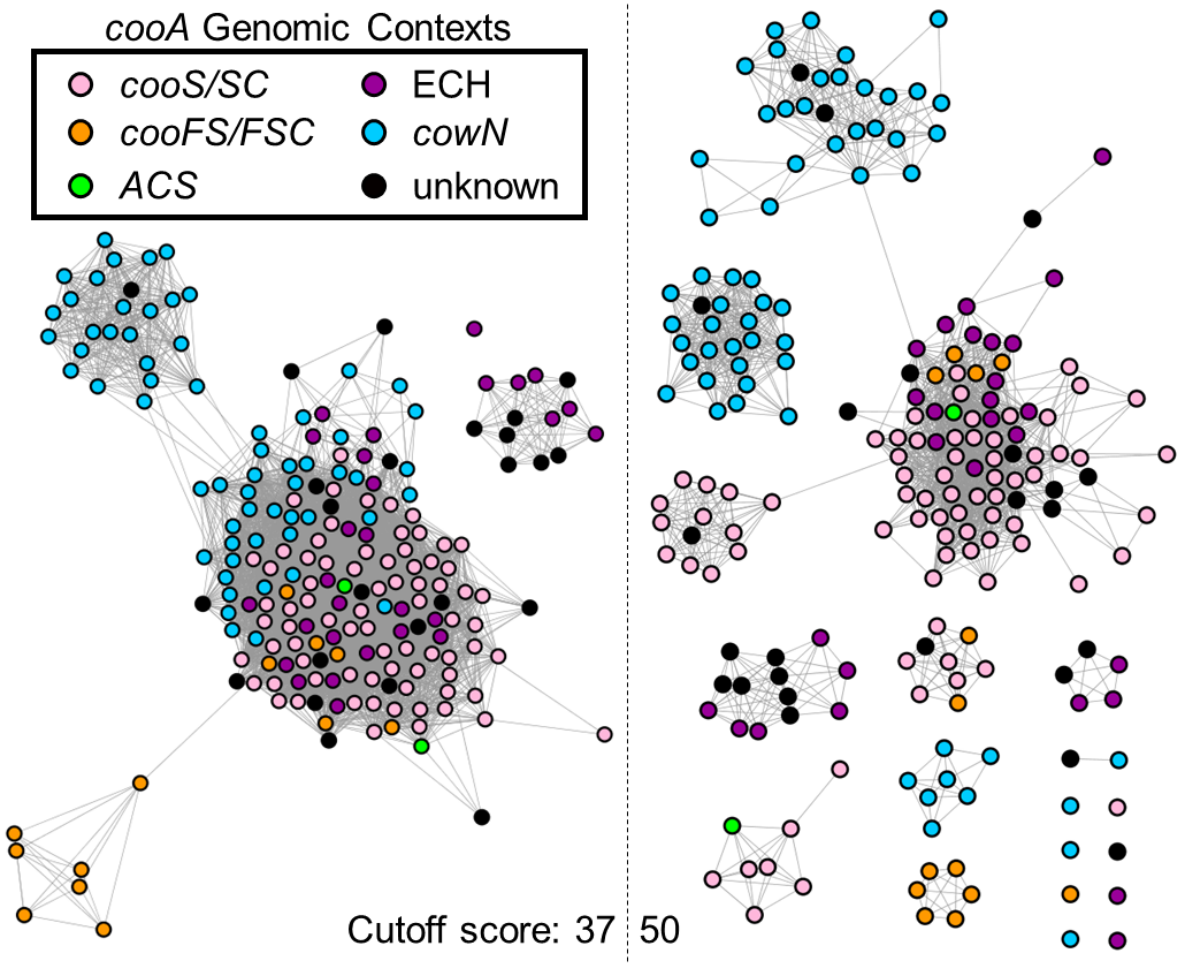


Figure 1.6 Genomic contexts of *cooA* genes mapped onto the CooA protein sequence similarity network. The same 100% identity representative node network is displayed at a lower (37) and higher (50) sequence similarity cutoffs. Node colors correspond to *cooA* genomic contexts, which are labeled according to the identity of genes in close proximity (within 20 genes) to *cooA*, as identified through genome neighborhood analysis. In the energy-conserving hydrogenase (ECH) genomic context, *cooA* is located in a *coo* operon containing both CODH and a hydrogenase enzymes.

an increased alignment cutoff score, where further sub-clustering is observed. In particular, two *cowN*-associated sub-clusters diverge from those associated with anaerobic CO metabolism.

The genome-specific divergence in CooA protein sequences likely results from the fact that this transcription factor regulates two different physiological processes: CO metabolism and protection of nitrogenase enzymes from inactivation by CO. As suggested above, the micromolar CO binding affinity of *coo*-associated *Rr*CooA appears to be tuned to approach that of its downstream CODH enzyme; however, one may expect that CO inactivates Mo-containing nitrogenase enzymes at much lower CO concentrations. As a result, CooA-controlled expression of *cowN* genes would need to occur at correspondingly low concentrations of CO. Thus, we predict that differences in sequence between *coo*- and *cowN*-associated CooA proteins lead to enhanced CO binding affinity for those associated with *cowN*. A comparative analysis of CooA sequences from different genomic contexts, as well as comparative biochemical characterization, will be needed to lend support to this speculative hypothesis.

rcoM genomic contexts

Although RcoM was originally characterized in the context of *cox* operons encoding aerobic CO metabolism genes, *rcoM* genes are also associated with *coo* and *cowN* operons. The genomic contexts of *rcoM* were recently explored using the same methodology applied to *cooA* above (*vide infra*). Genomic analysis revealed 99 *rcoM* genes with a discernable genomic context and 13 with unknown genomic contexts (Table 1.1). Two genomic contexts associated with anaerobic CO metabolism were identified in which a putative *rcoM* gene was found adjacent to: a standalone CODH-encoding *cooS* (*cooS*) or *cooS* and *cooF* (*cooFS*). *rcoM* genes were also identified adjacent to the aerobic CO metabolism operon, *coxMSLGDEF*. Finally, *rcoM* genes were observed in close proximity to *cowN* genes.

As was the case for CooA, proteins in the RcoM SSN cluster according to genomic context (Figure 1.7). At lower alignment cutoff scores, all RcoM proteins fall into the same cluster; however the amino acid sequences of RcoM proteins encoded in the context of *coo* operons diverge from those associated with *cox* and *cowN* operons. This observation becomes clearer at the higher alignment cutoff score, in which most *coo*-associated RcoM proteins fracture into two sub-clusters that are distinct from the main *cox*/*cowN* cluster. Additionally at the higher alignment score cutoff, divergence between *cox* and *cowN*-associated RcoM proteins becomes more apparent, as does divergence within the *cowN* sub-cluster. As was speculated for CooA, genome-specific differences between the protein sequences of *coo*, *cox*, and *cowN*-associated RcoM may shed light on how the same transcription factor may be tuned to differentially regulate specific physiological processes.

A remarkable feature of RcoM is that the same transcription factor regulates three very different physiological processes. Given the relatively high degree of sequence similarity between RcoM proteins involved in regulating *coo*, *cox*, and *cowN* genes, it is likely that all RcoM proteins share a common ancestor. The mechanisms of *rcoM* gene transfer between organisms and amongst physiological pathways are very poorly understood. Interestingly, two *coo*-associated RcoM proteins remain within the main *cox*/*cowN* cluster at the higher alignment cutoff score. It is possible that these proteins provide an ancestral link between RcoM proteins from different genetic contexts, and further investigation of the sequences and biochemical properties of these protein is warranted.

1.5 Putative Novel Regulators of Microbial CO Metabolism

Despite the fact that nearly 200 *cooA* and *rcoM* genes have been identified in association with operons encoding CO metabolism machinery, CooA and RcoM cannot possibly be the only CO-responsive transcription factors involved in regulating the thousands of putative operons associated

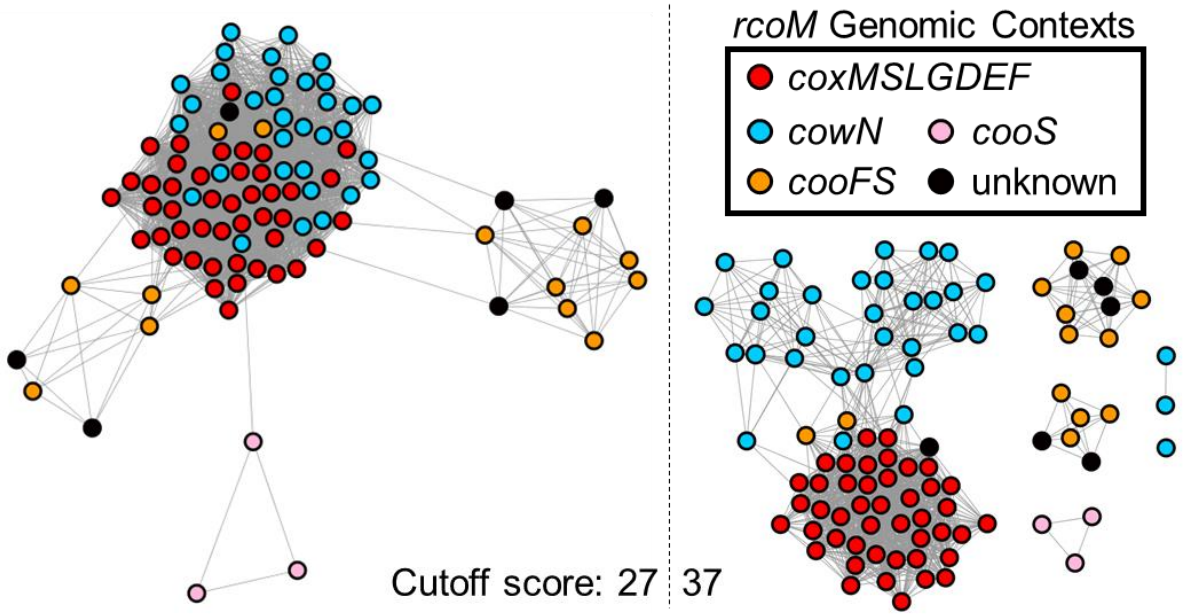


Figure 1.7 Genomic contexts of *rcoM* genes mapped onto the RcoM protein sequence similarity network. The same 95% identity representative node network is displayed at a lower (27) and higher (37) sequence similarity cutoffs. Node colors correspond to *rcoM* genomic contexts, which are labeled according to the identity of genes in close proximity (within 20 genes) to *rcoM*, as identified through genome neighborhood analysis.

with CO metabolism. Other regulatory pathways that modulate CO metabolism must therefore exist. This section briefly introduces a few putative regulators of CO metabolism that have not been characterized in detail.

Putative novel regulators of anaerobic metabolism

The recent genomic analysis of anaerobic CODHs by Inoue et al. led to the identification of several clusters of orthologous groups of proteins (COGs) associated with anaerobic CO metabolism that were annotated as transcriptional regulators.[23] In addition to CooA, several other COGs comprised of single-component transcriptional regulators were identified, including IscR/NsrR-, MarR-, AraC-, and TetR/AcrR-like COGs. IscR and NsrR are members of the Rrf2 structural superfamily of transcriptional repressors. IscR is a transcriptional repressor that downregulates Fe-S cluster assembly pathways when a [2Fe-2S] cluster binds to the transcription factor.[94,95] NsrR is a transcriptional repressor that mediates nitric oxide (NO) stress response, utilizing a [4Fe-4S] to sense nitric oxide.[96,97] Several members of the MarR (multiple antibiotic resistance regulator) family of transcription factors also bind Fe-S clusters,[98] including ChlR, which utilizes a [4Fe-4S] cluster to modulate expression of genes involved in pigment biosynthesis in response to oxygen.[99] Several AraC-like transcriptional regulators control expression of heme iron utilization pathways in bacteria by modulating gene expression in response to heme binding.[100,101] The transcriptional activator AcrR, a titular member of the TetR/AcrR structural superfamily, modulates expression of an aldehyde-alcohol dehydrogenase and is not known to bind any metallocofacor.[102]

Members of two-component signaling pathways have also been identified in the genomic context of anaerobic CO metabolism.[23] These regulators include EnzV/OmpR-like proteins involved in bacterial osmoregulation[103] and FixJ/NarR-like proteins, which are involved in heme-dependent oxygen sensing.[104] Additionally, a CO-responsive two-component transcriptional

regulator, CorQR, was recently identified in the genomic context of a *coo* operon in a strain of the carboxydotrophic hydrogenogenic archaeon *Thermococcus onnurineus*.[105] The sensory component, CorQ, contains a 4-vinyl reductase domain with a several conserved Cys residues that may bind an Fe-S cluster.[106] The presence of two-component regulators in the genomic context of anaerobic CO metabolism suggests that regulation of CO metabolism may occur as part of one or multiple signaling cascades.

Many of the putative transcriptional regulators associated with anaerobic CO metabolism are members of structural families with known metallocofactor-binding motifs. Given that CO exhibits a high affinity for biological transition metals, these uncharacterized metallocofactor-binding proteins, could represent new CO-sensing transcription factors. For example, members of the AraC family are known heme sensors. AraC-like transcriptional regulators associated with anaerobic CO metabolism genes could therefore comprise a third class heme-dependent CO-sensing transcription factors that is distinct from CooA and RcoM. Members of the IscR, NsrR, and MarR families of transcription factors bind Fe-S clusters. A large number of Fe-S cluster proteins have been identified that regulate transcription in response to small gaseous molecules, including NsrR and FNR, which sense NO and O_2, respectively.[96,107] Use of Fe-S clusters to sense CO would present a novel function for these cofactors, although the putative Fe-S cluster-binding residues in may instead bind other metallocofactors, such as heme.

Putative novel regulators of aerobic metabolism

Two putative CO-sensing transcription factors, CoxC and CoxH, were identified upstream and downstream of *coxMSL*, respectively, in the carboxidotrophic bacterium *oligotropha carboxidovorans*.[108,109] Both CoxC and CoxH possess an N-terminal transmembrane region annotated as an MHYT domain, named for four conserved residues identified in one of the

transmembrane domains.[110] Originally proposed as a copper-bonding site on the basis of the conserved His and Met residues in the MHYT domain, it is possible that this motif could also bind heme. In fact, when a low alignment cutoff score is applied to the RcoM SSN described above, RcoM proteins form a cluster that is linked to a cluster containing putative CoxC and CoxH proteins. This connection between RcoM and CoxC/H proteins in the SSN suggests that divergent evolution of a common ancestor may have given rise to these two distinct CO-sensing transcription factors with distinct N-terminal structural motifs. No biochemical characterization of CoxC or CoxH has been carried out to date.

1.6 Concluding Remarks

A diverse array of microorganisms have evolved elaborate biochemical systems that enable utilization of CO as a source of carbon and energy under anaerobic and aerobic conditions. Large energy inputs are needed to assemble these CO-metabolizing systems, which require specialized cofactors and accessory proteins in order to function. As a result, microorganisms have developed CO-responsive transcriptional regulators that ensure CO metabolizing pathways are only turned on under favorable conditions in which CO is abundant. Detailed biochemical characterization of two such transcription factors, CooA and RcoM, illustrate that CO-metabolizing microorganisms employ heme to sense and respond to the presence of CO. Genomic analyses suggest that widespread genetic variation exists in the enzymes (and associated accessory proteins) that carry out CO metabolism and in the transcription factors that regulate expression of these enzymes. In light of this genetic diversity, it is clear that there are many enzymes, accessory proteins, and regulatory pathways that have yet to be fully enumerated. In order to better understand these many components of microbial CO metabolism, it is important to simultaneously consider biological, genetic, and biochemical contexts. This approach has been critical in

developing detailed regulatory models in CooA and RcoM. The methodologies used to study these two CO-sensing transcription factors now serves as a guiding framework for exploring the many facets of microbial CO metabolism.

1.7 References

1. Khalil, M. A. K.; Rasmussen, R. A. The global cycle of carbon monoxide- trends and mass balance. *Chemosphere* **1990**, *20*, 227.
2. Stein, O.; Schultz, M. G.; Bouarar, I.; Clark, H.; Huijnen, V.; Gaudel, A.; George, M.; Clerbaux, C. On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations. *Atmos. Chem. Phys.* **2014**, *14*, 9295.
3. Liu, L. C.; Zhuang, Q. L.; Zhu, Q.; Liu, S. Q.; van Asperen, H.; Pihlatie, M. Global soil consumption of atmospheric carbon monoxide: an analysis using a process-based biogeochemistry model. *Atmos. Chem. Phys.* **2018**, *18*, 7913.
4. Prather, M.; Ehhalt, D. In *Climate Change, 2001: The Scientific Basis*; Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., Johnson, C. A., Eds.; Cambridge University Press: Cambridge, UK, 2001, p 239.
5. King, G. M.; Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* **2007**, *5*, 107.
6. Oelgeschlager, E.; Rother, M. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch. Microbiol.* **2008**, *190*, 257.
7. King, G. M.; Crosby, H. Impacts of plant roots on soil CO cycling and soil-atmosphere CO exchange. *Glob. Change Biol.* **2002**, *8*, 1085.
8. Delmoral, R.; Clampitt, C. A. Growth of native plant-species on recent volcanic substrated from Mount St. Helens. *Am. Midl. Nat.* **1985**, *114*, 374.
9. Kleiner, M.; Wentrup, C.; Lott, C.; Teeling, H.; Wetzel, S.; Young, J.; Chang, Y. J.; Shah, M.; VerBerkmoes, N. C.; Zarzycki, J.; Fuchs, G.; Markert, S.; Hempel, K.; Voigt, B.; Becher, D.; Liebeke, M.; Lalk, M.; Albrecht, D.; Hecker, M.; Schweder, T.; Dubilier, N. Metaproteomics

of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1173.

10. Kleiner, M.; Wentrup, C.; Holler, T.; Lavik, G.; Harder, J.; Lott, C.; Littmann, S.; Kuypers, M. M. M.; Dubilier, N. Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ. Microbiol.* **2015**, *17*, 5023.
11. Adam, P. S.; Borrel, G.; Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1166.
12. Lilley, M. D.; Deangelis, M. A.; Gordon, L. I. CH_4, H_2, CO and N_2O in submarine hydrothermal vent waters. *Nature* **1982**, *300*, 48.
13. Techtmann, S. M.; Colman, A. S.; Robb, F. T. 'That which does not kill us only makes us stronger': the role of carbon monoxide in thermophilic microbial consortia. *Environ. Microbiol.* **2009**, *11*, 1027.
14. Can, M.; Armstrong, F. A.; Ragsdale, S. W. Structure, Function, and Mechanism of the Nickel Metalloenzymes, CO Dehydrogenase, and Acetyl-CoA Synthase. *Chem. Rev.* **2014**, *114*, 4149.
15. Ragsdale, S. W.; Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO2 fixation. *Biochim. Biophys. Acta Proteins Proteom.* **2008**, *1784*, 1873.
16. Bowien, B.; Schlegel, H. G. Physiology and biochemistry of aerobic hydrogen-oxidizing bateria. *Annu. Rev. Microbiol.* **1981**, *35*, 405.
17. Dobbek, H.; Svetlitchnyi, V.; Gremer, L.; Huber, R.; Meyer, O. Crystal structure of a carbon monoxide dehydrogenase reveals a Ni-4Fe-5S cluster. *Science* **2001**, *293*, 1281.
18. Gong, W.; Hao, B.; Wei, Z.; Ferguson, D. J.; Tallant, T.; Krzycki, J. A.; Chan, M. K. Structure of the alpha(2)epsilon(2) Ni-dependent CO dehydrogenase component of the Methanosarcina

barkeri acetyl-CoA decarbonylase/synthase complex. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9558.

19. Svetlitchnyi, V.; Peschel, C.; Acker, G.; Meyer, O. Two membrane-associated NiFeS-carbon monoxide dehydrogenases from the anaerobic carbon-monoxide-utilizing eubacterium Carboxydothermus hydrogenoformans. *J. Bacteriol.* **2001**, *183*, 5134.
20. Techtmann, S. M.; Lebedinsky, A. V.; Colman, A. S.; Sokolova, T. G.; Woyke, T.; Goodwin, L.; Robb, F. T. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* **2012**, *3*, 16.
21. Jeoung, J.-H.; Dobbek, H. Carbon Dioxide Activation at the Ni,Fe-Cluster of Anaerobic Carbon Monoxide Dehydrogenase. *Science* **2007**, *318*, 1461.
22. Dobbek, H.; Gremer, L.; Kiefersauer, R.; Huber, R.; Meyer, O. Catalysis at a dinuclear [CuSMo(==O)OH] cluster in a CO dehydrogenase resolved at 1.1-A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15971.
23. Inoue, M.; Nakamoto, I.; Omae, K.; Oguro, T.; Ogata, H.; Yoshida, T.; Sako, Y. Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases. *Front. Microbiol.* **2019**, *9*.
24. Arendsen, A. F.; Soliman, M. Q.; Ragsdale, S. W. Nitrate-Dependent Regulation of Acetate Biosynthesis and Nitrate Respiration by *Clostridium thermoaceticum*. *J. Bacteriol.* **1999**, *181*, 1489.
25. Sundara Sekar, B.; Mohan Raj, S.; Seol, E.; Ainala, S. K.; Lee, J.; Park, S. Cloning and functional expression of *Citrobacter amalonaticus* Y19 carbon monoxide dehydrogenase in *Escherichia coli*. *Int. J. Hydrogen Energy* **2014**, *39*, 15446.

26. Jeon, W. B.; Cheng, J. J.; Ludden, P. W. Purification and characterization of membrane-associated CooC protein and its functional role in the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* **2001**, *276*, 38602.

27. Jeoung, J. H.; Giese, T.; Grunwald, M.; Dobbek, H. CooC1 from *Carboxydothermus hydrogenoformans* Is a Nickel-Binding ATPase. *Biochemistry* **2009**, *48*, 11505.

28. Jeoung, J. H.; Giese, T.; Grunwald, M.; Dobbek, H. Crystal Structure of the ATP-Dependent Maturation Factor of Ni,Fe-Containing Carbon Monoxide Dehydrogenases. *J. Mol. Biol.* **2010**, *396*, 1165.

29. Timm, J.; Brochier-Armanet, C.; Perard, J.; Zambelli, B.; Ollagnier-de-Choudens, S.; Ciurli, S.; Cavazza, C. The CO dehydrogenase accessory protein CooT is a novel nickel-binding protein. *Metallomics* **2017**, *9*, 575.

30. Watt, R. K.; Ludden, P. W. The identification, purification, and characterization of CooJ - A nickel-binding protein that is CO-regulated with the Ni-containing CO dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* **1998**, *273*, 10019.

31. Merrouch, M.; Benvenuti, M.; Lorenzi, M.; Leger, C.; Fourmond, V.; Dementin, S. Maturation of the Ni-4Fe-4S active site of carbon monoxide dehydrogenases. *J. Biol. Inorg. Chem.* **2018**, *23*, 613.

32. Ragsdale, S. W. Life with Carbon Monoxide. *Crti. Rev. Biochem. Mol. Biol.* **2004**, *39*, 165.

33. Singer, S. W.; Hirst, M. B.; Ludden, P. W. CO-dependent H2 evolution by *Rhodospirillum rubrum*: Role of CODH:CooF complex. *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 1582.

34. Meyer, O.; Rajagopalan, K. V. Molybdopterin in carbon monoxide oxidase from carboxydotrophic bacteria. *J. Bacteriol.* **1984**, *157*, 643.

35. Bell, J. M.; Colby, J.; Williams, E. CO Oxidoreductase from *Streptomyces strain G26* is a molybdenum hydroxylase. *Biochem. J.* **1988**, *250*, 605.

36. Hille, R. Molybdenum-containing hydroxylases. *Arch. Biochem. Biophys.* **2005**, *433*, 107.

37. Dobbek, H.; Gremer, L.; Meyer, O.; Huber, R. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8884.

38. Kaufmann, P.; Duffus, B. R.; Teutloff, C.; Leimkühler, S. Functional Studies on *Oligotropha carboxidovorans* Molybdenum−Copper CO Dehydrogenase Produced in *Escherichia coli*. *Biochemistry* **2018**, *57*, 2889.

39. Meyer, O.; Jacobitz, S.; Kruger, B. Biochemistry and physiology of aerobic carbon monoxide-utilizing bacteria. *FEMS Microbiol. Lett.* **1986**, *39*, 161.

40. Lorite, M. J.; Tachil, J.; Sanjuán, J.; Meyer, O.; Bedmar, E. J. Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Appl. Environ. Microbiol.* **2000**, *66*, 1871.

41. Pelzmann, A.; Gnida, M.; Meyer-Klaucke, W.; Maisel, T.; Meyer, O. The CoxD Protein of *Oligotropha carboxidovorans* Is a Predicted AAA+ ATPase Chaperone Involved in the Biogenesis of the CO Dehydrogenase [CuSMoO2] Cluster. *J. Biol. Chem.* **2009**, *284*, 9578.

42. Maisel, T.; Joseph, S.; Mielke, T.; Bürger, J.; Schwarzinger, S.; Meyer, O. The CoxD Protein, a Novel AAA+ ATPase Involved in Metal Cluster Assembly: Hydrolysis of Nucleotide Triphosphates and Oligomerization. *PLoS One* **2012**, *7*.

43. Pelzmann, A.; Mickoleit, F.; Meyer, O. Insights into the posttranslational assembly of the Mo-, S- and Cu-containing cluster in the active site of CO dehydrogenase of *Oligotropha carboxidovorans*. *J. Biol. Inorg. Chem.* **2014**, *19*, 1399.

44. Lanzilotta, W. N.; Schuller, D. J.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure of the CO sensing transcription activator CooA. *Nat. Struct. Biol.* **2000**, *7*, 876.

45. Leduc, J.; Thorsteinsson, M. V.; Gaal, T.; Roberts, G. P. Mapping CooA·RNA Polymerase Interactions: Identification of Activating Regions 2 and 3 in CooA, the CO-sensing Transcriptional Activator. *J. Biol. Chem.* **2001**, *276*, 39968.

46. Roberts, G. P.; Youn, H.; Kerby, R. L. CO-sensing mechanisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 453.

47. Kolb, A.; Busby, S.; Buc, H.; Garges, S.; Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **1993**, *62*, 749.

48. Fox, J. D.; He, Y. P.; Shelver, D.; Roberts, G. P.; Ludden, P. W. Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J. Bacteriol.* **1996**, *178*, 6200.

49. He, Y.; Shelver, D.; Kerby, R. L.; Roberts, G. P. Characterization of a CO-responsive Transcriptional Activator from *Rhodospirillum rubrum*. *J. Biol. Chem.* **1996**, *271*, 120.

50. Aono, S.; Nakajima, H.; Saito, K.; Okada, M. A novel heme protein that acts as a carbon monoxide-dependent transcriptional activator in *Rhodospirillum rubrum*. *Biochem. Biophys. Res. Comm.* **1996**, *228*, 752.

51. Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Chung, S. Y.; Roberts, G. P.; Reynolds, M. F.; Parks, R. B.; Burstyn, J. N. Identification of two important heme site residues (cysteine 75 and histidine 77) in CooA, the CO-sensing transcription factor of *Rhodospirillum rubrum*. *Biochemistry* **1999**, *38*, 2669.

52. Nakajima, H.; Nakagawa, E.; Kobayashi, K.; Tagawa, S.; Aono, S. Ligand-switching intermediates for the CO-sensing transcriptional activator CooA measured by pulse radiolysis. *J. Biol. Chem.* **2001**, *276*, 37895.

53. Youn, H.; Kerby, R. L.; Thorsteinsson, M. V.; Clark, R. W.; Burstyn, J. N.; Roberts, G. P. Analysis of the L116K variant of CooA, the heme-containing CO sensor, suggests the presence of an unusual heme ligand resulting in novel activity. *J. Biol. Chem.* **2002**, *277*, 33616.

54. Puranik, M.; Nielsen, S. B.; Youn, H.; Hvitved, A. N.; Bourassa, J. L.; Case, M. A.; Tengroth, C.; Balakrishnan, G.; Thorsteinsson, M. V.; Groves, J. T.; McLendon, G. L.; Roberts, G. P.; Olson, J. S.; Spiro, T. G. Dynamics of carbon monoxide binding to CooA. *J. Biol. Chem.* **2004**, *279*, 21096.

55. Youn, H.; Conrad, M.; Chung, S.-Y.; Roberts, G. P. Roles of the heme and heme ligands in the activation of CooA, the CO-sensing transcriptional activator. *Biochem. Biophys. Res. Comm.* **2006**, *348*, 345.

56. Borjigin, M.; Li, H.; Lanz, N. D.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure-based hypothesis on the activation of the CO-sensing transcription factor CooA. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2007**, *63*, 282.

57. Akiyama, S.; Fujisawa, T.; Ishimori, K.; Morishima, I.; Aono, S. Activation mechanisms of transcriptional regulator CooA revealed by small-angle X-ray scattering. *J. Mol. Biol.* **2004**, *341*, 651.

58. Yamamoto, K.; Ishikawa, H.; Takahashi, S.; Ishimori, K.; Morishima, I.; Nakajima, H.; Aono, S. Binding of CO at the Pro^2 side is crucial for the activation of CO-sensing transcriptional activator CooA. ^{1}H NMR spectroscopic studies. *J. Biol. Chem.* **2001**, *276*, 11473.

59. Kerby, R. L.; Youn, H.; Thorsteinsson, M. V.; Roberts, G. P. Repositioning about the dimer interface of the transcription regulator CooA: A major signal transduction pathway between the effector and DNA-binding domains. *J. Mol. Biol.* **2003**, *325*, 809.

60. Youn, H.; Kerby, R. L.; Roberts, G. P. The role of the hydrophobic distal heme pocket of CooA in ligand sensing and response. *J. Biol. Chem.* **2003**, *278*, 2333.

61. Clark, R. W.; Youn, H.; Parks, R. B.; Cherney, M. M.; Roberts, G. P.; Burstyn, J. N. Investigation of the role of the N-terminal proline, the distal heme ligand in the CO sensor CooA. *Biochemistry* **2004**, *43*, 14149.

62. Otomo, A.; Ishikawa, H.; Mizuno, M.; Kimura, T.; Kubo, M.; Shiro, Y.; Aono, S.; Mizutani, Y. A study of the dynamics of the heme pocket and C-helix in CooA upon CO dissociation using time-resolved visible and UV resonance Raman spectroscopy. *J. Phys. Chem. B* **2016**, *120*, 7836.

63. Tripathi, S.; Poulos, T. L. Testing the N-terminal velcro model of CooA carbon monoxide activation. *Biochemistry* **2018**.

64. Clark, R. W.; Lanz, N. D.; Lee, A. J.; Kerby, R. L.; Roberts, G. P.; Burstyn, J. N. Unexpected NO-dependent DNA binding by the CooA homolog from *Carboxydothermus hydrogenoformans*. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 891.

65. Youn, H.; Kerby, R. L.; Conrad, M.; Roberts, G. P. Functionally critical elements of CooA-related CO sensors. *Journal of Bacteriology* **2004**, *186*, 1320.

66. Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.

67. Tzeng, S.-R.; Kalodimos, C. G. Allosteric inhibition through suppression of transient conformational states. *Nat. Chem. Biol.* **2013**, *9*, 462.

68. Tzeng, S.-R.; Kalodimos, C. G. The role of slow and fast protein motions in allosteric interactions. *Biophys. Rev.* **2015**, *7*, 251.

69. Capdevila, D. A.; Braymer, J. J.; Edmonds, K. A.; Wu, H.; Giedroc, D. P. Entropy redistribution controls allostery in a metalloregulatory protein. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 4424.

70. Rubtsov, I. V.; Zhang, T.; Nakajima, H.; Aono, S.; Rubtsov, G. I.; Kumazaki, S.; Yoshihara, K. Conformational dynamics of the transcriptional regulator CooA protein studied by subpicosecond mid-infrared vibrational spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 10056.

71. Hines, J. P.; Dent, M. R.; Stevens, D. J.; Burstyn, J. N. Site-directed spin label electron paramagnetic resonance spectroscopy as a probe of conformational dynamics in the Fe(III) "locked-off" state of the CO-sensing transcription factor CooA. *Prot. Sci.* **2018**, *27*, 1670.

72. Jeoung, J. H.; Fesseler, J.; Goetzl, S.; Dobbek, H. In *Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*; Kroneck, P. M. H., Torres, M. E. S., Eds.; Springer: Dordrecht, 2014; Vol. 14, p 37.

73. Kerby, R. L.; Youn, H.; Roberts, G. P. RcoM: a new single-component transcriptional regulator of CO metabolism in bacteria. *J. Bacteriol.* **2008**, *190*, 3336.

74. Smith, A. T.; Marvin, K. A.; Freeman, K. M.; Kerby, R. L.; Roberts, G. P.; Burstyn, J. N. Identification of Cys94 as the distal ligand to the Fe(III) heme in the transcriptional regulator RcoM-2 from *Burkholderia xenovorans*. *J. Biol. Inorg. Chem.* **2012**, *17*, 1071.

75. Kerby, R. L.; Roberts, G. P. Burkholderia xenovorans RcoMBx-1, a Transcriptional Regulator System for Sensing Low and Persistent Levels of Carbon Monoxide. *J. Bacteriol.***2012**, *194*, 5803.

76. Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met104 is the CO-replaceable ligand at Fe(II) heme in the CO-sensing transcription factor BxRcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21*, 559.

77. Bouzhir-Sima, L.; Motterlini, R.; Gross, J.; Vos, M. H.; Liebl, U. Unusual Dynamics of Ligand Binding to the Heme Domain of the Bacterial CO Sensor Protein RcoM-2. *J. Phys. Chem. B* **2016**, *120*, 10686.

78. Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17*, 1282.

79. Miyatake, H.; Mukai, M.; Park, S.-Y.; Adachi, S.-i.; Tamura, K.; Nakamura, H.; Nakamura, K.; Tsuchiya, T.; Iizuka, T.; Shiro, Y. Sensory mechanism of oxygen sensor FixL from *Rhizobium meliloti*: crystallographic, mutagenesis and resonance raman spectroscopic studies11Edited by K. Nagai. *J. Mol. Biol.* **2000**, *301*, 415.

80. Hao, B.; Isaza, C.; Arndt, J.; Soltis, M.; Chan, M. K. Structure-Based Mechanism of O2 Sensing and Ligand Discrimination by the FixL Heme Domain of *Bradyrhizobium japonicum*. *Biochemistry* **2002**, *41*, 12952.

81. Kurokawa, H.; Lee, D.-S.; Watanabe, M.; Sagami, I.; Mikami, B.; Raman, C. S.; Shimizu, T. A Redox-controlled Molecular Switch Revealed by the Crystal Structure of a Bacterial Heme PAS Sensor. *J. Biol. Chem.* **2004**, *279*, 20186.

82. Sidote, D. J.; Barbieri, C. M.; Wu, T.; Stock, A. M. Structure of the *Staphylococcus aureus* AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with an Unusual Mode of Binding. *Structure* **2008**, *16*, 727.

83. Nikolskaya, A. N.; Galperin, M. Y. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* **2002**, *30*, 2453.

84. Ween, O.; Gaustad, P.; Havarstein, L. S. Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **1999**, *33*, 817.

85. Cheung, J. K.; Rood, J. I. The VirR Response Regulator from *Clostridium perfringens* Binds Independently to Two Imperfect Direct Repeats Located Upstream of the *pfoA* Promoter. *J. Bacteriol.* **2000**, *182*, 57.

86. de Saizieu, A.; Gardès, C.; Flint, N.; Wagner, C.; Kamber, M.; Mitchell, T. J.; Keck, W.; Amrein, K. E.; Lange, R. Microarray-Based Identification of a Novel *Streptococcus pneumoniae* Regulon Controlled by an Autoinduced Peptide. *J. Bacteriol.* **2000**, *182*, 4696.

87. Diep, D. B.; Håvarstein, L. S.; Nes, I. F. Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. *J. Bacteriol.* **1996**, *178*, 4472.

88. Risoen, P. A.; Johnsborg, O.; Diep, D. B.; Hamoen, L.; Venema, G.; Nes, I. F. Regulation of bacteriocin production in *Lactobacillus plantarum* depends on a conserved promoter arrangement with consensus binding sequence. *Mol. Genetics Genom.* **2001**, *265*, 198.

89. Knutsen, E.; Ween, O.; Havarstein, L. S. Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. *J. Bacteriol.* **2004**, *186*, 3078.

90. Reyes, D.; Andrey, D. O.; Monod, A.; Kelley, W. L.; Zhang, G.; Cheung, A. L. Coordinated Regulation by AgrA, SarA, and SarR To Control *agr* Expression in *Staphylococcus aureus*. *J. Bacteriol.* **2011**, *193*, 6020.

91. Straume, D.; Johansen, R. F.; Bjoras, M.; Nes, I. F.; Diep, D. B. DNA binding kinetics of two response regulators, PlnC and PlnD, from the bacteriocin regulon of *Lactobacillus plantarum* C11. *BMC Biochem.* **2009**, *10*.

92. Kerby, R. L.; Roberts, G. P. Sustaining N_2-Dependent Growth in the Presence of CO. *J. Bacteriol.* **2011**, *193*, 774.

93. Hoffmann, M. C.; Pfander, Y.; Fehringer, M.; Narberhaus, F.; Masepohl, B. NifA- and CooA-Coordinated cowN Expression Sustains Nitrogen Fixation by *Rhodobacter capsulatus* in the Presence of Carbon Monoxide. *J. Bacteriol.* **2014**, *196*, 3494.

94. Schwartz, C. J.; Giel, J. L.; Patschkowski, T.; Luther, C.; Ruzicka, F. J.; Beinert, H.; Kiley, P. J. IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins. *Proc. Nat. Acad. Sci. U. S. A.* **2001**, *98*, 14895.

95. Giel, J. L.; Nesbit, A. D.; Mettert, E. L.; Fleischhacker, A. S.; Wanta, B. T.; Kiley, P. J. Regulation of iron–sulphur cluster homeostasis through transcriptional control of the Isc pathway by [2Fe–2S]–IscR in *Escherichia coli*. *Mol. Microbiol.* **2013**, *87*, 478.

96. Kommineni, S.; Lama, A.; Popescu, B.; Nakano, M. M. Global Transcriptional Control by NsrR in *Bacillus subtilis*. *J. Bacteriol.* **2012**, *194*, 1679.

97. Volbeda, A.; Dodd, E. L.; Darnault, C.; Crack, J. C.; Renoux, O.; Hutchings, M. I.; Le Brun, N. E.; Fontecilla-Camps, J. C. Crystal structures of the NO sensor NsrR reveal how its iron-sulfur cluster modulates DNA binding. *Nat. Comm.* **2017**, *8*, 15052.

98. Perera, I. C.; Grove, A. Molecular Mechanisms of Ligand-Mediated Attenuation of DNA Binding by MarR Family Transcriptional Regulators. *J. Mol. Cell Biol.* **2010**, *2*, 243.

99. Ludwig, M.; Pandelia, M.-E.; Chew, C. Y.; Zhang, B.; Golbeck, J. H.; Krebs, C.; Bryant, D. A. ChlR Protein of Synechococcus sp. PCC 7002 Is a Transcription Activator That Uses an Oxygen-sensitive [4Fe-4S] Cluster to Control Genes involved in Pigment Biosynthesis. *J. Biol.Chem.* **2014**, *289*, 16624.

100. Vanderpool, C. K.; Armstrong, S. K. Heme-Responsive Transcriptional Activation of *Bordetella bhu* Genes. *J. Bacteriol.* **2003**, *185*, 909.

101. Amarelle, V.; Koziol, U.; Fabiano, E. Highly conserved nucleotide motifs present in the 5′UTR of the heme-receptor gene shmR are required for HmuP-dependent expression of shmR in *Ensifer meliloti*. *BioMetals* **2019**, *32*, 273.

102. Yang, X.; Teng, K.; Su, R.; Li, L.; Zhang, T.; Fan, K.; Zhang, J.; Zhong, J. AcrR and Rex Control Mannitol and Sorbitol Utilization through Their Cross-Regulation of Aldehyde-Alcohol Dehydrogenase (AdhE) in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* **2019**, *85*, e02035.

103. Cai, S. J.; Inouye, M. EnvZ-OmpR Interaction and Osmoregulation in *Escherichia coli*. *J. Biol. Chem.* **2002**, *277*, 24155.

104. Wright, G. S. A.; Saeki, A.; Hikima, T.; Nishizono, Y.; Hisano, T.; Kamaya, M.; Nukina, K.; Nishitani, H.; Nakamura, H.; Yamamoto, M.; Antonyuk, S. V.; Hasnain, S. S.; Shiro, Y.; Sawai, H. Architecture of the complete oxygen-sensing FixL-FixJ two-component signal transduction system. *Sci. Signal.* **2018**, *11*, eaaq0825.

105. Kim, M.-S.; Choi, A. R.; Lee, S. H.; Jung, H.-C.; Bae, S. S.; Yang, T.-J.; Jeon, J. H.; Lim, J. K.; Youn, H.; Kim, T. W.; Lee, H. S.; Kang, S. G. A Novel CO-Responsive Transcriptional Regulator and Enhanced H2 Production by an Engineered *Thermococcus onnurineus* NA1 Strain. *Appl. Environ. Microbiol.* **2014**, *81*, 1708.

106. Anantharaman, V.; Koonin, E. V.; Aravind, L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol.Biol.* **2001**, *307*, 1271.

107. Crack, J. C.; Green, J.; Thomson, A. J.; Brun, N. E. L. Iron–Sulfur Clusters as Biological Sensors: The Chemistry of Reactions with Molecular Oxygen and Nitric Oxide. *Acc. Chem. Res.* **2014**, *47*, 3196.

108. Santiago, B.; Schübel, U.; Egelseer, C.; Meyer, O. Sequence analysis, characterization and CO-specific transcription of the cox gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. *Gene* **1999**, *236*, 115.

109. Fuhrmann, S.; Ferner, M.; Jeffke, T.; Henne, A.; Gottschalk, G.; Meyer, O. Complete nucleotide sequence of the circular megaplasmid pHCG3 of *Oligotropha carboxidovorans*: function in the chemolithoautotrophic utilization of CO, H-2 and CO2. *Gene* **2003**, *322*, 67.

110. Galperin, M. Y.; Gaidenko, T. A.; Mulkidjanian, A. Y.; Nakano, M.; Price, C. W. MHYT, a new integral membrane sensor domain. *FEMS Microbiol. Lett.* **2001**, *205*, 17.

1.8 Appendix 1: CooA Bioinformatics Methods

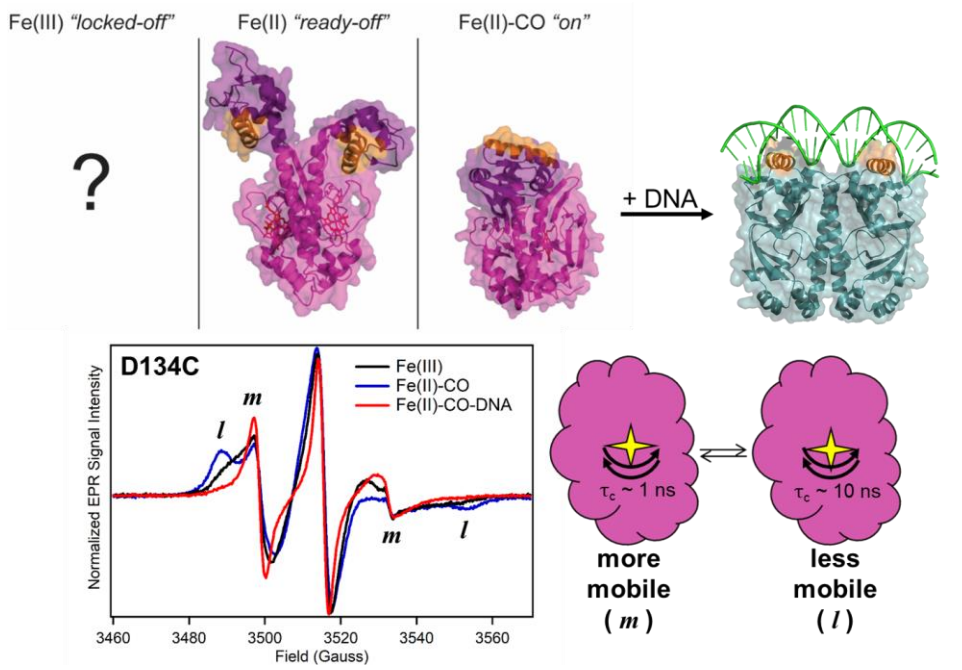
Bioinformatics analyses of CooA

A sequence similarity network (SSN) of putative CooA proteins was generated using the single-sequence query function of the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) and providing *Rr*CooA (UniProt ID: P72322) as the query sequence. The top 1,000 hits with an E-value less than 10^{-5} were included in the preliminary data processing step. An initial 100% representative node SSN containing 767 unique protein sequences was generated using an alignment score of 24 and subsequently processed using Cytoscape 3.6.0. Using the Enzyme Function Initiative-Genome Neighborhood Tool (EFI-GNT), genome neighborhoods of each protein within the network were accessed and queried for proteins annotated with Pfam descriptions of Prismane, Fer4_11, and Proton_antiporter_M, corresponding to the CODH, CooF, and CooM family of proteins, respectively. These queried proteins were associated with only the second and third largest clusters of CooA proteins (N = 202 nodes, 219 sequences) in the initial SSN. These clusters were extracted and used for all further analysis.

Utilizing the genome neighborhoods generated by the EFI-GNT, putative CooA sequences were classified by their association with CODH (*cooS*), *cooF*, ECH proteins (*cooMKLXUH*), ACS, and *cowN*. These classifications were mapped onto the SSN in cytoscape, and the network nodes were filtered to sequence alignment cutoff scores of 37 and 50 to demonstrate the relationships between the putative CooA proteins within the SSN. Within the network, 75 putative CooA nodes were classified in the context of CODH only, 62 in the context of *cowN*, and 13 in the context of *cooF*, 28 in the context of ECH proteins, and 2 in the context of ACS. In total, 24 putative CooAs showed no association with any of the aforementioned proteins and were classified as within an unknown genomic context.

Chapter Two

Exploring the Role of Protein Dynamics in the CO-Dependent Allosteric Activation Mechanism of *Rr*CooA



A previous version of this chapter was published as:

*Hines, J.P.; *Dent, M.R.; Stevens, D.J.; Burstyn, J.N. Site-directed spin label electron paramagnetic resonance spectroscopy as a probe of conformational dynamics in the Fe(III) "locked-off" state of the CO-sensing transcription factor CooA. *Prot. Sci.* **2018**, *27*, 1670.

*These authors contributed equally to this work

Matthew Dent performed experiments and wrote the manuscript. Judy Hines wrote a preliminary version of the manuscript. Daniel Stevens developed the protein variants and designed the initial site-directed spin label electron paramagnetic resonance spectroscopy experiments.

2.1 Introduction

The CRP/FNR (cAMP receptor protein/fumarate and nitrate reductase) superfamily encompasses a large number of bacterial transcriptional regulators involved in sensing a diverse range of environmental stimuli.[1] An important question that remains unanswered for most members of this structural superfamily is how interaction between the protein and a specific environmental stimulus results in activation/inactivation of protein function (i.e. DNA binding). While this question has been addressed in detail for CRP (also known as catabolite activator protein, CAP), a transcriptional activator that senses the catabolite cAMP, mechanisms of allosteric activation for most other members of the CRP/FNR superfamily are largely unknown. The transcriptional activator CooA (CO oxidation activator) is one such member of the superfamily. Unlike CRP, CooA regulates anaerobic carbon monoxide (CO) metabolism in facultative and obligate anaerobes in response to CO.[2-5] Although a structural mechanism for the allosteric activation of CooA exists, the role of dynamics, which has been shown to be critical in regulating function in CRP, has not been explored for CooA.[6-10] In order to compare the role of dynamics in allosteric activation between CooA and CRP, we developed a means to probe conformational dynamics in CooA and demonstrated that CooA samples multiple conformations in the poorly-characterized Fe(III) *"locked-off"* state.

For CRP, changes in protein dynamics account for the negative cooperativity of cAMP binding while an effector-driven structural transition adequately explains allosteric activation. cAMP binding to CRP is negatively cooperative: binding of the first cAMP molecule at one monomer activates slow (μs-ms) motions across the entire protein without affecting changes to fast (ps-ns) dynamics; binding of the second cAMP molecule suppresses fast dynamics and thereby imparts an entropic binding penalty.[6,11] In the structural activation model for wild-type (WT) CRP,

binding of two cAMP molecules induces a coil-to-helix transition that results in repositioning of the DNA-binding domains for favorable interaction with DNA.[11-13] This coil-to-helix transition is driven by specific interactions between the adenine base and phosphate groups of cAMP with the residues surrounding the effector binding pocket.[13]

Importantly, there are a number of CRP variants in which altered allosteric activation is attributed to changes in dynamics *without* changes in structure.[7,8,10,14] Using a combination of NMR chemical shift and relaxation dispersion analyses, Tzeng and Kalodimos found no correlation between the population of active species and DNA binding affinity for eleven different CRP variants with single- or double-point mutations.[8] To better understand the factors that contribute to favorable DNA binding in these variants, the authors quantified thermodynamic parameters ΔH and $-T\Delta S$ for DNA binding using isothermal titration calorimetry (ITC). Variants with high DNA binding affinities that did not significantly populate the active DNA binding conformation exhibited a large, favorable change in entropy upon DNA binding. The origin of this favorable change in entropy was attributed to an increase in fast (ns-ps) dynamics upon DNA binding, as determined by monitoring the NMR order parameter, S^2. These observations clearly demonstrate that effector-driven changes in fast dynamics and conformational entropy can completely account for allosteric activation without invoking changes in structure.

Despite being a structural homologue of CRP, the transcriptional activator CooA employs a different strategy for sensing its allosteric effector, CO. Like CRP, CooA exists as a homodimer in solution, and each monomer contains an effector binding domain (residues 2-130) and a DNA-binding domain (residues 131-213).[15] In place of a cAMP binding site, the effector binding domain of CooA contains a heme cofactor. In the facultative anaerobe *Rhodospirillum rubrum* (*Rr*), the heme serves dual functions as a redox sensor and a CO sensor: CO binds to the Fe(II) heme of

each monomer following a reversible redox-mediated ligand switch (Figure 2.1).[5,16-22] Cooperative binding of CO to the Fe(II) heme of each monomer allosterically activates the protein to bind DNA.[20,23] Notable are the differences in effector binding cooperativity between CRP and CooA: binding of the cAMP molecules to CRP is negatively cooperative,[11] while binding of the CO molecules to CooA is positively cooperative.[20]

The current allosteric activation model for CooA invokes a structural change reminiscent of that of WT CRP. Evidence for this structural model includes a comparison of crystallographic data between inactive Fe(II) *"ready-off" Rr* CooA and a constitutively active variant of CooA from *Carboxydothermus hydrogenoformans* (*Ch*) (Figure 2.1).[15,24] In the model, CO binding induces a structural transition that results in exposure of the DNA-binding helices; however, asymmetry in the crystal structures precludes a complete understanding of this structural transition.[25,26] Small angle X-ray scattering (SAXS) data revealed no significant changes in the radius of gyration between Fe(II) *"ready-off"* and Fe(II)–CO *"on" Rr* CooA, consistent with the conclusion that the folded-over monomer of the asymmetric Fe(II) *"ready-off" Rr* CooA crystal structure is the major species in solution.[27] These data were interpreted to suggest a "swinging hinge" model in which the DNA binding domains undergo a propeller-like rotation to expose the DNA-binding F-helices upon CO binding. This rigid-body rotation resembles the changes observed in the effector-driven activation of CRP. [28]

Propagation of the allosteric CO binding signal in CooA is not completely understood. Citing constitutive activity of a covalently cross-linked variant of *Ch* CooA, Tripathi and Poulos recently suggested that, in *Ch* CooA, a structural transition to the active conformation may be facilitated by stabilizing interactions between the N-terminal tail of CooA and two other regions: the heme domain and the DNA-binding domain.[28] In this model, CO replaces an N-terminal

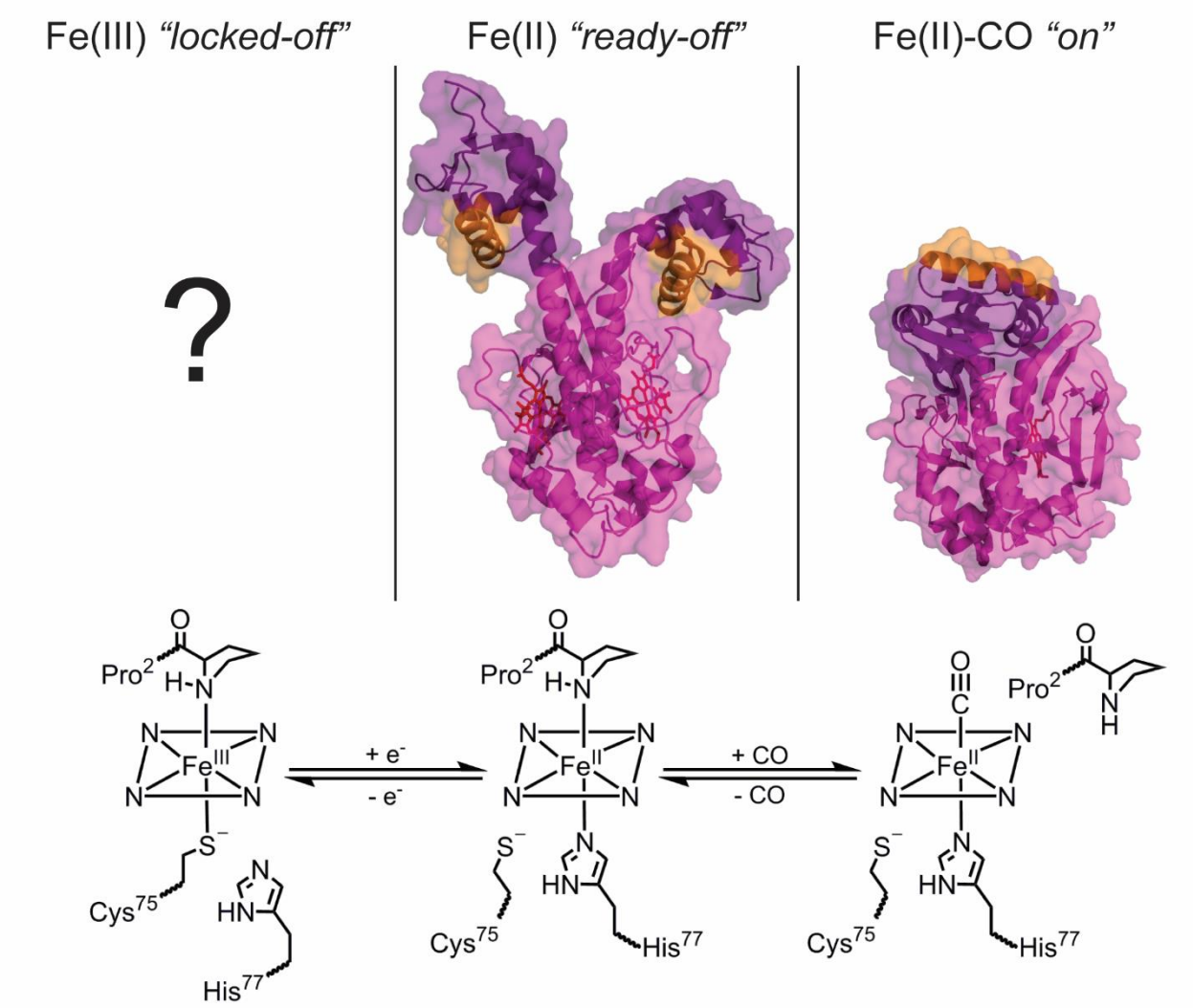


Figure 2.1 The three functional states of CooA and corresponding structural data for the Fe(II) "*ready-off*" state in *Rr* CooA (PDB 1FT9)[15] and Fe(II)–CO *"on"* state of *Ch* CooA (PDB 2HKX).[24] For each structure, the effector binding domain is shown in magenta, DNA binding domain in dark purple, and heme cofactors in red. The DNA-binding F-helices are highlighted in orange. Images were created using PyMOL Molecular Graphics System (Version 1.3 Schrödinger, LLC.). The unique heme ligation/oxidation state associated with each functional state is depicted below the structural data.

residue as an axial heme ligand, freeing the N-terminus to stabilize the active conformation. The allosteric mechanisms of *Ch* and *Rr* CooA may or may not be the same; the N-terminal amino acid sequence of *Ch* CooA differs from that of *Rr* CooA, *Ch* CooA does not appear to undergo redox-mediated ligand switching, and it has yet to be determined whether the N-terminus coordinates to heme in *Ch* CooA.[26,29] Other studies probed the propagation mechanism in *Rr* CooA. The effects of mutation of N-terminal residues[21,25,30] and recent UV resonance Raman data[31] suggest that CO-dependent allosteric activation occurs through loss of the Pro^2 heme ligand and subsequent rotation of the C-helices in the dimer interface. Without a direct structural comparison between activity states in the same CooA homologue, these structural hypotheses remain untested.

Neither the structural allosteric activation model nor models for allosteric signal propagation in CooA account for changes in protein dynamics; therefore, these models may be incomplete. Several lines of evidence suggest that protein dynamics may play a role in CO-dependent modulation of DNA binding. CO recombination following flash photolysis exhibited nonexponential behavior when probed by sub-picosecond mid-IR spectroscopy.[32] Biphasic kinetics for CO association and dissociation were observed using a combination of rapid mixing electronic absorbance, flash photolysis, and time-resolved resonance Raman spectroscopy.[20] Taken together, these observations provide spectroscopic evidence that Fe(II)–CO *"on"* CooA samples multiple conformational substates on fast timescales. The presence of such substates suggests that allosteric regulation of CooA may rely on modulation of fast dynamics. Additionally, the SAXS data summarized above does not eliminate the possibility that CooA samples both active and inactive conformations. It is therefore possible that CO binding allosterically modulates slow dynamics by increasing the fraction of protein sampling the active conformation. We seek to

understand the role of dynamics in the allosteric activation of CooA and to compare dynamic changes in CooA to those of its structural homologue, CRP.

To better understand how changes in slow dynamics may regulate CO-dependent DNA binding activity in CooA, we have employed site-directed spin label electron paramagnetic resonance (SDSL-EPR) spectroscopy. Of the comprehensive investigations of dynamically driven allostery present in the literature, most have utilized NMR experiments to probe motions on slow and fast timescales.[7,33-39] The size of CooA (50 kDa) and presence of a paramagnetic heme in the Fe(III) *"locked-off"* state render such NMR experiments challenging; however, these constraints do not limit the use of site-specifically incorporated nitroxide spin labels that report on local protein dynamics on the μs-ms timescale.[40-43] SDSL-EPR spectroscopy, which has been utilized to study conformational exchange and structural transitions in a number of soluble and membrane-bound proteins, is well-suited to probe conformational dynamics in CooA.[44-48] In this study, we compare conformational dynamics of five cysteine substitution variants of *Rr* CooA in the Fe(III) *"locked-off"* state using two nitroxide spin labels: MTSL and MAL-6 (Figure 2.2). There are no structural data available for Fe(III) *"locked-off" Rr* CooA; consequently, this activity state has not been considered in prior allosteric activation models. We demonstrate that MAL-6 exhibits solution additive-dependent spectral changes that are nearly identical to those of MTSL when bound to CooA. Through this observation, we reveal that Fe(III) *"locked-off"* CooA is conformationally flexible and establish MAL-6 as a nitroxide label with a redox-stable linkage appropriate for use in comparing conformational dynamics under the reducing conditions necessary to probe the other activity states of CooA.

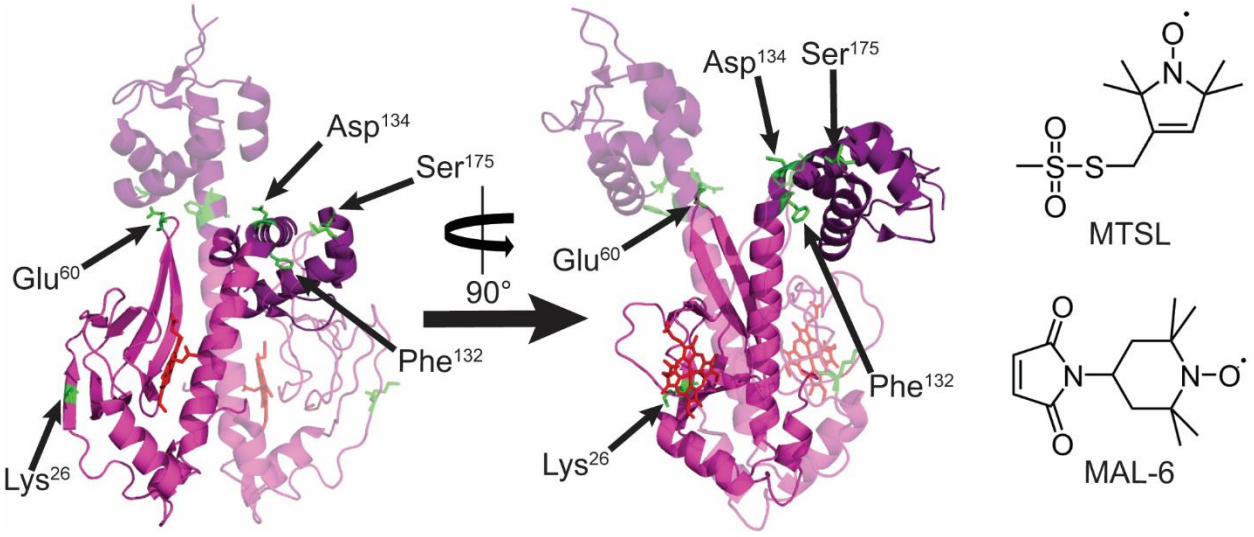


Figure 2.2 (Left) Crystal structure of WT *Rr* CooA (PDB 1FT9) in the inactive, Fe(II) "*ready-off*" state[15] highlighting the positions of Cys substitution sites. For clarity, one monomer is opaque and the other is translucent. The Cys substitution sites in the opaque monomer are labeled and shown in green. The effector binding domain is shown in magenta, DNA binding domain in dark purple, and heme cofactors in red. Images were created using PyMOL Molecular Graphics System (Version 1.3 Schrödinger, LLC.). (Right) Nitroxide spin labels used in this study.

2.2 Materials and Methods

Materials. MTSL (2,2,5,5-tetramethyl-1-oxyl-3-methyl methanethiosulfonate spin label) was purchased from Toronto Research Chemicals, and MAL-6 (*N*-(2,2,6,6-tetramethylpiperidin-4-yl-1-oxyl)maleimide) was purchased from Sigma Aldrich. All spin labels were used without further purification. CO(g) (≥ 99.0% purity) was obtained from Linde Gas LLC, and Ar(g) (100% purity) was obtained from Airgas. Gases were used as received. All other chemicals were obtained from Sigma Aldrich or Fisher Scientific.

Vector construction. Plasmids encoding site-specific Cys substitutions in *R. rubrum* cooA were created in a pEXT-20-based vector containing the gene for ΔCys$_4$CooA (C35A C80A C105Y C123I). The source vector for *ΔCys$_4$cooA* was created by Dr. Hwan Youn and was a gift from Prof. Gary Roberts, formerly of the UW-Madison Department of Bacteriology. The *ΔCys$_4$cooA* vector maintains coding for Cys^{75} that binds to the Fe(III) heme, but encodes no other Cys residues. The *ΔCys$_4$cooA* vector also bears a non-cleavable 6-His tag at the C-terminus of the protein. Five site-specific cysteine substitutions were created in this background: K26C, E60C, F132C, D134C and S175C. Appropriate site-specific mutations were made using a Stratagene QuickChange II Site-Directed Mutagenesis kit. Mutagenesis was confirmed by sequencing, and the resulting vectors encoding K26C-, E60C-, F132C-, D134C- and S175C-ΔCys$_4$CooA were maintained and propagated in *E. coli* DH5α.

Protein expression and purification. Expression of variant CooA proteins followed the protocol developed in the Roberts lab using the bacterial strain UQ2892 and low concentration IPTG induction to ensure maximal heme incorporation.[26,49] All cells were grown in Lennox-LB media supplemented with 20 mg/L ferric citrate and 3 g/L nutrient broth. Briefly, a dense 2 mL inoculum was added to 25 mL media and incubated for 6 h at 28°C. Cells were diluted to an OD_{550} of 0.06

and expression was induced with 7 μM IPTG, followed by growth for 20 h at 28°C. Following expression, cells were harvested by centrifugation and lysed in a French press. The lysate was cleared by centrifugation and added to a Ni-NTA column (Qiagen). Following batch adsorption, the column was washed with 10 mM imidazole, 50 mM MOPS (pH 7.5), 500 mM KCl, 0.5 mM DTT to remove non-specifically bound proteins. Bound CooA was eluted with the addition of 220 mM imidazole in 50 mM MOPS (pH 7.5), 500 mM NaCl. Eluted CooA protein was precipitated with 55% saturated ammonium sulfate to remove excess imidazole and to concentrate the protein. Precipitated CooA was re-dissolved in 100 mM MOPS (pH 7.4) buffer and bound to a Q-FF anion exchange column (GE Healthcare) 100 mM MOPS (pH 7.4) buffer. Protein was eluted off the column in buffer containing 100 mM MOPS (pH 7.4), 400 mM NaCl. Purity of the isolated protein was assessed by SDS-PAGE (13.5% (v/v) acrylamide). Isolated protein was stored in 100 mM MOPS (pH 7.4), 500 mM NaCl at -80°C until use.

Electronic Absorption Spectroscopy. All isolated proteins, ΔCys_4CooA (control), K26C-, E60C-, F132C-, D134- and S175C-ΔCys_4CooA, were characterized by visible spectroscopy to determine heme content and to confirm correct heme coordination. Total protein concentration was determined using the bicinchoninic acid (BCA) assay (Pierce BCA Protein Assay Kit). Heme concentration was determined using the pyridine hemochrome assay (Supporting Information).[50,51] Correct folding of the heme pocket was evaluated by recording visible spectra of the Fe(III), Fe(II) and Fe(II)–CO forms of each variant protein, and comparing them to WT (Supporting Information). All electronic absorption spectroscopy was performed on a Varian Cary 4 Bio spectrophotometer at room temperature.

Fluorescence Polarization Assay. The DNA binding affinity of CooA variants *in vitro* was measured using a fluorescence polarization assay as described previously (Supporting

Information).[52] The resulting binding curves were fit to a simple 1:1 binding model using nonlinear least squares analysis to obtain the dissociation constant, K_d (Supporting Information).[53] For each protein, K_d values were determined in triplicate, and the average value was reported.

Site-Directed Spin Labeling of CooA Variants. K26C-, E60C-, F132C-, D134C- and S175C-ΔCys_4CooA proteins were labeled by reacting the Fe(III) CooA variants with a 5- to 10-fold excess of spin label MTSL or MAL-6 (Figure 2.2). All labeling reactions were carried out in 100 mM MOPS, 500 mM NaCl, pH 7.4. For MTSL labeling, MTSL (10 mg/mL in ethanol) was added to 12.5-15 nmol CooA in 1250 – 5000 µL buffer and allowed to react at room temperature for 2-4 hours. For MAL-6 labeling, MAL-6 (1 mg/mL in acetonitrile) was added to 5 – 30 nmol CooA in 500 – 5000 µL buffer and allowed to react at 4°C for 16-24 hours. Labeled samples were dialyzed against 10 mM MOPS, 500 mM NaCl, pH 7.4 overnight to remove excess free label (Slide-A-lyzer, Thermo 20,000 MWCO). The resulting labeled CooA proteins were concentrated and buffer exchanged using a 30,000 MWCO spin concentrator (Amicon Ultra); the specific buffer employed for each sample is noted in the figure legends.

EPR Spectroscopy. Labeled variants were concentrated to 100-200 µM and loaded into capillary tubes (Drummond Scientific Company, 50 µL). EPR data were acquired on a Bruker Elexsys 500 system at 1.0 mW or 10.0 mW power, 100 G scan width, 2 G modulation amplitude and 100 kHz modulation frequency. The number of scans ranged from 10-100 based on protein concentration.

2.3 Results

Heme characteristics of ΔCys_4CooA and Cys substitution variants

Because CooA contains four potentially reactive Cys residues, the variant ΔCys_4CooA lacking these Cys residues was created to facilitate site-specific spin labeling. This protein contains only

one of the five Cys residues present in WT CooA: Cys^{75} is a ligand to the heme iron in the Fe(III) state and is important for CooA function. Cys^{75} is bound to the heme iron and therefore will not react with the spin label probe in the Fe(III) state. The ΔCys_4CooA variant (C35A, C80A, C105Y, C123I) was generated by random mutagenesis and selection in a cell-based activity screen.[52] Five single-point Cys substitution variants were created within the ΔCys_4CooA background: K26C-, E60C-, F132C-, D134C- and S175C-ΔCys_4CooA. Hereafter we will refer to these Cys substitution variants as K26C, E60C, F132C, D134C and S175C, although all bear the four additional variations of ΔCys_4CooA. Each Cys substitution serves to represent a key functional region within ΔCys_4CooA: the heme-containing effector domain (K26C), the top of the 4/5 loop (E60C), the hinge (F132C and D134C) and the DNA-binding domain (S175C) (Figure 2.2).

The heme reactivities of ΔCys_4CooA and Cys substitution variants appear unchanged from WT CooA. The positions and band shapes of the heme absorption features of all variants in the Fe(III), Fe(II), and Fe(II)–CO states were essentially indistinguishable from those of WT CooA (Table 2.S1, Figures 2.S1 – 2.S6). This observation suggests that neither the removal of the four native Cys residues nor substitution of Cys residues at the five locations listed results in significant perturbation of the heme pocket and that the redox-mediated ligand switching and small molecule sensing functions of CooA are maintained in all variants.

DNA binding activity of ΔCys_4CooA and Cys substitution variants

Fluorescence polarization data revealed that all CooA cysteine variants, with the exception of F132C, exhibit DNA binding affinities similar to that of WT CooA (Table 2.1 and Figures 2.S7 – 2.S12). The fluorescence polarization assay monitors changes in anisotropy of Texas Red-labeled, double-stranded P_{cooF} oligomer as a function of protein concentration.[53,54] (P_{cooF} is the 26-base pair-long CooA consensus recognition sequence.) Importantly, all variants exhibited negligible

Table 2.1 DNA binding affinities expressed as dissociation equilibrium constants for Fe(II)–CO WT and ΔCys_4CooA variants as determined by *in vitro* fluorescence polarization assay. Values represent the average of three individual experiments plus or minus one standard deviation.

CooA Variant	K_d **(nM)**
WT	6.0 ± 0.05
ΔCys_4	3.2 ± 1.6
K26C	3.7 ± 0.6
E60C	22.9 ± 1.1
F132C	2,790 ± 210
D134C	10.2 ± 2.2
S175C	47.1 ± 3.1

concentration-dependent changes in anisotropy in the Fe(III) and Fe(II) states (data not shown). This observation demonstrates that none of the Cys substitutions results in constitutive DNA binding activity. For each variant, affinity values (K_d) for the protein-DNA binding equilibrium were measured in the Fe(II)–CO state. The DNA binding affinities for variants ΔCys_4CooA and K26C were slightly greater than that of WT CooA, while variants E60C and S175C exhibited affinities that were slightly lower. The two variants located in the hinge region, F132C and D134C, exhibited dramatically different DNA binding affinities relative to one another. The affinity of F132C was two orders of magnitude lower than that of WT CooA, while the affinity of D134C was nearly indistinguishable from that of WT CooA. This discrepancy strongly suggests that Phe^{132} plays a role in propagating the allosteric CO binding signal from the effector binding domain to the DNA-binding domain as discussed below.

EPR spectra of MTSL-labeled Fe(III) Cys substitution variants

EPR spectra of MTSL-labeled Cys substitution variants in the Fe(III) *"locked-off"* state exhibited multicomponent signals (Figure 2.3). In EPR spectra of spin-labeled proteins, a more mobile spin label population is represented by a narrower signal distribution while a less mobile spin label population is represented by a broader signal distribution. Multicomponent SDSL-EPR signals (derived from labels with differing mobilities) may arise from the presence of 1) protein

rotary diffusion on the EPR timescale, 2) multiple nitroxide spin label rotamers, and/or 3) multiple protein conformational states.[55] To account for rotary diffusion and nitroxide label rotameric sampling, two confounding effects irrelevant to studying conformational dynamics, we assessed perturbations to EPR spectra in the presence of solution additives. Two solution additives were employed: sucrose, a protecting osmolyte that favors more compact protein conformations, and Ficoll-70, a water-soluble polymer that alters solution viscosity without altering protein structure

due to osmolytic effects.[45,55-57] Extensive SDSL-EPR studies have investigated the effects of these solution additives in several MTSL-labeled soluble proteins. These studies found that 1) effects due to protein rotary diffusion are minimized by increasing solvent viscosity, 2) the population distribution of multiple spin label rotamers is insensitive to the presence of an osmolyte, and 3) the population distribution of multiple protein conformational states is sensitive to the presence of an osmolyte.[55]

To understand the contribution of protein rotary diffusion to multicomponent EPR spectra in CooA, we compared SDSL-EPR spectra of MTSL-labeled Fe(III) Cys substitution variants in the presence and absence of 25% (w/w) Ficoll-70 (Figure 2.3, left). K26C and D134C showed a small increase in the population of the less mobile component upon addition of Ficoll-70, while F132C and S175C showed a more pronounced increase in the population of the less mobile component upon addition of Ficoll-70. In contrast, E60C exhibited a large increase in the more mobile component upon addition of Ficoll-70. We attribute Ficoll-70-induced changes in signal distribution to the suppression of effects due to protein rotary diffusion: the increase in solution viscosity imparted by Ficoll-70 slows protein rotary diffusion enough to eliminate the contribution of this effect to the EPR spectrum.

Multicomponent EPR spectra of all MTSL-labeled Fe(III) CooA variants in this study report on multiple protein conformations. To identify SDSL-EPR spectral features that report on protein conformational dynamics, we compared the SDSL-EPR signal distribution of Fe(III) CooA Cys substitution variants in buffer with 25% (w/w) Ficoll-70 and in buffer with 30% (w/w) sucrose (Figure 2.4, left).These two buffers have nearly identical viscosities, and Ficoll-70 is known to interact minimally with MTSL-bound protein;[45,55] as a result, differences in the distribution of

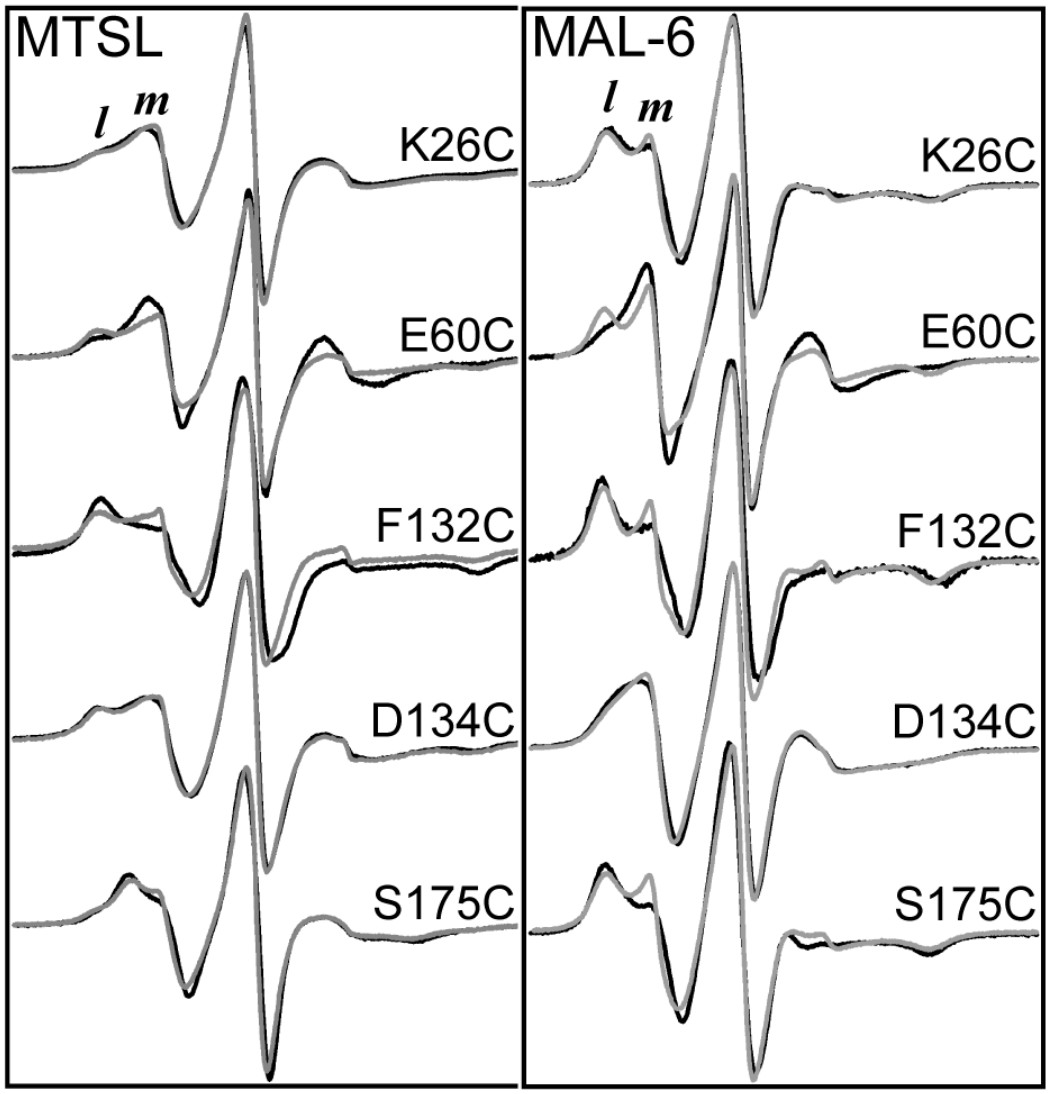


Figure 2.3 EPR spectra of MTSL- and MAL6-labeled Fe(III) ΔCys$_4$CooA variants in 100 mM MOPS, 500 mM NaCl, pH 7.4 (grey) and 100 mM MOPS, 500 mM NaCl, pH 7.4 with 25% (w/w) Ficoll-70 (black). The spectra are scaled to equivalent central line amplitude. More mobile (***m***) and less mobile (***l***) components of the $m_I = +1$ line are labeled for clarity.

signals in multicomponent EPR spectra arise from the presence or absence of the protecting osmotlye sucrose. For all MTSL-labeled Fe(III) CooA variants, addition of sucrose increased the intensity of the less mobile EPR signal with a concomitant decrease in the intensity of the more mobile EPR signal. This sucrose-dependent signal response demonstrates that the multicomponent EPR spectra observed in MTSL-labeled CooA variants report on multiple protein conformations. Furthermore, these results demonstrate that Fe(III) CooA exhibits conformational heterogeneity in the heme-binding domain, 4/5 loop, hinge region, and DNA-binding domain.

Comparison of EPR spectra between MTSL- and MAL-6-labeled Fe(III) Cys substitution variants

The relative population of EPR signal components and response to solution additives in MAL-6-labeled Fe(III) CooA variants were similar to those of MTSL-labeled variants. EPR spectral response to Ficoll-70-induced increase in solution viscosity in MAL-6-labeled samples parallels that observed in MTSL-labeled samples: K26C, F132C, D134C, and S175C exhibited an increase in the population of the less mobile component upon addition of Ficoll-70, while E60C exhibited an increase in the more mobile component upon addition of Ficoll-70 (Figure 2.3, right). Like MTSL-labeled samples, all MAL-6-labeled Fe(III) CooA variants exhibited increased intensity of the less mobile EPR signal with a concomitant decrease in the intensity of the more mobile EPR signal in the presence of sucrose (Figure 2.4, right). Changes in the distribution of more and less mobile EPR components in response to the presence of solution additives were somewhat more pronounced for MAL-6-labeled samples than for MTSL-labeled samples. The differences in chemical structure between MTSL and MAL-6 likely give rise to the observed subtle changes in signal behavior; however, both labels exhibit sucrose-dependent changes in EPR signal intensity in Fe(III) CooA variants, demonstrating that 1) that both labels report on conformational

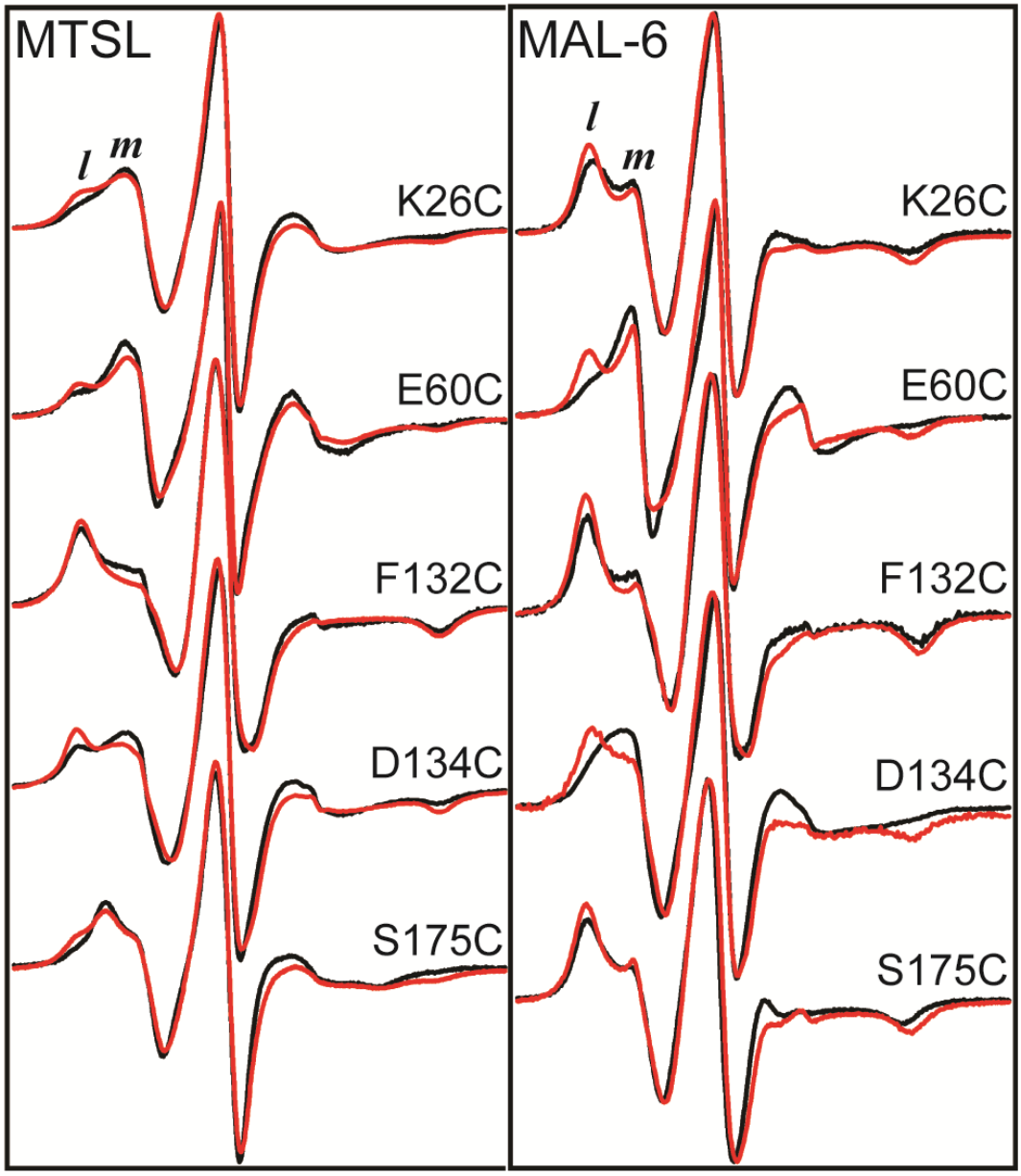


Figure 2.4 EPR spectra of MTSL- and MAL-6-labeled Fe(III) ΔCys$_4$CooA variants in 100 mM MOPS, 500 mM NaCl, pH 7.4 with 30% (w/w) sucrose (red) and 100 mM MOPS, 500 mM NaCl, pH 7.4 with 25% (w/w) Ficoll-70 (black). The spectra are scaled to equivalent central line amplitude. More mobile (***m***) and less mobile (***l***) components of the m_I = +1 line are labeled for clarity.

dynamics and 2) that Fe(III) CooA exhibits global conformational flexibility on the μs-ms timescale.

2.4 Discussion

EPR spectra of MTSL- and MAL-6-labeled Cys substitution variants reveal conformational heterogeneity in five functionally-important regions of CooA in the Fe(III) *"locked-off"* state. Previous crystallographic and SAXS studies provided structural information about CooA in the Fe(II) *"ready-off"* and CO-bound *"on"* states;[15,24,27] however, no data exist that describe the global structure of CooA in the Fe(III) *"locked-off"* state. Additionally, limited data exist that describe protein dynamics in CooA. In this study, we observed multicomponent EPR spectra that exhibited osmolyte-dependent changes in signal distribution for five MTSL- and MAL-6-labeled CooA Cys substitution variants in the Fe(III) *"locked-off"* state. These data demonstrate that the effector-binding domain (represented by K26C), 4/5 loop (represented by E60C), hinge region (represented by F132C and D134C), and DNA-binding domain (represented by S175C) sample multiple protein-derived conformational states on the μs-ms timescale in Fe(III) CooA. We cannot currently say whether these protein-derived conformational states are confined to the local protein environment of each spin label or are indicative of a global conformational change; however, it is unequivocal that CooA exhibits conformational flexibility on the slow (μs-ms) timescale.

Slow dynamics in Fe(III) *"locked-off"* CooA differ from those of the analogous apo-CRP, suggesting that there may be differences in the allosteric activation mechanism in these structural homologues. As demonstrated in the above SDSL-EPR experiments, several functional regions of Fe(III) CooA exhibit conformational flexibility on the μs-ms timescale. In contrast, apo-CRP does not exhibit conformational exchange, as evidenced by a single set of narrow resonances observed

using solution NMR.[6,13] In CRP, anti-cooperative binding of cAMP effector molecules activates protein motions on the μs-ms timescale and induces a coil-to-helix transition that repositions the DNA-binding domains.[11-13] We plan to utilize SDSL-EPR as a tool to probe how conformational flexibility in CooA changes upon activation of the protein by reduction and cooperative CO binding at the heme.

Phe[132] plays a critical role in allosteric communication between the CooA effector binding and DNA-binding domains. A functional role for Phe[132] was previously inferred from the Fe(II) *"ready-off"* crystal structure of *Rr* CooA.[15] In the crystal structure, the aromatic ring of Phe[132] directly interacts with the amide backbone of Thr[182] in the DNA recognition helix. Additionally, Phe is highly conserved in this position for members of the CRP/FNR superfamily. Based on these observations, it was proposed that Phe[132] is involved in propagating the CO binding signal from the regulatory domain to the DNA binding domain. Despite this initial prediction, no current model for propagation of the CO binding signal suggests a role for Phe[132].[28,31] We observe that, in the variant F132C, a Phe-to-Cys mutation abolishes CO-dependent DNA binding activity, while in D134C an Asp-to-Cys mutation at residue 134 has minimal effect on DNA binding activity. Introduction of the F132C variation disrupts allosteric communication between the heme and the DNA-binding domain; CO binding to the heme appears normal, but the CO-bound protein does not bind to its DNA target, confirming that Phe[132] makes a *specific* allosteric contact that is essential for signal propagation across CooA.

Importantly, this study establishes MAL-6 as a spin label with a redox-stable, covalent linkage that reports on protein conformational dynamics in CooA. Solution additives have been employed to extract conformational dynamics information from multicomponent EPR spectra of soluble, MTSL-labeled proteins;[45,55] however, MTSL is covalently attached to proteins through a

redox-sensitive, cysteine-linked disulfide bond.[40-42] As a result, the label is lost under reducing conditions, making the study of MTSL-labeled CooA in the Fe(II) *"ready-off"* and Fe(II)–CO *"on"* states infeasible. An important outcome of this study was the observation of similar EPR signal distributions and solution additive responses in Cys substitution variants labeled with MTSL and MAL-6. These observations allow us to employ solution additives to extract conformational dynamics information from multicomponent EPR spectra of MAL-6-labeled CooA containing a redox-stable linkage between the protein and spin label.

Local protein dynamics in CooA change as a function of protein activity state. We have recently developed a labeling strategy that has enabled us to compare conformational dynamics in Fe(II)–CO *"on"*, and Fe(II)–CO *"DNA-bound"* CooA (see Appendix 2, Section 2.8). Activation of CooA upon reduction and CO binding leads to a redistribution of more and less mobile components in nearly all variants, suggesting that local conformational dynamics throughout the protein are altered upon changes in heme oxidation state and ligation. The largest change in signal was observed in D134C, where there was a significant increase in the fraction of the more mobile component upon reduction and CO binding. Remarkably, upon mixing Fe(II)-CO D134C to the CooA P_{cooF} promoter site, the broader EPR signal almost completely disappears, indicative a redistribution of mobile populations from an even mixture of more and less mobile to almost entirely more mobile. This observation may be indicative of relaxation of the hinge region upon DNA binding, which may occur in order to compensate for motional restriction in the DNA binding domain. Further investigation of the thermodynamics driving DNA binding, using techniques such as isothermal titration calorimetry, will be required in order to connect local changes in protein dynamics with the global allosteric activation model.

2.5 Conclusion

EPR spectra of five MTSL- and MAL-6-labeled Cys substitution variants report on multiple protein conformations in Fe(III) *"locked-off" Rr* CooA. These observations suggest that CooA exhibits global conformational flexibility in the Fe(III) state, in direct contrast with the structural homologue CRP, which is rigid on the μs-ms timescale in the effector-free, off state. Given this difference in conformational flexibility, it is possible that the allosteric activation mechanism of CooA differs from that of CRP. Fluorescence polarization and electronic absorption studies of F132C show that this variant binds CO identically to WT, but is not activated by CO binding; these data demonstrate that the allosteric CO binding signal, which activates DNA binding, is propagated by Phe^{132} in the hinge region. Most importantly, the spin labels MTSL and MAL-6 report on similar dynamics and conformational behaviors of Fe(III) CooA. Because the protein-MAL-6 linkage is redox stable (unlike MTSL), MAL-6 labeling of CooA in the Fe(II) *"ready-off"*, Fe(II)–CO *"on"*, and Fe(II)–CO *"DNA-bound"* states is feasible. Comparison of dynamics between these functional states reveals significant modulation of protein dynamics in the hinge region; however, the influence of these local changes in protein dynamics on the global allosteric activation mechanism is not yet understood.

2.6 References

1. Körner, H.; Sofia, H. J.; Zumft, W. G. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* **2003**, *27*, 559.
2. Shelver, D.; Kerby, R. L.; He, Y.; Roberts, G. P. Carbon monoxide-induced activation of gene expression in *Rhodospirillum rubrum* requires the product of *cooA*, a member of the cyclic AMP receptor protein family of transcriptional regulators. *J. Bacteriol.* **1995**, *177*, 2157.
3. Aono, S.; Nakajima, H.; Saito, K.; Okada, M. A novel heme protein that acts as a carbon monoxide-dependent transcriptional activator in *Rhodospirillum rubrum*. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 752.
4. Shelver, D.; Kerby, R. L.; He, Y.; Roberts, G. P. CooA, a CO-sensing transcription factor from *Rhodospirillum rubrum*, is a CO-binding heme protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 11216.
5. Aono, S.; Ohkubo, K.; Matsuo, T.; Nakajima, H. Redox-controlled ligand exchange of the heme in the CO-sensing transcriptional activator CooA. *J. Biol. Chem.* **1998**, *273*, 25757.
6. Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.
7. Tzeng, S.-R.; Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **2009**, *462*, 368.
8. Tzeng, S.-R.; Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **2012**, *488*, 236.
9. Tzeng, S.-R.; Kalodimos, C. G. The role of slow and fast protein motions in allosteric interactions. *Biophys. Rev.* **2015**, *7*, 251.

10. Tzeng, S.-R.; Kalodimos, C. G. Allosteric inhibition through suppression of transient conformational states. *Nat. Chem. Biol.* **2013**, *9*, 462.

11. Harman, J. G. Allosteric regulation of the cAMP receptor protein. *Biochim. Biophys. Acta* **2001**, *1547*, 1.

12. Passner, J. M.; Schultz, S. C.; Steitz, T. A. Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1Å resolution. *J. Mol. Biol.* **2000**, *304*, 847.

13. Popovych, N.; Tzeng, S.-R.; Tonelli, M.; Ebright, R. H.; Kalodimos, C. G. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6927.

14. Cooper, A.; Dryden, D. T. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **1984**, *11*, 103.

15. Lanzilotta, W. N.; Schuller, D. J.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure of the CO sensing transcription activator CooA. *Nat. Struct. Biol.* **2000**, *7*, 876.

16. Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Chung, S. Y.; Roberts, G. P.; Reynolds, M. F.; Parks, R. B.; Burstyn, J. N. Identification of two important heme site residues (cysteine 75 and histidine 77) in CooA, the CO-sensing transcription factor of *Rhodospirillum rubrum*. *Biochemistry* **1999**, *38*, 2669.

17. Yamamoto, K.; Ishikawa, H.; Takahashi, S.; Ishimori, K.; Morishima, I.; Nakajima, H.; Aono, S. Binding of CO at the Pro^2 side is crucial for the activation of CO-sensing transcriptional activator CooA. ^{1}H NMR spectroscopic studies. *J. Biol. Chem.* **2001**, *276*, 11473.

18. Nakajima, H.; Nakagawa, E.; Kobayashi, K.; Tagawa, S.; Aono, S. Ligand-switching intermediates for the CO-sensing transcriptional activator CooA measured by pulse radiolysis. *J. Biol. Chem.* **2001**, *276*, 37895.

19. Youn, H.; Kerby, R. L.; Thorsteinsson, M. V.; Clark, R. W.; Burstyn, J. N.; Roberts, G. P. Analysis of the L116K variant of CooA, the heme-containing CO sensor, suggests the presence of an unusual heme ligand resulting in novel activity. *J. Biol. Chem.* **2002**, *277*, 33616.

20. Puranik, M.; Nielsen, S. B.; Youn, H.; Hvitved, A. N.; Bourassa, J. L.; Case, M. A.; Tengroth, C.; Balakrishnan, G.; Thorsteinsson, M. V.; Groves, J. T.; McLendon, G. L.; Roberts, G. P.; Olson, J. S.; Spiro, T. G. Dynamics of carbon monoxide binding to CooA. *J. Biol. Chem.* **2004**, *279*, 21096.

21. Clark, R. W.; Youn, H.; Parks, R. B.; Cherney, M. M.; Roberts, G. P.; Burstyn, J. N. Investigation of the role of the N-terminal proline, the distal heme ligand in the CO sensor CooA. *Biochemistry* **2004**, *43*, 14149.

22. Youn, H.; Conrad, M.; Chung, S. Y.; Roberts, G. P. Roles of the heme and heme ligands in the activation of CooA, the CO-sensing transcriptional activator. *Biochem. Biophys. Res. Commun.* **2006**, *348*, 345.

23. Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P. Altering the specificity of CooA, the carbon monoxide-sensing transcriptional activator: characterization of CooA variants that bind cyanide in the Fe(II) form with high affinity. *Biochemistry* **2000**, *39*, 8284.

24. Borjigin, M.; Li, H.; Lanz, N. D.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure-based hypothesis on the activation of the CO-sensing transcription factor CooA. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2007**, *63*, 282.

25. Kerby, R. L.; Youn, H.; Thorsteinsson, M. V.; Roberts, G. P. Repositioning about the dimer interface of the transcription regulator CooA: A major signal transduction pathway between the effector and DNA-binding domains. *J. Mol. Biol.* **2003**, *325*, 809.

26. Youn, H.; Kerby, R. L.; Conrad, M.; Roberts, G. P. Functionally critical elements of CooA-related CO sensors. *J. Bacteriol.* **2004**, *186*, 1320.

27. Akiyama, S.; Fujisawa, T.; Ishimori, K.; Morishima, I.; Aono, S. Activation mechanisms of transcriptional regulator CooA revealed by small-angle X-ray scattering. *J. Mol. Biol.* **2004**, *341*, 651.

28. Tripathi, S.; Poulos, T. L. Testing the N-terminal velcro model of CooA carbon monoxide activation. *Biochemistry* **2018**.

29. Clark, R. W.; Lanz, N. D.; Lee, A. J.; Kerby, R. L.; Roberts, G. P.; Burstyn, J. N. Unexpected NO-dependent DNA binding by the CooA homolog from Carboxydothermus hydrogenoformans. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 891.

30. Youn, H.; Kerby, R. L.; Roberts, G. P. The role of the hydrophobic distal heme pocket of CooA in ligand sensing and response. *J. Biol. Chem.* **2003**, *278*, 2333.

31. Otomo, A.; Ishikawa, H.; Mizuno, M.; Kimura, T.; Kubo, M.; Shiro, Y.; Aono, S.; Mizutani, Y. A study of the dynamics of the heme pocket and C-helix in CooA upon CO dissociation using time-resolved visible and UV resonance Raman spectroscopy. *J. Phys. Chem. B* **2016**, *120*, 7836.

32. Rubtsov, I. V.; Zhang, T.; Nakajima, H.; Aono, S.; Rubtsov, G. I.; Kumazaki, S.; Yoshihara, K. Conformational dynamics of the transcriptional regulator CooA protein studied by subpicosecond mid-infrared vibrational spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 10056.

33. Boulton, S.; Melacini, G. Advances in NMR methods to map allosteric sites: From models to translation. *Chem. Rev.* **2016**, *116*, 6267.

34. Zhuravleva, A.; Gierasch, L. M. Substrate-binding domain conformational dynamics mediate Hsp70 allostery. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2865.

35. Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18249.

36. Campanello, G. C.; Ma, Z.; Grossoehme, N. E.; Guerra, A. J.; Ward, B. P.; DiMarchi, R. D.; Ye, Y.; Dann, C. E.; Giedroc, D. P. Allosteric inhibition of a zinc-sensing transcriptional repressor: Insights into the arsenic repressor (ArsR) family. *J. Mol. Biol.* **2013**, *425*, 1143.

37. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448*, 325.

38. Moorman, V. R.; Valentine, K. G.; Bédard, S.; Kasinath, V.; Dogan, J.; Love, F. M.; Wand, A. J. Dynamic and thermodynamic response of the Ras protein Cdc42Hs upon association with the effector domain of PAK3. *J. Mol. Biol.* **2014**, *426*, 3520.

39. Capdevila, D. A.; Braymer, J. J.; Edmonds, K. A.; Wu, H.; Giedroc, D. P. Entropy redistribution controls allostery in a metalloregulatory protein. *Proc Natl Acad Sci U S A* **2017**, *114*, 4424.

40. Hubbell, W. L.; McHaourab, H. S.; Altenbach, C.; Lietzow, M. A. Watching proteins move using site-directed spin labeling. *Structure* **1996**, *4*, 779.

41. Hubbell, W. L.; Cafiso, D. S.; Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nat. Struct. Mol. Biol.* **2000**, *7*, 735.

42. Klug, C. S.; Feix, J. B. In *Biophysical Tools for Biologists: In Vitro Techniques, Vol. 1*; Correia, J. J., Detrich, H. W., Eds.; Academic Press: Burlington, 2008, p 617.

43. Longhi, S.; Belle, V.; Fournel, A.; Guigliarelli, B.; Carrière, F. Probing structural transitions in both structured and disordered proteins using site-directed spin-labeling EPR spectroscopy. *J. Pept. Sci.* **2011**, *17*, 315.

44. Pirman, N. L.; Milshteyn, E.; Galiano, L.; Hewlett, J. C.; Fanucci, G. E. Characterization of the disordered-to-α-helical transition of IA3 by SDSL-EPR spectroscopy. *Protein Sci.* **2011**, *20*, 150.

45. López, C. J.; Oga, S.; Hubbell, W. L. Mapping molecular flexibility of proteins with ste-directed spin labeling: A case study of myoglobin. *Biochemistry* **2012**, *51*, 6568.

46. Cafiso, D. S. Identifying and quantitating conformational exchange in membrane proteins using site-directed spin labeling. *Acc. Chem. Res.* **2014**, *47*, 3102.

47. Gaffney, B. J. Connecting lipoxygenase function to structure by electron paramagnetic resonance. *Acc. Chem. Res.* **2014**, *47*, 3588.

48. Bradshaw, M. D.; Gaffney, B. J. Fluctuations of an exposed pi-helix involved in lipoxygenase substrate recognition. *Biochemistry* **2014**, *53*, 5102.

49. Stewart, V.; Lu, Y.; Darwin, A. J. Periplasmic nitrate reductase (NapABC Enzyme) supports anaerobic respiration by *Escherichia coli* K-12. *J. Bacteriol.* **2002**, *184*, 1314.

50. Antonini, E.; Brunori, M. *Hemoglobin and myoglobin in their reactions with ligands*; North-Holland Publishing Company: Amsterdam, 1971.

51. De Duve, C. A spectrophotometric method for the simultaneous determination of myoglobin and hemoglobin in extracts of human muscle. *Acta Chem. Scand.* **1948**, *2*, 264.

52. He, Y.; Gaal, T.; Karls, R.; Donohue, T. J.; Gourse, R. L.; Roberts, G. P. Transcription activation by CooA, the CO-sensing factor from *Rhodospirillum rubrum*. The interaction between CooA and the C-terminal domain of the a subunit of RNA polymerase. *J. Biol. Chem.* **1999**, *274*, 10840.

53. Lundblad, J. R.; Laurance, M.; Goodman, R. H. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol. Endocrinol.* **1996**, *10*, 607.

54. Thorsteinsson, M. V.; Kerby, R. L.; Conrad, M.; Youn, H.; Staples, C. R.; Lanzilotta, W. N.; Poulos, T. J.; Serate, J.; Roberts, G. P. Characterization of variants altered at the N-terminal proline, a novel heme-axial ligand in CooA, the CO-sensing transcriptional activator. *J. Biol. Chem.* **2000**, *275*, 39332.

55. López, C. J.; Fleissner, M. R.; Guo, Z.; Kusnetzow, A. K.; Hubbell, W. L. Osmolyte perturbation reveals conformational equilibria in spin-labeled proteins. *Protein Sci.* **2009**, *18*, 1637.

56. Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **1981**, *256*, 7193.

57. Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375.

2.7 Supplementary Information

Electronic Absorption

Pyridine hemochrome assay: Purified protein (5 µL) was diluted to 100 µL total volume with 4.3 M pyridine, 0.10 M NaOH, and mixed with 2-5 mg of sodium dithionite. The absorbance values at 557 and 525 nm were recorded of the resulting Fe(II)PPIX$(py)_2$, and the heme concentration calculated using ε_{557} = 32 $mM^{-1}cm^{-1}$ and ε_{525} = 16 $mM^{-1}cm^{-1}$.

Reduction and CO binding: Fe(III) CooA is the resting state of the protein: spectra were recorded of 5 µM as-isolated protein in 100 mM MOPS, 500 mM NaCl, pH 7.4 with a total volume of 300 µL. Fe(II) CooA was prepared by reduction of Fe(III) CooA with sodium dithionite. In a septum-sealed cuvette, Fe(III) CooA was anaerobically reduced under argon with 10 µL, 30 mM sodium dithionite in 100 mM MOPS, 500 mM NaCl, pH 7.4 and incubated at room temperature for 20 minutes to ensure complete conversion. Fe(II)–CO adducts were prepared by anaerobic addition of 100 µL of CO(*g*) to Fe(II) CooA. Conversion to the CO adduct was complete after several minutes.

Fluorescence Polarization Assay

Briefly, increasing amounts of a CooA variant were incubated with double-stranded DNA oligomer containing a 5’ Texas Red fluorescent label and the consensus CooA recognition sequence P_{cooF} (26 base-pair complementary duplex, 5′-ATAACTGTCATCTGGCCGACAGACGG-3′). When CooA binds to P_{cooF}, there is a change in the rotational correlation time of the fluorophore-bound oligomer, which is observed as a change in anisotropy. Wild-type CooA only binds to P_{cooF} in the Fe(II)–CO state: Fe(III)- and Fe(II)- CooA do not bind to DNA. Anisotropy values were recorded for all CooA variants in the Fe(III)-, Fe(II)-

and Fe(II)–CO states to determine whether any variants exhibited abnormal DNA binding behavior. The necessary range of CooA concentrations (0.4 – 3500 nM CooA dimer for the F132C variant, 1 – 2000 nM CooA dimer for all other proteins) were achieved through serial dilution of 50 µM CooA stock. All protein samples were assayed for P_{cooF} binding in 40 mM Tris·HCl (pH 8.0), 50 mM KCl, 5% (v/v) glycerol, 6.0 mM $CaCl_2$, 5.0 mM DTT and 6.4 nM P_{cooF}. The anisotropy values were recorded for 500 µL samples in glass culture tubes (6×50 mm dimensions, VWR). Fe(III) CooA was reduced by capping each tube with a septum, purging the headspace with Ar(*g*) for 10 minutes, and anaerobically adding 10 µL of a 100 mM dithionite solution in 40 mM Tris·HCl (pH 8.0), 50 mM KCl, 5% (v/v) glycerol. After 10 minutes to ensure complete reduction, Fe(II)–CO adducts were obtained by adding 100 µL CO(*g*) to each tube. After 10 minutes to ensure complete conversion to the Fe(II)–CO species, the anisotropy value was recorded for each sample. The resulting binding curve was fit to a simple 1:1 binding model using nonlinear least squares analysis:

$$\boldsymbol{A = A_f + (A_b - A_f) \times \left[\frac{1 + K_a[L_T] + K_a[R_T] - \sqrt{\{(1 + K_a[L_T] + K_a[R_T])^2 - 4[L_T][R_T]K_a^2\}}}{2K_a[R_T]} \right]}$$

where A is the observed fluorescence anisotropy, A_f is the fluorescence anisotropy in the absence of binding, A_b is the maximum fluorescence anisotropy when all DNA is bound to protein, K_a is the association equilibrium constant, L_T is the total concentration of protein, and R_T is the total concentration of DNA. Table S2 summarizes fitted values for A_f, A_b, and K_a for each variant. The association equilibrium constant, K_a is related to the dissociation equilibrium constant, K_d by:

$$\boldsymbol{K_d = \frac{1}{K_a}}$$

Table 2.S1 Compiled electronic absorption peak positions for ΔCys_4CooA variants.

CooA	Fe(III)				Fe(II)			Fe(II)−CO		
variant	δ	Soret	β	α	Soret	β	α	Soret	β	α
WT	360	424	541	569	425	529	559	423	538	568
ΔCys_4	362	424	538	568	425	529	558	422	540	568
K26C	360	424	539	570	426	529	559	423	539	568
E60C	361	422	538	571	425	528	559	422	539	567
F132C	361	424	541	569	425	529	559	423	541	569
D134C	361	424	540	568	425	529	559	422	540	568
S175C	360	424	540	569	425	529	559	422	540	569

Table 2.S2 Average fluorescence polarization fitting parameters for WT CooA and ΔCys_4CooA variants relative standard deviations (RSDs) expressed as a percentage of each average value.

CooA Variant	A_f (*RSD*)	A_b (*RSD*)	K_a (nM) (*RSD*)
WT	0.166 (*0.86*)	0.214 (*0.39*)	0.175 (*24*)
ΔCys_4	0.170 (*0.98*)	0.221 (*0.26*)	0.379 (*38*)
K26C	0.173 (*0.70*)	0.221 (*0.068*)	0.278 (*17*)
E60C	0.169 (*0.47*)	0.224 (*0.72*)	0.0422 (*9.8*)
F132C	0.173 (*0.55*)	0.274 (*4.2*)	0.000372 (*4.2*)
D134C	0.170 (*1.7*)	0.227 (*0.44*)	0.103 *(20*)
S175C	0.166 (*1.8*)	0.216 (*1.3*)	0.0213 *(6.4*)

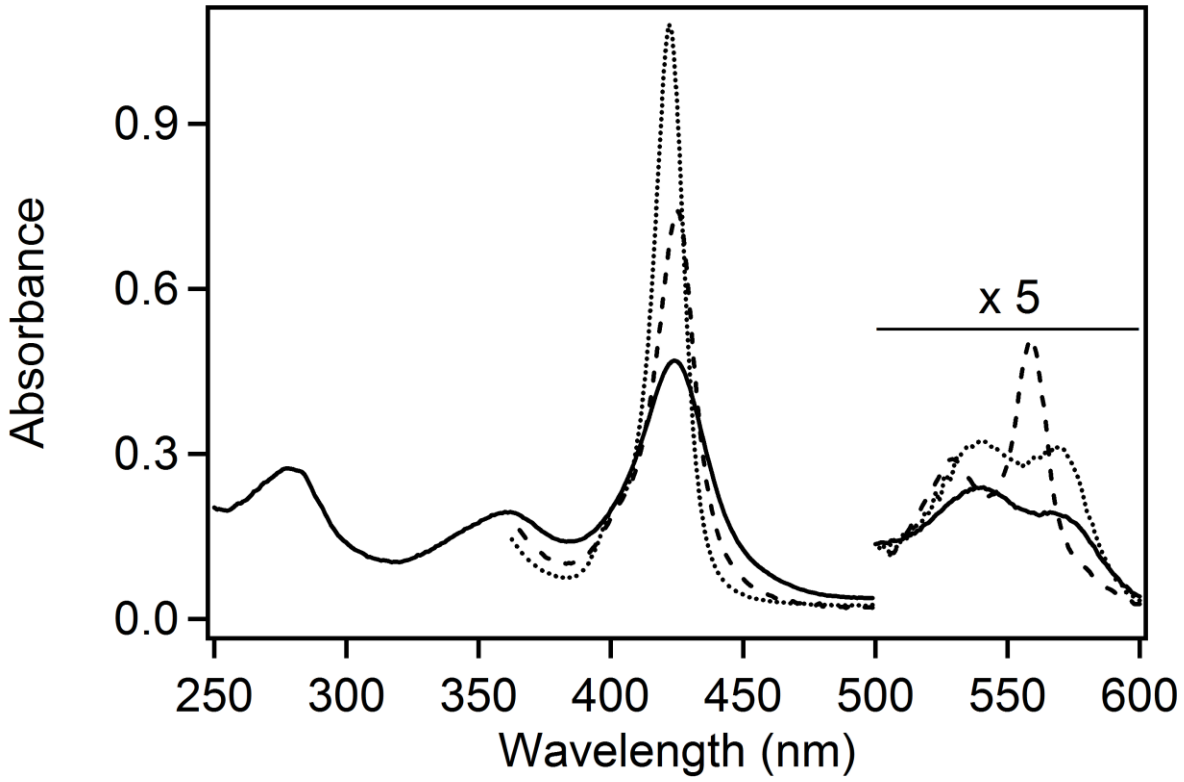


Figure 2.S1 Electronic absorption spectra of ΔCys$_4$CooA (2.7 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 100 mM MOPS, pH 7.4, 500 mM NaCl. Reduction and CO binding were carried out as described in the methods section.

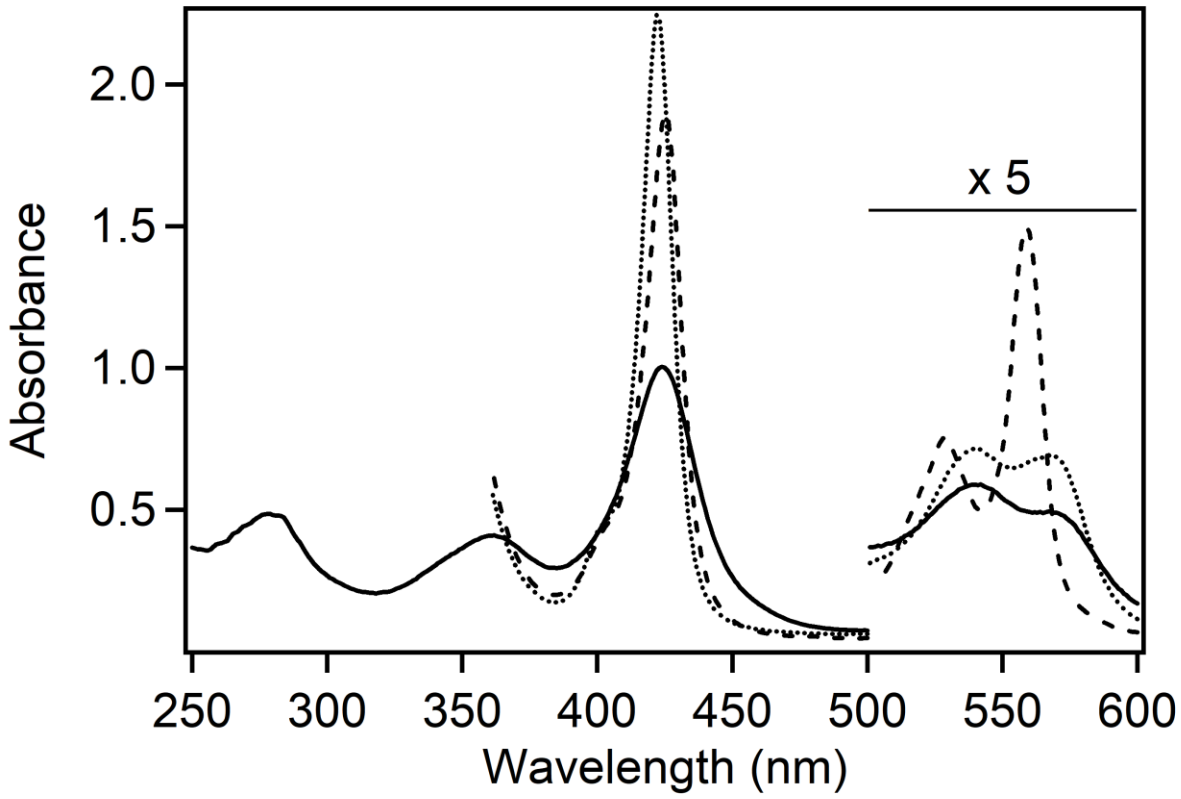


Figure 2.S2 Electronic absorption spectra of K26C CooA (12.8 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 100 mM MOPS, pH 7.4, 500 mM NaCl. Reduction and CO binding were carried out as described in the methods section.

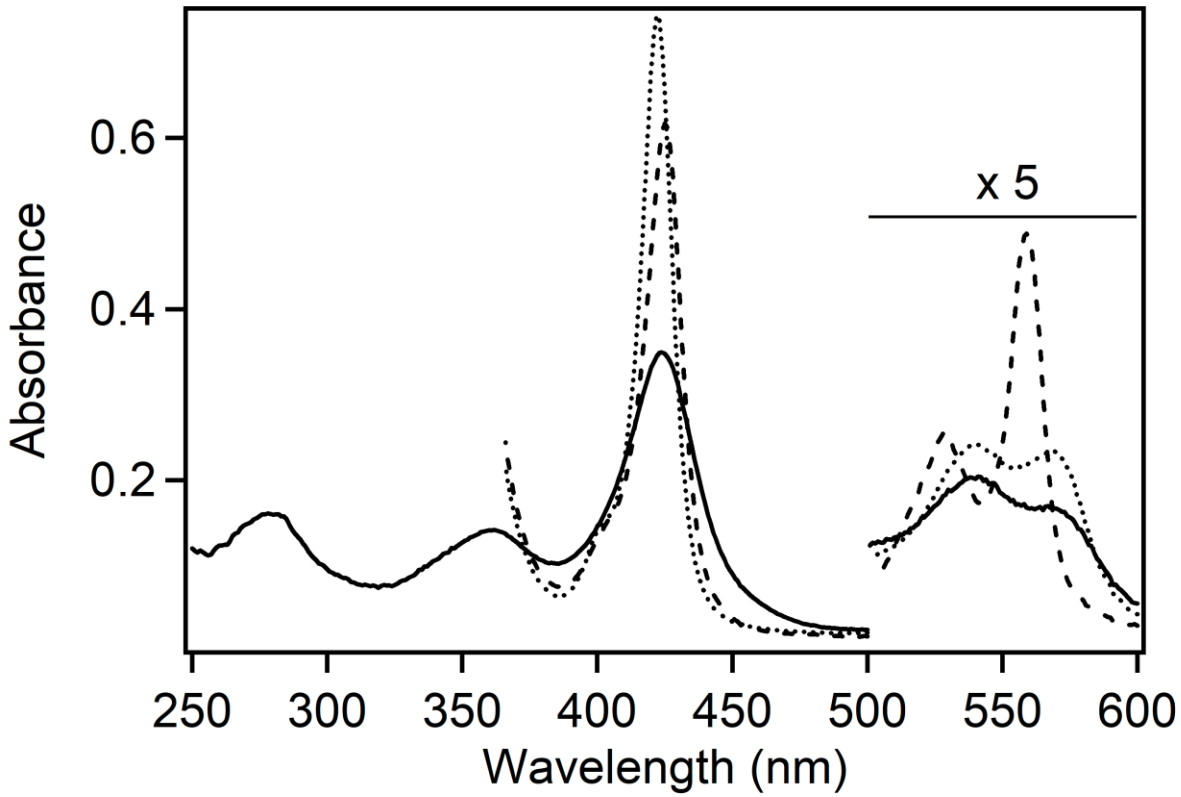


Figure 2.S3 Electronic absorption spectra of E60C CooA (4.5 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 100 mM MOPS, pH 7.4, 500 mM NaCl. Reduction and CO binding were carried out as described in the methods section.

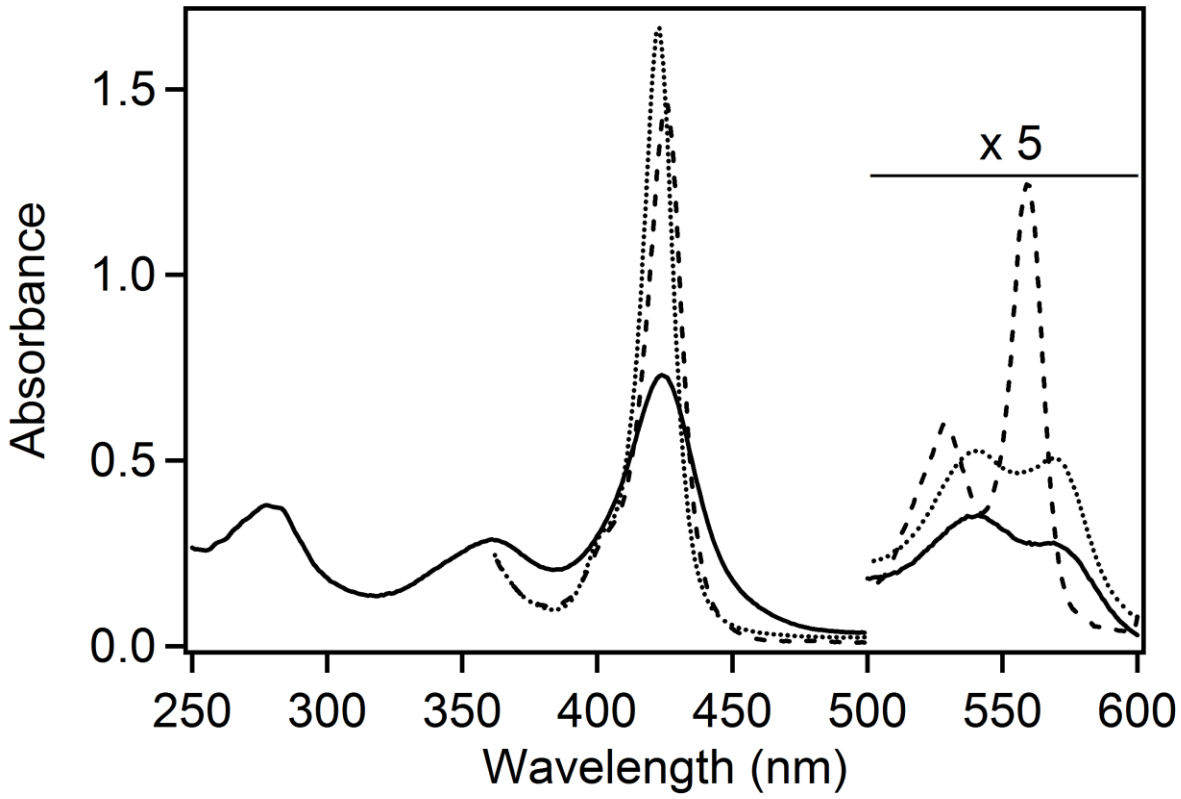


Figure 2.S4 Electronic absorption spectra of F132C CooA (10.6 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 100 mM MOPS, pH 7.4, 500 mM NaCl. Reduction and CO binding were carried out as described in the methods section.

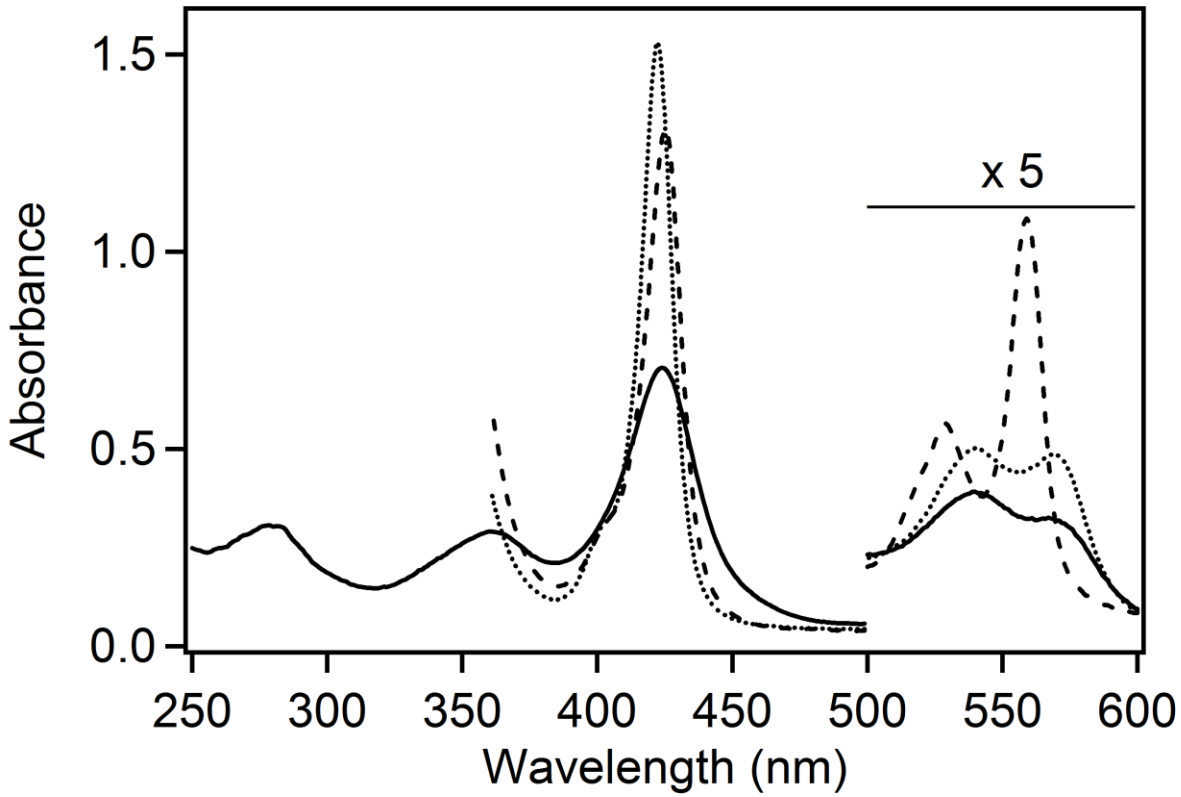


Figure 2.S5 Electronic absorption spectra of D134C CooA (4.6 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 25 mM MOPS, pH 7.4, 500 mM KCl. Reduction and CO binding were carried out as described in the methods section.

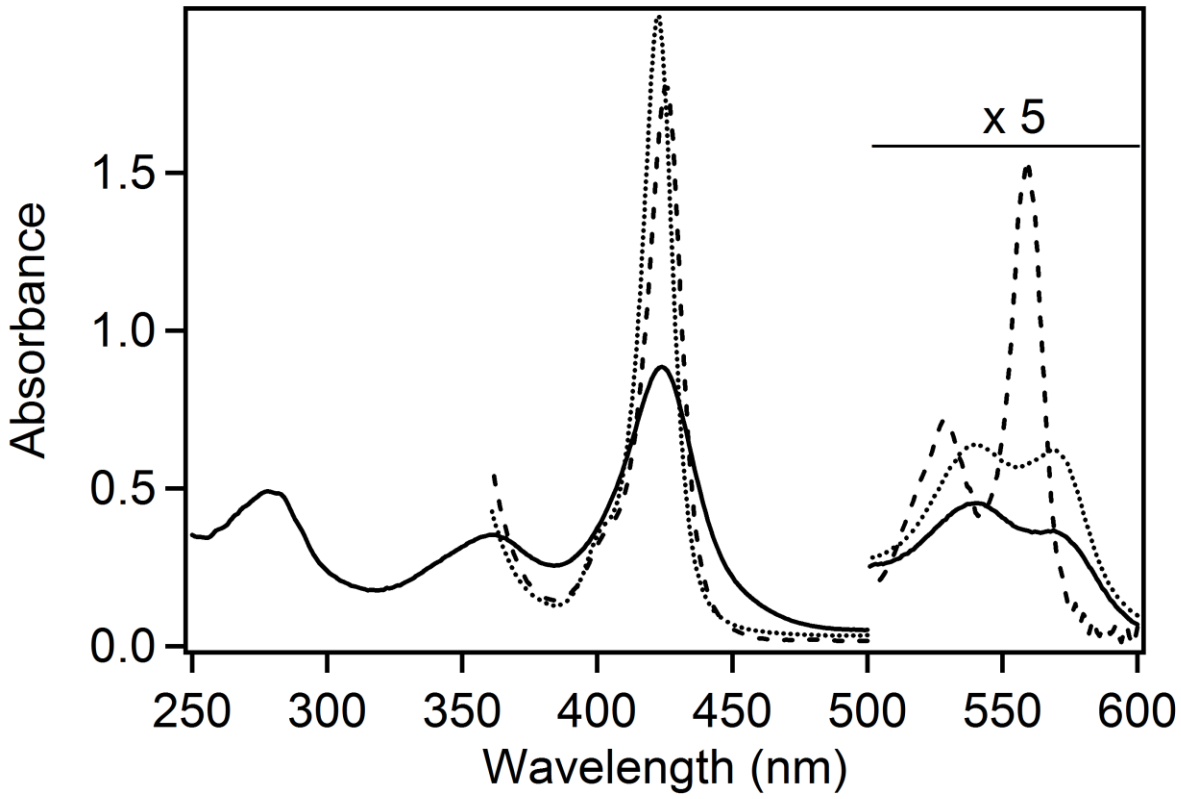


Figure 2.S6 Electronic absorption spectra of S175C CooA (8.2 μM dimer) in the Fe(III) (solid), Fe(II) (dotted), and Fe(II)-CO (dashed) states. Spectra were recorded at 25 °C, and samples were in 100 mM MOPS, pH 7.4, 500 mM NaCl. Reduction and CO binding were carried out as described in the methods section.

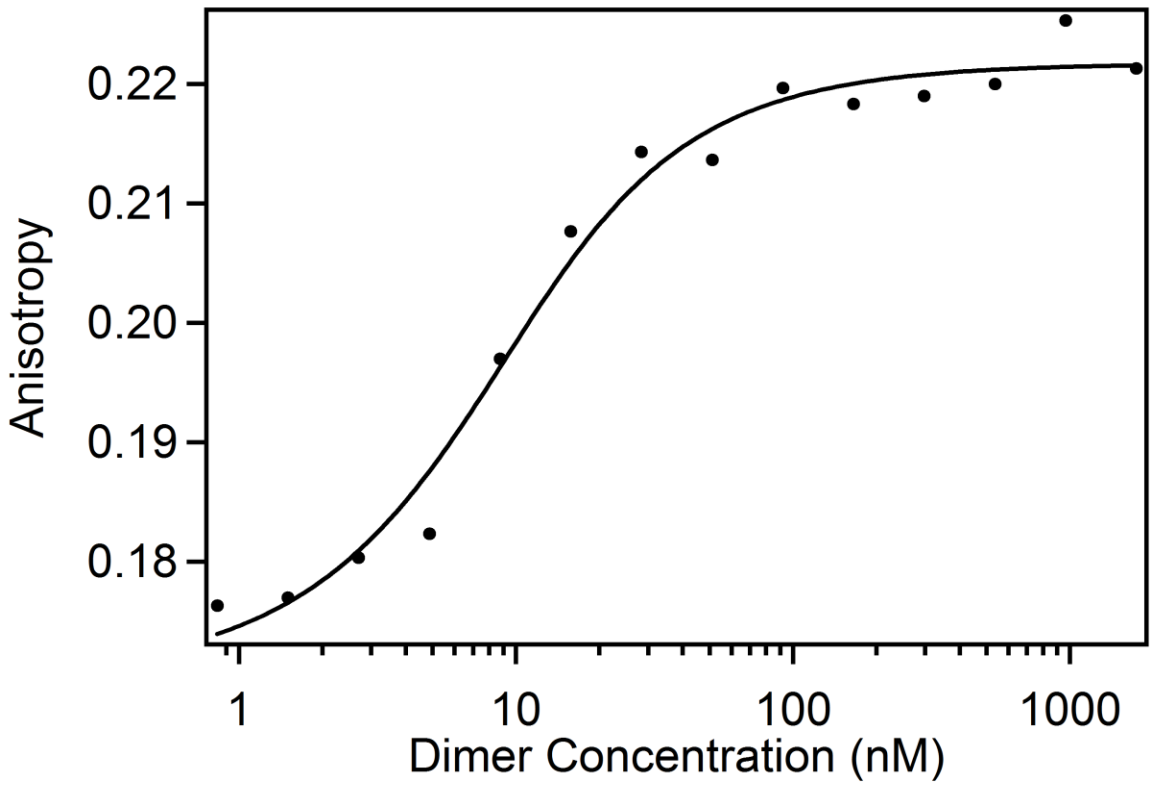


Figure 2.S7 Representative DNA binding curve for ΔCys$_4$CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of ΔCys$_4$CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

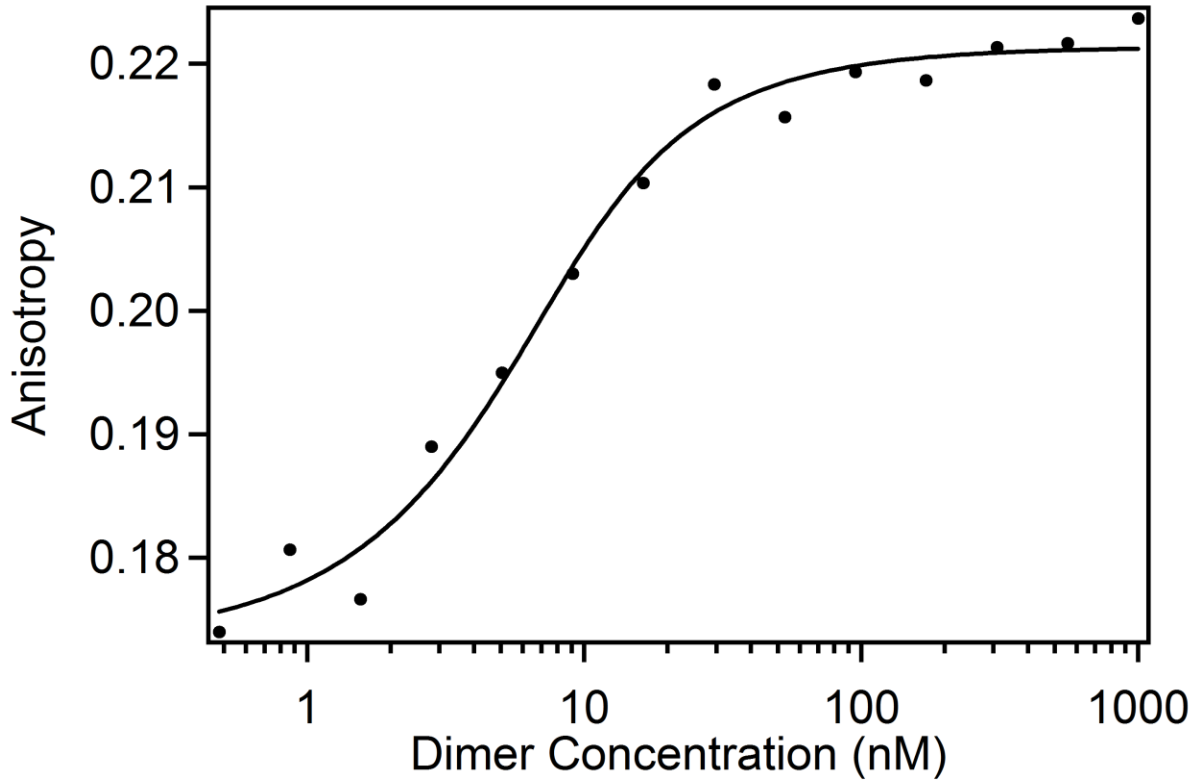


Figure 2.S8 Representative DNA binding curve for K26C CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of K26C CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

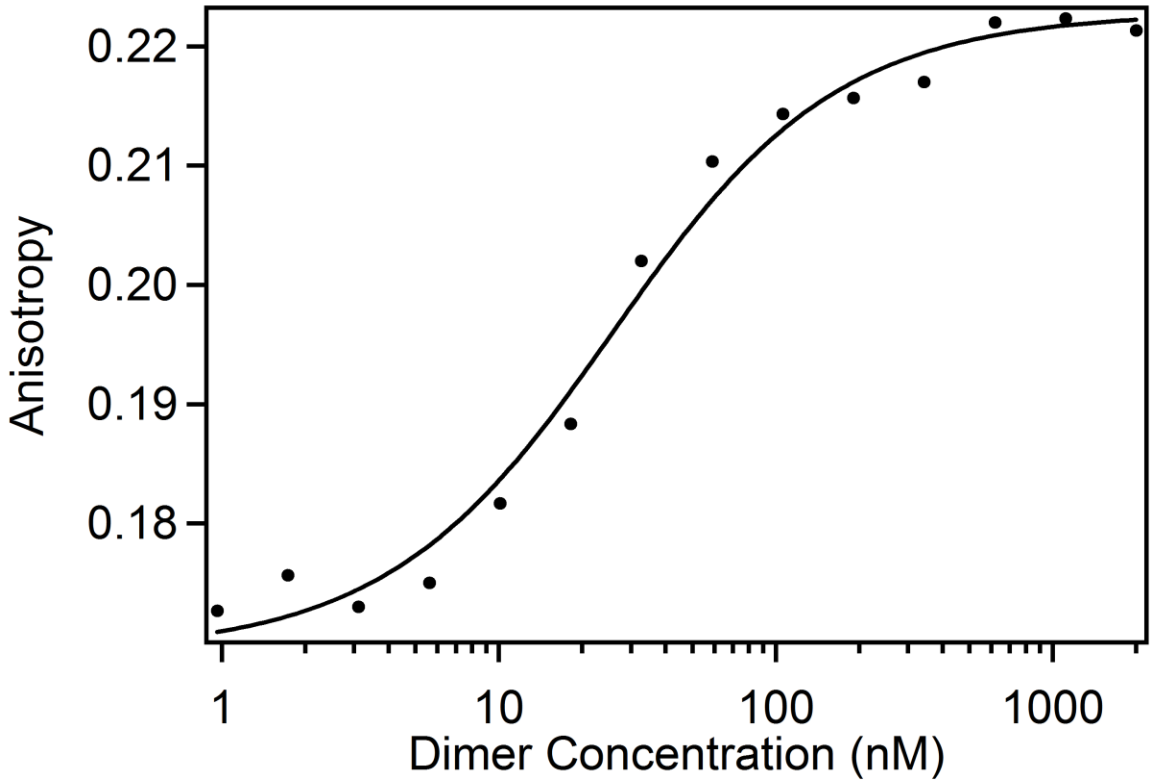


Figure 2.S9 Representative DNA binding curve for E60C CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of E60C CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

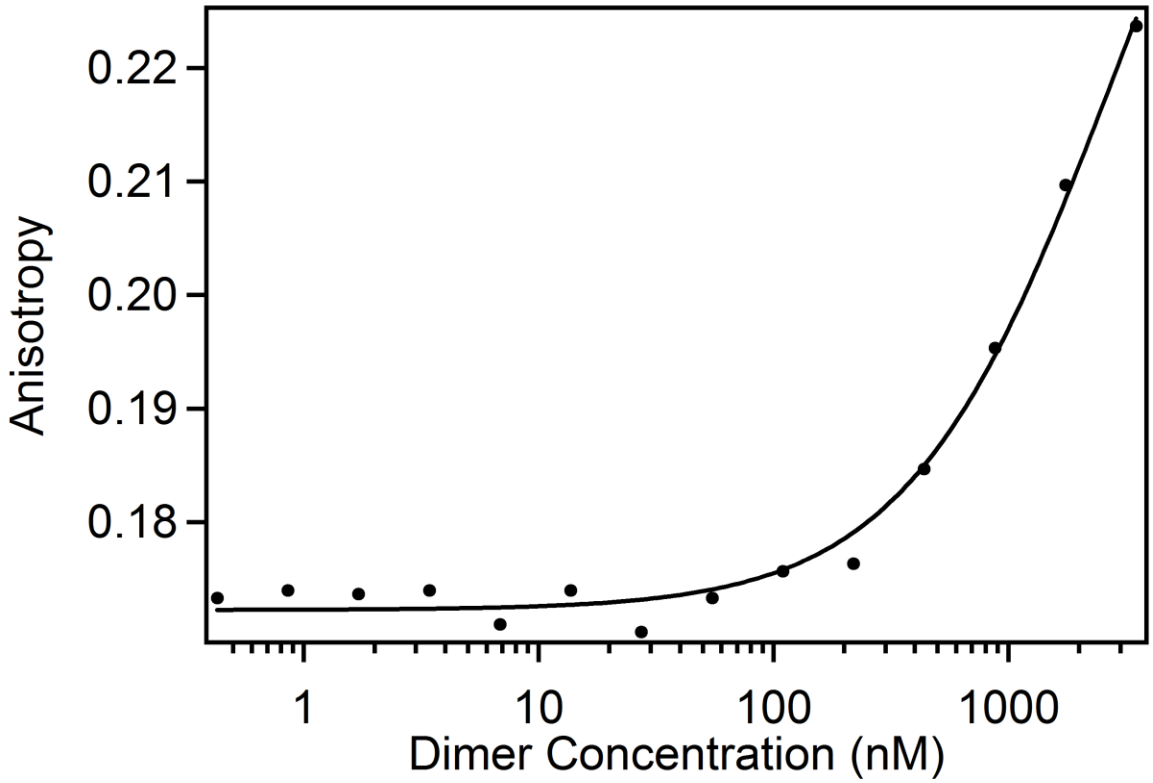


Figure 2.S10 Representative DNA binding curve for F132C CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of F132C CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

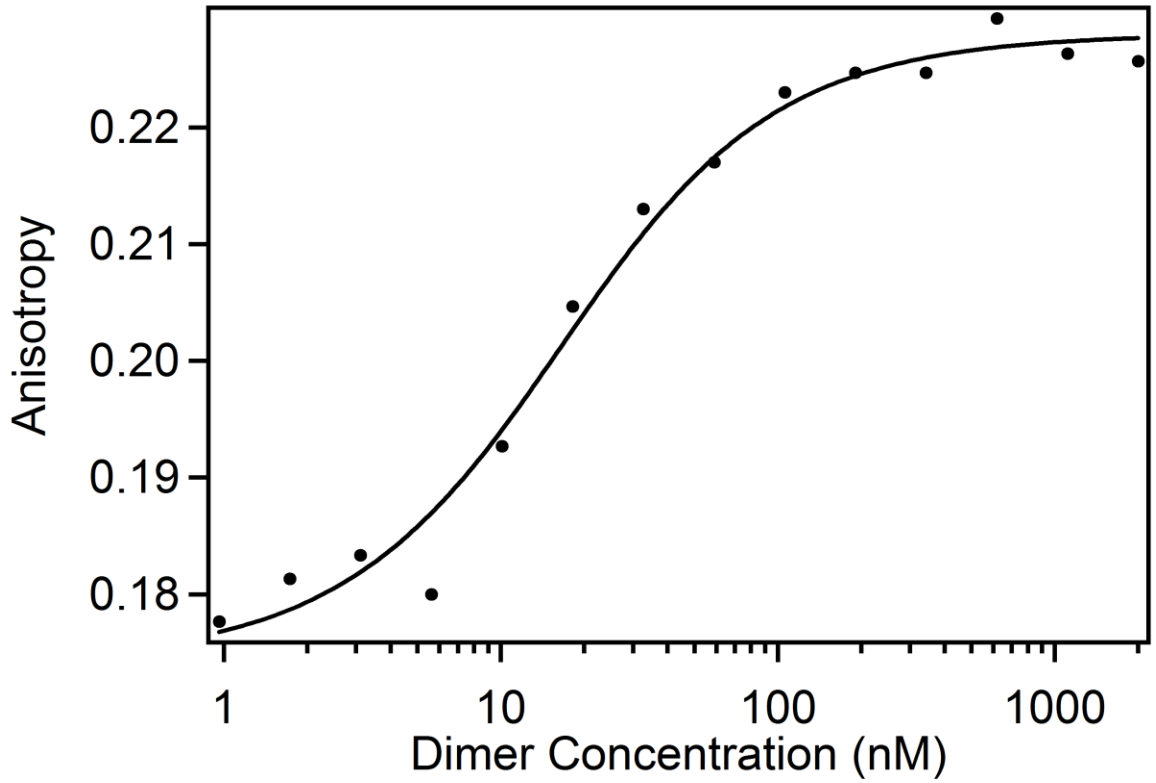


Figure 2.S11 Representative DNA binding curve for D134C CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of D134C CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

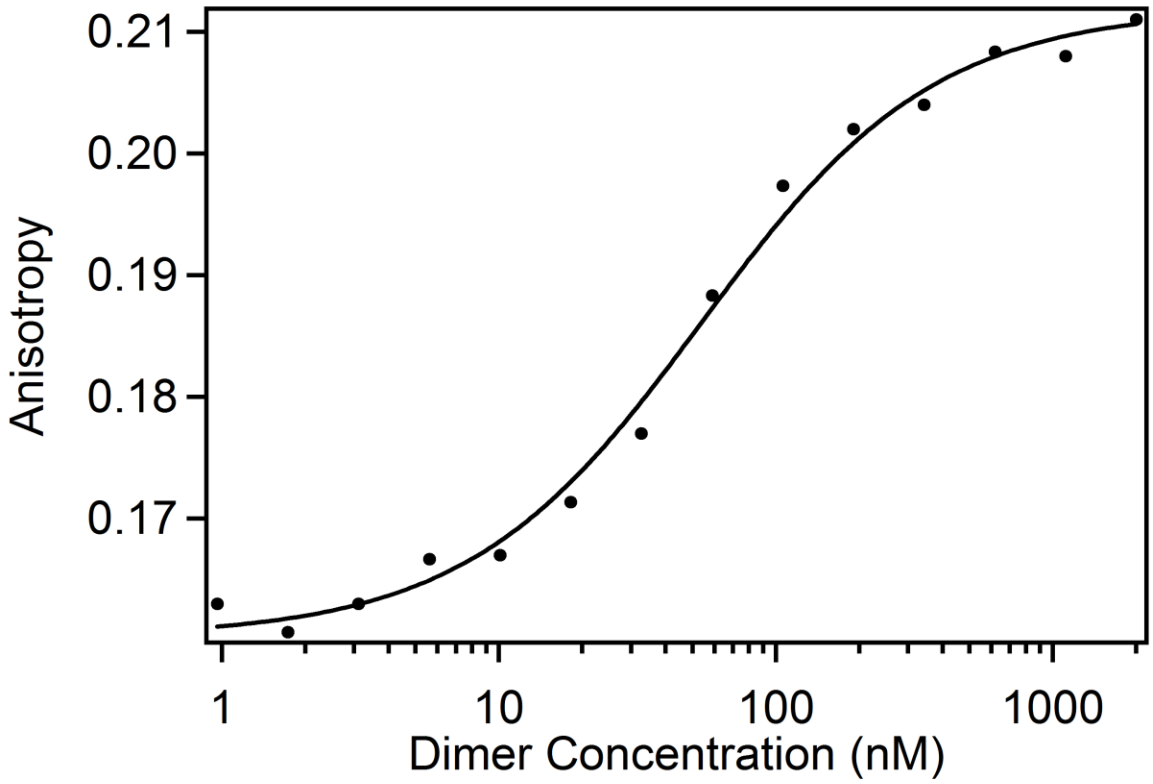


Figure 2.S12 Representative DNA binding curve for S175C CooA. Dots represent anisotropy values for double-stranded, Texas Red-labelled P_{cooF} (6.4 nM) in the presence of varying concentrations of S175C CooA in the Fe(II)–CO state. The line represents the best fit of the data to a simple binding model, as determined by nonlinear least squares analysis. DNA binding experiments were performed in triplicate, and the K_d value was determined for each curve.

2.8 Appendix 2: SDSL-EPR Data for Fe(II)–CO "*on*" and Fe(II)–CO "*DNA-bound*" CooA

Spin labelling of Fe(II)-CO and Fe(II)-CO-DNA-Bound CooA Variants. All labeling reactions were carried out in 100 mM MOPS, 500 mM NaCl, pH 7.4. MAL-6 (10 mg/mL in acetonitrile) was added to 15-25 nmol Fe(III) CooA variants in 2.5 mL buffer and allowed to react overnight at 4°C. To remove excess spin label, labeled protein samples were passed over Sephadex G-25 in a PD-10 desalting column using fresh labeling buffer (GE Life Sciences). Desalted protein samples were concentrated and buffer exchanged using a 30,000 MWCO spin concentrator (Amicon Ultra); all buffers used for this step were purged with $Ar_{(g)}$ and $CO_{(g)}$, and the exact buffer compositions are noted in the figure legends.

Each labeled protein sample (50-100 μL, 75-200 μM heme) was transferred to a septum-sealed PCR tube, and the headspace was purged with $Ar_{(g)}$ for 2 min, followed by $CO_{(g)}$ for 30 s. A solution of 10 mM sodium dithionite in $Ar_{(g)}$-purged buffer was then added to each protein sample via a syringe to a concentration three times higher than that of protein, and the solution was gently mixed for several minutes until the dark-red protein solution turned bright pink. To re-oxidize the MAL-6 spin label, the septum was then removed and air gently introduced into the tube with shaking for 1 min. Samples were then immediately loaded into a capillary tube and measured using EPR spectroscopy. For DNA-bound samples, the above procedure was repeated exactly, except that double-stranded P_{cooF} DNA oligomer was added to a final concentration equal to that of protein dimer.

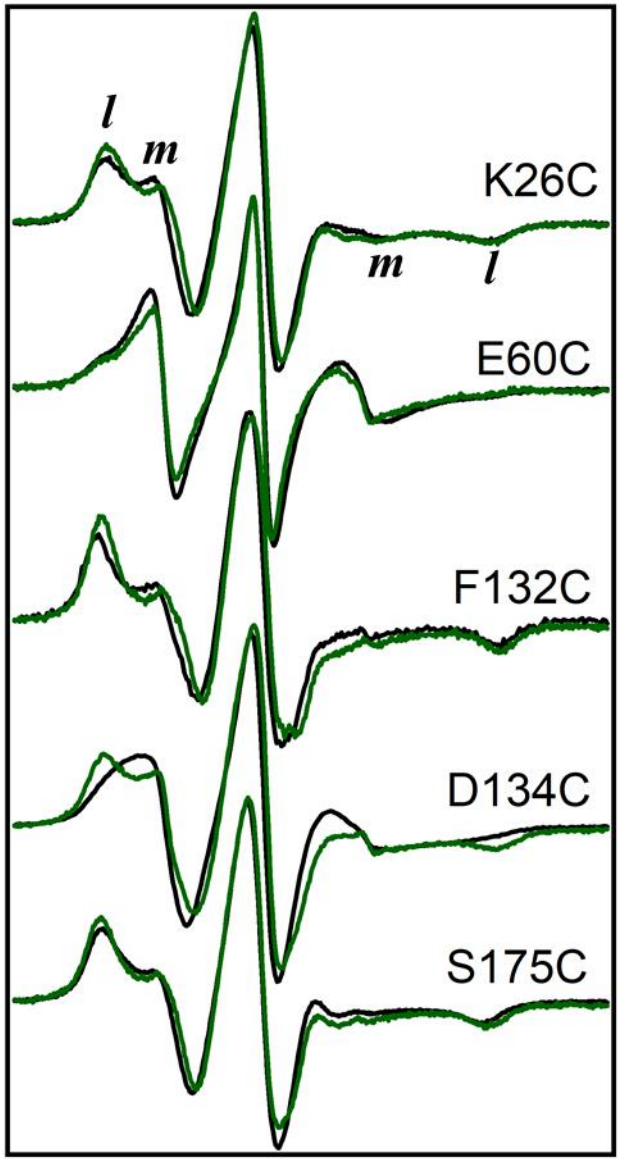


Figure A2.1 EPR spectra of MAL6-labeled Fe(III) ΔCys$_4$CooA variants (black) and labeled Fe(II)-CO ΔCys$_4$CooA variants (green) in 100 mM MOPS, 500 mM NaCl, pH 7.4 with 25% (w/w) Ficoll-70. The spectra are scaled to equivalent central line amplitude. More mobile (***m***) and less mobile (***l***) components of the m_I = +1 and m_I = -1 lines are labeled for clarity.

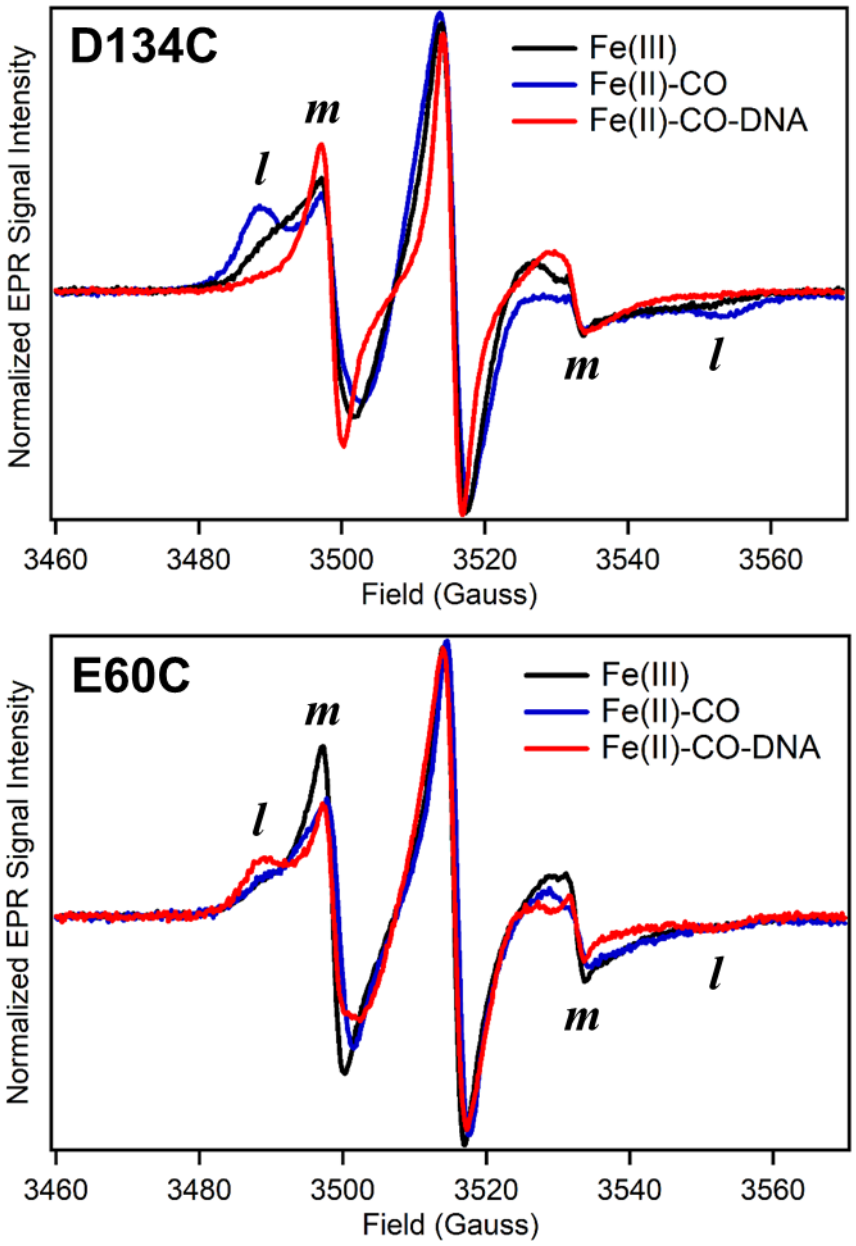
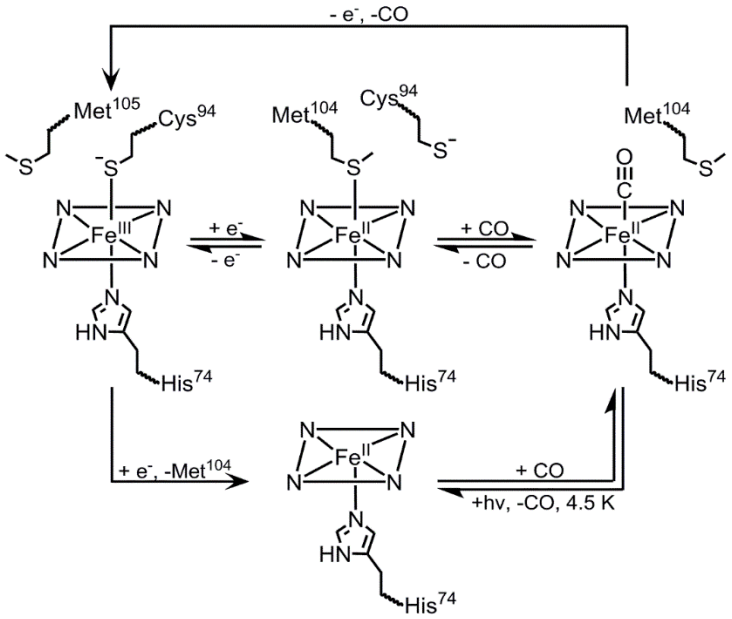


Figure A2.2 EPR spectra of MAL6-labeled Fe(III), Fe(II)-CO, and Fe(II)-CO-DNA-bound ΔCys_4 D134C and E60C CooA variants in 100 mM MOPS, 500 mM NaCl, pH 7.4 with 25% (w/w) Ficoll-70. The spectra are scaled to equivalent central line amplitude. More mobile (***m***) and less mobile (***l***) components of the $m_I = +1$ and $m_I = -1$ lines are labeled for clarity.

Chapter Three

Met^{104} is the CO-Replaceable Ligand at Fe(II) Heme in the CO-Sensing Transcription Factor *Px*RcoM-1



A previous version of this chapter was published as:

Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met^{104} is the CO-replaceable ligand at Fe(II) heme in the CO-sensing transcription factor *Bx*RcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21*, 559.

H.E.B. performed most of the protein overexpression and purification, and M.R.D. carried out the spectroscopic investigations. H.E.B. and M.R.D. wrote the manuscript.

3.1 Introduction

The ability of organisms to selectively express genes in response to environmental factors is a key regulatory component of metabolic pathways. Transcriptional control of gene expression, in which transcription factors recruit or repress RNA polymerase binding at promotor sequences, is one strategy employed for gene regulation.[1] Exogenous ligands, changes in pH, light, and other allosteric effectors activate or deactivate transcription factors for binding to specific DNA sequences in response to environmental changes.[2] Some prokaryotes, which have evolved to survive in diverse conditions, have the ability to sense the presence of small molecule gases such as carbon monoxide (CO), oxygen (O_2), and nitric oxide (NO) in order to switch between respiration and other anaerobic metabolic pathways.[3,4] Often, these gas-sensing proteins contain a heme co-factor because it makes an excellent allosteric binding site due to the high affinity of the iron center for these species. Additionally, the large surface area of the porphyrin allows for many points of contact to the protein causing subtle structural changes to be magnified.[5]

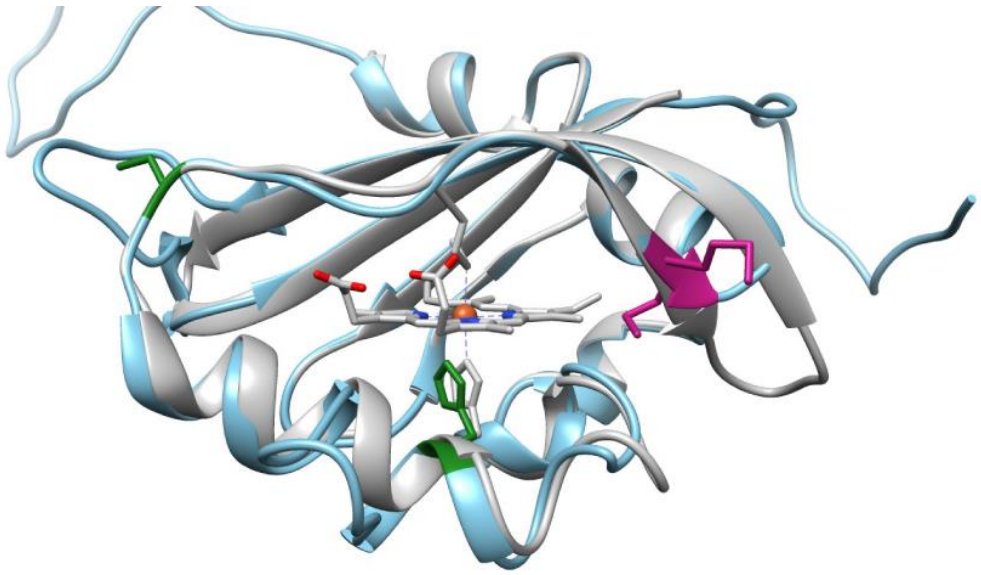
The most well studied heme-dependent gas sensing transcription factor is CooA (CO oxidation activator) from the purple photosynthetic bacterium *Rhodospirillum rubrum*.[6,7] CooA, as part of the FNR/CRP superfamily, is a homodimeric protein consisting of two domains: an N-terminal heme binding domain and a C-terminal helix-turn-helix DNA binding domain. A leucine zipper/coiled-coil region forms the dimer interface.[8] The heme iron undergoes a redox-mediated ligand switch as the mechanism for activation. Under oxidizing conditions, Cys^{75} and Pro^{2} from opposite monomers are axially bound to the Fe(III) heme. Upon reduction to Fe(II), the cysteine thiolate is replaced by His^{77}.[9] The presence of CO in the environment causes CO to replace the weakly bound Pro ligand and activates the protein to bind DNA through changes in dynamics and conformation.[10,11]

Another heme-containing transcription factor associated with CO metabolism genes was recently discovered in the non-pathogenic bacterium *Paraburkholderia xenovorans* (LB400).[12] This organism has generated interest for its ability to aerobically degrade polychlorinated biphenyl.[13] In addition to its xenobiotic metabolism, *P. xenovorans* performs dissimilatory nitrate reduction and aerobic CO oxidation.[14] The *cox* genes, which encode for the molybdo-flavoprotein complex required for CO oxidation, appear to be regulated by the transcription factors RcoM-1 and RcoM-2 (regulator of CO metabolism).[12] RcoM-1 is encoded on chromosome 1 while RcoM-2 is located on the megaplasmid, and each protein is upstream of a separate *cox* operon.[15] RcoM-1 and RcoM-2 share 88% sequence identity including the ligands for the heme co-factor.

The RcoM proteins are single-component transcriptional regulators consisting of two domains. The N-terminal domain binds to heme *b*, while the C-terminal domain is responsible for DNA binding. No structures yet exist for either protein, but homology modelling with the most closely related, structurally characterized protein, *Escherichia coli* direct oxygen sensor (*Ec*DOS), suggests that the N-terminal, heme-binding region adopts a PAS (Per-Arnt-Sim) fold.[12] PAS domains are usually found covalently linked to the N-terminus of an effector domain such as the histidine kinase or phosphodiesterase domains of sensor proteins involved in two-component signaling systems. A spatially conserved cleft provides a binding region for a variety of cofactors such heme, flavin, and metabolites.[16] The DNA binding region aligns with several known LytTR domain-containing transcription factors.[12] The LytTR domain is the second most commonly used DNA-binding motif, after the helix-turn-helix motif, in response regulator proteins. These proteins often control virulence pathways in response to phosphorylation.[17] Uniquely, the predicted structure of RcoM combines both sensor and response domains in one single-component signaling protein.

Despite differences between the predicted tertiary structure of RcoM and the known structure of CooA, the heme co-factor in RcoM appears to act as the allosteric effector site for CO in a manner similar to that of CooA. Previous work identified Cys^{94} and His^{74} as the axial ligands in the Fe(III) state of both RcoM proteins.[12,18] Reduction to the Fe(II) state results in loss of the cysteine thiolate and replacement by an unknown neutral donor ligand, which is subsequently replaced by CO. His^{74} remains bound to iron in all states.[19] The CO-bound state of RcoM has been shown to bind upstream from the *cox* gene cluster in a putative regulatory region, but the exact mechanism of transcriptional regulation has yet to be elucidated.[20]

The identity of the neutral ligand in the Fe(II) state of RcoM was hypothesized to be Met^{104} based on structural homology with the Fe(II) state of *Ec*DOS which has His/Met ligation (Figure 3.1).[21] However, the sequence of RcoM-1 (UniProt Q13YL3) contains two consecutive methionine residues at positions 104 and 105. In the homology model, neither residue adopts a particularly suitable geometry for heme binding, but Met^{104} is predicted to be in the heme pocket while Met^{105} is solvent exposed (Figure 3.1). Previous work with the M104L variant of RcoM-1 showed that replacing methionine with leucine, a residue with similar hydrophobicity and volume, caused perturbations to the electronic absorption spectrum of the Fe(II) state of the heme, namely a shoulder on the Soret peak and a lack of resolution in the α/β region.[12] The implication of these spectral changes is unclear, since the perturbations could be due to an electronic effect from the loss of the ligand or a steric effect in which the heme pocket has been disturbed. Furthermore, in the event that Met^{104} is the ligand, it is possible that Met^{105} could compensate by acting as a substitute ligand. The electronic absorption data alone do not provide conclusive evidence for the identity of the ligand. Further spectroscopic characterization of M104, M105, and M104/M105 variants is required to determine the identity of the Fe(II) ligand.



```
        10         20         30         40         50
MKSSEPASVS AAERRAETFQ HKLEQFNPGI VWLDQHGRVT AFNDVALQIL
        60         70         80         90        100
GPAGEQSLGV AQDSLFGIDV VQLHPEKSRD KLRFLLQSKD VGGCPVKSPP
       110        120        130        140        150
PVAMMINIPD RILMIKVSSM IAAGGACGTC MIFYDVTDLT TEPSGLPAGG

SAPS
```

Figure 3.1 *Top:* The sequence of the *Paraburkholderia xenovorans* RcoM-1 heme-binding domain (blue) threaded onto the structure of *Ec*DOS Fe(II) (grey, PDB 1V9Z) using Modeller via UCSF Chimera.[22-25] The two potential methionine ligands are highlighted in purple while the known His and Cys ligands are highlighted in green. *Bottom:* The sequence of the RcoM-1 heme-binding domain with the known heme ligands highlighted in green and the two potential methionine ligands highlighted in purple.

Herein, we identify Met^{104} as the ligand trans to His^{74} in Fe(II) RcoM-1 and show that Met^{105} may play a role in stabilizing Cys^{94}(thiolate) coordination in Fe(III) RcoM-1. Electron paramagnetic resonance (EPR) spectra of the wild type (WT) and the M104L variant demonstrate that that the Fe(III) heme pocket is not significantly altered by the change in this amino acid. Magnetic circular dichroism (MCD) and electronic absorption spectra of M104L and M104L/M105L RcoM-1 variants indicate that a five-coordinate, high-spin (S = 2) species forms upon reduction of the heme to the Fe(II) state. This evidence, coupled with the observation of a six-coordinate, low-spin Fe(II) species in the WT and M105L variant confirms Met^{104} as the sixth neutral donor ligand in the Fe(II) heme. Surprisingly, EPR spectra of oxidized M105L and M104L/M105L contain an additional low-spin signal. The high rhombicity of this signal indicates that a neutral ligand has replaced Cys^{94} (thiolate); although Met^{105} does not directly bind to Fe(II) heme, it does influence the ligand environment in Fe(III) RcoM-1.

3.2 Materials and Methods

Materials

All chemicals used in buffer and media preparation (99% purity or better) were purchased from Research Products International Corporation and used as received. Glycerol (greater than 99.5% purity) and ferric citrate (BioReagent grade) were purchased from Sigma-Aldrich and used as received. Certified ACS grade potassium ferricyanide – $K_3[Fe(CN)_6]$ – was purchased from ThermoFisher Scientific and used as received. Sodium dithionite ($Na_2S_2O_4$; 85% purity) was purchased from Fluka and stored under argon gas at -20 °C until used. A CO gas (99.5% purity) cylinder was obtained from Air Gas.

Isolation and purification of RcoM-1

RcoM-1 and its variants were isolated and purified in a manner similar to that previously described.[12,18,19] Briefly, *E. coli* VJS6737 [26] containing the pEXT20 expression vector [12,27] for each *Paraburkholderia xenovorans* RcoM-1 variant was grown in Lennox-LB medium supplemented with ferric citrate (20 mg/L) and nutrient broth (3 g/L). Dense overnight starter cultures were diluted 1:10 into fresh media and grown to an OD_{550} greater than 4 at 30 ˚C and 220 rpm. These cultures were diluted to an OD_{550} of 0.06 in fresh media containing isopropyl β-D-thiogalactopyranoside (50 μM) to induce expression for 19-20 h at 28 °C. Cells were pelleted by centrifugation (6,000 x g, 20 min), resuspended in lysis buffer [100 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.4, 500 mM NaCl, 1 mM dithiothreitol (DTT)], and lysed by sonication. A HisTrap FF column (GE Lifesciences, 2x5 mL column volume) was pre-equilibrated on an AKTA Prime FPLC with 10 mM imidazole, 100 mM MOPS, pH 7.4, 400 mM NaCl. The cell supernatant was applied to the column and washed until the 280 nm absorbance reached baseline. The protein was eluted off the column at 5 mL/min using a gradient from 10 mM to 350 mM imidazole. The red fractions were pooled, brought to 45%-saturated $(NH_4)_2SO_4$ (v/v) and 0.5 mM DTT, and precipitated by incubation on ice for 15 minutes followed by brief centrifugation. The protein pellet was resuspended in 25 mM MOPS, pH 7.4, 50 mM NaCl and applied to a HiTrap Heparin HP column (GE Lifesciences, 5 mL column volume) pre-equilibrated in the same buffer. The protein was eluted off the column using a gradient from 50 mM to 1 M NaCl applied by the FPLC. Fractions were again pooled and buffer exchanged back into lysis buffer using Spin-X UF concentrators (Corning, 10 kDa MWCO, 500 μL volume) for storage at -80 °C. Total protein concentrations were determined using the bicinchoninic acid method (Pierce-Thermo Fisher, Rockford, IL, USA); sodium dodecyl sulfate polyacrylamide gel electrophoresis

verified protein purity was greater than 80%. Heme content was determined using the pyridine hemochromogen assay.[28]

Oxidation of purified RcoM-1

Wild type and variant RcoM-1 proteins were isolated as mixtures of Fe(III) and Fe(II)-CO states. To ensure homogeneity of protein oxidation/coordination state, selected samples were oxidized using potassium ferricyanide. A stock solution of potassium ferricyanide (30 mM) was prepared by dissolving 5 mg of solid potassium ferricyanide in 500 μL of 25 mM MOPS buffer, pH 7.4, 500 mM KCl. This stock solution was added to each protein sample to a final concentration of 2-3 mM and allowed to react on ice, until there was no evidence of the Fe(II)-CO species. The solution was then loaded onto a 10 kDa MWCO spin concentrator (Corning) and washed 3-4 times by concentration and dilution into fresh buffer using a tabletop centrifuge (14,100 x g) to remove excess potassium ferricyanide.

EPR spectroscopy

X-band electron paramagnetic resonance (EPR) spectral data were collected at 10 K on a Bruker ELEXSYS E500 spectrometer equipped with an Oxford ESR 900 continuous flow liquid helium cryostat and an Oxford ITC4 temperature controller. The microwave frequency was monitored using an EIP model 625 CW microwave frequency counter. When so denoted in the text and figure legends, protein samples were oxidized as described above. All other samples were used as isolated. Prior to loading into EPR tubes, samples were concentrated and exchanged into the buffers denoted in figure legends. Samples were slowly transferred to a quartz EPR tube, via a wide needle to minimize denaturation, and frozen in liquid nitrogen. Samples of approximately 150 μL had a final concentration range of 80-220 μM.

Electronic absorption spectroscopy

Electronic absorption spectra were recorded at room temperature with a double-beam Varian Cary 4 Bio spectrophotometer set to a spectral bandwidth of 0.5 nm. Protein samples, 2.2-16.5 μM in heme, were prepared in 25 mM MOPS, pH 7.4 buffer with 500 mM KCl. Previously oxidized proteins, purged with argon gas for at least 10 minutes, were reduced by addition of a few crystals of sodium dithionite under an argon atmosphere. Conversion to the Fe(II)-CO state was accomplished by injecting a 100 μL bolus of CO into the headspace of a sealed, purged cuvette containing Fe(II) protein.

MCD spectroscopy

MCD spectra were recorded at temperatures ranging from 4.5-50 K on a Jasco J-715 CD spectropolarimeter equipped with an Oxford Instruments SM-4000-8T magnetocryostat. To correct for the CD contribution to each spectrum, the MCD signal at -7 T was subtracted from the signal at +7 T, and the difference was divided by two. MCD samples were prepared in 25 mM MOPS pH 7.4, 500 mM KCl, and 60% (v/v) glycerol. Fe(III) samples were prepared by direct addition of oxidized protein mixed with glycerol to an MCD cell. All manipulations to generate samples in the Fe(II) state were carried out in an anaerobic chamber. A small amount of solid sodium dithionite was added to oxidized protein samples prior to the addition of degassed glycerol. Once glycerol was well incorporated, samples were transferred to an MCD cell via a gas tight syringe inside the glove box. Fe(II)-CO protein was prepared in a 2 mL auto-sampler vial (VWR) in the same manner as for electronic absorption. A small volume of glycerol was purged with argon in a second vial. Upon complete conversion to the Fe(II)-CO adduct, the protein was transferred to the glycerol vial via gas-tight syringe and mixed well. The resulting solution was transferred to an argon-purged MCD cell. A variable temperature, variable field (VTVH) experiment was carried

out on a Fe(II) sample of M104L RcoM-1. The signal with maximum positive intensity in the Soret region (441 nm) was monitored as the field was varied from +7 T to -7 T at constant temperature (2.2 K, 4.5 K, 8.0 K, 15 K, and 25 K). The signal of each curve was averaged as described above to remove the CD background.

3.3 Results

RcoM-1 is isolated as an admixture of heme states

An appreciable population of as-isolated RcoM-1 exists as the Fe(II)-CO adduct. A representative MCD spectrum of as-isolated M104L RcoM-1 exhibits an intense, temperature-dependent feature in the Soret region at low temperatures under 15 K (Figure 3.2a). This signal is characteristic of a five-coordinate, high-spin (S = 2) Fe(II) heme center formed upon photodissociation of CO.[29] The focused light source used to measure MCD induces this photodissociation and at low temperatures the "CO off" state is trapped. The intense high-spin Fe(II) signal dominates the low temperature spectra, precluding identification of peaks due to other heme states. As the temperature is increased, the rate of CO rebinding also increases with a concomitant reduction in the paramagnetic signal. Above 15 K, a temperature-independent, diamagnetic signal identical to that of Fe(II)-CO RcoM-2 is observed as the dominant signal in all RcoM-1 variants (Figure 3.2a, inset).[19]

As-isolated WT and M105L samples could be reduced and bound additional CO, as observed by electronic absorption (data not shown); however, the Fe(II) "CO off" signal still dominates the low temperature MCD spectra. This observation suggests that WT and M105L are isolated as mixtures of the Fe(III) and Fe(II)-CO states. Furthermore, we observe that the relative amounts of Fe(III) and Fe(II)-CO protein in as-isolated WT and M105L samples are highly variable. In contrast, as-isolated M104L and M104L/M105L do not exhibit significant changes to

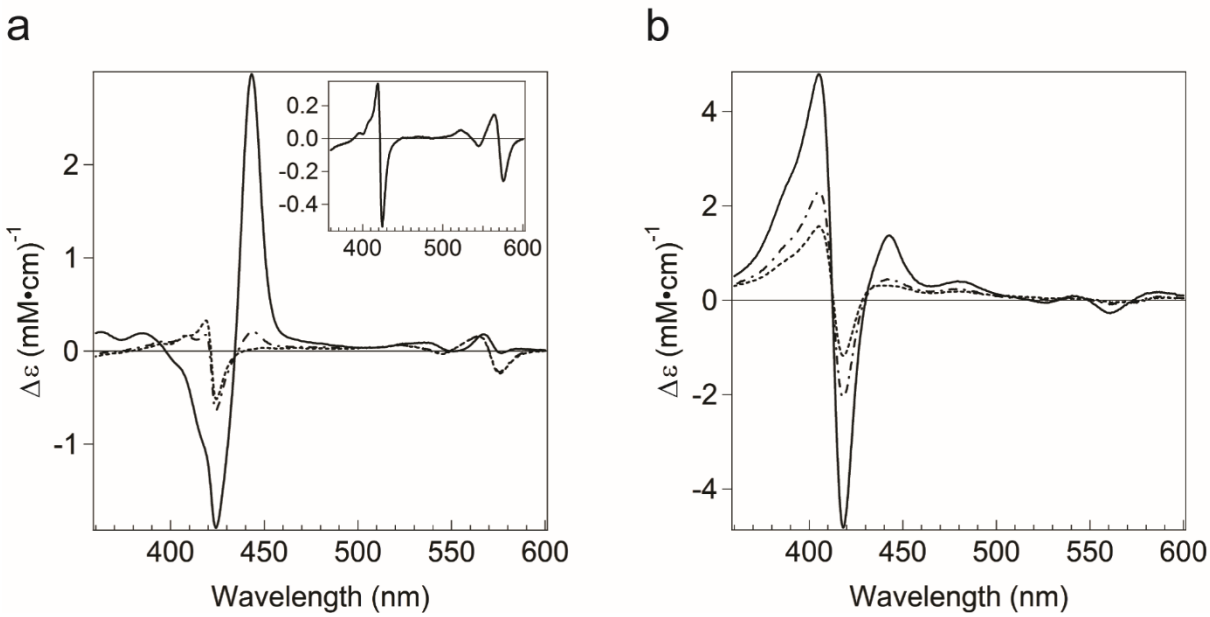


Figure 3.2 (a) MCD spectrum of as isolated M104L RcoM-1. The MCD sample was prepared in 25 mM MOPS and 500 mM KCl buffer (pH 7.4) with 60% (v/v) glycerol at a final concentration of 10 μM heme. Spectra shown are the average of three scans collected at constant magnetic field (7 T) and: 4.5 K (solid), 15 K (dot-dashed), 25 K (dashed) and 50 K (inset). (b) MCD spectrum of M104L RcoM-1 after oxidation with 3 mM potassium ferricyanide. Spectra shown are the average of three scans collected at constant magnetic field (7 T) and: 4.5 K (solid), 15 K (dot-dashed), and 25 K (dashed).

electronic absorption spectra upon reduction and addition of CO, suggesting that these samples are isolated almost entirely in the Fe(II)-CO state. These differences in behavior between as-isolated WT and M105L on one hand, and as-isolated M104L and M104L/M105L on the other hand, are born out by EPR experiments; the former exhibit Fe(III) signals while the latter do not (data not shown).

To obtain homogenous RcoM-1 samples, we developed a method that fully converts the protein to the Fe(III) state. Previous work with RcoM showed that photolysis by white light could be used to obtain a homogenous Fe(III) heme population[12,19]; however, the process requires an overnight incubation at 4 °C (277 K) to achieve full conversion, and the protein is highly susceptible to denaturation during photolysis. A faster and more consistent method is chemical oxidation of the heme with potassium ferricyanide, which requires only 30-120 min incubation at 4 °C (277 K) with less protein denaturation (Figure 3.2b, Table 3.1). This improved method was used to form homogeneous Fe(III) heme in subsequent experiments as noted in the text and figures.

Met^{105} influences the thiolate ligand in Fe(III) RcoM-1

When oxidized, WT RcoM-1 and M104L largely retain the native His/Cys ligation. The EPR spectrum of as-isolated WT RcoM-1 contains a single rhombic signal (g_x = 1.88, g_y = 2.26, g_z = 2.48) that is indicative of a low-spin (S = ½), six-coordinate Fe(III) heme (Figure 3.3a). This narrow signal, which is very similar to that observed in RcoM-2 (g_x = 1.88, g_y = 2.28, g_z = 2.52), is characteristic of His/Cys ligation.[19,30,31] WT RcoM-1 that is oxidized with potassium ferricyanide exhibited a major signal with the same EPR spectral features as the as-isolated WT RcoM-1 (Figure 3.3b). The variant M104L also exhibits this low-spin rhombic signal (Figure 3.3c) suggesting that a majority fraction of oxidized WT and M104L contain six-coordinate heme with His/Cys(thiolate) ligation. Two small signals due to high-spin rhombic (g = 4.30) and axial (g =

Table 3.1 List of MCD features for low-spin, Fe(III) RcoM variants following oxidation.

Variant	Peak	Crossover	Trough	Peak	Crossover	Trough
WT	411	417	423	545	555	566
M104L	405	412	418	541	552	561
M105L	406	412	419	541	553	562
M104L/M105L	405	412	418	542	551	562

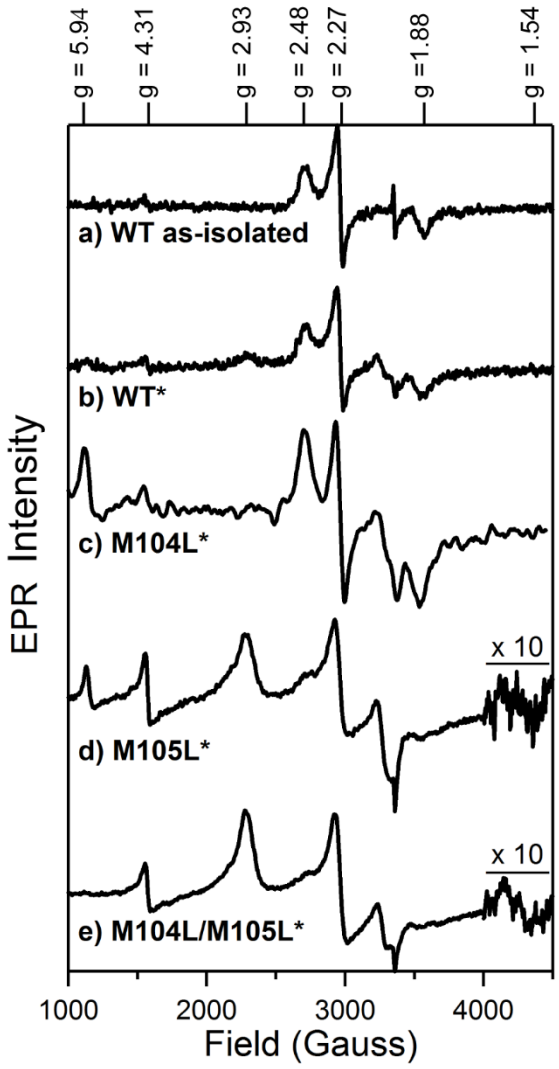


Figure 3.3 X-band EPR spectra of RcoM-1 variants. Asterisks indicate that samples were oxidized with potassium ferricyanide prior to analysis. Spectra were recorded with samples in 50 mM borate buffer, pH 7.4, and 500 mM KCl. The spectra were recorded at 10 K, 9.38 GHz microwave frequency, 1.002 mW microwave power, 60 dB receiver gain, 8.310 G modulation amplitude, 100 kHz modulation frequency, 20.48 ms time constant, and 20.48 ms conversion time. Between 20 and 45 scans were averaged for each sample depending upon heme concentration [a) 189 μM b) 149 μM c) 100 μM d) 88 μM e) 218 μM] and 2048 data points were collected during each scan.

5.94) Fe(III) are observed to some extent in most protein samples (Figure 3.3). The rhombic signal is often attributed to "junk iron". The axial signal suggests that a fraction of the protein bears a five-coordinate, high-spin heme, implicating loss 1 of a heme ligand. These signals are present in varying amounts and appear to be batch-dependent (Figures 3.3a vs 3.4a).

EPR spectroscopy revealed that when Met^{105} is replaced by leucine, a neutral donor replaces the native thiolate ligand upon oxidation. Oxidized Fe(III) M105L exhibits two distinct low-spin signals that differ in their *g* anisotropy: a broad signal with $g_x = 1.54$, $g_y = 2.27$, $g_z = 2.93$ and a narrow signal with $g_x = 1.89$, $g_y = 2.27$, $g_z = 2.48$ (Figure 3.3d). The *g* values of the broader signal are similar to those observed for Fe(III) hemes bearing two neutral donor ligands (Table 3.2)[32,33] and are significantly more widely spaced than those of WT RcoM-1 and M104L. The presence of this broad low-spin signal implies coordination by two neutral ligands. The *g* values of the narrower low-spin signal are comparable to those of WT RcoM-1 and M104L, implying that this species retains the native His/Cys(thiolate) coordination. The oxidized variant M104L/M105L exhibits two low-spin signals that are comparable to those of Fe(III) M105L. Since native thiolate ligation is disrupted only in oxidized variants where Met^{105} is altered, these observations suggest that Met^{105} plays a role in stabilizing Cys^{94}(thiolate) coordination in Fe(III) RcoM-1.

The broad, low-spin EPR signal observed in oxidized Fe(III) M105L does not appear in as-isolated Fe(III) samples of that variant. As mentioned above, WT and M105L RcoM-1 are sometimes (and unpredictably) isolated with a large fraction of the protein in the Fe(III) state. The majority fraction of such as-isolated Fe(III) WT RcoM-1 exhibits a narrow low-spin, rhombic signal ($g_x = 1.89$, $g_y = 2.26$, $g_z = 2.48$) characteristic of His/Cys ligation (Figure 3.4a). The low-spin signal observed in EPR spectra of the as-isolated variant M105L is nearly identical to those seen in as-isolated WT RcoM-1 (Figure 3.4b). Interestingly, the broad, low-spin signal observed

Table 3.2 Comparison of *g*-values for low-spin, Fe(III) RcoM variants to those of related heme proteins.

protein	low rhombicity signal			high rhombicity signal			ref.
	g_z	g_y	g_x	g_z	g_y	g_x	
WT RcoM-1, as-isolated	2.48	2.26	1.88	--	--	--	this work
WT RcoM-1, oxidized	2.48	2.27	1.89	--	--	--	this work
M104L RcoM-1	2.49	2.26	1.89	--	--	--	this work
M105L RcoM-1	2.48	2.27	1.89	2.93	2.27	1.54	this work
M104L/M105L RcoM-1	2.47	2.26	1.89	2.95	2.26	1.54	this work
RcoM-2	2.52	2.28	1.88	--	--	--	19
CooA (major signal)	2.46	2.25	1.89	--	--	--	30
CooA (minor signal)	2.58	2.25	1.54	--	--	--	30
nitric oxide synthase	2.45	2.29	1.90	--	--	--	31
nitric oxide reductase	--	--	--	2.97	2.25	~1.4	32
cytochrome *c*	--	--	--	3.06	2.25	1.25	33

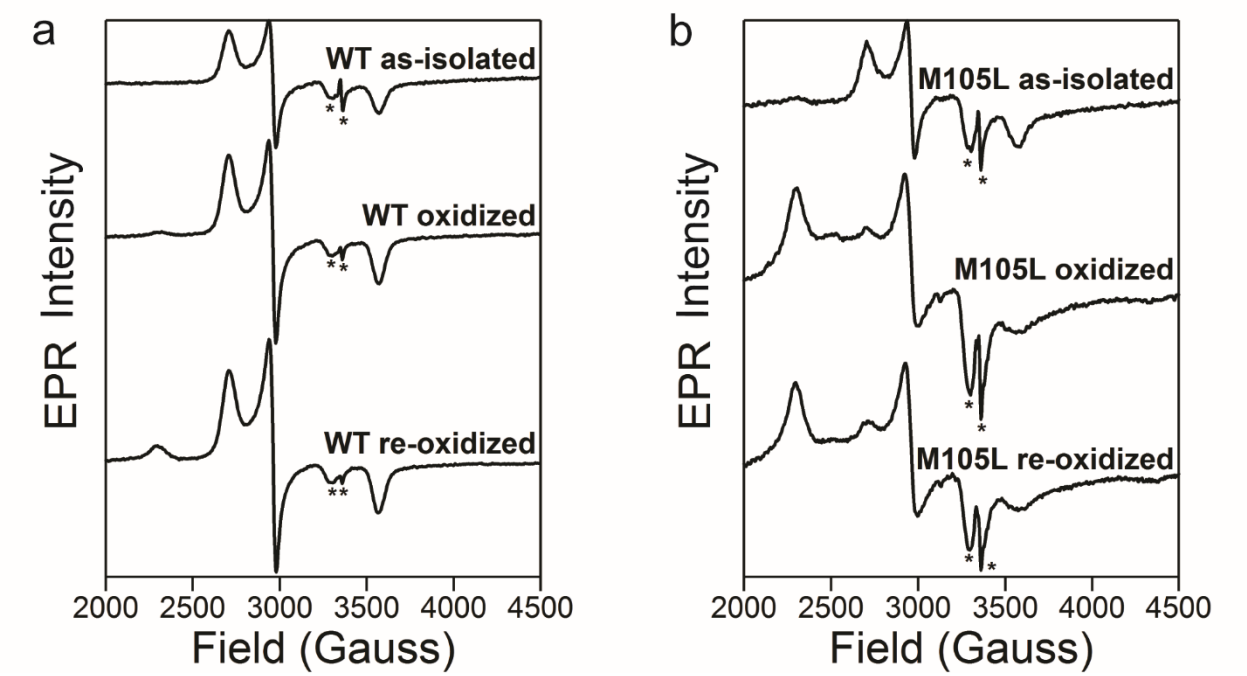


Figure 3.4 X-band EPR spectra of low-spin, Fe(III) heme in WT (a) and M105L (b) RcoM-1 as a function of oxidation. As-isolated samples were prepared for EPR with no modifications. Oxidized samples were prepared by treatment of as-isolated protein with potassium ferricyanide. Re-oxidized samples were prepared by reduction and CO-binding of as-isolated protein as described for Fe(II)-CO MCD samples, followed by treatment with potassium ferricyanide. Spectra were recorded with samples 67-193 μM heme for WT and 48-71 μM heme for M105L) in 10 mM phosphate buffer, pH 7.4, and 150 mM NaCl. The spectra were recorded at 10 K, 9.38 GHz microwave frequency, 0.5024 mW microwave power (0.6325 mW for WT as-isolated sample), 60 dB receiver gain, 8.310 G modulation amplitude, 100 kHz modulation frequency, 20.48 ms time constant, and 20.48 ms conversion time. Scans (15 or 20) were averaged for each sample and 2048 data points were collected during each scan.

in oxidized Fe(III) M105L is not observed in the as-isolated Fe(III) sample, suggesting that this signal arises as a result of oxidation.

The broad, low-spin signal seen upon oxidation of M105L and M104L/M105L does not arise from heme bearing a neutral Cys^{94}(thiol). Protein isolated in the Fe(II)-CO state likely contains Cys^{94} in the protonated form since it is not bound to the heme. To determine whether the bound Cys^{94}(thiolate) could be protonated, we recorded EPR spectra of as-isolated WT and M105L RcoM-1 over a pH range from 6.8 to 8.6 (Figure 3.4). This pH range overlaps with known pK_a values of heme-coordinated thiols, which range from 6 to 7.[34] A small increase in the axial, high-spin signal occurs concomitant with a small decrease in the rhombic, low-spin signal as the pH is lowered. This observation suggests that in a minor fraction, Cys^{94} may be protonated and lost as a ligand, forming a species analogous to that observed in the RcoM-2 variant C94S.[18] No evidence for a broad low-spin signal bearing a Cys^{94}(thiol) ligand was observed in as-isolated samples at lower pH. We therefore conclude that the presence of such signals in oxidized samples is a consequence of the oxidation process. These observations suggest that Met105 is necessary for reversible redox-mediated ligand switching at the RcoM-1 heme.

Met^{104} is an axial ligand in Fe(II) RcoM-1

Electronic absorption and MCD spectral data illustrate that Met^{105} is not a heme-coordinating ligand in the Fe(II) state of RcoM-1. Reduction of Fe(III) M105L RcoM-1 with sodium dithionite results in an electronic absorption spectrum indicative of low-spin (S = 0), six-coordinate Fe(II) heme. The electronic absorption spectrum of Fe(II) M105L exhibits a Soret band maximum at 426 nm and sharp α and β bands at 562 and 532 nm; this spectrum is essentially identical to that of WT RcoM-1 (Figure 3.5a, 3.5b top panels). The spectral similarity suggests that Fe(II) M105L retains the native His/Met axial heme ligation present in WT RcoM-1 and RcoM-2.[12,19] The MCD

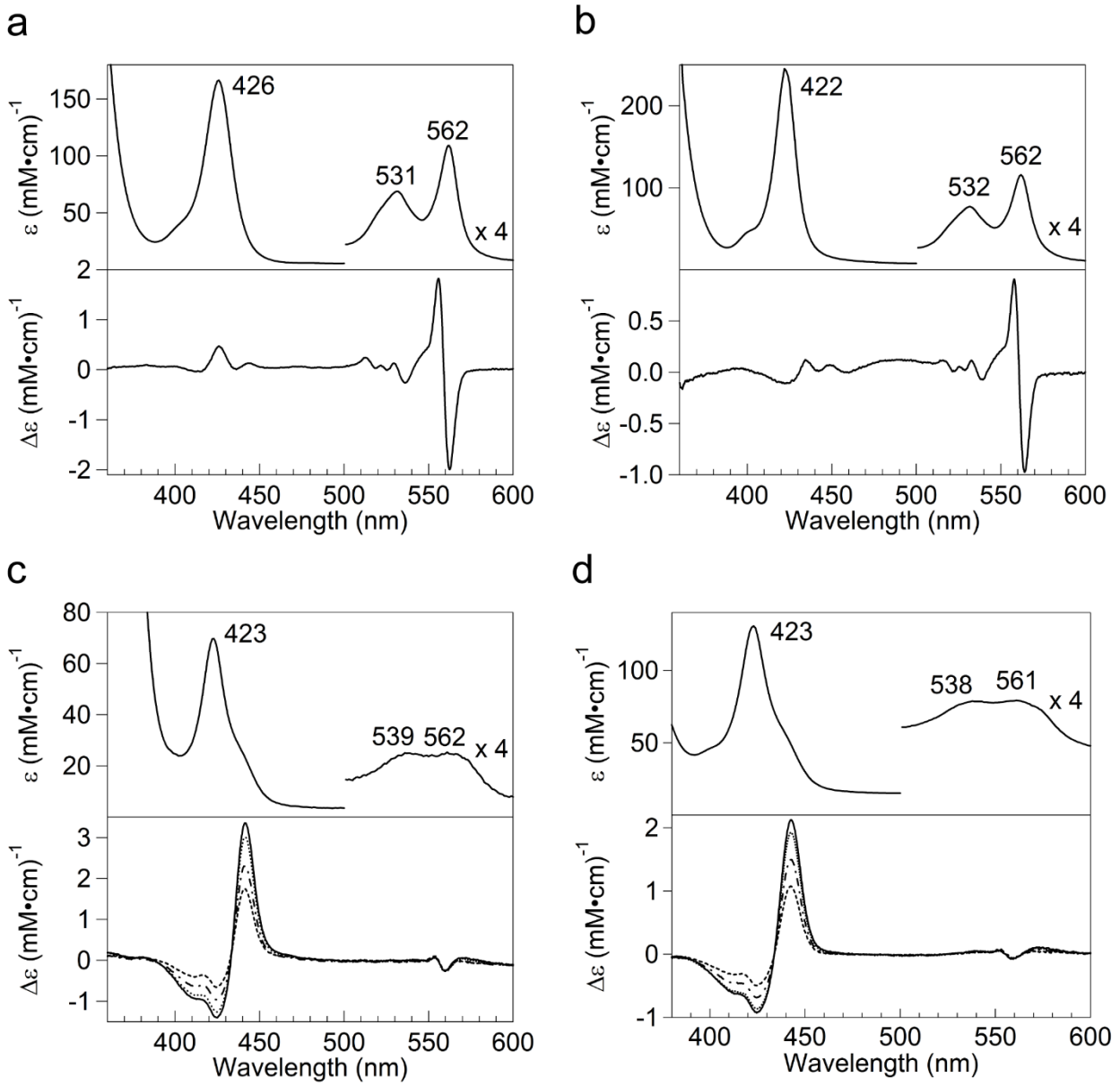


Figure 3.5 Electronic absorption (top panels) and MCD (bottom panels) spectra of oxidized RcoM-1 variants: (a) WT (2.2 μM heme), (b) M105L (16.5 μM heme), (c) M104L (7.5 μM heme) and (d) M104L/M105L (12 μM heme). For electronic absorption at 25 °C, samples in 25 mM MOPS with 500 mM KCl (pH 7.4) were reduced with a few grains of solid sodium dithionite under Ar. The MCD samples were prepared under N_2 in 25 mM MOPS and 500 mM KCl buffer (pH 7.4) with 60% (v/v) glycerol at a final concentration of 10 μM heme. Samples were reduced with solid sodium dithionite prior to the addition of glycerol. Spectra shown are the average of three scans collected at constant magnetic field (7 T) and: 4.5 K, 8 K, 15 K and 25 K (c,d) or 50 K (a,b).

spectrum of M105L RcoM-1 supports the assignment of the low-spin, Fe(II) heme coordination environment, and again is indistinguishable from that of WT RcoM-1 (Figures 3.4a, 3.4b bottom panels). Both spectra exhibit an intense, temperature-independent *A* term in the α-β region with a crossover at 559 nm and a weak, temperature-independent feature in the Soret region. Observation of these features implies that the Fe(II) heme is low-spin ($S = 0$) and six-coordinate. The *A* term crossover position at 559 nm, which is similar to that of several other His/Met ligated Fe(II) hemes including RcoM-2 (560 nm) and *Ec*Dos (561 nm), is consistent with His/Met axial ligation in M105L RcoM-1.[19,21,35] Maintenance of native His/Met coordination in Fe(II) M105L demonstrates that Met^{105} is not a heme-coordinating ligand in RcoM-1.

The electronic absorption and MCD spectral data for RcoM-1 variants in which Met^{104} is removed (M104L and M104L/M105L) reveal a five-coordinate, high-spin heme, confirming that Met^{104} is an axial ligand in Fe(II) RcoM-1. Electronic absorption spectra of Fe(II) M104L and M104L/M105L variants contain a Soret band at 423 nm which is blue shifted relative to those of WT and M105L. Weak α and β bands are also observed around 561 and 539 nm, respectively, for both M104L and M104L/M105L (Figures 3.5c, 3.5d top panels). These observations suggest that the Fe(II) heme environment in M104L and M104L/M105L RcoM-1 is perturbed. MCD data demonstrate that the Fe(II) heme is five-coordinate, high-spin in these variants (Figures 3.4c, 3.4d bottom panels). The M104L-containing variants both exhibit a strong, temperature-dependent Soret signal that includes a trough at 428 nm with a shoulder at 425 nm, a crossover point at 433 nm, and a peak at 441 nm. The temperature-dependent magnetization saturation behavior for the feature at 441 nm in Fe(II) M104L is indicative of a metal center with $S > ½$, which would only occur in a high-spin ($S = 2$) Fe(II) heme (Figure 3.6).[36] The anaerobically reduced a_3 heme center of cytochrome *c* oxidase exhibits low-temperature MCD spectra and magnetization saturation

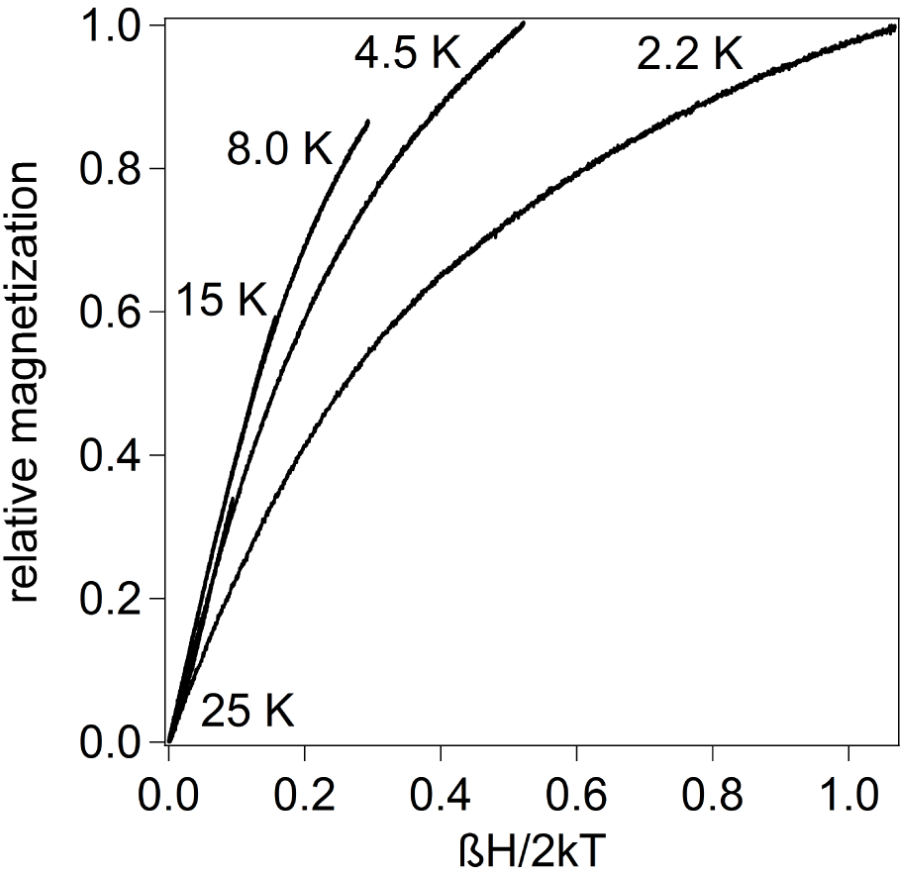


Figure 3.6 Magnetization saturation curves for the intense, positive MCD feature in the Soret region of high-spin (S = 2), Fe(II) M104L RcoM-1. Each curve plots the signal intensity at 441 nm as the magnetic field was varied from 0 to +7 T at the temperature noted. All signals were normalized to the most intense signal observed at 7 T and 2.2 K. Sample conditions matched those in Figure 3.5c.

behavior similar to those of M104L RcoM-1.[37] These similarities suggest that the Fe(II) heme in M104L RcoM-1 is also a high-spin (S = 2), five-coordinate heme with an axial histidine ligand.[37] The fact that removal of Met^{104} results in formation of this five-coordinate, high-spin species, allows for unequivocal identification of Met^{104} as the sixth axial ligand to in Fe(II) RcoM-1.

3.4 Discussion

The RcoM proteins from *P. xenovorans* represent unique, single-component signaling proteins that contain both PAS-like heme binding and LytTR-like DNA-binding domains. RcoM, unlike other proteins using these architectures, incorporates both gas-sensing and DNA-binding functions in one protein. The regulatory domain most likely resembles a PAS domain and contains a heme b cofactor.[12] In other known heme-containing PAS domain proteins found in prokaryotes, gas binding triggers either histidine kinase or phosphodiesterase activity as part of a two-component signaling pathway.[38] However, in RcoM, the expected output signal is DNA binding through a LytTR domain. The majority of LytTR-containing transcription factors are involved in two-component signaling and, as such, are activated upon phosphorylation by the partner sensor kinase.[39,40] In contrast, the suggested mechanism for RcoM is that CO binding to the heme activates the LytTR domain to bind to the DNA promoter region of DNA. [12,20]

The heme in RcoM undergoes a redox-mediated ligand switch that allows for CO binding under reducing conditions. Like the functionally similar CooA, Fe(III) RcoM-1 is axially ligated by a cysteine thiolate and a neutral donor ligand.[19] In CooA, the trans ligand is an N-terminal proline from the opposite monomer; whereas in RcoM, the trans ligand is His^{74}, presumably derived from the same polypeptide as the coordinating Cys^{94}.[41] The thiolate is lost under reducing conditions and is replaced by a second neutral donor, histidine in CooA and a previously unknown

ligand in RcoM.[9,19] The more weakly coordinating residue, Pro^2 in CooA or the neutral donor in RcoM, is replaced by CO.

Homology modelling of RcoM-1 based on *Ec*DOS in the Fe(II) state suggested that the ligand was Met^{104}.[12] *Ec*DOS, another protein with a heme-bound PAS domain, senses O_2 through a redox-mediated ligand-switch. In Fe(II) *Ec*DOS, Met^{95} coordinates trans to His^{77} and is replaced by O_2. Phosphodiesterase activity is observed for both Fe(II) and Fe(II)-O_2 *Ec*DOS. Oxidation to Fe(III) *Ec*DOS, in which water replaces Met^{95}, inactivates the protein.[42] Even eight years after the original homology model of RcoM-1 was generated, *Ec*DOS continues to be predicted as the closest structural relative of RcoM using the I-TASSER server.[43-45] We re-implemented the model using Modeller in UCSF Chimera for further analysis.[22-25] Threading of the RcoM-1 sequence on the *Ec*DOS Fe(II) structure (Figure 3.1) gives good secondary structure agreement (RMSD 0.627 Å for 107 residues); however, only His^{74} obtains a position near the heme. Cys^{94} and the Met^{104}/Met^{105} pair are located on opposite sides of the heme pocket. Of the two methionines, Met^{104} appears to be the more likely candidate for the Fe(II) ligand as it is predicted to be in the pocket whereas Met^{105} is solvent exposed.

In this work, we determined that Met^{104} is the CO-replaceable neutral donor ligand trans to His^{74} in the Fe(II) heme of RcoM-1. Replacing Met^{105} with a non-coordinating leucine residue results in electronic absorption and MCD spectra that are nearly identical to those arising from the six-coordinate, low-spin Fe(II) heme of WT RcoM-1 (Figure 3.5a,b). This result suggests that Met^{105} is not an axial ligand in Fe(II) RcoM-1. In contrast, replacing Met^{104} with a non-coordinating leucine residue results in electronic absorption and MCD spectra that are indicative of a high-spin (S = 2), five-coordinate species (Figure 3.5c,d). Specifically, Fe(II) M104L and M104L/M105L exhibit blue-shifted Soret bands and the loss of distinct features in the α/β region.

A strong, temperature-dependent Soret feature was observed in the low-temperature MCD spectra of both variants without Met^{104}. Magnetization saturation behavior of this feature in Fe(II) M104L supports the assignment of a high-spin ($S > ½$) heme (Figure 3.6). Similar MCD signals and magnetization saturation behavior were observed for the Fe(II) a_3 heme center in anaerobically reduced cytochrome c oxidase, which exists as a five-coordinate, high-spin species axially ligated by a histidine residue.[37] The formation of such a five-coordinate, high-spin heme in Fe(II) RcoM-1 variants without Met^{104}, confirms that this residue is an axial ligand in Fe(II) RcoM-1.

A complete picture of the RcoM-1 ligand-switching mechanism may now be formed (Figure 3.7). Similar to *Ec*DOS, the histidine ligand in RcoM-1 is retained throughout the ligand switch process.[46] This mechanism is subtly different than that of CooA, where a ligand switch occurs on both faces of the heme and neither of the original ligands are bound after activation upon binding CO.[47] In Fe(III) RcoM-1, the Cys^{94}(thiolate) acts as the ligand opposite His^{74}. Met^{104} replaces Cys^{94} concomitant with reduction of the heme iron. CO replaces the more weakly bound Met^{104} ligand while His^{74} remains bound. All of the Met variants bind CO and exhibit the same spectral features as WT RcoM-1 in the CO adducts (data not shown). Many batches of WT RcoM-1 and its variants are isolated in the Fe(II)-CO state; photolysis and chemical oxidation may be employed to convert the CO-adduct to the Fe(III) state. Upon loss of Met^{105}, the chemical oxidation does not necessarily result in coordination of the Cys^{94}(thiolate).

EPR data suggest that Met^{105} stabilizes Cys^{94}(thiolate) ligation to the Fe(III) heme of RcoM-1. When isolated in the Fe(III) state, WT RcoM-1 and the M105L variant exhibit a single narrow, rhombic low-spin signal characteristic of His/Cys(thiolate) ligation.[19,30,31] When WT RcoM-1 is isolated in the Fe(II)-CO form and oxidized to Fe(III) with ferricyanide, the proteins exhibit the same narrow signal. In contrast, when variants bearing alterations at Met^{105} (M105L

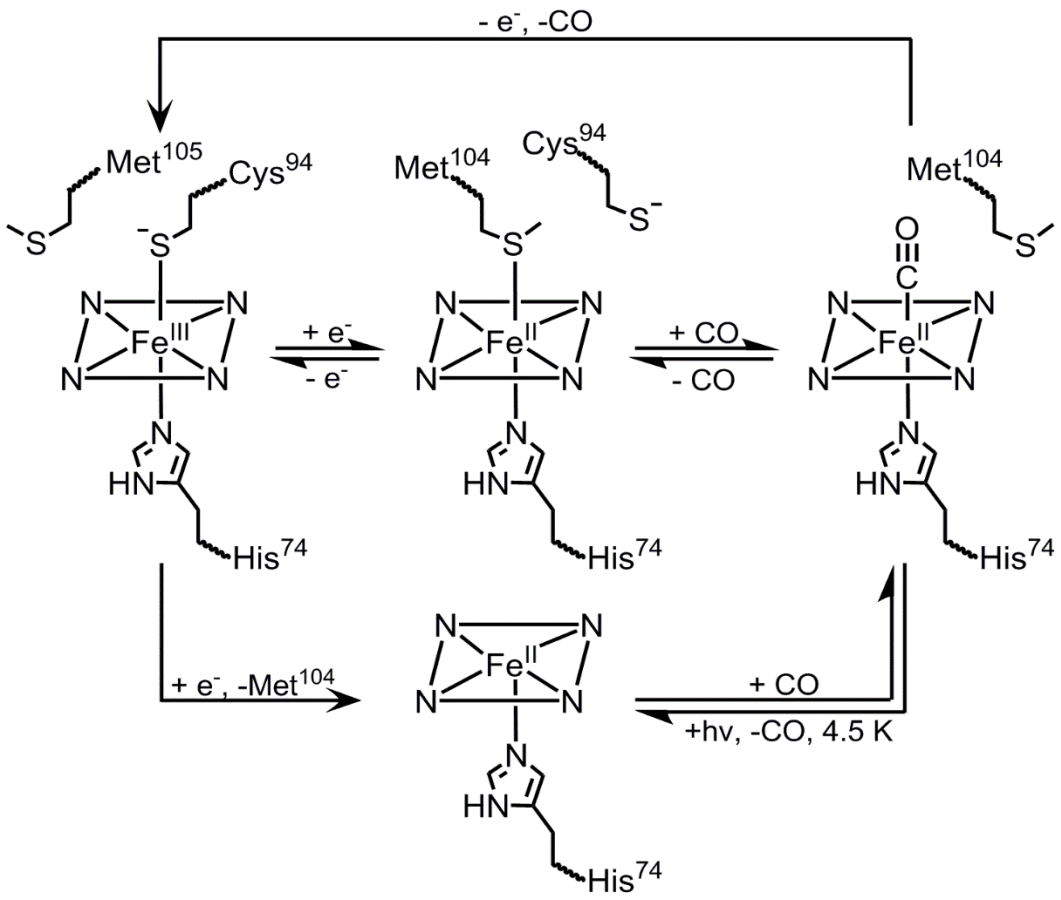


Figure 3.7 Proposed ligand switch mechanism and other observed heme coordination states for RcoM-1. His^{74} is retained as a ligand throughout. As-isolated RcoM-1 contains an admixture of ferric and ferrous-CO heme. Oxidation by ferricyanide converts the CO adduct back to Fe(III) heme. Under reducing conditions, Met^{104} replaces Cys^{94} in WT RcoM-1. Met^{104} is then replaced upon binding CO. Reduction of M104L variants or photolysis of any Fe(II)-CO adduct at liquid helium temperatures generates a high-spin, five-coordinate heme in which His^{74} is retained. While not a heme ligand, Met^{105} plays a role in stabilizing the Fe(III) coordination environment.

and M104L/M105L) are oxidized, a broad, rhombic low-spin signal appears as the major component. The narrow, rhombic low-spin signal is present as a minor component. The broad, rhombic low-spin signal exhibits *g* values comparable to those of other heme proteins axially coordinated by two neutral ligands; therefore, we postulate that this signal arises from such an Fe(III) heme.[32,33] These data lead to two distinct conclusions. First, as-isolated Fe(III) protein may differ from chemically oxidized Fe(III) protein. Second, w 1 hen chemically oxidized in the absence of Met105, a significant fraction of the resulting Fe(III) heme lacks the Cys94(thiolate). Thus, when Met105 is absent, the ligand switch between Met104 and Cys94(thiolate) is not reversible, implicating Met105 in stabilizing the heme-coordinating Cys94(thiolate) ligand.

Second sphere residues are known to stabilize thiolate coordination in many other heme-thiolate proteins. For example, in cytochrome P450cam from P. putida, a hydrogen bond exists between the backbone amide proton of residue Leu358 and the coordinating thiolate.[48,49] Deletion of this hydrogen bond results in changes to the EPR spectrum. In the L358P variant, a broad, rhombic low-spin signal appears in addition to the narrow, rhombic low-spin signal observed in the WT protein. This result parallels what we observe for oxidized Met105 variants in RcoM-1, suggesting that there may be a secondary interaction between Met105 and the heme coordination sphere. The aforementioned structural model of RcoM-1 does not predict a role for Met105 as a second-sphere residue, because in that model Met105 is solvent exposed on the surface of the protein. Based on these data, we conclude that further studies are needed to refine the RcoM structural model and to define the heme pocket interactions.

RcoM shares similar heme coordination with CooA and a similar structural fold with *Ec*DOS; however, it appears likely that RcoM functions in a distinct manner. A key difference between RcoM and CooA is that as-isolated RcoM-1 contains a significant amount of CO-bound

heme, whereas CooA purifies purely in the Fe(III) state. This difference implies that RcoM-1 must have a higher affinity for CO, since the adduct forms even under aerobic expression conditions where ambient CO levels are reported to be 60-300 ppb.[50] A key difference between RcoM-1 and *Ec*DOS is that RcoM-1 directly couples PAS domain sensory activity to DNA-binding, whereas *Ec*DOS regulates transcription through phosphodiesterase-mediated two-component signaling.[42] It remains unclear how changes in heme coordination in RcoM affect DNA-binding through the LytTR domain, as this type of one-component protein has not previously been seen. Further studies are required to determine which heme coordination states are relevant to RcoM-1 transcriptional regulation. Additionally, the importance of RcoM in regulating aerobic CO metabolism has yet to be demonstrated. A recent DNAase footprinting study showed that RcoM-1 interacts with up to six repeat motifs found upstream of the *cox* operon.[20] The proposed RcoM binding sites do not bear resemblance to those seen for other LytTR-containing transcription factors.[51] We observed that in *in vitro* fluorescence polarization assays, RcoM does not bind to the sequence that was previously reported by Kerby and Roberts to have the highest activity. Further studies are required to determine whether and how RcoM-1 regulates expression of the *cox* genes in response to CO.

3.5 References

1. Browning, D. F.; Busby, S. J. W. The regulation of bacterial transcription initiation. *Nat. Rev. Micro.* **2004**, *2*, 57.

2. del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure* **2009**, *17*, 1042.

3. Bunn, H. F.; Poyton, R. O. Oxygen sensing and molecular adaptation to hypoxia. *Physiol. Rev.* **1996**, *76*, 839.

4. *Bacterial Metabolism*; 4th ed.; Jurtshuk Jr., P., Ed.; University of Texas Medical Branch at Galveston: Galveston, TX, 1996.

5. Gilles-Gonzalez, M.-A.; Gonzalez, G. Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *J. Inorg. Biochem.* **2005**, *99*, 1.

6. Shelver, D.; Kerby, R. L.; He, Y.; Roberts, G. P. Carbon monoxide-induced activation of gene expression in Rhodospirillum rubrum requires the product of cooA, a member of the cyclic AMP receptor protein family of transcriptional regulators. *J. Bacteriol.* **1995**, *177*, 2157.

7. Shelver, D.; Kerby, R. L.; He, Y.; Roberts, G. P. CooA, a CO-sensing transcription factor from Rhodospirillum rubrum, is a CO-binding heme protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 11216.

8. Lanzilotta, W. N.; Schuller, D. J.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure of the CO sensing transcription activator CooA. *Nat. Struct. Biol.* **2000**, *7*, 876.

9. Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Chung, S.-Y.; Roberts, G. P.; Reynolds, M. F.; Parks, R. B.; Burstyn, J. N. Identification of two important heme site residues (Cysteine 75 and Histidine 77) in CooA, the CO-sensing transcription factor of *Rhodospirillum rubrum*. *Biochemistry* **1999**, *38*, 2669.

10. Kerby, R. L.; Youn, H.; Thorsteinsson, M. V.; Roberts, G. P. Repositioning about the dimer interface of the transcription regulator CooA: A major signal transduction pathway between the effector and DNA-binding domains. *J. Mol. Biol.* **2003**, *325*, 809.

11. Puranik, M.; Nielsen, S. B.; Youn, H.; Hvitved, A. N.; Bourassa, J. L.; Case, M. A.; Tengroth, C.; Balakrishnan, G.; Thorsteinsson, M. V.; Groves, J. T.; McLendon, G. L.; Roberts, G. P.; Olson, J. S.; Spiro, T. G. Dynamics of carbon monoxide binding to CooA. *J. Biol. Chem.* **2004**, *279*, 21096.

12. Kerby, R. L.; Youn, H.; Roberts, G. P. RcoM: A new single-component transcriptional regulator of CO metabolism in bacteria. *J. Bacteriol.* **2008**, *190*, 3336.

13. Chain, P. S. G.; Denef, V. J.; Konstantinidis, K. T.; Vergez, L. M.; Agulló, L.; Reyes, V. L.; Hauser, L.; Córdova, M.; Gómez, L.; González, M.; Land, M.; Lao, V.; Larimer, F.; LiPuma, J. J.; Mahenthiralingam, E.; Malfatti, S. A.; Marx, C. J.; Parnell, J. J.; Ramette, A.; Richardson, P.; Seeger, M.; Smith, D.; Spilker, T.; Sul, W. J.; Tsoi, T. V.; Ulrich, L. E.; Zhulin, I. B.; Tiedje, J. M. Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15280.

14. King, G. M. Nitrate-dependent anaerobic carbon monoxide oxidation by aerobic CO-oxidizing bacteria. *FEMS Microbiol. Ecol.* **2006**, *56*, 1.

15. Winsor, G. L.; Khaira, B.; Van Rossum, T.; Lo, R.; Whiteside, M. D.; Brinkman, F. S. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics* **2008**, *24*, 2803.

16. Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17*, 1282.

17. Galperin, M. Y. Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. *J. Bacteriol.* **2006**, *188*, 4169.

18. Smith, A. T.; Marvin, K. A.; Freeman, K. M.; Kerby, R. L.; Roberts, G. P.; Burstyn, J. N. Identification of Cys(94) as the distal ligand to the Fe(III) heme in the transcriptional regulator RcoM-2 from Burkholderia xenovorans. *J. Biol. Inorg. Chem.* **2012**, *17*, 1071.

19. Marvin, K. A.; Kerby, R. L.; Youn, H.; Roberts, G. P.; Burstyn, J. N. The transcription regulator RcoM-2 from Burkholderia xenovorans is a cysteine-ligated hemoprotein that undergoes a redox-mediated ligand switch. *Biochemistry* **2008**, *47*, 9016.

20. Kerby, R. L.; Roberts, G. P. Burkholderia xenovorans RcoM(Bx)-1, a Transcriptional Regulator System for Sensing Low and Persistent Levels of Carbon Monoxide. *J. Bacteriol.* **2012**, *194*, 5803.

21. Hirata, S.; Matsui, T.; Sasakura, Y.; Sugiyama, S.; Yoshimura, T.; Sagami, I.; Shimizu, T. Characterization of Met95 mutants of a heme-regulated phosphodiesterase from Escherichia coli. *Eur. J. Biochem.* **2003**, *270*, 4771.

22. Fiser, A.; Sali, A. In *Macromolecular Crystallography, Pt D*; Carter, C. W., Sweet, R. M., Eds. 2003; Vol. 374, p 461.

23. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2014**, *47*, 5.6.1.

24. Yang, Z.; Lasker, K.; Schneidman-Duhovny, D.; Webb, B.; Huang, C. C.; Pettersen, E. F.; Goddard, T. D.; Meng, E. C.; Sali, A.; Ferrin, T. E. UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **2012**, *179*, 269.

25. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605.

26. Stewart, V.; Lu, Y.; Darwin, A. J. Periplasmic Nitrate Reductase (NapABC Enzyme) Supports Anaerobic Respiration by Escherichia coli K-12. *J. Bacteriol.* **2002**, *184*, 1314.

27. Dykxhoorn, D. M.; StPierre, R.; Linn, T. A set of compatible tac promoter expression vectors. *Gene* **1996**, *177*, 133.

28. De Duve, C. A spectrophotometric method for the simultaneous determination of myoglobin and hemoglobin in extracts of human muscle. *Acta Chem. Scand.* **1948**, *2*, 264.

29. Brittain, T.; Greenwood, C.; Springall, J. P.; Thomson, A. J. The nature of ferrous haem protein complexes prepared by photolysis. *Biochim. Biophys. Acta- Prot. Struct. Mol. Enzymol.* **1982**, *703*, 117.

30. Reynolds, M. F.; Shelver, D.; Kerby, R. L.; Parks, R. B.; Roberts, G. P.; Burstyn, J. N. EPR and Electronic Absorption Spectroscopies of the CO-Sensing CooA Protein Reveal a Cysteine-Ligated Low-Spin Ferric Heme. *J. Am. Chem. Soc.* **1998**, *120*, 9080.

31. Brunel, A.; Wilson, A.; Henry, L.; Dorlet, P.; Santolini, J. The Proximal Hydrogen Bond Network Modulates Bacillus subtilis Nitric-oxide Synthase Electronic and Structural Properties. *J. Biol. Chem.* **2011**, *286*, 11997.

32. Cheesman, M. R.; Zumft, W. G.; Thomson, A. J. The MCD and EPR of the heme centers of nitric oxide reductase from Pseudomonas stutzeri: evidence that the enzyme is structurally related to the heme-copper oxidases. *Biochemistry* **1998**, *37*, 3994.

33. Brautigan, D. L.; Feinberg, B. A.; Hoffman, B. M.; Margoliash, E.; Preisach, J.; Blumberg, W. E. Multiple low spin forms of the cytochrome c ferrihemochrome. EPR spectra of various eukaryotic and prokaryotic cytochromes c. *J. Biol. Chem.* **1977**, *252*, 574.

34. Sono, M.; Andersson, L. A.; Dawson, J. H. Sulfur donor ligand binding to ferric cytochrome P-450-CAM and myoglobin. Ultraviolet-visible absorption, magnetic circular dichroism, and electron paramagnetic resonance spectroscopic investigation of the complexes. *J. Biol. Chem.* **1982**, *257*, 8308.

35. Weber, C.; Michel, B.; Bosshard, H. R. Spectroscopic analysis of the cytochrome c oxidase-cytochrome c complex: circular dichroism and magnetic circular dichroism measurements reveal change of cytochrome c heme geometry imposed by complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 6687.

36. Rudowicz, C. Spectroscopic Properties of High-Spin Ferrous Ion in Complexes with Rhombic Distortions. *Acta Phys. Pol., A* **1977**, *51*, 515.

37. Thomson, A. J.; Johnson, M. K. Mangnetization curves of hemoproteins measured by low-temperature magnetic circular-dichroism spectroscopy. *Biochem. J* **1980**, *191*, 411.

38. Gilles-Gonzalez, M. A.; Gonzalez, G. Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *J. Inorg. Biochem.* **2005**, *99*, 1.

39. Galperin, M. Y. Telling bacteria: Do not LytTR. *Structure* **2008**, *16*, 657.

40. Gao, R.; Mack, T. R.; Stock, A. M. Bacterial response regulators: versatile regulatory strategies from common domains. *Trends Biochem. Sci.* **2007**, *32*, 225.

41. Thorsteinsson, M. V.; Kerby, R. L.; Conrad, M.; Youn, H.; Staples, C. R.; Lanzilotta, W. N.; Poulos, T. J.; Serate, J.; Roberts, G. P. Characterization of variants altered at the N-terminal

proline, a novel heme-axial ligand in CooA, the CO-sensing transcriptional activator. *J. Biol. Chem.* **2000**, *275*, 39332.

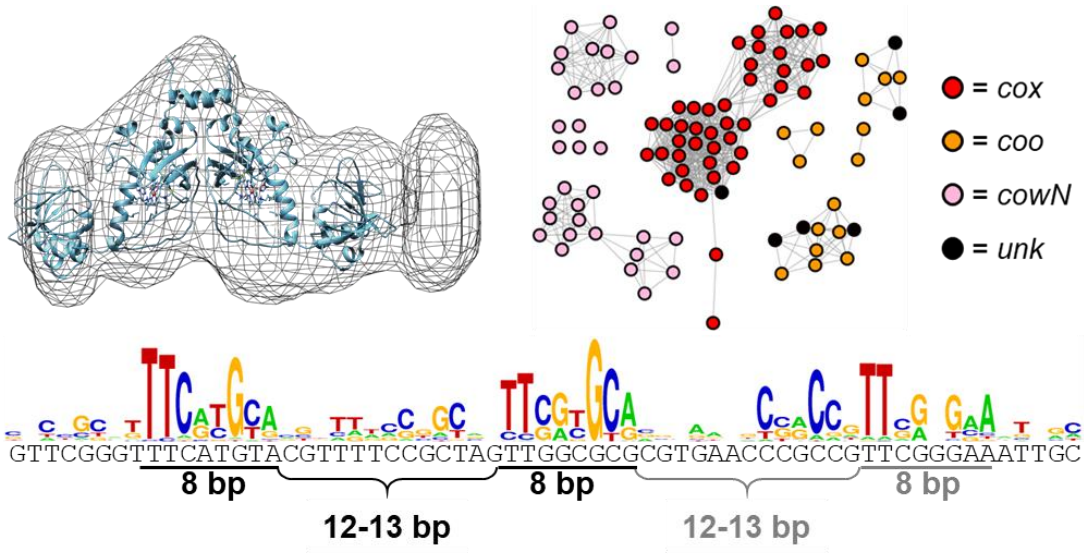
42. Sasakura, Y.; Yoshimura-Suzuki, T.; Kurokawa, H.; Shimizu, T. Structure-function relationships of EcDOS, a heme-regulated phosphodiesterase from Escherichia coli. *Accounts Chem. Res* **2006**, *39*, 37.
43. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725.
44. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Met.* **2015**, *12*, 7.
45. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **2008**, *9*.
46. Gonzalez, G.; Dioum, E. M.; Bertolucci, C. M.; Tomita, T.; Ikeda-Saito, M.; Cheesman, M. R.; Watmough, N. J.; Gilles-Gonzalez, M.-A. Nature of the Displaceable Heme-Axial Residue in the EcDos Protein, a Heme-Based Sensor from Escherichia coli. *Biochemistry* **2002**, *41*, 8414.
47. Reynolds, M. F.; Parks, R. B.; Shelver, D.; Thorsteinsson, M. V.; Kerby, R. B.; Roberts, G. P.; Burstyn, J. N. Spectroscopy of CO-sensing CooA reveals a ligand displacement mechanism for CO activation. *J. Inorg. Biochem.* **1999**, *74*, 276.
48. Galinato, M. G. I.; Spolitak, T.; Ballou, D. P.; Lehnert, N. Elucidating the role of the proximal cysteine hydrogen-bonding network in ferric cytochrome P450cam and corresponding mutants using magnetic circular dichroism spectroscopy. *Biochemistry* **2011**, *50*, 1053.
49. Yoshioka, S.; Tosha, T.; Takahashi, S.; Ishimori, K.; Hori, H.; Morishima, I. Roles of the proximal hydrogen bonding network in cytochrome P450cam-catalyzed oxygenation. *J. Am. Chem. Soc.* **2002**, *124*, 14571.

50. King, G. M. Molecular and Culture-Based Analyses of Aerobic Carbon Monoxide Oxidizer Diversity. *Appl. Environ. Microbiol.* **2003**, *69*, 7257.

51. Sidote, D. J.; Barbieri, C. M.; Wu, T.; Stock, A. M. Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure* **2008**, *16*, 727.

Chapter Four

Organization and DNA Binding Properties of the Heme-Dependent, CO-Sensing Transcriptional Regulator *Px*RcoM-1



A version of this chapter will be submitted as:

Dent, M.R.*; Bowman, H.E.*; Rodriguez-Martinez, J.; Cai, K; McCaslin, D. R.; Young, B.H.; Weaver, B.R.; Bhimsaria, D.; Berndsen, C.H.; Ansari, A.Z.; Burstyn, J.N. Organization and DNA Binding Properties of the Heme-Dependent, CO-Sensing Transcriptional Regulator *Px*RcoM-1.

*These authors contributed equally to this work

K.C. ran SAXS experiments at NMRFAM. J.R.M. performed SELEX and D.B. analyzed data in the lab of A.Z.A. D.R.M. performed AUC-SE experiments at the UW-Madison BIF. M.R.D. and H.E.B. carried out all other laboratory-based experimental procedures in the lab of J.N.B. B.H.Y. and C.H.B. developed the RcoM homology model. B.R.W. and M.R.D. carried out bioinformatics analyses.

4.1 Introduction

A diverse array of microorganisms found in oceans, soils, and freshwater environments utilize carbon monoxide (CO) as a source of energy and/or carbon.[1,2] In order to regulate expression of the complex (and in many cases oxygen-sensitive) molecular machinery required for CO metabolism, these organisms employ heme-dependent, CO-sensing transcription factors.[3,4] Only two such prokaryotic transcription factors have been identified and characterized to date: CooA (CO oxidation activator) and RcoM (regulator of CO metabolism).[5,6] CooA, a member of the well-studied CRP/FNR structural superfamily, represents the paradigm in transcriptional regulation via heme-based CO-sensing.[3,7,8] CO binding to Fe(II) heme in the CooA sensory domain allosterically activates DNA binding via a helix-turn-helix motif.[9-11] Binding of CooA to its consensus site upstream of the *coo* operon enhances the transcription of genes linked to anaerobic, oxidative CO metabolism.[5,12-14] RcoM, like CooA, utilizes heme to sense CO;[6,15-17] however, RcoM exhibits a much higher affinity for CO, and the structure of RcoM is thought to differ drastically from that of CooA.[18,19] As a result, RcoM is predicted to exhibit unique DNA binding properties and may undergo unique CO-driven allosteric regulation. By developing a more complete understanding of the structure and DNA binding properties of RcoM, we hope to better understand small molecule sensing in bacteria.

Identified in a growing number of prokaryotes, *rcoM* genes exist predominantly in three distinct genomic contexts. Initially, *rcoM* genes were identified through genetic analysis based on the sequence of *Px*RcoM-1, one of two RcoM orthologues from *P. xenovorans*.[6] As is the case for CooA, RcoM-encoding genes can be found adjacent to *cooS* genes, which encode oxygen-sensitive, Fe-Ni CO dehydrogenase (CODH) enzymes.[2,13] All *coo*-adjacent *rcoM* genes were identified in anaerobes that do not contain a *cooA* gene, an observation that is consistent with RcoM being a regulator of oxygen-sensitive CODH. Unlike CooA, *rcoM* genes were also

identified adjacent to *cox* operons, which encode oxygen-tolerant, molybdo-iron-sulfur-flavoprotein CODH enzymes.[6,20,21] Thus far, RcoM is the only known single-component transcriptional regulator of *cox* genes; however, two other putative CO-sensing transcriptional regulators, CoxC and CoxH, have been identified in *O. carboxidovorans*.[20,22] Unlike RcoM, CoxC and CoxH are proposed to be transmembrane proteins, although neither of these proteins have been studies in detail. Both RcoM- and CooA-encoding genes have been identified directly upstream of *cowN*, a gene that encodes a small (100 amino acid) accessory protein that mitigates CO toxicity in nitrogenase enzymes.[23] Specifically, CowN protects Mo-containing nitrogenases against inhibition by CO binding; however, the molecular mechanism underlying this protection is unkown.[24]

Based on sequence homology, RcoM is predicted to have a unique domain architecture comprised of an N-terminal sensory domain that adopts a PAS (Per-Arnt-Sim) fold and a C-terminal DNA binding domain that adopts a LytTR fold. The PAS fold is a sensory motif that is found in eukaryotes and prokaryotes; in bacteria, proteins with an N-terminal PAS domain and a C-terminal histidine kinase or phosphodiesterase domain are often utilized in two-component signal transduction pathways.[25] Several well-studied bacterial oxygen sensors possess a sensory PAS domain bearing heme, including FixL and *Ec*DOS, which have been structurally characterized.[4,26,27] In both of these proteins, changes in redox state and oxygen binding status at the heme directly modulate activity of the C-terminal enzymatic domain.[28,29] The LytTR fold exhibits a DNA binding motif distinct from the classic helix-turn-helix motif and is found in a number of virulence-associated transcriptional regulators in pathogenic bacteria, including AlgR from *P. aeruginosa*, VirR from *C. perfringens*, and AgrA from *S. aureus*.[30-33] A crystal structure of the LytTR domain of AgrA bound to a 15 base pair (bp) DNA oligomer uncovered interactions

between loops of the protein β-β-β fold and nucleobases in the DNA, and demonstrated that LytTR binding results in significant DNA bending.[34] While RcoM is composed of two well-characterized domains, this transcription factor is unique among PAS- and LytTR-containing proteins as it is the only known single-component fusion of these two motifs. As such, little is known about how RcoM interacts with DNA or how CO binding to heme modulates DNA binding activity.

A consensus DNA binding sequence has been identified for a number of bacterial transcription factors that contain a LytTR domain. The canonical DNA binding site for such proteins is a pair of direct repeats upstream of the -35 region of the relevant operon.[30,35-38] These repeats are typically imperfect, 9 bp in length, and spaced 10-13 bp apart; importantly, the spacing between repeats is critical in determining transcription factor binding strength.[39,40] One notable exception is the transcription factor AlgR, which regulates expression of *algD*, a gene involved in the production of the virulence factor alginate in *P. aeruginosa*.[41] This transcription factor binds two high affinity direct repeats spaced 66 bp apart and located far (~400 bp) upstream of *algD*, in addition to binding a third, low affinity site located 40 bp upstream of *algD*.[42-44]

The proposed consensus sites in RcoM-1 from *P. xenovorans* differ drastically from previously-identified LytTR DNA binding motifs. A combination of DNA footprinting, *in vivo* reporter, and *in vitro* fluorescence anisotropy data, led to the identification of two *Px*RcoM-1 binding regions spaced in the intergenic region between the $rcoM_1$ and $coxM_1$ genes.[18] One of these two binding regions is located directly upstream of the -10/-35 extension region preceding the $coxM_1$ gene, while the second binding region is ~100 bp further upstream. Both putative *Px*RcoM-1 binding regions contain three direct repeats with a 5'-TTnnnG-3' motif that modestly resembles repeat motifs observed in other LytTR-containing transcription factors.[30,33,35,45] Not only do the proposed number of direct repeats differ between RcoM (a triplet motif) and other

LytTR (a repeat pair) binding sites, but the spacing between direct repeats is nearly doubled for RcoM sites (21 bp) compared to other LytTR sites (10-13 bp).

In light of the discrepancies between the consensus binding site proposed for RcoM and those known for other LytTR-containing transcription factors, we sought to better characterize the structure of *Px*RcoM-1 and re-examine the proposed DNA binding site of this CO-sensing transcription factor. In this study, we demonstrate that *Px*RcoM-1 exists primarily as an elongated homodimer in solution through size-exclusion chromatography, analytical ultracentrifugation sedimentation equilibrium, and small-angle X-ray scattering experiments. Using this experimental insight in conjunction with homology modeling and molecular dynamics, we developed a structural model for *Px*RcoM-1. Using bioinformatics, we explore the relationship between RcoM protein sequence and genomic context. For *cox*-regulating RcoM proteins, we identified a consensus DNA binding motif that differs slightly from that originally proposed. Comparison of this consensus motif with an 8 bp cognate motif identified through *in vitro* SELEX-CSI experiments reveals that the $coxM_1$ promoter from *P. xenovorans* is not optimized for *Px*RcoM-1 binding.

4.2 Materials and Methods

Materials

All chemicals used in buffer and media preparation (99% purity or greater, Research Products International Corporation), sodium dithionite (85% purity, Fluka), and CO gas (99.5% purity, Air Gas) were used as received. All other chemical reagents were purchased from Sigma-Aldrich and used as received. Oligonucleotides for site-directed mutagenesis were synthesized by the Biotechnology Center (University of Wisconsin-Madison). Cell stocks of *E. coli* containing wild

type (WT), C94S, and heme binding domain (HBD) truncate *Px*RcoM-1 were originally provided by Dr. Robert Kerby (University of Wisconsin-Madison, Department of Bacteriology).[6]

Site-directed mutagenesis

The *Px*RcoM-1 variant H74A was generated by mutagenesis of the cloned H74Y *Px*RcoM-1 variant using the QuickChange II Protocol (Agilent/Stratagene). Primers for the Y74A mutations were created using Agilent's online tool (forward: 5'-cggcttttctccggggccagttgcaccacgt-3'; reverese: 5'-acgtggtgcaactggccccggagaaaagccg-3'). After 16 cycles of polymerase chain reaction, the resulting products were treated with DpnI to remove methylated template DNA and transformed by heat shock into chemically-competent DH5α *E. coli* cells as well as VJS6737 expression cells. Sanger sequencing of all resulting mutant *Px*RcoM-1 genes was carried out by Functional Biosciences (Madison, WI) to verify the presence of the appropriate mutations.

Protein expression and purification

*Px*RcoM-1 and variants were isolated and purified in a manner similar to that previously described.[6,15-18] *E. coli* VJS6737[46] containing the pEXT20 expression vector[6,47] for each *P. xenovorans* RcoM-1 variant was grown in Lennox-LB medium supplemented with ferric citrate (20 mg/L) and nutrient broth (3 g/L). Dense overnight starter cultures were diluted 1:10 into fresh media and grown to an OD_{550} greater than 4 at 30 °C with shaking. These cultures were diluted to an OD_{550} of 0.06 in fresh media containing 50 uM isopropyl β-D-thiogalactopyranoside (IPTG) to induce expression for 19-20 h at 28 °C with shaking. Cells were harvested by centrifugation, resuspended in lysis buffer [100 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.4, 500 mM NaCl, 1 mM dithiothreitol (DTT)], and lysed by sonication. A HisTrap FF column (GE Lifesciences, 2x5 mL column volume) was pre-equilibrated on an AKTA Prime FPLC with binding buffer (10 mM imidazole, 100 mM MOPS, pH 7.4, 400 mM NaCl). The cell supernatant

was applied to the column and washed with binding buffer until the 280 nm absorbance reached baseline. Additional washing and elution of bound RcoM was achieved by adding elution buffer (350 mM imidazole, 100 mM MOPS, pH 7.4, 400 mM NaCl) to the column using a step gradient. The red-colored fractions were pooled and further purified by ammonium sulfate precipitation. The pooled fractions were incubated in 45%-saturated ammonium sulfate (v/v) and 0.5 mM DTT on ice for 15 min. Aggregated protein was harvested by centrifugation, and the resulting protein pellet resuspended in storage buffer (50 mM MOPS, pH 7.4, 150 mM NaCl, 1 mM DTT). To ensure complete desalting, resuspended protein was passed through a Sephadex G-25 gravity column pre-equilibrated with storage buffer. Dilute, desalted protein was concentrated using Amicon spin concentrators (Millipore, 10 kDa MWCO, 500 μL volume) and stored at -80 °C.

Total protein concentrations were determined using the bicinchoninic acid micro-assay or Pierce 660nm assay (Pierce-Thermo Fisher) using bovine serum albumin as a calibration standard; sodium dodecyl sulfate polyacrylamide gel electrophoresis verified that the protein purity was greater than 90%. Heme content was determined using the pyridine hemochromagen assay.[48] Protein concentration was confirmed by measuring absorbance at 280 nm using molar absorptivity values calculated using ExPASy (ε_{WT} = 14,440 $M^{-1}cm^{-1}$; ε_{HBD} = 6,990 $M^{-1}cm^{-1}$).[49,50]

Heme reconstitution

A stock solution of hemin chloride (100 μM in 1 mM NaOH) was added to protein in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT such that the final concentration of NaOH was less than or equal to 100 μM. Equimolar amounts of hemin chloride and protein monomer (as determined using the Pierce 660 assay) were mixed, resulting in an immediate color change from light pink to dark red. After incubating on ice for 5 min, excess hemin chloride was removed by

three successive rounds of concentration using a 600 µL 30K MWCO spin concentrator and dilution into 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT.

Electronic absorption spectroscopy

Electronic absorption spectra were recorded at room temperature with a double-beam Varian Cary 4 Bio spectrophotometer set to a spectral bandwidth of 0.5 nm. Protein samples were prepared in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT. The spectrophotometer was blanked with this same buffer to remove absorbance due to oxidized DTT. After purging the cuvette headspace with $Ar_{(g)}$, heme reduction was achieved by adding a few granules of solid sodium dithionite to the cuvette. Following heme reduction, 100-300 µL CO were added to the headspace, followed by gentle mixing to form the Fe(II)-CO heme adduct.

Size exclusion chromatography

Size exclusion chromatography (SEC) was performed on an AKTA Purifier FPLC equipped with a Superdex 200 Increase 10/300 column (GE Lifesciences) at a flow rate of 0.6 mL/min. For each experiment, the column was pre-equilibrated with the running buffer stated in the figure legends. Each 100 µL sample (41-280 µM) was injected using a 1 mL sample loop. Molecular weight standards (Sigma-Aldrich, catalog no. MWGF200) were prepared according to the manufacturer's instructions in running buffer. A calibration curve was created by plotting K_{av} vs. log (M_r) for the standards and fitting a linear function to the data. When indicated, data were normalized to the most intense peak for ease of plotting.

Analytical ultracentrifugation sedimentation equilibrium

The analytical ultracentrifugation sedimentation equilibrium (AUC-SE) experiment and data analysis was performed at the Biophysical Instrumentation Facility (University of Wisconsin-Madison) using a Beckman Coulter XL-A analytical ultracentrifuge. Heme absorption at 417 nm

without a cut-off filter was used to monitor the radial position of the samples. The samples (2.2 and 3.3 μM heme in 50 mM MOPS, pH 7.4, 150 mM NaCl, 1 mM DTT) were spun at 12.0k, 15.0k, 19.0k, and 26.0k rpm with gradients checked every 3-4 hours until the gradient super-imposed indicating that the samples had reached equilibrium. After data was collected at 26.0k rpm, the speed was reduced to 15.0k rpm to look for irreversible material loss. High speed depletion of protein material resulted in non-sedimentable absorbance and those values were used a baseline optical density correction during global fitting. A sequence molecular weight (M_s) of 29,451 Da and a partial specific volume of 0.743 mL/g were used based on the Uniprot sequence (Q13YL3). The density, as calculated by an increment table for NaCl and adjusted to 4 °C, was 1.01 g/mL.

Small-angle X-ray scattering

WT *Px*RcoM-1 was prepared for small-angle X-ray scattering (SAXS) experiments by first performing SEC using a 0.6 mL/min flow rate and 50 mM MOPS, pH 7.4, 150 mM NaCl, 1 mM DTT running buffer. The dimer peak was collected and concentrated with an Amicon 10k MWCO spin-concentrator (Millipore) followed by dilution to concentrations of 2, 4, and 8 mg/mL. The resulting solutions were passed through 0.26 μm spin filters to remove particulate matter. Samples were stored at 4 °C until run. No significant interparticle interactions were observed at any of the concentrations used in the SAXS studies. SAXS experiments were carried out on a Nanostar benchtop SAXS system (Bruker AXS) equipped with a rotating Cu anode Turbo X-ray Source and a VÅNTEC-2000 (2048 × 2048 pixel) detector. The sample-to-detector distance was set at ~1 m, allowing for the detection range: $0.012 > q > 0.240$ $Å^{-1}$. Protein and buffer samples (40 μL) were loaded separately into a capillary cell with 1 mm diameter, and scattering data were collected for 3 h with frames recorded every hour. Each frame was compared to check for radiation damage,

and none was detected over the course of the experiments. The SAXS data sets were then averaged and converted to 1D scattering profiles for further analysis.

The ATSAS software suite was used to process the SAXS data.[51] The radius of gyration (R_g) for each protein was determined using the Guinier approximation in the q range ($q_{max} \cdot R_g$) < 1.3. Pairwise distance distribution functions (P_r) were obtained using the GNOM software package. The output from GNOM was then used in conjunction with DAMMIF to generate 30 independent *ab initio* dummy atom models to assess the molecular shape of each sample.[52] As the protein is known to form a homodimer, P2 symmetry was enforced during the DAMMIF runs. Most of the models exhibited excellent agreement with experimental data and had a normalized spatial discrepancy (NSD) <1. Molecular weights were estimated from the correlation volume (V_c).[53]

Homology models and molecular dynamics simulations

Homology modeling for the plasmid sequence was performed in YASARA using the default settings.[54-56] The homology model was built using template structures 1D06, 1V9Y, and 1LT0 for the heme binding domain (HBD), and 3D6W for the LytTR domain to produce models of the heme-binding and LytTR domains independently.[27,28,57] Heme placement was determined by structural alignment of 1V9Y to the HBD homology model. Upon heme placement, molecular dynamics (MD) was performed on the HBD model in YASARA using the AMBER ff03. The simulation was equilibrated in explicit solvent at a temperature of 298 K, 0.9% NaCl, and pH 7.4, using the AMBER ff14 force field in a simulation cell with periodic boundaries.[55] To enforce heme positioning and coordination geometries, springs were added between the delta nitrogen of His[74] and the iron of the heme as well as between the carbon of carbon monoxide and the iron of the heme. Homology modeling was then run again with the modeled HBD and LytTR domains as

template structures to produce the final full-length model. Molecular dynamics was performed on the full length homology model in YASARA using the AMBER ff14 force field in order to improve the quality and determine the stability of the structure. Each simulation was equilibrated in explicit solvent at a temperature of 298 K, 0.9% NaCl, and pH 7.4, using the AMBER ff14 force field in a simulation cell with periodic boundaries.[55] Simulations were run with time step of 1.25 fs and temperature adjusted using a Berendsen thermostat as described by Krieger, et al..[54]

Systematic enhancement of ligands by exponential enrichment sequencing (SELEX-seq) and Cognate Sequence Identity (CSI) analysis

Three rounds of SELEX-seq were performed using WT *Px*RcoM-1 according to the method reported by Gu et al.[58] The following 40 bp randomized region, flanked by two 20 bp regions, was used:

5'-CTGATCCTACCATCCGTGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGACACTTCTGCCCAGGCGAG-3'

The cognate sequence identifier (CSI) approach was utilized to interpret results from SELEX-seq and determine the cognate 8 bp motif.[59,60]

Bioinformatics analyses

A sequence similarity network (SSN) containing putative *Px*RcoM proteins was generated using the single-sequence query function of the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) and visualized using Cytoscape 3.6.0.[61,62] The *Px*RcoM-1 protein sequence (UniProt ID: Q13YL3) was utilized as the query input, and the top 1,000 sequences with E-value scores of 10^{-5} or lower were included in the all-by-all BLAST. An initial 95% identity representative node network, generated using an *E*-value cutoff of 24, displayed three nodal clusters and two unassociated nodes. The second largest of the three clusters, which contained the *Px*RcoM-1 query sequence, exhibited two distinct sub-clusters. Both sub-clusters contained proteins annotated as RcoMs and a C-terminal LytTR domain; however, only proteins in the larger, more divergent sub-

cluster contained an annotated N-terminal PAS domain characteristic of RcoM. This sub-cluster, which also contained both *P. xenovorans* RcoM orthologues, was extracted and utilized in subsequent analyses.

The EFI Genome Neighborhood Tool (EFI-GNT) was used to identify the 20 genes upstream and downstream of each RcoM-encoding gene in the SSN. The EFI-GNT successfully identified genome neighborhoods (a 40 gene window centered on each *rcoM*) for 105 of the 112 putative RcoM proteins in the SSN. Inspection of genome neighborhoods allowed us to classify each RcoM protein based on its association with *cox*, *cowN*, or *coo* genes. Six *rcoM* genes, which did not contain *cox*, *cowN*, or *coo* genes within their genome neighborhood, were classified as unknown. Of the 99 successfully classified RcoMs, 44 were identified in the genomic context of a *cox* operon.

The European Nucleotide Archive IDs, provided by the EFI-GNT, were used to procure the nucleotide sequence of each putative *cox*-associated RcoM promoter region, identified as either the *rcoM*/*coxM* intergenic region or 200 base pairs (bp) upstream of *coxM*. These putative promoter regions, which varied in length from 147 to 547 nucleotides, were aligned using M-Coffee via the T-Coffee online web server and visualized using WebLogo.[63,64] An initial alignment revealed two conserved repeats upstream of the -10/-35 sequence in each putative promoter region; however, the disparately-sized promoter regions also gave rise to a large number of gaps in the sequence alignment. To eliminate these gaps, an additional alignment was carried out in which the size of the promoter region was restricted to 100 bp, centered on the second conserved repeat. To improve the clarity of the final alignment, four sequences containing inserts in the conserved direct repeat region were eliminated from the alignment.

4.3 Results

PxRcoM-1 is homodimeric and dimerization is mediated by the heme-binding domain

Size exclusion chromatography (SEC) data demonstrate that WT *Px*RcoM-1 exists as a homodimer in solution. Based on the *Px*RcoM-1 amino acid sequence (Uniprot ID: Q13YL3), a monomer should exhibit a monomeric molecular weight of 29.4 kDa and a dimeric molecular weight of 58.8 kDa. The majority *Px*RcoM-1 species observed in solution by SEC has an apparent mass of 70 kDa (Figure 4.1). While this mass is 11.2 kDa greater than expected for a homodimeric species, SEC assumes globular protein shape; thus, an apparent molecular weight larger than the expected molecular weight indicates that the protein adopts a non-spherical configuration. Several small peaks were observed at lower elution volumes than the dimeric fraction, suggesting that at least a small portion of the protein exists as higher order oligomers in solution. This observation is unsurprising given the tendency of prokaryotic PAS domains to form homo-oligomers.[65]

Analytical ultracentrifugation sedimentation equilibrium (AUC-SE) confirmed that *Px*RcoM-1 exists as a dimer (Figure 4.S1). At holoprotein concentrations of 2 to 3 μM (in the linear absorption range for the heme Soret feature detected at 417 nm), samples behaved as a single, ideal species with minimal loss of material due to aggregation. Global fits to the absorbance gradients gave a reduced molecular weight average (M_w) of 16,542 ± 40 Da. Comparing M_w to M_s (the sequence molecular weight) gave a M_w/M_s of 2.25 ± 0.01, confirming that heme-bound *Px*RcoM-1 is dimeric. In contrast to SEC, no higher-order species were observed in AUC-SE experiments. This discrepancy is best explained by considering that higher order oligomers likely form in a concentration-dependent manner and that protein concentrations are significantly lower in AUC-SE (0.1 mg/mL) than in SEC (5 mg/mL).

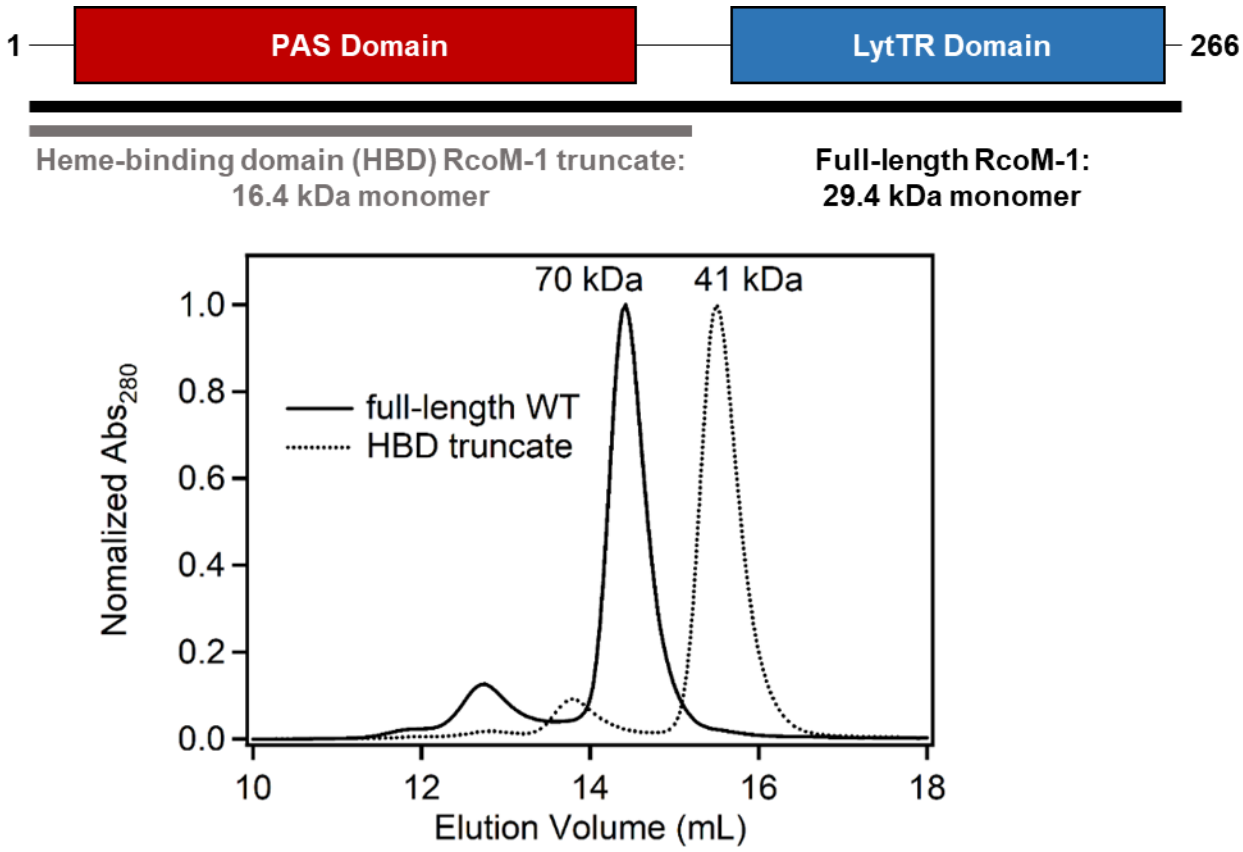


Figure 4.1 (Above) *Px*RcoM-1 domain architecture and molecular weights of full-length and heme-binding domain (HBD) truncate *Px*RcoM-1. (Below) Size exclusion chromatograms of full-length WT and HBD truncate *Px*RcoM-1. Protein samples (5.58 mg/mL monomer for WT and 3.90 mg/mL monomer for HBD truncate) in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT (100 µL) were injected via a 1 mL sample loop at a flow rate of 0.6 mL/min. The traces are normalized to the highest intensity peak.

A truncate of *Px*RcoM-1 containing residues 1-153 and encompassing the heme-binding domain (HBD) also predominantly exists as a non-spherical homodimer in solution. The calculated monomer molecular weight for the *Px*RcoM-1 HBD truncate is 16.4 kDa, while the majority truncate species runs with an apparent molecular weight of 44 kDa by SEC (Figure 4.1). The apparent molecular weight of the homodimeric truncate is somewhat (11.2 kDa) larger than that predicted by sequence alone, suggesting that, like full-length *Px*RcoM-1, the truncated HBD homodimer is not perfectly spherical. Importantly, these data demonstrate that dimerization in *Px*RcoM-1 occurs via the heme-binding domain. Several small peaks were observed at lower elution volumes than the dimeric fraction. As was true in the case of the full-length protein, the presence of this minority species suggests that a fraction of the protein exists as higher order oligomers in solution. The SEC data suggest that dimerization occurs via the heme-binding domain.

PxRcoM-1 binds one heme per monomer, but heme is not required for dimerization

Each *Px*RcoM-1 monomer contains one heme prosthetic group; however, the protein is not isolated fully heme loaded. Comparison of total protein concentration to concentration of heme for as-isolated protein reveals sub-stoichiometric (53.2 ± 1.2 % assuming a 1:1 monomer to heme ratio) heme loading per protein monomer (Table 4.1). Upon mixing of equimolar amounts of as-isolated protein monomer and hemin chloride, the percentage heme loading, assuming a 1:1 monomer to heme ratio, increased to 113.3 ± 7.8 %. The observation of heme loading near 100% suggests that each monomer binds a single heme cofactor. Mixing of additional equivalents of hemin chloride did not further increase heme loading (data not shown), corroborating the 1:1 protein to heme stoichiometry. The percent heme loading was slightly greater than 100%, suggesting that a small amount of heme may bind adventitiously to *Px*RcoM-1. This hypothesis is supported by EPR

Table 4.1 Heme quantitation of WT and HBD *Px*RcoM-1. Total protein concentration values were determined using the Pierce 660 assay, and heme concentration values were determined using the pyridine hemochromagen assay. Percentage heme loading assumes a 1:1 ratio of protein monomer to heme, and errors represent ± one standard deviation.

Protein	Trial	[protein] (μM)	[heme] (μM)	Average % Heme Loading
as-isolated WT	1	210.3	114.5	53.2 ± 1.2
	2	252.8	131.5	
reconstituted WT	1	138.4	146	113.3 ± 7.8
	2	57.1	69.2	
as-isolated HBD	1	111.9	26.3	27.6 ± 4.1
	2	119.2	37.8	
reconstituted HBD	1	138.7	130	99.0 ± 5.3
	2	94.3	98.3	

spectra comparing as-isolated and reconstituted protein (Figure 4.S2). Reconstituted *Px*RcoM-1 exhibits an extra EPR signal at $g = 4.7$, indicative of an axial high-spin, Fe(III) species, while the as-isolated sample exhibits no such signal. This axial signal is characteristic of heme that is adventitiously bound to the protein surface. Heme reconstitution does not appear to alter the electronic absorption features nor the apparent size/oligomeric status of *Px*RcoM-1 (Figure 4.S3).

As is the case for full length RcoM-1, each HBD truncate monomer contains a single heme binding site that is not fully occupied as-isolated. Heme electronic absorption features for the HBD truncate are nearly identical to those of the full-length protein, demonstrating that heme-binding function is maintained with loss of the LytTR domain (Figure 4.S4). Heme loading dramatically increased between as-isolated truncate, which exhibited heme loading of 27.6 ± 4.1 % (assuming a 1:1 monomer to heme ratio), and truncate reconstituted with hemin chloride, which exhibited heme loading of 99.0 ± 5.3 % (Table 4.1). Unlike full-length *Px*RcoM-1, the HBD truncate binds almost exactly one equivalent of heme per monomer, suggesting that potential adventitious heme binding sites may be localized to the C-terminal LytTR domain.

*Px*RcoM-1 homodimerization does not require the presence of heme. Two full-length *Px*RcoM-1 variants with perturbed heme environments, C94S and H74A, were characterized by electronic absorption and SEC. Previous studies of RcoM-2 demonstrated that the variant C94S binds heme in a five-coordinate, high-spin state, likely through His^{74}.[16] We find that C94S *Px*RcoM-1 also binds heme, as indicated by the presence of Soret (423 nm) and α/β bands (569 nm and 540 nm) in the electronic absorption spectrum (Figure 4.2A). Previous characterization of H74A *Px*RcoM-1 revealed that this variant does not bind any appreciable amount of heme,[6] an observation corroborated in this study by a lack of heme-derived peaks in the visible spectrum (Figure 4.2A). By SEC, the majority fractions of C94S (68.9 kDa) and H74A (66.8 kDa) were

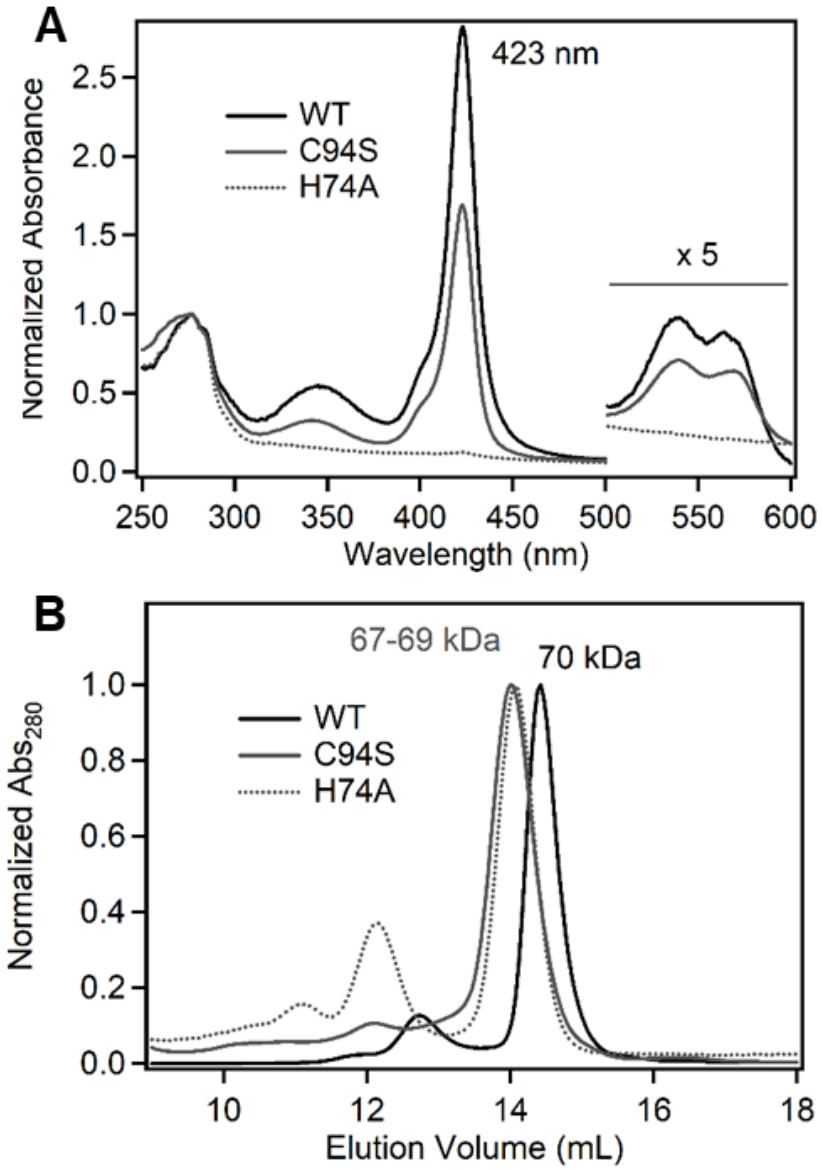


Figure 4.2 (A) Electronic absorption spectra of WT (0.48 mg/mL monomer), C94S (1.46 mg/mL), and H74A (0.515 mg/mL) *Px*RcoM-1 in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT. Traces are normalized to absorbance at 280 nm. (B) Size exclusion chromatograms of full-length WT, C94S, and H74A *Px*RcoM-1. Protein samples of 100 µL (5.58 mg/mL monomer for WT, 1.46 mg/mL monomer for C94S, and 1.19 mg/mL monomer for H74A *Px*RcoM-1) were injected via a 1 mL sample loop at a flow rate of 0.6 mL/min. WT and C94S *Px*RcoM-1 were run in 50 mM MOPS pH 7.4, 150 mM NaCl, and H74A *Px*RcoM-1 was run in 20 mM Tris pH 8.0, 400 mM NaCl, 1 mM DTT. Traces are normalized to the highest intensity peak.

dimeric; however, the dimeric variants exhibited slightly smaller apparent molecular weights compared to WT (72 kDa) indicating that these proteins adopt a shape that is slightly more spherical/more compact than that of the WT protein (Figure 4.2B).

Solution structure of PxRcoM-1

Solution small-angle X-ray scattering (SAXS) data confirm that full-length WT *Px*RcoM-1 exists as an elongated homodimer in solution. The molecular weight of *Px*RcoM-1, estimated from the SAXS-derived correlation volume (V_c),[53] was 70 ± 10 kDa, consistent with the full-length protein forming a dimer in solution (Figure 4.3A). The radius of gyration (R_g), determined by Guinier analysis, was 36.1 ± 0.55 Å, and the maximum end-to-end distance (D_{max}), calculated from the pairwise distribution function plot (Pr), was 126 ± 3 Å (Figure 4.3B). Theoretical scattering profiles based on monomeric and dimeric *Px*RcoM-1 homology models were generated using the FoXS server and compared to the experimental profile obtained from SAXS (Figure 4.3C). The dimeric model agreed well with the experimental data (Chi = 7.26), while the monomeric model did not agree with the experimental data (Chi = 24.44). An *ab initio* dummy-atom beads model, constructed from SAXS data using the DAMMIF software program,[52] suggests that the *Px*RcoM-1 homodimer adopts an elongated, cylindrical conformation (Figure 4.3D). We sought to compare the shape of full-length *Px*RcoM-1 with that of the HBD truncate; however, SAXS data reveal that the truncate forms higher order oligomers in a concentration-dependent manner (Supplemental Information, Figure 4.S5). A mixture of dimeric and tetrameric HBD truncate was observed at concentrations as low as 1.07 mg/mL, precluding structural analysis of the truncate by SAXS.

In the absence of crystallographic data for *Px*RcoM-1, we developed a new structural model using a combination of homology modeling and molecular dynamics simulations. This model takes into account the observation that dimerization occurs via the heme-binding domain;

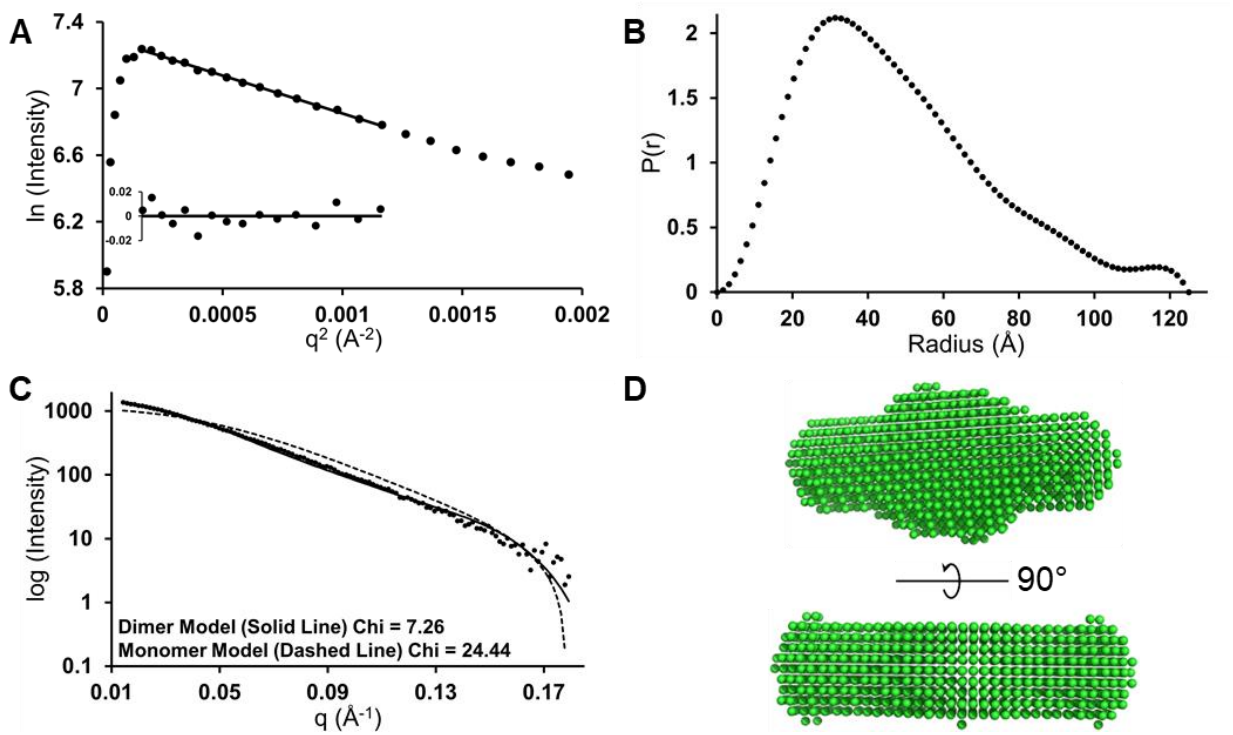


Figure 4.3 (A) Guinier fit to the WT *Px*RcoM-1 SAXS data (R_g: 37.0 ± 0.981 Å, *I*(0): 1489.6 ± 24.1). (B) P(r) plot for the *Px*RcoM-1 SAXS data. (R_g 37.70 ± 0.143 Å, *I*(0) = 1484 ± 4.05). (C) Dimer (solid line, chi = 7.26) and monomer (dashed line, chi = 24.44) models fitted to the experimental data (dots) using FoXS server. (D) Dummy atoms beads *ab initio* model showing the average of 30 runs on DAMMIF with enforced P2 symmetry (NSD = 0.706).

however, we did not impose any geometric constraints based on the solution SAXS model of *Px*RcoM-1. We used YASARA to create a monomeric homology model based on three template structures.[54,55] The PAS and LytTR domains were modelled independently and then linked together in YASARA as described in the methods. A model of the *Px*RcoM-1 dimer was constructed by aligning two monomers with the dimer interface of *Ec*DOS (PDB: 1V9Y)[27] and docking using HADDOCK (Figure 4.4A).[66] The buried surface area of the dimer interface between the PAS domains is 2,538 $Å^2$, and the dimer was stable over 125 ns of molecular dynamics using the AMBERff14 force field (Figure 4.4B). Fluctuations in RMSD observed during the simulation resulted from movement of one of the LytTR domains relative to the rest of the molecule. This observation is corroborated by a comparison of root mean square fluctuation (RMSF) values between residues in the heme-binding domain, which range from 3-12 Å, and residues in the LytTR domain, which range from 4-23 Å (Figure 4.4C).

SAXS data are consistent with the structure of the *Px*RcoM-1 homology model. Over the course of the molecular dynamics simulation, the calculated r_g for the dimeric homology model was 33.6 ± 1.9 Å, consistent with the experimentally determined r_g of 36.1 ± 0.55 Å. An overlay of the relaxed, dimeric homology model and the *ab initio* dummy atom model further demonstrates good agreement between the computationally- and experimentally-derived models (Figure 4.4D). Taken together, these observations present further evidence that *Px*RcoM-1 is dimeric in solution and suggest that our dimeric homology model accurately represents the general shape of the *Px*RcoM-1 homodimer.

PxRcoM sequence similarity network reflects genomic context

Using the protein sequence of *Px*RcoM-1 from *P. xenovorans* as the query input, we developed an RcoM sequence similarity network (SSN) and identified an isofunctional cluster with over 100

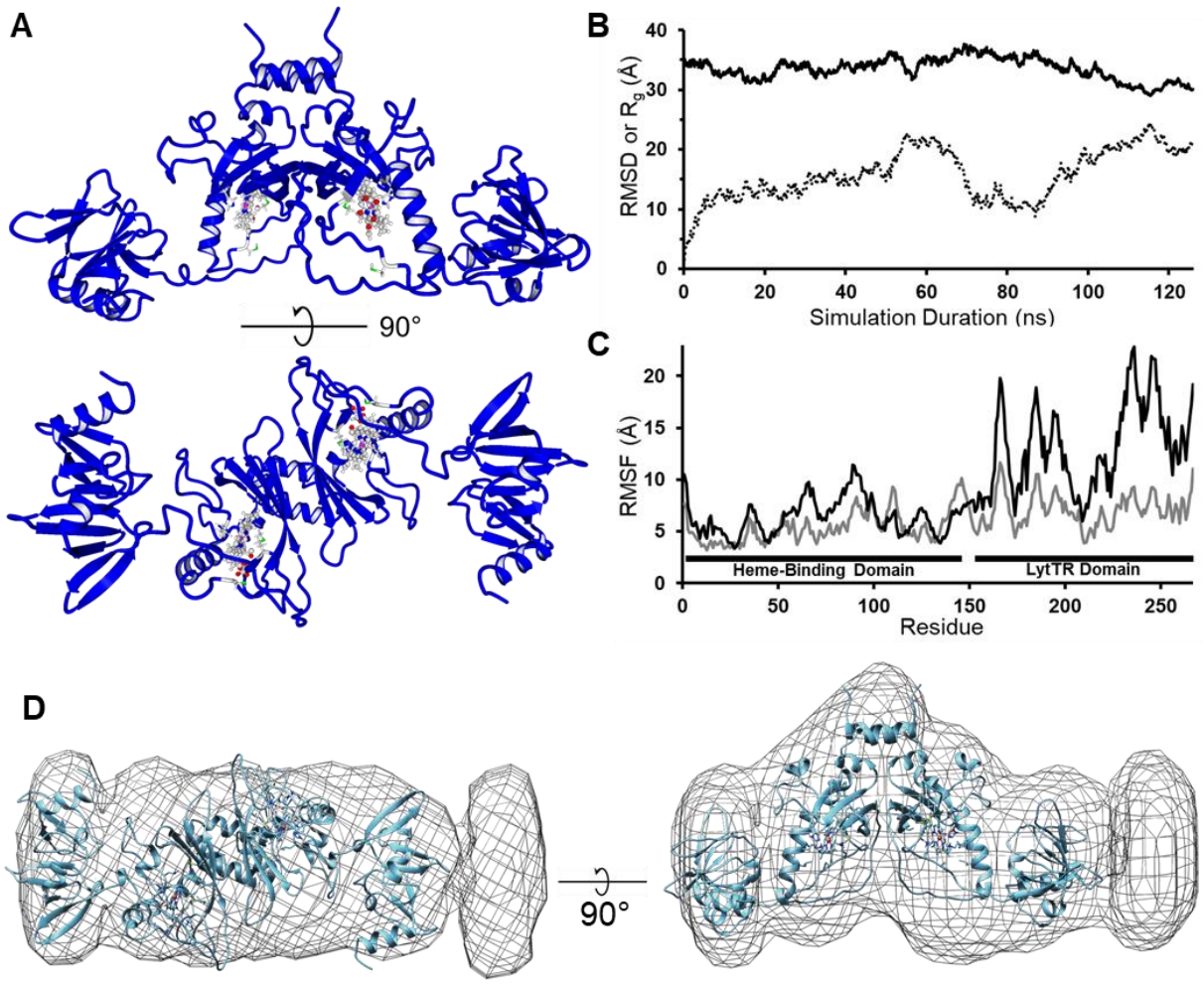


Figure 4.4 (A) Energy minimized, dimeric *Px*RcoM-1 homology model structure used for SAXS fitting. Residues His74, Cys94, and Met104 are shown in gray. (B) Plot of RMSD (dotted line) and R_g (solid line) over the course of the molecular dynamics simulation of the *Px*RcoM-1 model. (C) Root mean square fluctuation (RMSF) for each residue in both Chain A (black) and Chain B (gray) over the course of the molecular dynamics simulation. (D) Alignment of *Px*RcoM-1dimer structure to the DAMMIN-predicted shape (mesh).

putative RcoM proteins. In the initial 95% representative node network (alignment cutoff score = 24), we observed three nodal clusters (Figure 4.S6, Figure 4.S7). A common feature of nearly all proteins in this network was the presence of an annotated LytTR domain; however, only proteins in cluster 2 were also annotated as putative regulators of CO metabolism. This cluster contained two distinct sub-clusters. The larger ($N_{proteins}$ = 112), more divergent of these sub-clusters contained both RcoM orthologs from *P. xenovorans*, in addition to several other proteins annotated as heme-binding regulators of CO metabolism (Figure 4.5A). Proteins in the smaller ($N_{proteins}$ = 68), less divergent sub-cluster contained an N-terminal, membrane-associated MHYT domain, characteristic of CoxC and CoxH proteins, another class of putative CO-sensing transcription factors.[20] The connection between RcoM and CoxC/H proteins in the SSN suggests that divergent evolution of a common ancestor may have given rise to these CO-sensing transcription factors with distinct N-terminal structural motifs.

By applying more stringent cutoff scores to the RcoM sub-cluster, we observed further clustering based on the identity of the adjacent, RcoM-regulated operon. Using the EFI Genome Neighborhood Tool (EFI-GNT), we successfully identified the genomic context of 99 of the 112 proteins within the putative RcoM cluster. Of the 99 *rcoM* genes associated with these proteins, we identified 44 adjacent to a *cox* gene cluster, 39 adjacent to a *cowN* gene, and 16 adjacent to a *coo* gene cluster. The observation of *rcoM* genes adjacent to *cox*, *cowN*, and *coo* operons is consistent with previous studies.[6,18] Remarkably, upon increasing the stringency of the alignment cutoff score, we observed sub-clustering in the SSN that correlates with the identity of the associated operon (Figure 4.S8). At an alignment cutoff score of 50, we observe complete separation between *cox*-, *cowN*-, and *coo*-associated nodes (Figure 4.5B). The *cowN*- and *coo*-associated cluster exhibit significant fracturing at this cutoff score, while the *cox*-associated cluster

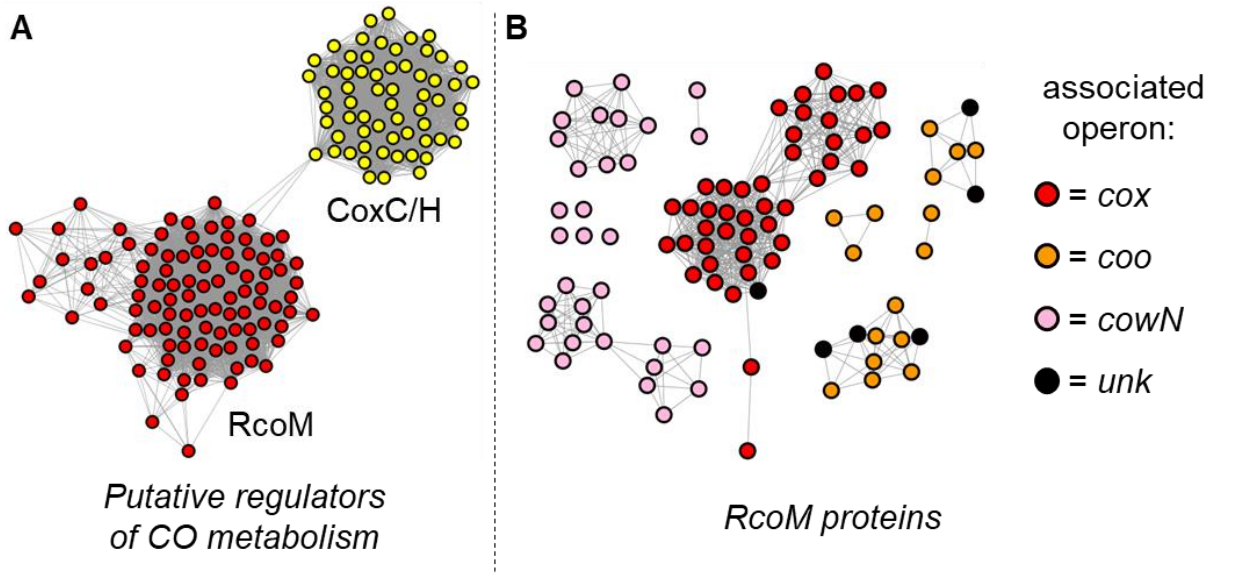


Figure 4.5 (A) Cluster 2 from sequence similarity network based on *Px*RcoM-1 at an alignment score cutoff of 24. (B) *Px*RcoM-containing sub-clusters shown at an alignment score cutoff of 50. Nodes are labeled based the identity of the *rcoM*-adjacent operon identified in each genome neighborhood.

remains intact. Within the *cox*-associated cluster, two sub-clusters are evident: one sub-cluster contains RcoM proteins whose genes are transcribed upstream and *parallel to* the associated *cox* operon, while the other sub-cluster contains RcoM proteins whose genes are transcribed upstream and *antiparallel to* the associated *cox* operon. The observation of sub-clustering based on genomic context in the RcoM SSN, which is constructed based on protein sequences alone, suggests a connection between the structure of RcoM and its regulatory DNA target.

Re-examining DNA binding in RcoM

Through alignment of 40 putative *cox*-associated gene promoter regions, we identified conserved DNA repeats that may reflect an RcoM consensus binding motif. An important initial observation was that the far upstream "a + b + c" triplet repeat was not conserved amongst *cox*-associated gene promoters. As a result, we narrowed the region defined as the putative *coxM* promoter to encompass a 100 bp window centered on a conserved repeat identified just upstream of the -10/-35 region of *coxM* (Figure 4.S9). This window encompassed the "d + e + f" triplet motif site originally identified as a *Px*RcoM-1 binding site.[18] The sequence alignment revealed a highly conserved 8 bp motif, TT[CG][GA][TC]G[CT][AG], that encompasses the "TTnnnG" motif originally proposed for *Px*RcoM-1 (Figure 4.6A). This motif is strongly conserved at the "d" and "e" sites and weakly conserved at the "f" site, and each motif is separated by a 12-13 bp linker.

An 8 bp *Px*RcoM-1 bindng motif was identified through selective enhancement of dsDNA oligonucleotides bearing a 40 bp randomized region flanked by two 20 bp regions with a fixed sequence (SELEX). The randomized oligonucleotides were incubated with *Px*RcoM-1, and the consensus sequence of bound oligonucleotides was determined using the cognate sequence identifier (CSI) high throughput approach.[59] After three successive rounds of enhancement, an 8 bp motif with sequence CAnGGACG (and its reverse complement) was identified as the cognate motif (Figure 4.6B).

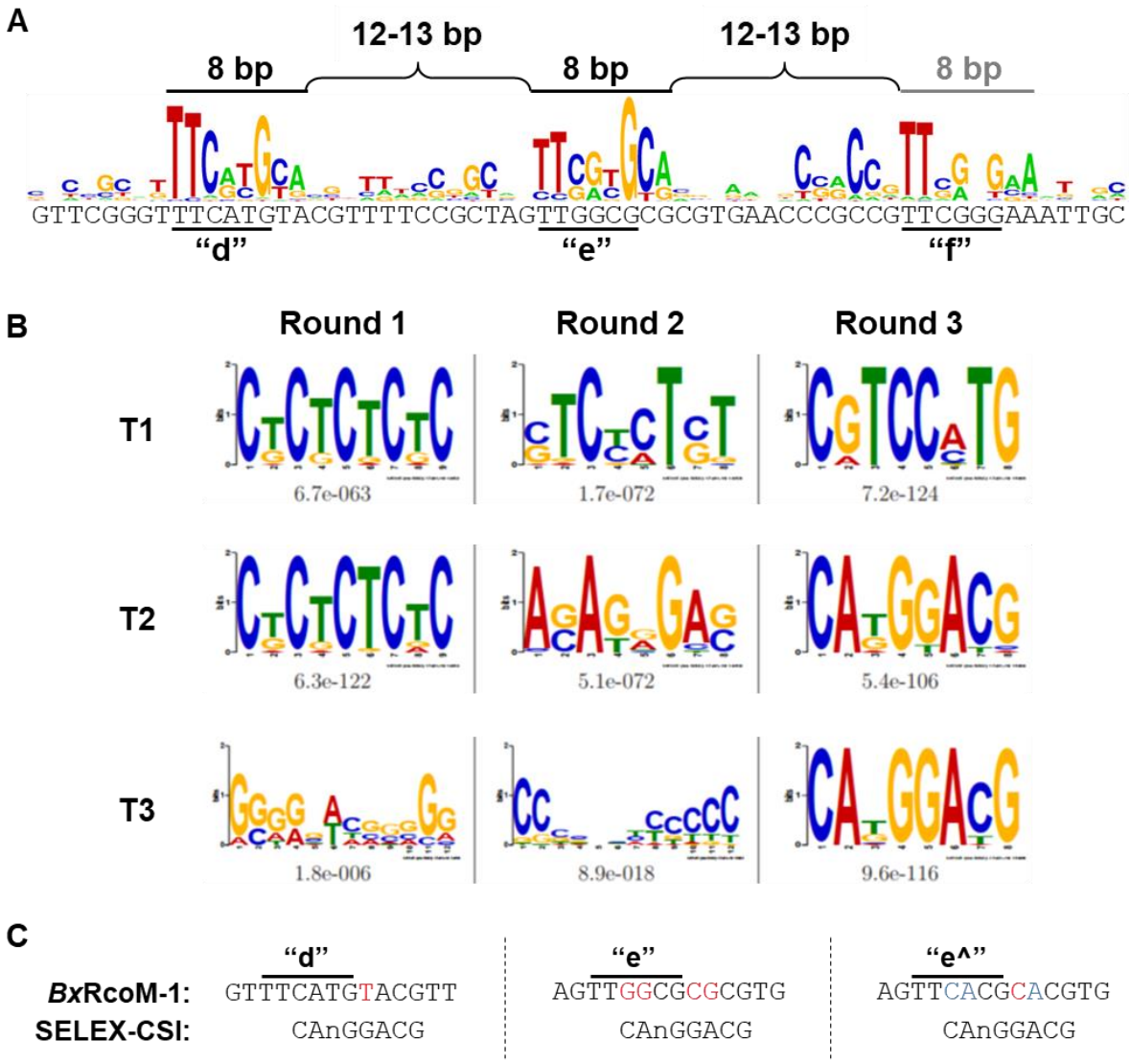


Figure 4.6 (A) Logo plot generated from multiple sequence alignment of 40 putative *cox*-associated RcoM promoter regions. The three 8 bp repeats, identified through sequence alignment, are annotated above the logo plot. The *Px*RcoM-1 promoter region is shown below the logo plot with the originally-proposed 6 bp "d", "e", and "f" sites highlighted in red. (B) SELEX-CSI data for three separate trials (T1, T2, and T3) in which *Px*RcoM-1 was incubated with the oligo library, as described in the Materials and Methods Section. Logo plots report on the sequence-specific base pair identities observed after each round of selective oligomer enhancement. (C) Comparison of highly-conserved upstream regions of *Px*RcoM-1 corresponding to "d", "e", and "e^" sites to the 8 bp high-affinity sequence identified from SELEX. Base pair mismatches between the *Px*RcoM-1 sequences and SELEX-derived sequence are denoted in red, and base pair changes between the e and e^ sites are denoted in blue.

This 8 bp cognate *Px*RcoM-1 binding motif is similar, but not identical, to those identified in previous studies of *Px*RcoM-1 DNA binding and through sequence alignment of putative *coxM* promoter regions. A perfect match of the SELEX-CSI cognate motif could not be identified anywhere in the intergenic region between $rcoM_1$ and $coxM_1$; however, this motif could be mapped onto sites in the two highly-conserved regions corresponding to "d" and "e" sites (Figure 4.6C). A single base pair mismatch is observed between the SELEX-CSI cognate motif and base pairs near the conserved "d" site, while four mismatches are observed between the SELEX-CSI cognate motif and base pairs near the less well-conserved "e" site. Interestingly, the "e^" sequence, which gives rise to enhanced *Px*RcoM-1-dependent expression *in vivo*,[18] more closely resembles the SELEX-CSI cognate sequence. This observation is consistent with the assertion that the 8 bp cognate sequence CAnGGACG exhibits maximal *Px*RcoM-1 binding affinity despite the fact that this sequence is not observed in the between $rcoM_1$ and $coxM_1$ intergenic region.

4.4 Discussion

In this study, we begin to enumerate the structure of the heme-depndent, single-component transcriptional regulator *Px*RcoM-1. Through sequence homology, RcoM proteins are predicted to exhibit a unique domain architecture consisting of an N-terminal PAS domain, which binds heme, and a C-terminal LytTR DNA binding domain. We have unequivocally demonstrated that *Px*RcoM-1 is homodimeric and dimerizes via the PAS domain, an observation that is consistent with other PAS domain-containing proteins.[67] Heme does not appear to mediate dimerization, and the 1:1 heme:monomer stoichiometry is consistent with that of other heme-dependent gas sensors.

Our experimentally-validated *Px*RcoM-1 homology model may provide insight into the protein's functional mechanism. Each heme cofactor is solvent-accessible, allowing for diffusion of CO into the heme pocket, and the heme-coordinating ligands Cys^{94} and Met^{104} are within 13.5

Å and 5.2 Å of the heme, respectively. The loop linking the PAS domain to the LytTR domain runs adjacent to the loop that interacts directly with the heme and contains residue Cys[94]. As such, the structure and dynamics in this domain-linker region may be altered upon redox-mediated ligand switching at the heme. This flexibility in the LytTR domain relative to the heme-binding domain suggests that a large conformational changes may occur in *Px*RcoM-1 in the activation mechanism.

Divergence amongst RcoM protein sequences appears to occur in a manner that is dependent upon genomic context (Figure 4.6). *rcoM* genes have been identified in the genomic context of *cox*, *coo*, and *cowN* operons, which encode proteins associated with three distinct physiological processes: anaerobic CO metabolism, aerobic CO metabolism, and protection of nitrogenase enzymes from inactivation by CO, respectively.[1,6,12,14,23,24] The CO binding affinities of CODH enzymes that carry out aerobic (K_M = 11.7-0.52 μM)[68,69] and anaerobic (K_M = 18 μM)[70] CO metabolism differ, suggesting that these pathways are activated under different CO concentration regimes. It is therefore possible that the differences in RcoM protein sequences, observed as a function of genetic context, reflect differences in either CO binding affinity or DNA binding affinity that give rise to differential gene expression as a function of CO concentration. Further study of RcoM proteins encoded in the context of *coo* and *cowN* genes is required to substantiate this hypothesis.

The $coxM_1$ promoter region of *P. xenovorans* is likely not optimized to bind *Px*RcoM-1. Kerby and Roberts initially demonstrated that *Px*RcoM-1 binding in this promoter region is readily improved upon making changes to the promoter sequence.[18] A comparison of the SELEX-CSI cognate motif identified in our work to proposed promoter sites in *Px*RcoM-1 support this finding. The cognate motif, which represents the DNA sequence with the highest affinity for *Px*RcoM-1 *in*

vitro, can only be mapped onto the $coxM_1$ promoter sites imperfectly. Furthermore, preliminary genomescape analysis using the methods recently reported by Bhimsaria et al. reveals that *none* of the putative *cox* promoter sites examined in this study contain sequences that match the cognate motif identified using SELEX-CSI. Taken together, these observations suggest that *cox* promoter sites have evolved such that RcoM binding is sub-optimal.

Aerobic CO metabolizers may exploit changes in the DNA sequence of the *coxM* promoter region to attenuate CO-dependent gene expression. The CO binding affinities of RcoM proteins from *P. xenovorans* approach picomolar levels;[19] however, aerobic CODH enzymes exhibit CO binding affinities in the low micromolar regime.[68,69] If RcoM, which is activated upon CO binding, exhibited a high affinity for its promoter site, than *cox* genes would be expressed at CO concentrations far too low to give rise to significant CODH activity. To account for the discrepancy in CO affinities, microorganisms may have evolved their *coxM* promoter sites to give rise to weak RcoM binding. This attenuation strategy would allow RcoM orthologues to exhibit CO-dependent transcriptional regulation in drastically different CO concentration regimes and may enable this versatile transcription factor to regulate three different physiological processes.

4.5 References

1. King, G. M.; Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* **2007**, *5*, 107.

2. Oelgeschlager, E.; Rother, M. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch. Microbiol.* **2008**, *190*, 257.

3. Roberts, G. P.; Youn, H.; Kerby, R. L. CO-sensing mechanisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 453.

4. Shimizu, T.; Huang, D.; Yan, F.; Stranava, M.; Bartosova, M.; Fojtíková, V.; Martínková, M. Gaseous O2, NO, and CO in Signal Transduction: Structure and Function Relationships of Heme-Based Gas Sensors and Heme-Redox Sensors. *Chem.Rev.* **2015**, *115*, 6491.

5. Shelver, D.; Kerby, R. L.; He, Y. P.; Roberts, G. P. Carbon monoxide-induced activation of gene expression in *Rhodospirillum rubrum* requires the product of *cooA*, a member of the cyclic AMP receptor family of transcriptional regulators. *J. Bacteriol.* **1995**, *177*, 2157.

6. Kerby, R. L.; Youn, H.; Roberts, G. P. RcoM: a new single-component transcriptional regulator of CO metabolism in bacteria. *J. Bacteriol.* **2008**, *190*, 3336.

7. Lanzilotta, W. N.; Schuller, D. J.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure of the CO sensing transcription activator CooA. *Nat. Struct. Biol.* **2000**, *7*, 876.

8. Borjigin, M.; Li, H.; Lanz, N. D.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. Structure-based hypothesis on the activation of the CO-sensing transcription factor CooA. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2007**, *63*, 282.

9. Aono, S.; Nakajima, H.; Saito, K.; Okada, M. A novel heme protein that acts as a carbon monoxide-dependent transcriptional activator in *Rhodospirillum rubrum*. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 752.

10. Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Chung, S. Y.; Roberts, G. P.; Reynolds, M. F.; Parks, R. B.; Burstyn, J. N. Identification of two important heme site residues (cysteine 75 and histidine 77) in CooA, the CO-sensing transcription factor of *Rhodospirillum rubrum*. *Biochemistry* **1999**, *38*, 2669.
11. Reynolds, M. F.; Parks, R. B.; Burstyn, J. N.; Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Vogel, K. M.; Spiro, T. G. Electronic absorption, EPR, and resonance raman spectroscopy of CooA, a CO-sensing transcription activator from R. rubrum, reveals a five-coordinate NO-heme. *Biochemistry* **2000**, *39*, 388.
12. Fox, J. D.; Kerby, R. L.; Roberts, G. P.; Ludden, P. W. Characterization of the CO-induced, CO-tolerant hydrogenase from *Rhodospirillum rubrum* and the gene encoding the large subunit of the enzyme. *J. Bacteriol.* **1996**, *178*, 1515.
13. Bonam, D.; Ludden, P. W. Purification and characterization of carbon-monoxide dehydrogenase, a nickel, zinc, iron-sulfur protein from *Rhodospirillum rubrum*. *J. Biol. Chem.* **1987**, *262*, 2980.
14. Fox, J. D.; He, Y. P.; Shelver, D.; Roberts, G. P.; Ludden, P. W. Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J. Bacteriol.* **1996**, *178*, 6200.
15. Marvin, K. A.; Kerby, R. L.; Youn, H.; Roberts, G. P.; Burstyn, J. N. The transcription regulator RcoM-2 from *Burkholderia xenovorans* is a cysteine-ligated hemoprotein that undergoes a redox-mediated ligand switch. *Biochemistry* **2008**, *47*, 9016.
16. Smith, A. T.; Marvin, K. A.; Freeman, K. M.; Kerby, R. L.; Roberts, G. P.; Burstyn, J. N. Identification of Cys94 as the distal ligand to the Fe(III) heme in the transcriptional regulator RcoM-2 from *Burkholderia xenovorans*. *J. Biol. Inorg. Chem.* **2012**, *17*, 1071.

17. Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met104 is the CO-replaceable ligand at Fe(II) heme in the CO-sensing transcription factor BxRcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21*, 559.

18. Kerby, R. L.; Roberts, G. P. Burkholderia xenovorans RcoMBx-1, a Transcriptional Regulator System for Sensing Low and Persistent Levels of Carbon Monoxide. *J. Bacteriol.* **2012**, *194*, 5803.

19. Bouzhir-Sima, L.; Motterlini, R.; Gross, J.; Vos, M. H.; Liebl, U. Unusual Dynamics of Ligand Binding to the Heme Domain of the Bacterial CO Sensor Protein RcoM-2. *J. Phys. Chem. B* **2016**, *120*, 10686.

20. Santiago, B.; Schübel, U.; Egelseer, C.; Meyer, O. Sequence analysis, characterization and CO-specific transcription of the cox gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. *Gene* **1999**, *236*, 115.

21. Meyer, O.; Gremer, L.; Ferner, R.; Ferner, M.; Dobbek, H.; Gnida, M.; Meyer-Klaucke, W.; Huber, R. The role of Se, Mo and Fe in the structure and function of carbon monoxide dehydrogenase. *Biol. Chem.* **2000**, *381*, 865.

22. Fuhrmann, S.; Ferner, M.; Jeffke, T.; Henne, A.; Gottschalk, G.; Meyer, O. Complete nucleotide sequence of the circular megaplasmid pHCG3 of *Oligotropha carboxidovorans*: function in the chemolithoautotrophic utilization of CO, H-2 and CO2. *Gene* **2003**, *322*, 67.

23. Kerby, R. L.; Roberts, G. P. Sustaining N_2-Dependent Growth in the Presence of CO. *J. Bacteriol.* **2011**, *193*, 774.

24. Hoffmann, M. C.; Pfander, Y.; Fehringer, M.; Narberhaus, F.; Masepohl, B. NifA- and CooA-Coordinated cowN Expression Sustains Nitrogen Fixation by *Rhodobacter capsulatus* in the Presence of Carbon Monoxide. *J. Bacteriol.* **2014**, *196*, 3494.

25. Henry, J. T.; Crosson, S. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu. Rev. Microbiol.* **2011**, *65*, 261.

26. Gong, W.; Hao, B.; Mansy, S. S.; Gonzalez, G.; Gilles-Gonzalez, M. A.; Chan, M. K. Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 15177.

27. Kurokawa, H.; Lee, D.-S.; Watanabe, M.; Sagami, I.; Mikami, B.; Raman, C. S.; Shimizu, T. A Redox-controlled Molecular Switch Revealed by the Crystal Structure of a Bacterial Heme PAS Sensor. *J. Biol. Chem.* **2004**, *279*, 20186.

28. Hao, B.; Isaza, C.; Arndt, J.; Soltis, M.; Chan, M. K. Structure-Based Mechanism of O2 Sensing and Ligand Discrimination by the FixL Heme Domain of *Bradyrhizobium japonicum*. *Biochemistry* **2002**, *41*, 12952.

29. Park; Suquet, C.; Satterlee, J. D.; Kang, C. Insights into Signal Transduction Involving PAS Domain Oxygen-Sensing Heme Proteins from the X-ray Crystal Structure of *Escherichia Coli* Dos Heme Domain (Ec DosH). *Biochemistry* **2004**, *43*, 2738.

30. Nikolskaya, A. N.; Galperin, M. Y. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* **2002**, *30*, 2453.

31. Cody, W. L.; Pritchett, C. L.; Jones, A. K.; Carterson, A. J.; Jackson, D.; Frisk, A.; Wolfgang, M. C.; Schurr, M. J. *Pseudomonas aeruginosa* AlgR Controls Cyanide Production in an AlgZ-Dependent Manner. *J. Bacteriol.* **2009**, *191*, 2993.

32. Cheung, J. K.; Keyburn, A. L.; Carter, G. P.; Lanckriet, A. L.; Van Immerseel, F.; Moore, R. J.; Rood, J. I. The VirSR Two-Component Signal Transduction System Regulates NetB Toxin Production in *Clostridium perfringens*. *Infect. Immun.* **2010**, *78*, 3064.

33. Reyes, D.; Andrey, D. O.; Monod, A.; Kelley, W. L.; Zhang, G.; Cheung, A. L. Coordinated Regulation by AgrA, SarA, and SarR To Control *agr* Expression in *Staphylococcus aureus*. *J. Bacteriol.* **2011**, *193*, 6020.

34. Sidote, D. J.; Barbieri, C. M.; Wu, T.; Stock, A. M. Structure of the *Staphylococcus aureus* AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with an Unusual Mode of Binding. *Structure* **2008**, *16*, 727.

35. Ween, O.; Gaustad, P.; Havarstein, L. S. Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **1999**, *33*, 817.

36. Cheung, J. K.; Rood, J. I. The VirR Response Regulator from *Clostridium perfringens* Binds Independently to Two Imperfect Direct Repeats Located Upstream of the *pfoA* Promoter. *J. Bacteriol.* **2000**, *182*, 57.

37. de Saizieu, A.; Gardès, C.; Flint, N.; Wagner, C.; Kamber, M.; Mitchell, T. J.; Keck, W.; Amrein, K. E.; Lange, R. Microarray-Based Identification of a Novel *Streptococcus pneumoniae* Regulon Controlled by an Autoinduced Peptide. *J. Bacteriol.* **2000**, *182*, 4696.

38. Diep, D. B.; Håvarstein, L. S.; Nes, I. F. Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum C11*. *J. Bacteriol.* **1996**, *178*, 4472.

39. Risoen, P. A.; Johnsborg, O.; Diep, D. B.; Hamoen, L.; Venema, G.; Nes, I. F. Regulation of bacteriocin production in *Lactobacillus plantarum* depends on a conserved promoter arrangement with consensus binding sequence. *Mol. Genet. Genomics* **2001**, *265*, 198.

40. Knutsen, E.; Ween, O.; Havarstein, L. S. Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. *J. Bacteriol.* **2004**, *186*, 3078.

41. Deretic, V.; Dikshit, R.; Konyecsni, W. M.; Chakrabarty, A. M.; Misra, T. K. The algR gene, which regulates mucoidy in *Pseudomonas aeruginosa*, belongs to a class of environmentally responsive genes. *J. Bacteriol.* **1989**, *171*, 1278.

42. Mohr, C. D.; Martin, D. W.; Konyecsni, W. M.; Govan, J. R.; Lory, S.; Deretic, V. Role of the far-upstream sites of the algD promoter and the algR and rpoN genes in environmental modulation of mucoidy in *Pseudomonas aeruginosa*. *J. Bacteriol.* **1990**, *172*, 6576.

43. Mohr, C. D.; Hibler, N. S.; Deretic, V. AlgR, a response regulator controlling mucoidy in *Pseudomonas aeruginosa*, binds to the FUS sites of the algD promoter located unusually far upstream from the mRNA start site. *J. Bacteriol.* **1991**, *173*, 5136.

44. Mohr, C. D.; Leveau, J. H.; Krieg, D. P.; Hibler, N. S.; Deretic, V. AlgR-binding sites within the algD promoter make up a set of inverted repeats separated by a large intervening segment of DNA. *J. Bacteriol.* **1992**, *174*, 6624.

45. Straume, D.; Johansen, R. F.; Bjoras, M.; Nes, I. F.; Diep, D. B. DNA binding kinetics of two response regulators, PlnC and PlnD, from the bacteriocin regulon of *Lactobacillus plantarum C11*. *Bmc Biochem.* **2009**, *10*.

46. Stewart, V.; Lu, Y.; Darwin, A. J. Periplasmic Nitrate Reductase (NapABC Enzyme) Supports Anaerobic Respiration by *Escherichia coli* K-12. *J. Bacteriol.* **2002**, *184*, 1314.

47. Aicart-Ramos, C.; Valhondo Falcón, M.; Ortiz de Montellano, P. R.; Rodriguez-Crespo, I. Covalent Attachment of Heme to the Protein Moiety in an Insect E75 Nitric Oxide Sensor. *Biochemistry* **2012**, *51*, 7403.

48. Berry, E. A.; Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **1987**, *161*, 1.

49. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005, p 571.

50. Ernst, O.; Zor, T. Linearization of the Bradford Protein Assay. **2010**, e1918.

51. Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2012**, *45*, 342.

52. Franke, D.; Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009**, *42*, 342.

53. Rambo, R. P.; Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **2013**, *496*, 477.

54. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009**, *77*, 114.

55. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comp. Chem.* **2003**, *24*, 1999.

56. Qiu, J.; Elber, R. SSALN: An alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. *Proteins* **2006**, *62*, 881.

57. Miyatake, H.; Mukai, M.; Park, S.-Y.; Adachi, S.-i.; Tamura, K.; Nakamura, H.; Nakamura, K.; Tsuchiya, T.; Iizuka, T.; Shiro, Y. Sensory mechanism of oxygen sensor FixL from

Rhizobium meliloti: crystallographic, mutagenesis and resonance raman spectroscopic studies11Edited by K. Nagai. *J. Mol. Biol.* **2000**, *301*, 415.

58. Gu, G.; Wang, T.; Yang, Y.; Xu, X.; Wang, J. An Improved SELEX-Seq Strategy for Characterizing DNA-Binding Specificity of Transcription Factor: NF-κB as an Example. *PLoS One* **2013**, *8*, e76109.
59. Warren, C. L.; Kratochvil, N. C. S.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N.; Ansari, A. Z. Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 867.
60. Bhimsaria, D.; Rodríguez-Martínez, J. A.; Pan, J.; Roston, D.; Korkmaz, E. N.; Cui, Q.; Ramanathan, P.; Ansari, A. Z. Specificity landscapes unmask submaximal binding site preferences of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E10586.
61. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta Proteins Proteom.* **2015**, *1854*, 1019.
62. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498.
63. Notredame, C.; Higgins, D. G.; Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *J. Mol. Biol.* **2000**, *302*, 205.
64. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14*, 1188.

65. Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17*, 1282.

66. van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720.

67. Ayers, R. A.; Moffat, K. Changes in Quaternary Structure in the Signaling Mechanisms of PAS Domains. *Biochemistry* **2008**, *47*, 12078.

68. Dobbek, H.; Gremer, L.; Kiefersauer, R.; Huber, R.; Meyer, O. Catalysis at a dinuclear [CuSMo(==O)OH] cluster in a CO dehydrogenase resolved at 1.1-A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15971.

69. Kaufmann, P.; Duffus, B. R.; Teutloff, C.; Leimkühler, S. Functional Studies on *Oligotropha carboxidovorans* Molybdenum−Copper CO Dehydrogenase Produced in *Escherichia coli*. *Biochemistry* **2018**, *57*, 2889.

70. Svetlitchnyi, V.; Peschel, C.; Acker, G.; Meyer, O. Two membrane-associated NiFeS-carbon monoxide dehydrogenases from the anaerobic carbon-monoxide-utilizing eubacterium *Carboxydothermus hydrogenoformans*. *J. Bacteriol.* **2001**, *183*, 5134.

4.6 Supplementary Information

Solution SAXS of BxRcoM-1 heme-binding domain truncate

The heme-binding domain (HBD) truncate forms higher order oligomers in a concentration-dependent manner (Figure S4a). When SAXS was carried out at a concentration of 1 mg/mL, the R_g, D_{max} and molecular weight decreased to 30.5 ± 0.4 Å, 100 ± 5 Å and 53 ± 3 kDa, respectively. The molecular weight calculated from SAXS data using *Vc* approach was 64 ± 5 kDa, which is larger than a dimer (theoretical molecular weight 32.8 kDa) and close to a tetramer (theoretical molecular weight 49.2 kDa). Guinier plot analysis of the HDB sample yielded an R_g value of 32.6 ± 0.6 Å (Figure S4). The maximum end-to-end distance, D_{max}, determined from pairwise distribution function plot, was 110 ± 5 Å (Figure S4). The SAXS *ab initio* dummy-atom beads model calculated by DAMMIF exhibited a large void that could not be filled by the structural model of a HBD truncate dimer. It appears that the HBD truncate has a strong tendency to form higher order oligomeric species in a concentration dependent. The tetramerization of the HBD truncate was also observed by SEC (Figure S4). Our hypothesis is that the HBD truncate has a higher tendancy to form higher order of oligomeric species than full-length *Bx*RcoM-1.

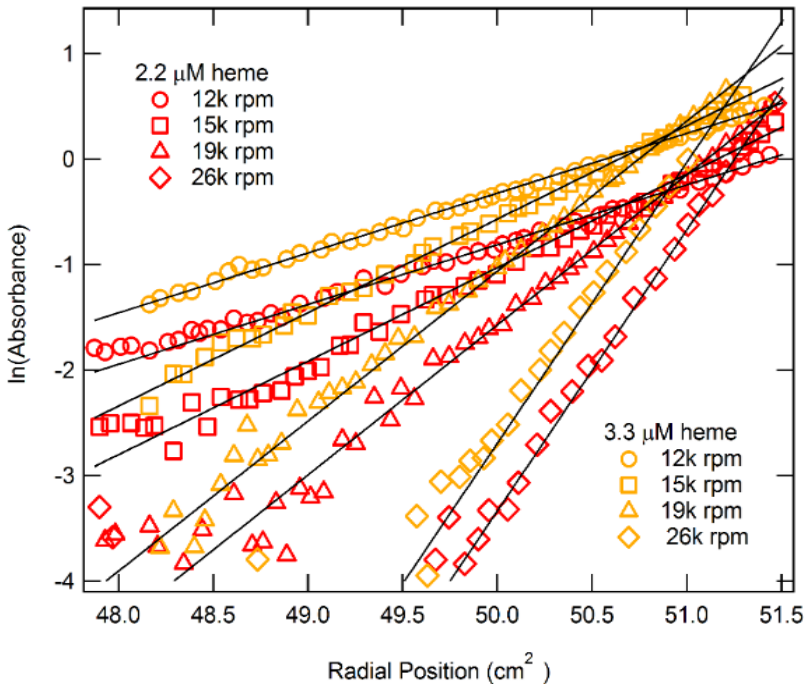


Figure 4.S1 Analytical ultracentrifugation data (every 2nd data point) plotted as ln(absorbance) vs. radial position for two concentrations at four speeds. The linear fits for different concentrations at same speed are parallel, demonstrating that the protein exists as a single species. Linear slopes are proportional to reduced M_w, speed, and temperature (4 °C). These data correspond to a M_w/M_s ratio of 2.25 or a single, dimeric species.

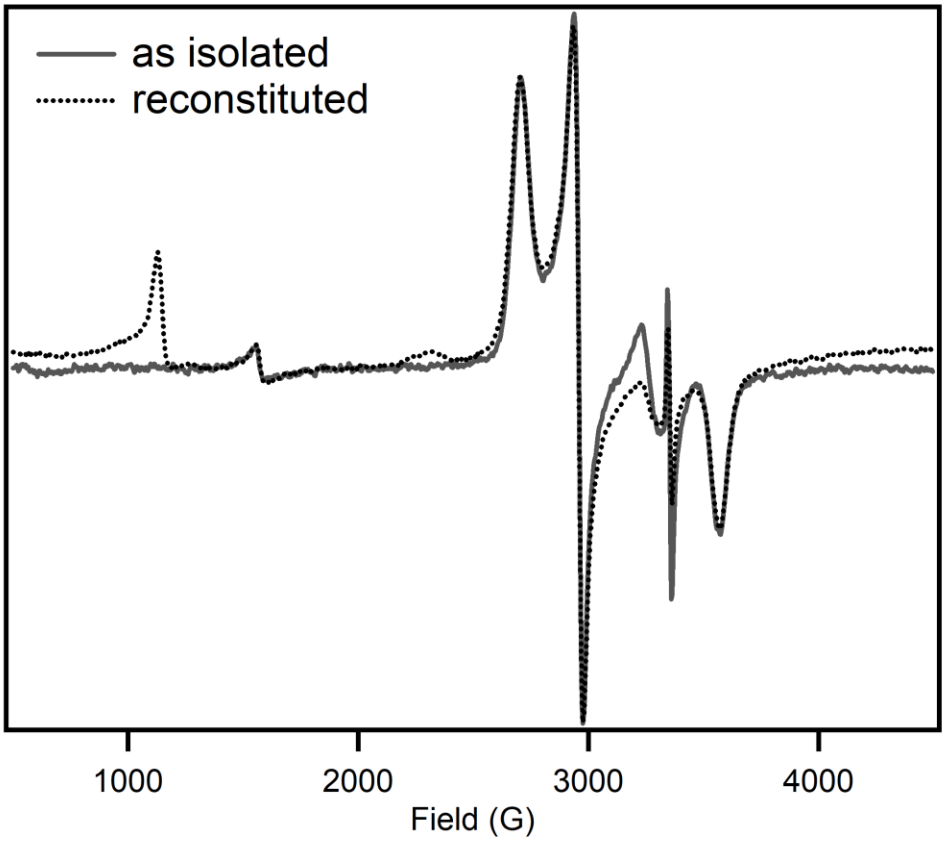


Figure 4.S2 X-band EPR spectra of WT *Bx*RcoM-1 as isolated after purification and reconstituted with hemin chloride. The asterisk denotes the EPR cavity signal, while the triangle denotes the axial, high-spin Fe(III) signal indicative of a small amount of adventitiously-bound heme.

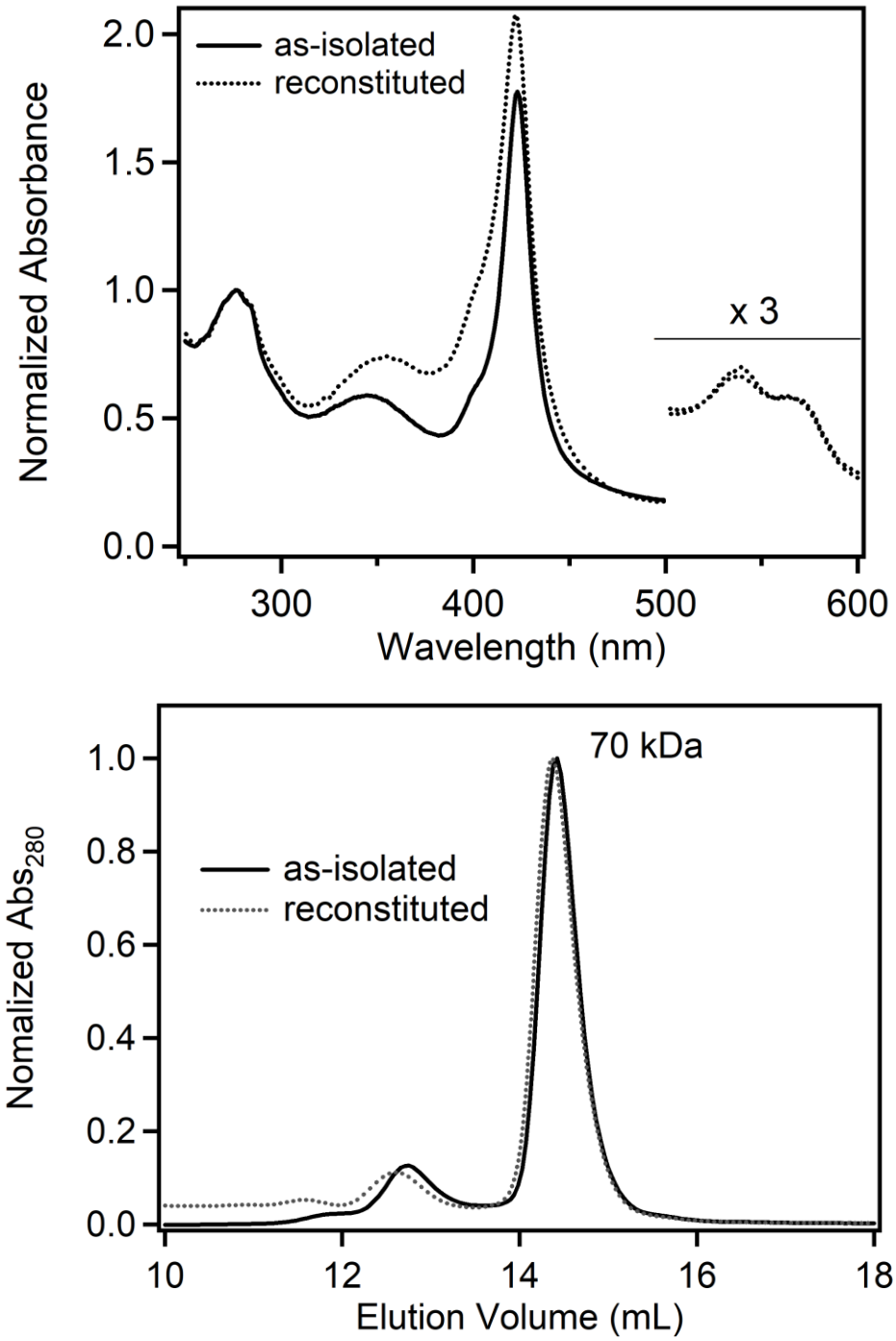


Figure 4.S3 A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography traces (lower panel) for as-isolated and reconstituted full-length WT *Bx*RcoM-1. Electronic absorption spectra are normalized to the intensity at 280 nm, and SEC traces are normalized to the most intense peak.

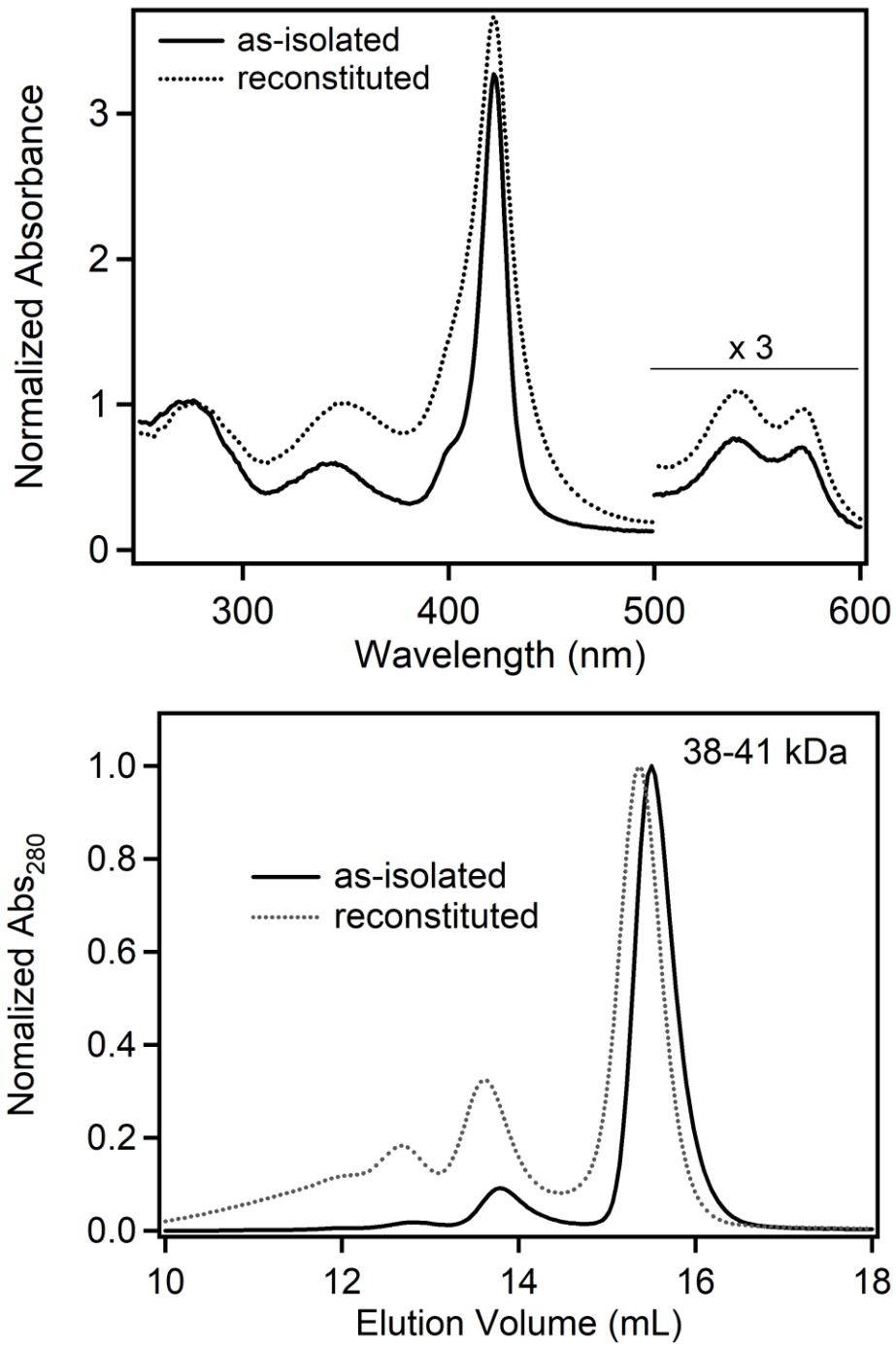


Figure 4.S4 A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography traces (lower panel) for the *Bx*RcoM-1 heme binding domain truncate. Electronic absorption spectra are normalized to the intensity at 280 nm, and SEC traces are normalized to the most intense peak.

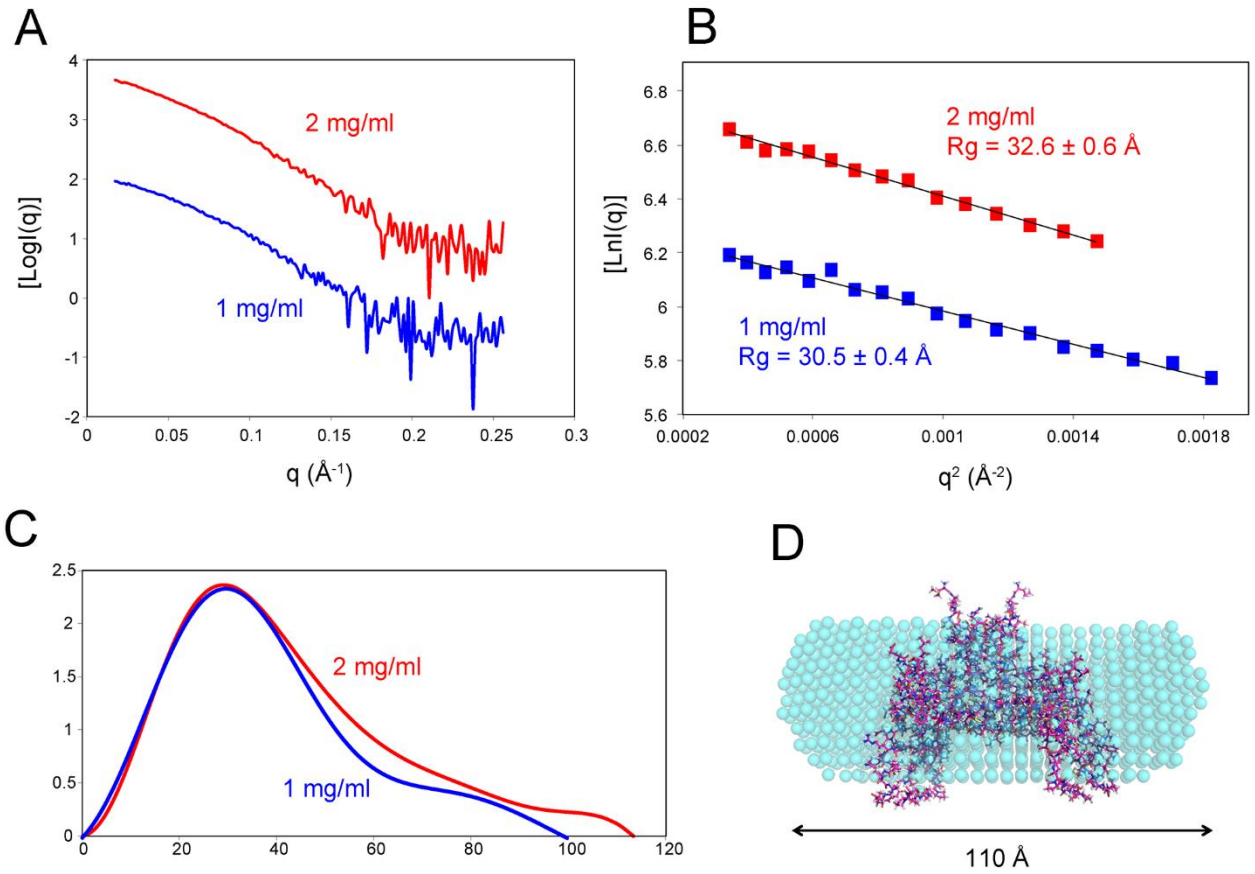


Figure 4.S5 Small-angle X-ray scattering data for RcoM-1 HBD truncate. (A) 1D SAXS profile of sample at 2 mg/mL and 1 mg/mL. (B) Guinier plot of the SAXS data. (C) Pairwise distribution function plot of the SAXS data. (D) The dummy atoms beads *ab initio* model showing the average of 30 runs on DAMMIF with enforced P2 symmetry.

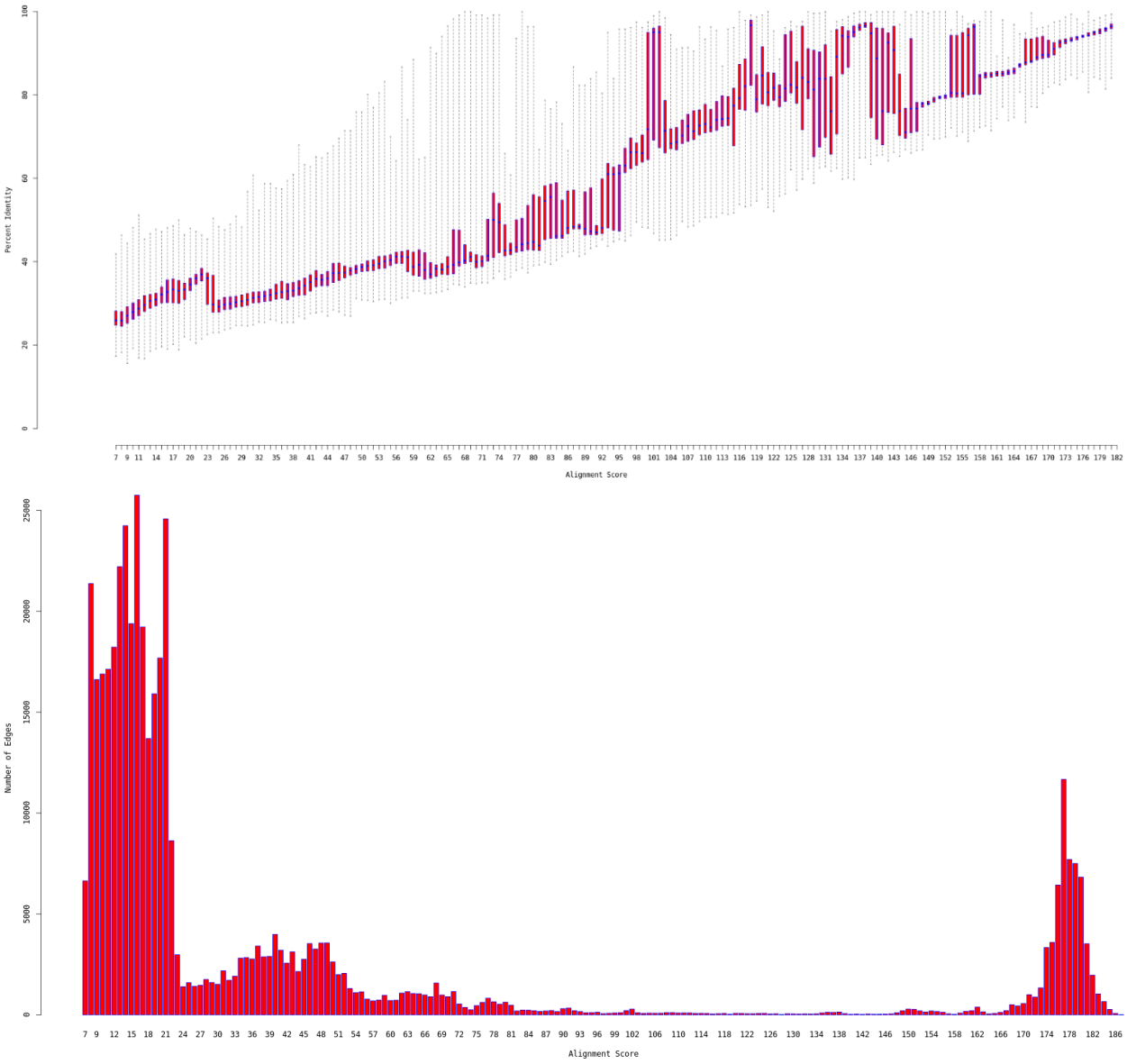


Figure 4.S6 RcoM sequence similarity network statistics. The quartile plot (upper panel) depicts the percent identity between the query sequence and each protein in the network as a function of alignment score. The histogram plot (lower panel) depicts the number of edges present in the sequence similarity network as a function of alignment score. An initial alignment cutoff score of 24 was chosen, which corresponded to an average percent identity of 35% and the first significant drop off in number of edges.

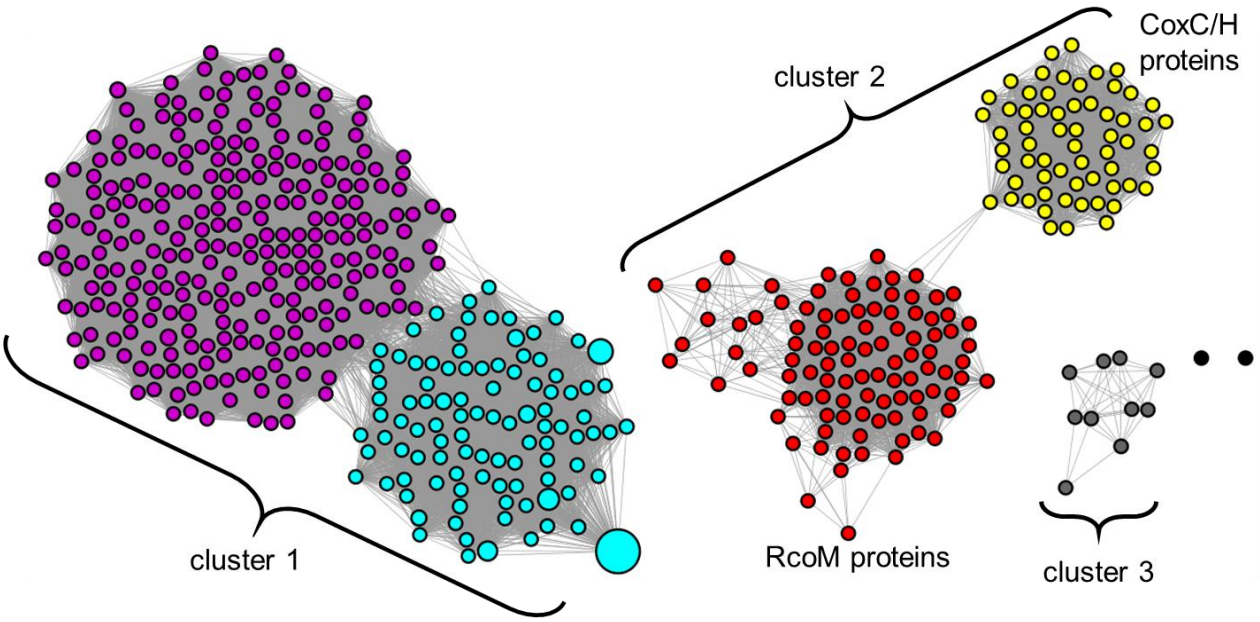


Figure 4.S7 RcoM sequence similarity network. The network is a 95% representative node network with an alignment score cutoff of 24. Nodes are colored to denote sub-clustering within nodal clusters.

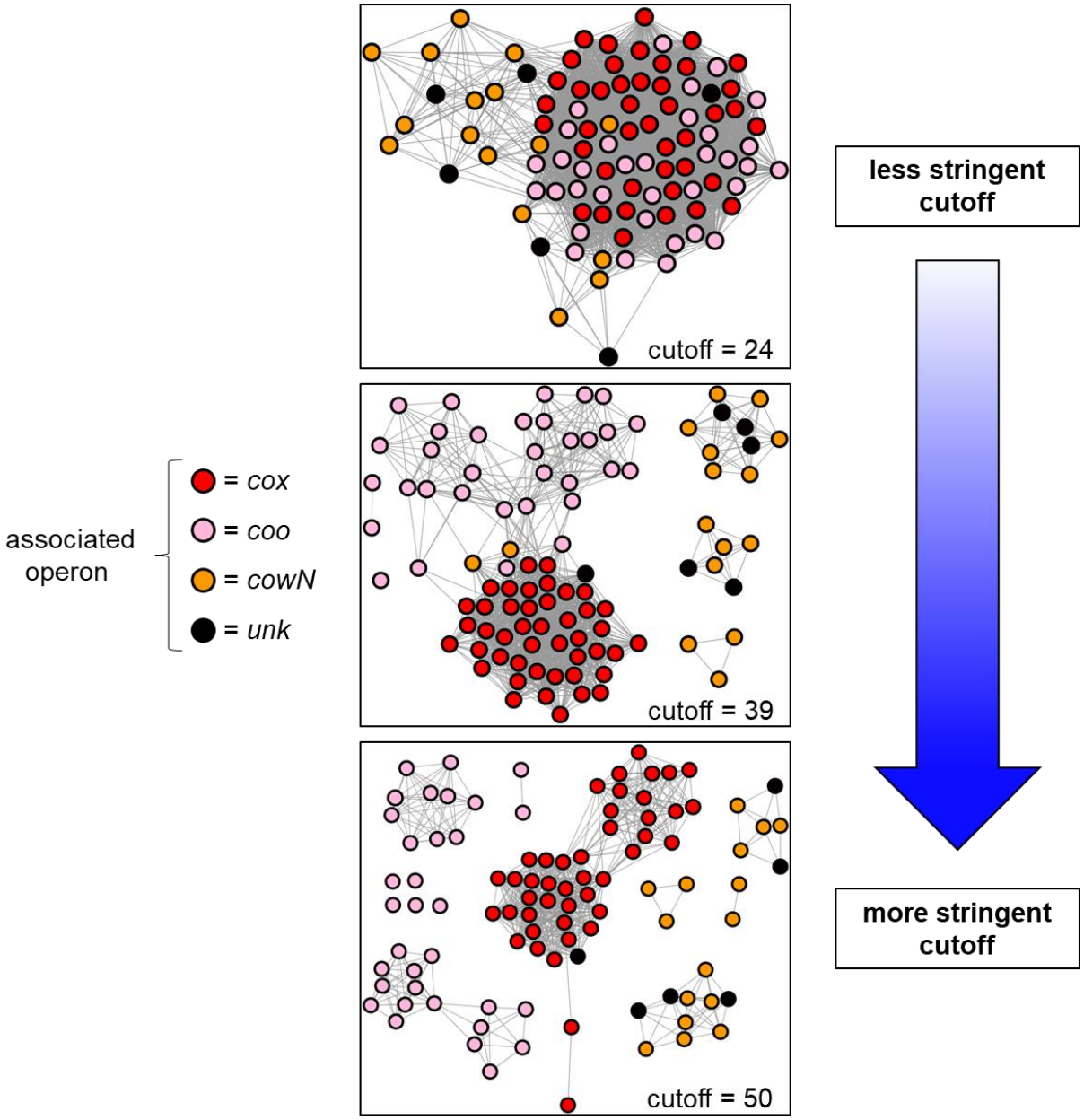
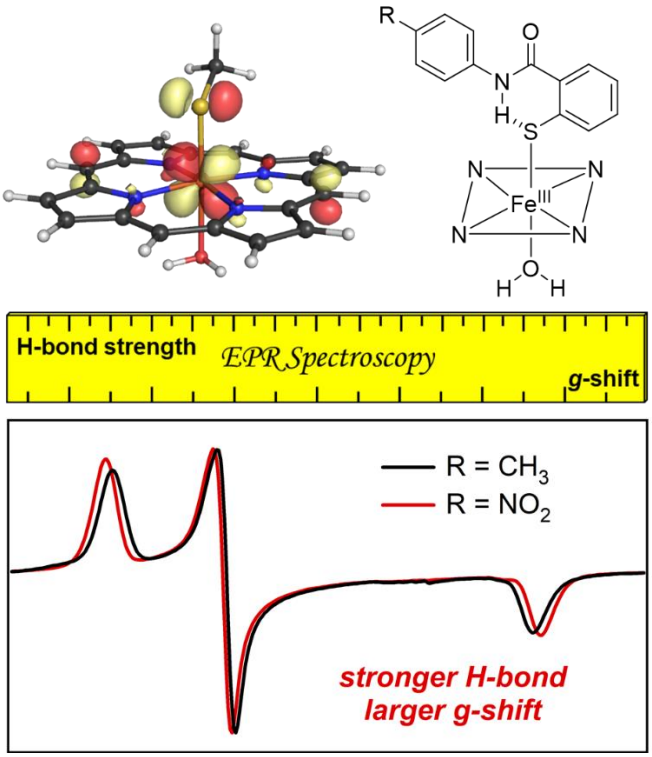


Figure 4.S8 RcoM sequence similarity network shown as a function of alignment score cutoff. The operon associated with each *rcoM* gene is annotated, and clustering based on genetic context is observed at an alignment cutoff score of 50.

Figure 4.S9 Alignment of putative promoter regions for *cox*-associated RcoM proteins. Colored nucleotides denote consensus of 65% or greater.

Chapter Five

Electron Paramagnetic Resonance Spectroscopy as a Probe of Hydrogen Bonding in Heme-Thiolate Proteins



A version of this chapter will be submitted as:

Dent, M.R.; Milbauer, M.W.; Hunt, A.P.; Aristov, M.M.; Guzei, I.; Lehnert, N.; Burstyn, J.N. Hydrogen bonding in heme-thiolate protein model complexes.

M.R.D. carried out computational analyses and wrote the manuscript. M.W.M. synthesized and characterized all ligands and ligand precursors in the lab of J.N.B. A.P.H. developed the methodology for the synthesis of 5-c aryl-thiolate complexes, and A.P.H. in the lab of N.L. A.P.H., M.W.M., and M.R.D. synthesized and characterized porphyrin complexes. M.M.A. and I.G. carried out single crystal X-ray diffraction experiments.

5.1 Introduction

Hemoproteins bearing an axial, cysteine-derived thiolate ligand (heme-thiolate proteins) are an important class of metalloproteins with diverse biological functions ranging from drug metabolism to transcriptional regulation. Importantly, properties of the *b*-type heme cofactor that give rise to diverse biological functions in heme-thiolate proteins also impart distinct spectroscopic characteristics. These properties include sixth axial ligand identity, iron coordination number and spin state, solvent exposure, and the propensity to react with substrates or bind small gaseous molecules. Two classes of heme-thiolate proteins emerge based on these functional and spectroscopic features: type-1 heme-thiolates, which act as *small-molecule activators*, and type-2 heme-thiolates, which act primarily as *small molecule sensors*.[1] Small molecule-sensing, type-2 heme-thiolate proteins include CooA and RcoM, CO-sensing transcription factors studied in detail throughout this dissertation.

We hypothesize that a key contributing factor to functional divergence in heme-thiolate proteins is the structure of the hydrogen-bonding (H-bonding) environment surrounding the coordinating Cys(thiolate) ligand. A well-defined H-bonding pocket envelops the heme-bound Cys(thiolate) in cytochrome P450 enzymes (a large group of type-1 heme-thiolate proteins), and a growing body of work supports functions for this H-bonding network in maintaining stability of the Fe-S bond and modulating reactivity of the heme.[2-10] In a number of type-2 heme-thiolate proteins, H-bonding interactions have been proposed to facilitate ligand switching at the heme,[11-14] although structural characterization of these interactions is limited.

Type-1 heme-thiolate proteins, which include a number of small-molecule-activating enzymes such as cytochromes P450 (Cyt P450s), chloroperoxidases, and nitric oxide synthases/reductases, possess a well-characterized H-bonding network that stabilizes thiolate

coordination and modulates heme reactivity. In the exemplar camphor-hydroxylating $P450_{cam}$ from *P. putida*, three amide N-H groups from three amino acid residues (Leu^{358}, Gly^{359}, and Gln^{360}) act as H-bond donors to the Cys(thiolate) ligand (Cys^{357}) (Figure 5.1, left).[15-18] Additionally, the side chain amide of Gln^{360} serves as a H-bond donor to the carbonyl O atom of Cys^{357}. Disruptions to this H-bonding pocket weaken the Fe-S bond and give rise to increased formation of the inactive P420 species.[4-6] Thus, an important role of the thiolate H-bonding network is to maintain thiolate ligation, which is critical in 1) facilitating O-O bond cleavage of the hydroperoxo intermediate via the "thiolate-push" effect and 2) increasing the basicity of the protonated ferryl species in compound II.[2,3,19-23] Hunt and Lehnert recently enumerated the electronic origin of the thiolate-push effect in synthetic models of cytochrome P450 nitric oxide reductase.[10] The authors found that the thiolate ligand exhibits a strong σ-*trans* effect, which mediates N-O bond cleavage (and O-O bond cleavage by analogy to Cyt P450s) through population of an Fe-N-O σ*-antibonding orbital. Importantly, this study demonstrated that thiolate donor strength, a property that may be tuned via changes in the thiolate H-bonding environment, directly modulates the strength of this σ-*trans* effect.

The thiolate H-bonding environment in type-2 heme-thiolate proteins likely facilitates redox-mediated ligand switching and may differ significantly from the environment of type-1 heme thiolate proteins. Type-2 heme-thiolate proteins are often small-molecule sensors involved in signal-dependent transcriptional regulation.[1] The type-2 heme exists as a low-spin species with two axial ligands; however, a "redox-mediated ligand switch" occurs in which the thiolate ligand is replaced with another protein-derived ligand upon reduction from Fe(III) to Fe(II).[11,12,24,25] A second change in heme coordination, in which a small molecule replaces an axial protein-derived ligand, occurs at the six-coordinate, Fe(II) heme *only after* the thiolate ligand is replaced. Binding

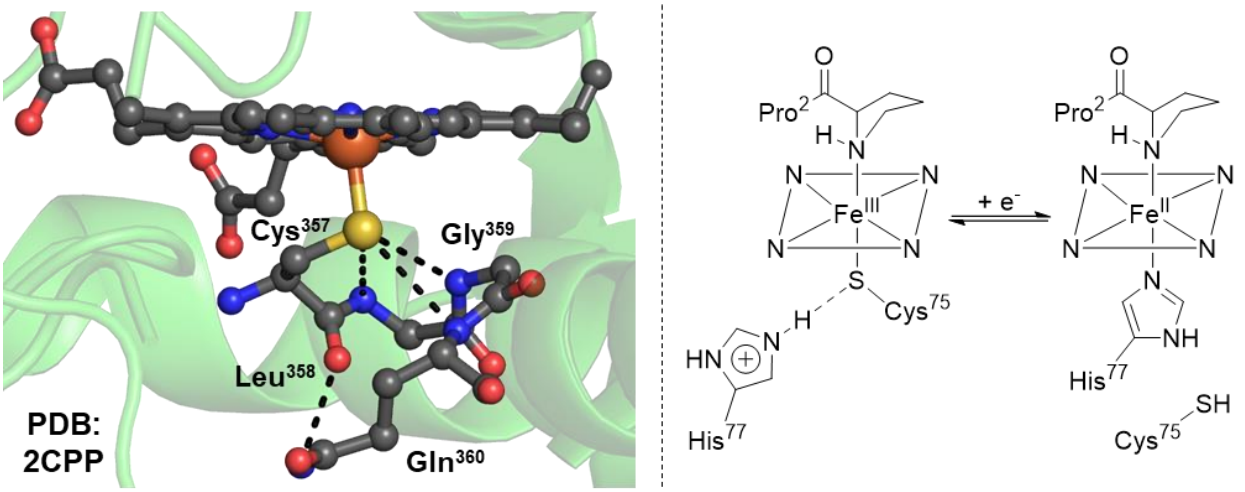


Figure 5.1 Comparison of heme-thiolate H-bonding environments in archetypical type-1 and type-2 heme-thiolate proteins. *Left*: Crystal structure of the thiolate H-bonding pocket in Cyt $P450_{cam}$ from *P. putida*, a type-1 heme-thiolate.[15] Residues involved in the thiolate H-bonding pocket are labelled, and important H-bonds are depicted with black dashes. Donor acceptor distances range from 3.08-3.56 Å. Structure was visualized using Pymol (v1.3). *Right*: Proposed H-bonding environment in *Rr* CooA, a type-2 heme-thiolate.

of a small gaseous molecule (such as carbon monoxide or nitric oxide) to heme allosterically modulates protein function, often leading to enhanced binding to a DNA promoter site.[26-28] In the archetypical type-2 heme-thiolate protein, CooA (CO oxidation activator), we speculate that a strong, directional H-bond may exist between the coordinating Cys^{75} and the protonated imidazolium ring of His^{77} (Figure 5.1, right); in support of this hypothesis, distinct changes in the EPR spectrum of the low-spin, Fe(III) heme are observed upon mutation of His^{77} to Tyr.[12] No crystal structure of CooA with heme in the Fe(III) state currently exists, possibly due to the dynamic nature of this protein in the inactive state.[29] There are limited crystallographic data for type-2 heme-thiolate proteins in general; therefore, it is difficult to compare and contrast thiolate H-bonding networks between type 1 and type 2 proteins. This limitation precludes a comprehensive assessment of the role of thiolate H-bonding in the functional differentiation between these protein types.

Ferric heme-thiolate proteins exhibit a rhombic, low-spin (S = ½), electron paramagnetic resonance (EPR) signal with characteristically small g-shifts. Table 5.1 summarizes the set of g-values for a number of hemeproteins with different axial ligands. Hemoproteins with axial His/His (cytochrome b_5) and His/Met ligation (cytochrome c) exhibit a broad rhombic signal with g-values ranging from $g \approx 3.1$ to $g \approx 1.35$.[30,31] For heme-thiolate proteins, the three g-values are observed in a very narrow range between $g \approx 2.5$ and $g \approx 1.9$.[4,25,27,32,33] Blumberg and Peisach noted the unique features of heme-thiolate proteins in their analysis of low-spin, Fe(III) hemoprotein EPR spectra in the 1970's, and the characteristic EPR signal facilitated the identification of Cys(thiolate) as an axial heme ligand when Cyt P450s were originally characterized.[34-36] The uniquely narrow rhombic signal continues to be useful in identifying new heme-thiolate proteins;[37]

Table 5.1 Comparison of experimental *g*-values recorded for low-spin, rhombic EPR signals observed in Fe(III) hemoproteins. Note: g_z, g_y, and g_x are given the historical assignments of g_{max}, g_{mid}, and g_{min} in this table.

Protein	**Axial Heme Ligands**	g_z	g_y	g_x	**Ref.**
cytochrome b_5	His/His	3.03	2.21	1.40	30
cytochrome *c*	His/Met	3.08	2.14	1.35	31
hCBS	Cys(S^-)/His	2.49	2.31	1.87	32
*Bx*RcoM-1	Cys(S^-)/His	2.48	2.26	1.88	25
*Rr*CooA	Cys(S^-)/Pro	2.46	2.25	1.89	27
cyt $P450_{cam}$	Cys(S^-)/H_2O	2.41	2.26	1.93	4
cyt $P450_{cam}$ + imidazole (Im)	Cys(S^-)/Im	2.56	2.27	1.87	33

however, the underlying features of the heme electronic structure that give rise to the narrow spread of *g*-values in heme-thiolate proteins have not been elucidated.

Herein, we uncover the electronic origins of the characteristically small *g*-shifts observed in rhombic EPR spectra of low-spin, Fe(III) heme-thiolate proteins. Additionally, we establish a relationship between the magnitude of the *g*-shift and heme-thiolate H-bond strength. We characterize a series of six-coordinate, aryl-thiolate-ligated Fe(III) porphyrin models bearing a tunable intramolecular H-bond using computational and spectroscopic methods. Our findings demonstrate that EPR spectroscopy, which is extremely sensitive to changes in Fe-S bonding, is particularly well-suited to probe second coordination sphere effects in heme-thiolate proteins. Specifically, we show that EPR spectroscopy reports on the strength of the thiolate H-bonding environment and therefore represents an important tool for understanding how this critical second coordination sphere effect influences heme-thiolate protein function.

5.2 Experimental Section

5.2.1 General Methods

Unless otherwise specified, reagents were used as received from commercial suppliers (Sigma-Aldrich, Acros Organics, Alfa Aesar, and Cambridge Isotope Laboratories Inc.). The preparation and handling of all O_2 and H_2O sensitive materials was carried out under inert conditions (N_2 or Ar gas) using standard Schlenk techniques, or in an N_2-atmosphere MBraun glovebox equipped with a circulating purifier (O_2, H_2O < 0.1 ppm). For procedures involving the synthesis of thiol ligand precursors, solvents were dried by elution through alumina. For procedures involving porphyrin complexes, all dry solvents were distilled from CaH_2 under N_2, degassed via five freeze-pump-thaw cycles, and stored over appropriately sized (3 or 4 Å) activated molecular sieves in a glovebox until used, unless otherwise stated.

5.2.2 Physical Measurements

Electronic Absorption Spectroscopy. An Analytik Jena Specord S600 spectrometer was used to record electronic absorption spectra of solutions in anaerobic screw-cap quartz cuvettes, which were prepared in a glovebox.

NMR Spectroscopy. Proton and fluorine NMR spectra were recorded on a Varian MR 400 MHz instrument or a Varian NMRS 500 or 700 MHz spectrometer at room temperature (20-22 °C). All spectra were referenced to internal solvent peaks (e.g. CD_2Cl_2: 5.32 ppm).

EPR Spectroscopy. Electron paramagnetic resonance (EPR) spectra were recorded on a Bruker X-band EMX or ELEXSYS E500 spectrophotometer equipped with an Oxford Instruments liquid helium cryostat. EPR spectra were obtained on frozen solution samples (0.5-2 mM) in a 1:1 mixture of CH_2Cl_2 and toluene, using 1-20 mW microwave power and 100 kHz field modulation with the amplitude set to 3 or 8.3 G. The *g*-values for each EPR spectrum were extracted from simulations performed using EasySpin (v5.2.23).[38]

Elemental Analysis. Elemental analyses (C, H, and N) were performed at Atlantic Microlab.

Mass Spectrometry. High-resolution mass spectrometry (HRMS) experiments were carried out using a Thermo Q Exactive Plus electrospray ionization-quadrupole-ion trap mass spectrometer.

5.2.3 Syntheses

*General Procedure for the Synthesis of 2,2'-disulfanediylbis(N-phenylbenzamide) (**SS**-H) and Derivatives*

Oxalyl chloride (4 mL, 47 mmol, 2.6 eq) was added to a suspension of 2,2'-dithiodibenzoic acid (5.5 g, 18.0 mmol) and catalytic dimethylformamide in CH_2Cl_2 (75 mL) at room temperature. The reaction mixture was heated to reflux and stirred for 4h until no solids remained. The reaction mixture was then hot filtered, and solvent was removed from the collected filtrate *in vacuo*. The

resulting amber, crystalline solid was used without further purification. Aniline (0.8 mL, 8.7 mmol, 3 eq) was added slowly to a suspension of 2,2'-disulfanediyldibenzoyl chloride (1.00 g, 2.9 mmol) and $NaHCO_3$ (0.979 g, 11.7 mmol, 4 eq) in CH_2Cl_2 (30 mL), and the reaction mixture was stirred overnight at room temperature. The precipitate was collected by vacuum filtration, suspended in $HCl_{(aq)}$ (3.7%, 40 mL), and stirred for 15 min. Solids were collected by vacuum filtration and washed with Et_2O (40 mL) to afford 2,2'-disulfanediylbis(*N*-phenylbenzamide) as an off-white solid. ^{1}H NMR (DMSO-d_6) δ 10.57 (s, 2H), 7.81 – 7.73 (m, 8H), 7.52 (td, *J* = 7.7, 1.5 Hz, 2H), 7.43 – 7.36 (m, 6H), 7.17 – 7.10 (m, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 166.17, 139.33, 136.92, 135.17, 131.91, 129.22, 128.95, 126.84, 126.74, 124.46, 120.60. HRMS (+ESI) m/z: $[M + Na]^+$ Calculated for $C_{26}H_{20}N_2O_2S_2Na$: 479.0864; Found: 479.0825. Yield = 1.17 g (2.6 mmol), 87%.

*2,2'-disulfanediylbis(N-(p-tolyl)benzamide (**SS**-CH₃)*: *p*-Toluidine (0.937 g, 8.7 mmol, 3 eq) was used as a starting material following the general procedure above to afford 2,2'-disulfanediylbis(*N*-(*p*-tolyl)benzamide) as an off-white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.48 (s, 2H), 7.77 (dd, *J* = 7.7, 1.4 Hz, 2H), 7.74 (dd, *J* = 8.1, 1.1 Hz, 2H), 7.65 (d, *J* = 8.2 Hz, 4H), 7.51 (td, *J* = 7.7, 1.5 Hz, 2H), 7.39 (td, *J* = 7.4, 1.2 Hz, 2H), 7.18 (d, *J* = 8.2 Hz, 4H), 2.30 (s, 6H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 165.47, 136.45, 136.34, 134.71, 132.96, 131.34, 129.11, 128.40, 126.31, 126.20, 120.11, 20.52. ^{13}C NMR (126 MHz, DMSO-d_6) δ 165.95, 136.93, 136.82, 135.19, 133.43, 131.82, 129.59, 128.88, 126.79, 126.68, 120.59, 21.00. HRMS (+ESI) m/z: $[M + Na]^+$ Calculated for $C_{28}H_{24}N_2O_2S_2Na$: 507.1177; Found: 507.1152. Yield = 1.23 g (2.5 mmol), 85%.

*2,2'-disulfanediylbis(N-(4-chlorophenyl)benzamide) (**SS**-Cl)*: 4-Chloroaniline (1.16 g, 8.7 mmol, 3 eq) was used as a starting material following the general procedure above to afford 2,2'-disulfanediylbis(*N*-(4-chlorophenyl)benzamide) as an off-white solid . ^{1}H NMR (DMSO-d_6) δ

10.71 (s, 2H), 7.86 – 7.77 (m, 6H), 7.75 (dd, *J* = 8.1, 1.1 Hz, 2H), 7.53 (td, *J* = 7.7, 1.5 Hz, 2H), 7.47 – 7.43 (m, 4H), 7.40 (td, *J* = 7.5, 1.1 Hz, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 166.22, 138.29, 136.96, 134.84, 132.07, 129.15, 129.04, 128.07, 126.90, 126.82, 122.12. HRMS (+ESI) m/z: $[M + Na]^+$ Calculated for $C_{26}H_{18}Cl_2N_2O_2S_2Na$: 547.0084; Found: 547.0078.Yield = 1.39 g (2.7 mmol), 91%.

*2,2'-disulfanediylbis(N-(4-(trifluoromethyl)phenyl)benzamide) (**SS**-CF_3)*: 4-(Trifluoromethyl) aniline (1.1 mL, 8.7 mmol, 3 eq) was used as a starting material following the general procedure above to afford 2,2'-disulfanediylbis(*N*-(4-(trifluoromethyl)phenyl)benzamide) as an off-white solid. ^{1}H NMR (DMSO-d_6) δ 10.92 (s, 2H), 7.99 (d, *J* = 8.6 Hz, 4H), 7.83 (dd, *J* = 7.7, 1.4 Hz, 2H), 7.78-7.75 (m, 6H), 7.55 (td, *J* = 7.7, 1.4 Hz, 2H), 7.42 (td, *J* = 7.5, 1.1 Hz, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 166.16, 142.47, 136.56, 134.19, 131.81, 128.77, 126.53, 126.49, 126.08 (q, *J* = 3.9 Hz), 124.37 (q, *J* = 271.3 Hz), 123.97 (q, *J* = 32.0 Hz), 120.04. HRMS (+ESI) m/z: $[M + Na]^+$ Calculated for $C_{28}H_{18}F_6N_2O_2S_2Na$: 615.0612; Found: 615.0598.Yield = 1.61 g (2.7 mmol), 93%.

*2,2'-disulfanediylbis(N-(4-nitrophenyl)benzamide) (**SS**-NO_2)*: The general procedure was modified by heating the reaction to reflux and stirring for 2 days following the addition of 4-nitroaniline (1.21 g, 8.7 mmol, 3 eq). 2,2'-disulfanediylbis(*N*-(4-nitrophenyl)benzamide) was obtained as an off-white solid. ^{1}H NMR (DMSO-d_6) δ 11.13 (s, 2H), 8.32 – 8.27 (m, 4H), 8.03 (d, *J* = 9.3 Hz, 4H), 7.86 (dd, *J* = 7.6, 1.5 Hz, 2H), 7.81 – 7.76 (m, 2H), 7.57 (td, *J* = 7.8, 1.5 Hz, 2H), 7.47 – 7.41 (m, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 166.31, 145.04, 142.74, 136.66, 133.90, 132.03, 128.92, 126.62, 126.60, 124.93, 119.81. HRMS (+ESI) m/z: $[M + Na]^+$ Calculated for $C_{26}H_{18}N_4O_6S_2Na$: 569.0565; Found: 569.0561. Yield = 1.25 g (2.3 mmol), 77%.

*2-mercapto-N-phenylbenzamide (**HS**-H, Method A)*: 2,2'-disulfanediylbis(*N*-phenylbenzamide) (456 mg, 1 mmol) and triphenylphosphine (1.18 g, 4.5 mmol) were suspended in a 50/50 mixture of CH_3CN/H_2O and stirred at reflux for one hour. The suspension was then added to a separatory funnel and mixed with $NaCl_{(aq)}$ brine (15 mL) and hexanes (15 mL). The middle CH_3CN layer was collected and dried with $MgSO_4$, and solvent was removed *in vacuo*. The crude solid was purified by silica column chromatography (100% CH_2Cl_2) to afford 2-mercapto-*N*-phenylbenzamide as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.77 (s, 1H), 7.61 (d, *J* = 8.0 Hz, 2H), 7.58 (dd, *J* = 7.7, 1.5 Hz, 1H), 7.39 – 7.34 (m, 3H), 7.30 (td, *J* = 7.6, 1.5 Hz, 1H), 7.22 – 7.14 (m, 2H), 4.56 (s, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.59, 137.61, 133.55, 132.98, 131.39, 131.01, 129.14, 128.03, 125.50, 124.87, 120.25. HRMS (+ESI) m/z: $[M + H]^+$ Calculated for $C_{13}H_{12}NOS$: 228.0489; Found: 228.0489. Yield = 395 mg (1.7 mmol), 86%.

*2-mercapto-N-(p-tolyl)benzamide (**HS**-CH_3)*: 2,2'-disulfanediylbis(*N*-(*p*-tolyl)benzamide) (485 mg, 1 mmol) was used as a starting material following Method A above to afford 2-mercapto-*N*-(*p*-tolyl)benzamide as a white solid. ^{1}H NMR ($CDCl_3$) δ 7.65 (s, 1H), 7.58 (d, *J* = 7.6 Hz, 1H), 7.50 (d, *J* = 7.9 Hz, 2H), 7.38 (dd, *J* = 7.8, 1.3 Hz, 1H), 7.31 (td, *J* = 7.6, 1.5 Hz, 1H), 7.22 (dd, *J* = 7.6, 1.3 Hz, 1H), 7.19 (t, *J* = 7.1 Hz, 2H), 4.62 (s, 1H), 2.35 (s, 3H).). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 166.69, 135.15, 134.64, 133.66, 133.04, 131.36, 130.97, 129.67, 128.14, 125.50, 120.49, 21.03. HRMS (+ESI) m/z: $[M + H]^+$ Calculated for $C_{14}H_{14}NOS$: 242.0645; Found: 242.0645. Yield = 482 mg (2.0 mmol), 99%.

*2-mercapto-N-(4-chlorophenyl)benzamide (**HS**-Cl)*: 2,2'-disulfanediylbis(*N*-(4-chlorophenyl)benzamide) (525 mg, 1 mmol) was used as a starting material following Method A above to afford 2-mercapto-*N*-(4-chlorophenyl)benzamide as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.79 (s, 1H), 7.59 – 7.53 (m, 3H), 7.37 (dd, *J* = 7.9, 1.3 Hz, 1H), 7.35 – 7.29 (m, 3H),

7.21 (td, *J* = 7.5, 1.3 Hz, 1H), 4.48 (s, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.50, 136.19, 133.20, 133.01, 131.55, 131.22, 129.90, 129.16, 128.05, 125.59, 121.46. HRMS (+ESI) m/z: $[M + H]^+$ Calculated for $C_{13}H_{11}ClNOS$: 262.0099; Found: 262.0097. Yield = 384 mg (1.5 mmol), 73%.

*2-mercapto-N-(4-(trifluoromethyl)phenyl)benzamide (**HS**-CF_3, Method B)*: 2,2'-disulfanediylbis(*N*-(4-(trifluoromethyl)phenyl)benzamide) (593 mg, 1 mmol) and $NaBH_4$ (378 mg, 10 mmol) were placed in an oven-dried flask and evacuated and refilled with N_2 three times. Methanol (15 mL) was added slowly to prevent vigorous bubbling, and the resulting reaction mixture was stirred for one hour at ambient temperature. Methanol was then removed *in vacuo*, and the resulting solid was redissolved in ethyl acetate (20 mL). This solution was stirred with $HCl_{(aq)}$ (7%, 20 mL) for one hour. The compound of interest was extracted from this mixture into ethyl acetate (2 x 10 mL). Combined organic layers were dried over $MgSO_4$ and solvent was removed *in vacuo*. The crude solid was purified by silica column chromatography (100% CH_2Cl_2) to afford 2-mercapto-*N*-(4-(trifluoromethyl)phenyl)benzamide as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.88 (s, 1H), 7.76 (d, *J* = 8.4 Hz, 2H), 7.63 (dd, *J* = 9.0, 7.4 Hz, 3H), 7.41 (dd, *J* = 7.8, 1.4 Hz, 1H), 7.35 (td, *J* = 7.6, 1.5 Hz, 1H), 7.28 – 7.23 (m, 2H), 4.44 (s, 1H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 166.79, 140.82, 133.22, 133.10, 131.80, 131.56, 128.25, 126.72 (q, *J* = 32.9 Hz), 126.52, 125.80, 124.16 (q, *J* = 271.7 Hz), 119.93. ^{19}F NMR (471 MHz, Chloroform-*d*) δ -62.19. HRMS (+ESI) m/z: $[M + H]^+$ Calculated for $C_{14}H_{11}F_3NOS$: 298.0508; Found: 298.0504. Yield = 164 mg (0.55 mmol), 55%.

*2-mercapto-N-(4-nitrophenyl)benzamide (**HS**-NO_2)*: 2,2'-disulfanediylbis(*N*-(4-nitrophenyl)benzamide) (547 mg, 1 mmol) was used as a starting material following Method B above to afford 2-mercapto-*N*-(4-nitrophenyl)benzamide as a pale yellow microcrystalline solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.28 – 8.22 (m, 2H), 8.18 (s, 1H), 7.85 – 7.80 (m, 2H), 7.63 (dd, *J*

= 7.7, 1.5 Hz, 1H), 7.41 (dd, *J* = 7.9, 1.4 Hz, 1H), 7.36 (td, *J* = 7.6, 1.5 Hz, 1H), 7.28 – 7.22 (m, 1H), 4.35 (s, 1H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 166.63, 144.09, 143.55, 133.37, 132.71, 132.08, 131.91, 128.36, 125.99, 125.33, 119.64. HRMS (-ESI) m/z: $[M - H]^-$ Calculated for $C_{13}H_9N_2O_3S$: 273.0339; Found: 273.0640. Yield = 343 mg (1.3 mmol), 62%.

General Procedure for the Synthesis of Five-Coordinate, Fe(III) Thiolate-Ligated Porphyrin Complexes

All iron-containing tetraphenylporphyrinate (TPP^{2-}) complexes were prepared using the same general procedure recently described for the synthesis of high-spin, five-coordinate [Fe(TPP)(**S**-R)] complexes, by mixing the desired thiol ligand precursor with [Fe(TPP)(OCH_3)].[10] All complexes were recrystallized by dissolution in a minimal amount of toluene, followed by layering of the solutions with hexanes and/or methanol.

*[Fe(TPP)(**S**-CH$_3$)]*: Electronic Absorption (CH_2Cl_2): 380 (shoulder), 407, 515, 576, 719 nm. ^{1}H-NMR (CD_2Cl_2): 71.4 (br, β-pyrrole H), 64.1 and 56.4 (s, *meta*-SPh's), 12.7 (br-s, *meta*-Ph TPP), 9.6 (br-s, *ortho*-Ph TPP), 6.9 (br-s, *para*-Ph TPP), 4.1 (s, SPh-NHPh-*para*-CH_3), -85.2 (s, *para*-SPh), -92.6 (br, *ortho*-SPh) ppm. Other SPh-NHPh-*p*CH_3 signals observed: 8.3 (br), 4.8 (br) and -109.9 (br) ppm. Anal. Calc'd for $C_{58}H_{40}FeN_5OS$: C, 76.48; H, 4.43; N, 7.69. Found: C, 76.35; H, 4.47; N, 7.61. Yield = 47.9 mg, 78.8 % (starting [Fe(TPP)(OCH_3)] mass = 46.66 mg).

EPR (1:1 CH_2Cl_2:toluene): g_x =1.95, g_y =1.88, g_z =2.04; *E*/*D* = 0.040.

*[Fe(TPP)(**S**-H)]*: Electronic Absorption (CH_2Cl_2): 380 (shoulder), 407, 515, 576, 717 nm. ^{1}H-NMR (CD_2Cl_2): 71.6 (br, β-pyrrole H), 63.7 and 56.5 (s, *meta*-SPh's), 12.8 (br-s, meta-Ph TPP), 9.8 (br-s, *ortho*-Ph TPP), 6.9 (br-s, para-Ph TPP), 6.5 (s, SPh-NHPh-para-H), -85.2 (s, *para*-SPh), -92.8 (br, *ortho*-SPh) ppm. Other SPh-NHPh signals observed: 8.3 (br), 4.9 (br) and -111.3 (br) ppm. Anal. Calc'd for $C_{57}H_3FeN_5OS$: C, 76.34; H, 4.27; N, 7.81. Found: C, 75.66; H, 4.29; N,

7.62. Yield = 117.6 mg, 77.8 % (starting [Fe(TPP)(OCH_3)] mass = 101.4 mg). EPR (1:1 CH_2Cl_2:toluene): g_x =1.95, g_y =1.88, g_z =2.085; *E*/*D* = 0.034.

*[Fe(TPP)(**S**-Cl)]*: Electronic Absorption (CH_2Cl_2): 380 (shoulder), 408, 515, 578, 720 nm. ^{1}H-NMR (CD_2Cl_2): 71.5 (br, β-pyrrole H), 62.5 and 55.6 (s, *meta*-SPh's), 12.8 (br-s, *meta*-Ph TPP), 9.6 (br-s, *ortho*-Ph TPP), 6.8 (br-s, *para*-Ph TPP), -84.5 (s, *para*-SPh), -91.9 (br, *ortho*-SPh) ppm. Other SPh-NHPh-*p*Cl signals observed: 8.3 (br), 4.8 (br) and -111.6 (br) ppm. Anal. Calc'd for $C_{57}H_{37}ClFeN_5OS$: C, 73.51; H, 4.00; N, 7.52; S, 3.44. Found: C, 72.21; H, 4.07; N 7.42, S, 3.60. Yield = 32.8 mg, 41.7 % (starting [Fe(TPP)(OCH_3)] mass = 59.1 mg). EPR (1:1 CH_2Cl_2:toluene): g_x =1.95, g_y =1.88, g_z =2.06; *E*/*D* = 0.034.

*[Fe(TPP)(**S**-CF_3)]*: Electronic Absorption (CH_2Cl_2): 380 (shoulder), 408, 515, 578, 721 nm. ^{1}H-NMR (CD_2Cl_2): 68.4 (br, β-pyrrole H), 61.2 and 54.8 (s, *meta*-SPh's), 12.8 (br-s, *meta*-Ph TPP), 9.9 (br-s, *ortho*-Ph TPP), 6.9 (br-s, *para*-Ph TPP), -83.2 (s, *para*-SPh), -90.4 (br, *ortho*-SPh) ppm. Other SPh-NHPh-*p*CF_3 signals observed: 8.0 (br), 5.0 (br) and -110.5 (br) ppm. ^{19}F-NMR (CD_2Cl_2): -60.3 ppm. Anal. Calc'd for $C_{58}H_{37}F_3FeN_5OS$: C, 72.20; H, 3.87; N, 7.26. Found: C, 71.66; H, 3.86; N, 7.26. Yield = 34.8 mg, 46.7 % (starting [Fe(TPP)(OCH_3)] mass = 54.0 mg). EPR (1:1 CH_2Cl_2:toluene): g_x =1.93, g_y =1.87, g_z =2.06; *E*/*D* = 0.037.

*[Fe(TPP)(**S**-NO_2)]*: Electronic Absorption (CH_2Cl_2): 336 (shoulder), 380 (shoulder), 408, 516, 578, 722 nm. ^{1}H-NMR (CD_2Cl_2): 70.5 (br, β-pyrrole H), 69.9 and 54.3 (s, *meta*-SPh's), 12.9 (br-s, *meta*-Ph TPP), 10.5 (br-s, *ortho*-Ph TPP), 6.9 (br-s, *para*-Ph TPP), -83.7 (s, *para*-SPh), -90.2 (br, *ortho*-SPh) ppm. Other SPh-NHPh-*p*NO_2 signals observed: 8.2 (br), 4.9 (br) and -112.4 (br) ppm. Anal. Calc'd for $C_{57}H_{37}FeN_6O_3S$: C, 72.69; H, 3.96; N, 8.92. Found: C, 72.35; H, 3.75; N, 8.71. Yield = 87.5 mg, 64.0 % (starting [Fe(TPP)(OCH_3)] mass = 101.5 mg). EPR (1:1 CH_2Cl_2:toluene): g_x =1.94, g_y =1.95, g_z =2.00, *E*/*D* =0.048 .

Assembly of Six-Coordinate, Fe(III) Thiolate-Ligated Porphyrin Complexes for EPR Spectroscopy

Six-coordinate (6-c) low-spin, Fe(III) porphyrin-thiolate complexes were assembled with either 1-methylimidazole (1-MeIm) or water as a sixth axial ligand. All complexes were prepared in EPR tubes in a glove box. Porphyrin-thiolate complexes bound to 1-MeIm were prepared by mixing a solution of [Fe(TPP)(**S**-R)] (500 μM in 50:50 CH_2Cl_2:toluene) with three molar equivalents of 1-MeIm. Porphyrin-thiolate complexes axially ligated by a water molecule were prepared by dissolving each 5-c compound in solvent (50:50 CH_2Cl_2:toluene) that had been dried only by elution through alumina and *without* further drying. The concentration of water in the solvent ($\geq$ 2 mM, as determined by potentiometric titration) was such that it was in excess of the 5-c porphyrin species in each of these samples (500 μM).

5.2.4 X-Ray Crystallography

Data Collection. Crystals of [Fe(TPP)(**S**-NO_2)] were grown in a glovebox by vapor diffusion of dry pentanes into a solution of the compound dissolved in dry dichloromethane. A black crystal with approximate dimensions 0.098 x 0.050 x 0.013 mm^3 was selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount©. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. The crystal evaluation and data collection were performed on a Bruker Quazar SMART APEXII diffractometer with Mo K_α (λ = 0.71073 Å) radiation and the diffractometer to crystal distance of 4.96 cm.[39] The initial cell constants were obtained from three series of ω scans at different starting angles. Each series consisted of 12 frames collected at intervals of 0.5º in a 6º range about ω with the exposure time of 10 seconds per frame. The reflections were successfully indexed by an automated indexing routine built in the APEXII program suite. The final cell constants were calculated from a set of 9962 strong reflections from the actual data collection.

Diffraction data were collected by using the full sphere data collection routine to survey the reciprocal space to the extent of a full sphere to a resolution of 0.79 Å. A total of 99284 data points were harvested by collecting 6 sets of frames with 0.45º scans in ω and φ with exposure times of 70 sec per frame. These highly redundant datasets were corrected for Lorentz and polarization effects. The absorption correction was based on fitting a function to the empirical transmission surface as sampled by multiple equivalent measurements.[40]

Structure Solution and Refinement. The systematic absences in the diffraction data were uniquely consistent for the space groups $P2_1/c$ that yielded chemically reasonable and computationally stable results of refinement.[41-46] A successful solution by the direct methods provided most non-hydrogen atoms from the *E*-map. The remaining non-hydrogen atoms were located in an alternating series of least-squares cycles and difference Fourier maps. All non-hydrogen atoms were refined with anisotropic displacement coefficients. All hydrogen atoms except H5 were included in the structure factor calculation at idealized positions and were allowed to ride on the neighboring atoms with relative isotropic displacement coefficients. Atom H5 was found in the difference Fourier map and refined independently. This atom participates in the N5–H5···S1 interaction with a D-A distance of 3.0087(19) Å and a D–H···A angle of 149(2)° (Figure 5.S45). Atom H4a participates in a C4a–H4a···O2 [1-X,0.5+Y,0.5-Z] interaction with a D-A distance of 2.895(19) Å and a D–H···A angle of 159.9° (Figure 5.S46). The phenyl ring bonded at C2 was disordered over two positions. The occupancy of the major component was 86.7(8) %. The minor disordered component of the phenyl ring was refined with distance restraints. The final least-squares refinement of 646 parameters against 9382 data resulted in residuals *R* (based on F^2 for

Table 5.2 Crystal data and structural refinement parameters for [Fe(TPP)(**S**-NO_2)].

Empirical formula	$(C_{44}H_{28}N_4)Fe(C_{13}H_9SN_2O_3)$
Formula weight	941.83
Temperature/K	100.0
Crystal system	monoclinic
Space group	$P2_1/c$
a/Å	14.124(4)
b/Å	12.805(4)
c/Å	24.419(8)
α/°	90
β/°	90.916(13)
γ/°	90
Volume/$Å^3$	4416(2)
Z	4
$\rho_{calc}g/cm^3$	1.417
μ/mm^{-1}	0.445
F(000)	1948.0
Crystal size/mm^3	0.098 × 0.05 × 0.013
Radiation	MoKα (λ = 0.71073)
2Θ range for data collection/°	2.884 to 53.52
Index ranges	$-17 \leq h \leq 17$, $-16 \leq k \leq 16$, $-30 \leq l \leq 30$
Reflections collected	96775
Independent reflections	9382 [R_{int} = 0.0684, R_{sigma} = 0.0337]
Data/restraints/parameters	9382/12/646
Goodness-of-fit on F^2	1.039
Final R indexes [I>=2σ (I)]	R_1 = 0.0360, wR_2 = 0.0815
Final R indexes [all data]	R_1 = 0.0548, wR_2 = 0.0887
Largest diff. peak/hole / e $Å^{-3}$	0.36/-0.42

$I \geq 2\sigma$) and *wR* (based on F^2 for all data) of 0.0360 and 0.0887, respectively. The final difference Fourier map was featureless. Crystallographic parameters are summarized in Table 5.2.

5.2.5 Relationship between experimentally-determined *g*-values and ligand field parameters

Ligand field parameters were estimated from experimentally-determined *g*-values for low-spin, Fe(III) porphyrin models using the tetragonally-distorted, strong field d^5 model described by McGarvey.[47] In the strong field approximation, the wavefunction for the ground state Kramer's doublet is modeled in terms of three low-lying, metal-based *d*-orbitals, d_{xy}, d_{xz}, and d_{yz}. Wavefunction coefficients for each orbital (*A*, *B*, and *C* for d_{xy}, d_{yz}, and d_{xz}, respectively), as well as the orbital reduction factor *k*, are computed directly from experimentally-determined values of g_x, g_y, and g_z. Two additional ligand field parameters, rhombic and axial distortions, are then computed from these wavefunction coefficients. The rhombic distortion (V/ξ), scaled to the single electron spin-orbit constant for Fe^{3+} ($\xi = 460$ cm^{-1}),[48] approximates the energy difference between Fe $3d_{xz}$ and $3d_{yz}$ orbitals (i.e. $E_{xz} = V/2 = -E_{yz}$). The axial distortion (Δ/ξ), scaled to ξ, approximates the energy difference between Fe $3d_{xy}$ and the barycenter of the $3d_{xz}$ and $3d_{yz}$ orbitals (i.e. $E_{xy} = \Delta$).

All ligand field parameters were computed using the program written by McGarvey, which systematically assigns the sign and magnitude of each *g*-value assuming maximal distortion along the *z*-axis (i.e. $|\Delta| \geq 2|V|/3$). For most low-spin, Fe(III) porphyrin complexes, the axis of maximal distortion lies normal to the heme plane and thus aligns with the molecular *z*-axis. In this case, computed values for V/ξ and Δ/ξ were used directly. However, for thiolate-ligated porphyrins, the axis of maximal distortion lies *within* the heme plane, leading to an interchange of the *x* and *z* axes. To provide meaningful information about electronic structure in terms of the molecular axes (in

which z is normal to the heme plane), computed values for V/ξ and Δ/ξ in thiolate-ligated porphyrins were converted using Equations 1 and 2.

$$V' = 0.5V - \Delta \quad (1)$$

$$\Delta' = -0.5\Delta - 0.75V \quad (2)$$

5.2.5 Computational methods

Density functional theory calculations. For 6-c, low-spin Fe(III) porphyrin models, geometry optimizations, single-point calculations, and time-dependent density functional theory (TD-DFT) computations were carried out using ORCA v4.0,[49] and frequency calculations were carried out using Gaussian09.[50] Geometry optimizations utilized Becke's functional for exchange along with Perdew's functional for correlation (BP86),[51,52] and all atoms were described using Alrich's polarized triple ζ-valence (TZVP) basis set along with the def2/J auxiliary basis.[53,54] Frequency calculations utilized Becke's three-parameter hybrid functional for exchange and the Lee-Yang-Parr functional for correlation (B3LYP), with the 6-31G(d) basis set.[55] Spin-unrestricted single-point calculations were performed using the B3LYP functional along with the polarized split valence potential (SVP) basis set and def2/J auxiliary basis for all atoms except Fe and S, which were modeled using the TZVPP basis set, and Fe-coordinated N atoms, which were modeled using the TZVP basis set.[54,56] Energy differences between ground and excited states were calculated using TD-DFT with the camB3LYP functional along with the polarized SVP basis set and def2/J auxiliary basis for all atoms except Fe and S, which were modeled using the TZVPP basis set, and Fe-coordinated N atoms, which were modeled using the TZVP basis set.

Initial coordinates for archetype hemoprotein models were obtained from protein crystal structures for cytochrome b_5 (PBD 1CYO),[57] cytochrome *c* (PBD 1CCR),[58] and human

cystathionine β-synthase (PBD 1JBQ).[59] Histidine ligands were truncated at the β-methylene C atom, which was replaced with a methyl group, the methionine ligand was truncated at the γ-methylene C atom and replaced with a methyl group, and the cysteine ligand was truncated at the β-methylene C atom, which was replaced with a methyl group. Porphine (P^{2-}) was used to model all porphyrins by replacing porphyrin substituents H atoms. Initial coordinates for synthetic models were derived from the crystal structure of [Fe(TPP)(**S**-NO_2)]. To generate starting coordinates for each 6-c compound, the *N*-phenyl *para*-substituent was modeled using idealized geometries, and a second axial ligand was added based on an appropriate crystal structure of a six-coordinate heme or porphyrin model complex.[59-61] To ensure proper positioning of the Fe atom in each model, the Fe atom was shifted to achieve co-planarity with the coordinating porphyrin N atoms in starting coordinates. All structural modifications were performed using the molecular builder in WebMO (v19.0.009e).[62]

5.3 Results

5.3.1 Electronic Origins of the Uniquely Narrow Rhombic EPR Signal Associated with Low-Spin, Fe(III) Heme-Thiolate Proteins

We utilized density functional theory (DFT) computations to analyze metal-ligand bonding in three low-spin, Fe(III) hemoprotein models with His/His (cytochrome b_5), His/Met (cytochrome *c*), and His/Cys(S^-) (hCBS, *Px*RcoM) axial ligands. In all cases, the computed structures were consistent with those observed experimentally (Table 5.3). Axial metal-ligand bond distances were slightly (up to 0.1 Å) shorter than those observed crystallographically, which is unsurprising given that the computational models do not account for secondary coordination sphere effects of the protein heme pocket. Using geometry optimized model structures, we computed single-point energies and ground state wavefunctions using DFT. We compared metal-ligand bonding in each

Table 5.3 Summary of experimental and DFT-computed metal-ligand bond distances for low-spin, Fe(III) hemoprotein models.

protein	L_1	L_2	Experimental			Computed		
			$d_{Fe\text{-}Npor}$[1] (Å)	$d_{Fe\text{-}L1}$ (Å)	$d_{Fe\text{-}L2}$ (Å)	$d_{Fe\text{-}Npor}$[1] (Å)	$d_{Fe\text{-}L1}$ (Å)	$d_{Fe\text{-}L2}$ (Å)
cytochrome b_5	His(N)	His(N)	1.988	2.075	2.003	2.008	1.997	1.996
cytochrome *c*	His(N)	Met(S)	2.005	2.037	2.351	2.002	2.009	2.338
hCBS/*Px*RcoM	His(N)	Cys(S^-)	2.010	2.210	2.281	2.011	2.104	2.186

1. Average Fe-N bond distance for pyrrole N atoms

of these hemoprotein models by examining the energies and orbital compositions of the predominantly Fe 3*d*-based frontier molecular orbitals (MOs) (Figure 5.2).

The orbital contributions to iron-ligand bonding were similar to one another in the His/His and His/Met models (Figures 5.S1 and 5.S2). In the case of the His/His model, the lowest energy Fe 3*d*-based MO was primarily d_{xy} in character. For the His/Met model, the lowest energy orbital was also predominantly Fe 3d_{xy} in character; however, there was also a significant contribution from a sulfur 3*p* orbital. This filled-filled π-type interaction results in no net Fe-S bonding character. The next lowest energy orbitals observed are the nearly-degenerate Fe 3d_{xz}- and 3d_{yz}-based MOs, which both exhibit π-antibonding character with respect to the porphyrin p_z orbitals. The second-highest energy Fe 3*d*-based MO is largely d_{z2} in character. This MO is primarily σ-antibonding with respect to the two axial heme ligands, is slightly σ-antibonding with respect to the porphyrin N atoms, and is slightly higher in energy in the His/His model than in the His/Met model. The highest energy Fe 3*d*-based MO in both models is largely d_{x2-y2} in character and strongly σ-antibonding with respect to the porphyrin N atoms.

In the thiolate-bound His/Cys(S^-) model, the orbital contributions to axial iron-ligand bonding differ from those of the His/His and His/Met models. Specifically, one of the thiolate S 3*p* orbitals has the appropriate symmetry to engage in a p_π-d_π bonding interaction with the Fe 3d_{yz} orbital. This bonding interaction raises the energy of Fe 3d_{yz}-based MO and lifts the near-degeneracy of d_{xz} and d_{yz} (Figure 5.2). Additionally, the energy of Fe 3d_{xz}-based MO is lowered slightly, relative to those of the His/His and His/Met models.

Axial thiolate ligation to low-spin, Fe(III) heme increases the differences in energy between the ground state and two low-lying excited states. Using TD-DFT, we computed the differences in energy between the ground state and first two excited states in low-spin, Fe(III)

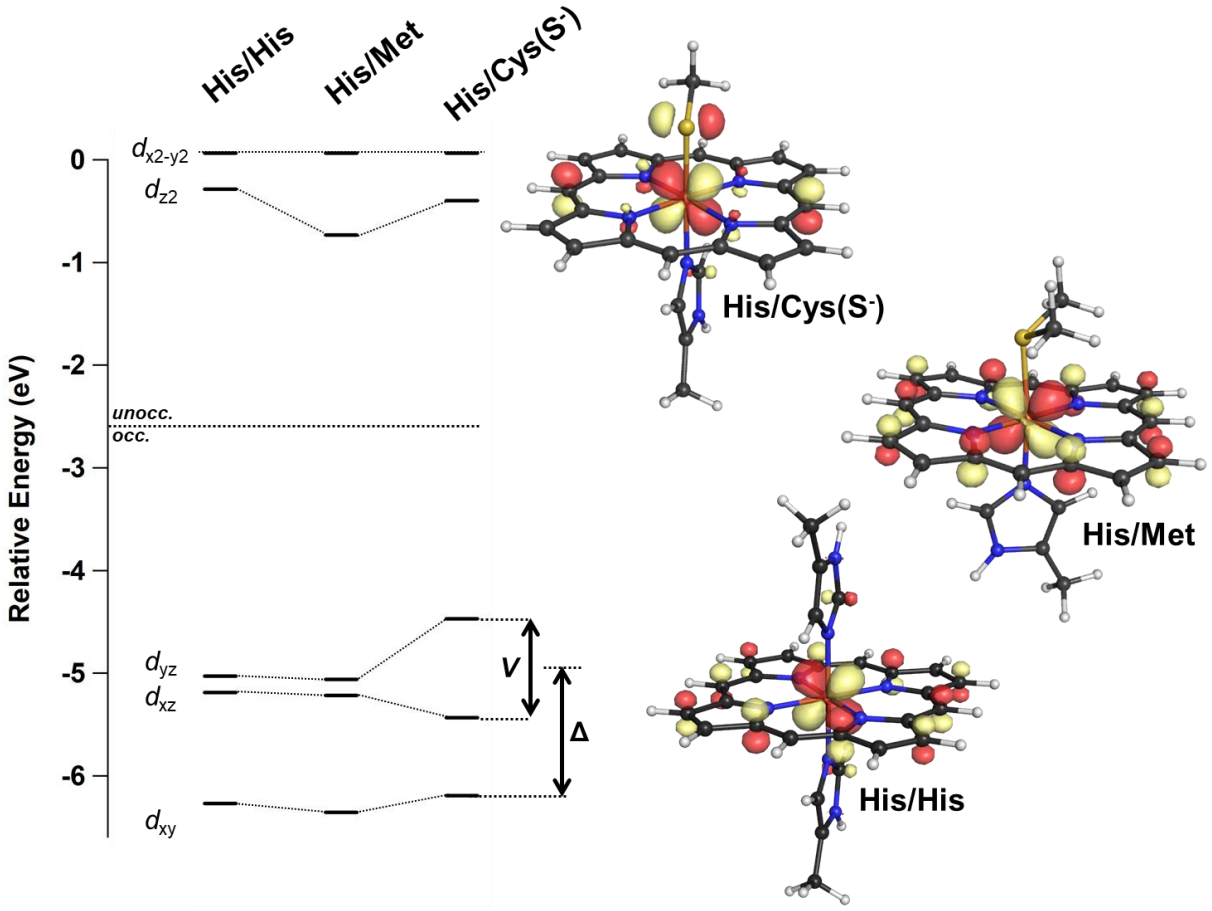


Figure 5.2 Electronic structure analysis of metal-ligand bonding in computational models of low-spin, Fe(III) porphyrins with different axial ligands. *Left*: The MO diagram depicts the relative energies of Fe 3*d*-based MOs (spin up, single-electron orbitals). Orbital energies are relative to the energy of the $3d_{x2-y2}$ orbital for each model. Two ligand field parameters, rhombic (*V*) and axial (Δ) splitting, are defined in terms of the relative energies of Fe $3d_{xy}$-, $3d_{xz}$-, and $3d_{yz}$-based MOs. *Right*: Geometry-optimized structures of each porphyrin model shown with an overlay of the Fe $3d_{yz}$-based, spin-down MO. Orbital plots were generated with an isosurface value of 0.05, and structures were visualized using Pymol (v1.3).

Table 5.4 Computational analysis of low-spin, Fe(III) hemoprotein models. Values for ΔE_1 and ΔE_2 were computed using TD-DFT, and Löewdin spin populations were determined from single-point calculations.

protein	Axial Heme Ligands	ΔE_1 (cm^{-1})	ΔE_2 (cm^{-1})	Spin Population			
				Fe $3d_{yz}$	Fe $3d_{xz}$	Fe $3d_{xy}$	S $3p$
cytochrome b_5	His/His	3785	4118	0.4666	0.445	0.0316	N/A
cytochrome *c*	His/Met	2072	2571	0.6091	0.2863	0.04704	--
hCBS	His/ $Cys(S^-)$	5865	6379	0.7787	0.01453	0.03369	0.1565

hemoprotein models. Consistent with greater energy separation between Fe 3d_{xz}- and 3d_{yz}-based MOs, the energy of the $(d_{xy})^2(d_{xz})^2(d_{yz})^1 \rightarrow (d_{xy})^2(d_{xz})^1(d_{yz})^2$ transition (ΔE_1) increased dramatically for the His/Cys(S^-) model (ΔE_1 = 5865 cm^{-1}) compared to His/His (ΔE_1 = 3785 cm^{-1}) and His/Met (ΔE_1 = 2072 cm^{-1}) models (Table 5.4). The energy of the $(d_{xy})^2(d_{xz})^2(d_{yz})^1 \rightarrow (d_{xy})^1(d_{xz})^2(d_{yz})^2$ transition (ΔE_2) also increased for the heme-thiolate model, consistent with a stronger overall axial ligand field.

An important connection exists between the magnitude of the *g*-shifts observed in EPR spectra of low-spin, Fe(III) hemoproteins and ground state-excited state energy differences. In EPR spectroscopy, the *g*-shift arises from spin-orbit coupling of low-lying excited states into the paramagnetic ground state. The extent of this spin-orbit coupling cannot be directly calculated using DFT; however, spin-orbit coupling is inversely proportional to the energy difference between the two coupled states. Thus, our computed values, ΔE_1 and ΔE_2, serve as an indirect measure of the extent of spin-orbit coupling observed in low-spin, Fe(III) hemoprotein models. From TD-DFT, we observe that thiolate ligation *increases* the magnitude of ΔE_1 and ΔE_2 by several thousand wavenumbers and thereby *diminishes* spin-orbit coupling in low-spin, Fe(III) heme centers bound by an axial Cys(S^-) ligand. Given that the magnitude of the *g*-shift is directly proportional to the extent of spin-orbit coupling, our model qualitatively recapitulates the experimental observation that thiolate-ligated hemes exhibit narrow EPR spectra. Thiolate ligation also appears to increase metal-ligand covalency, as indicated by the significant contribution of the sulfur 3*p* orbital to the unpaired spin population in the His/Cys(S^-) model (Table 5.4). Increased metal-ligand covalency also diminishes spin-orbit coupling and likely also contributes to the small *g*-shifts observed in EPR spectra of heme-thiolate proteins.

Using a tetragonally-distorted, strong field d^5 model, we estimated ligand field parameters for a number of hemoproteins with varying axial ligands. This model, introduced by Griffith in the late 1950's and refined through the 1990's, approximates energy differences between three low-lying, metal-based d_{xy}, d_{xz}, and d_{yz} orbitals from experimentally-determined g-values.[33,36,47,63-65] In low-spin, Fe(III) porphyrins, these orbital energies correspond to primarily Fe 3d-based MOs. The rhombic distortion, V/ξ, approximates the energy difference between d_{xz} and d_{yz}, while the axial distortion, Δ/ξ, approximates the energy difference between d_{xy} and the barycenter of the d_{xz} and d_{yz} (Figure 5.2). Table 5.5 summarizes g-value assignments and ligand field parameters for hemoproteins bearing a variety of axial ligands, computed using the program developed by McGarvey.[47] By McGarvey's convention, d_{xz} is stabilized relative to d_{yz} when $V/\xi < 0$, and d_{xy} is stabilized relative to d_{xz}/d_{yz} when $\Delta/\xi < 0$.

Ligand field parameters estimated from experimental g-values support the computational prediction that thiolate ligation increases splitting between the paramagnetic ground state and low-lying excited states in low-spin, Fe(III) porphyrins. Magnitudes for rhombic (V) and axial (Δ) splittings increased significantly for hemoproteins bearing an axial Cys(S^-) ligand ($|V'|$ = 1987-2751 cm^{-1} and $|\Delta'|$ = 1960-2765 cm^{-1}) compared to those without a thiolate ligand ($|V|$ = 685-750 cm^{-1} and $|\Delta|$ = 1477-1559 cm^{-1}). Cys(S^-) is a charged, strong field ligand that destabilizes d_{xz} and d_{yz} relative to d_{xy}, likely accounting for the increased axial splitting observed in thiolate-ligated porphyrins. Increased rhombic splittings in thiolate-ligated porphyrins were predicted by DFT: Fe-S p_π-d_π bonding destabilizes the Fe 3d_{yz}-based MO relative to the Fe 3d_{xz}-based MO. Thiolate-dependent increases in $|V'|$ and $|\Delta'|$ are consistent with TD-DFT computations, which predict an increase in ground/excited state energy differences (ΔE_1 and ΔE_2) upon thiolate coordination. This

Table 5.5 Summary of ligand field parameters computed from experimental *g*-values for hemoproteins with varying axial ligands.

Protein	**Axial Heme Ligands**	g_x	g_y	g_z	V/ξ [a]	Δ/ξ [b]	k [c]	V/Δ [d]
cytochrome b_5	His/His	1.40	-2.21	-3.03	-1.63	-3.21	1.01	0.51
cytochrome *c*	His/Met	1.35	-2.14	-3.08	-1.49	-3.39	0.99	0.44
hCBS	Cys(S^-)/His	-2.48	-2.26	1.88	-4.32	-5.03	1.05	0.86
*Bx*RcoM-1	Cys(S^-)/His	-2.49	-2.31	1.87	-4.39	-4.26	1.10	1.03
*Rr*CooA	Cys(S^-)/Pro	-2.46	-2.25	1.89	-4.51	-5.24	1.05	0.86
cyt $P450_{cam}$	Cys(S^-)/H_2O	-2.41	-2.26	1.93	-5.98	-6.01	1.22	0.99
cyt $P450_{cam}$ + imidazole (Im)	Cys(S^-)/Im	-2.56	-2.27	1.87	-4.06	-5.46	1.15	0.74

a. V/ξ = rhombic splitting, V'/ξ for thiolate-ligated species, $\xi = 460\ cm^{-1}$
b. Δ/ξ = axial splitting, Δ'/ξ for thiolate-ligated species
c. k = orbital reduction factor
d. V/Δ = rhombicity, V'/Δ' for thiolate-ligated species

qualitative agreement strengthens our connection between experimentally-observed *g*-shifts and computationally-predicted ground/excited state energy differences.

The *g*-shifts observed in EPR spectra of low-spin, Fe(III) porphyrins, as well as the resulting ligand field parameters, are sensitive to the identity of the second axial ligand *trans* to thiolate. Specifically, weaker axial ligands give rise to larger values for $|V'|$ and $|\Delta'|$ (e.g. $|V'| = 2751$ cm^{-1} and $|\Delta'| = 2765$ cm^{-1} for cyt P450$_{cam}$ with water *trans* to thiolate), while stronger axial ligands give rise to smaller values for $|V'|$ and $|\Delta'|$ (e.g. $|V'| = 1687$ cm^{-1} and $|\Delta'| = 2314$ cm^{-1} for hCBS with His-based imidazole ring *trans* to thiolate). This trend can be explained in terms of a *trans* influence: a weaker *trans* ligand allows for a stronger Fe-S bond and greater thiolate character, giving rise to larger values for $|V'|$ and $|\Delta'|$, as well as smaller *g*-shifts in the rhombic EPR signal. Interestingly, even heme-thiolate proteins bearing identical second axial ligands exhibit different ligand field parameters. For example, hCBS, *Px*RcoM-1, and cyt P450$_{cam}$ + imidzaole all exhibit axial Cys(S$^-$)/imidazole ligation; however, rhombic and axial splittings differ between these proteins. It is likely that protein-dependent variations to the second coordination sphere (including changes to the thiolate H-bonding environment) give rise to these variations in porphyrin electronic structure.

Protonation of the coordinating Cys(S$^-$) in a low-spin, Fe(III) heme reduces the energy differences between the three low-lying Kramer's doublets and should give rise to a rhombic EPR signal with larger *g*-shifts. We modeled this case by protonating the previously-optimized His/Cys(S$^-$) model and re-optimizing the geometry using DFT. We then compared ground state bonding and electronic transitions between the His/Cys(S$^-$) and His/Cys(SH) models in a manner analogous to that described above for other hemoprotein models. Protonation of the coordinating S atom weakens the donor strength of the coordinating cysteine in two ways. First, protonation

neutralizes the negative charge and eliminates a strong Coulombic attraction between the metal and ligand. Second, as was the case for the His/Met model, there is no Fe-S π-bonding interaction in the His/Cys(SH) model, and the near-degeneracy of the Fe $3d_{xz}$- and $3d_{yz}$-based MOs is restored upon protonation (Figure 5.3). Consequently, values for ΔE_1 and ΔE_2 are reduced by ~2,000 cm^{-1}. Smaller values for ΔE_1 and ΔE_2 reflect greater spin-orbit coupling, and thiolate protonation should give rise to larger magnitude *g*-shifts. H-bonding represents an intermediate case between a fully deprotonated thiolate and a protonated thiol. Therefore, we hypothesize that the breadth of the rhombic, low-spin EPR signal observed for a six-coordinate, Fe(III) heme-thiolate protein should reflect the strength of the thiolate H-bonding network.

5.3.2 Design and Synthesis of Intramolecularly Hydrogen-Bonded Aryl Thiolate Ligands

To test our hypothesis that EPR spectroscopy can be used to probe changes in the H-bonding environment in heme-thiolate proteins, we designed a series of aryl thiolate ligands in which the strength of an intramolecular hydrogen bond (H-bond) to the Fe-coordinating thiolate may be systematically tuned. Our ligand design takes inspiration from the substituted 2-acylaminobenzenethiolate ligands reported by Ueyama *et al.*,[66] whose syntheses we have been unable to successfully reproduce. The newly-designed ligand series is based on a 2-mercaptobenzamide core, in which an intramolecular N-H···S H-bond exists between the amide and coordinating aryl thiolate S atom. This intramolecular H-bond results in the formation of a six-membered ring, an improvement over the original ligand design in which a five-membered ring was formed. Expansion of the ring by one atom allows the D-H···A angle to further approach 180°, the ideal angle for this H-bonding interaction. To systematically tune the strength of the intramolecular H-bond, we varied the electronics of the amide by changing the identity of an *N*-phenyl *para* substituent. We chose this *para* substitution strategy for three reasons. First, the

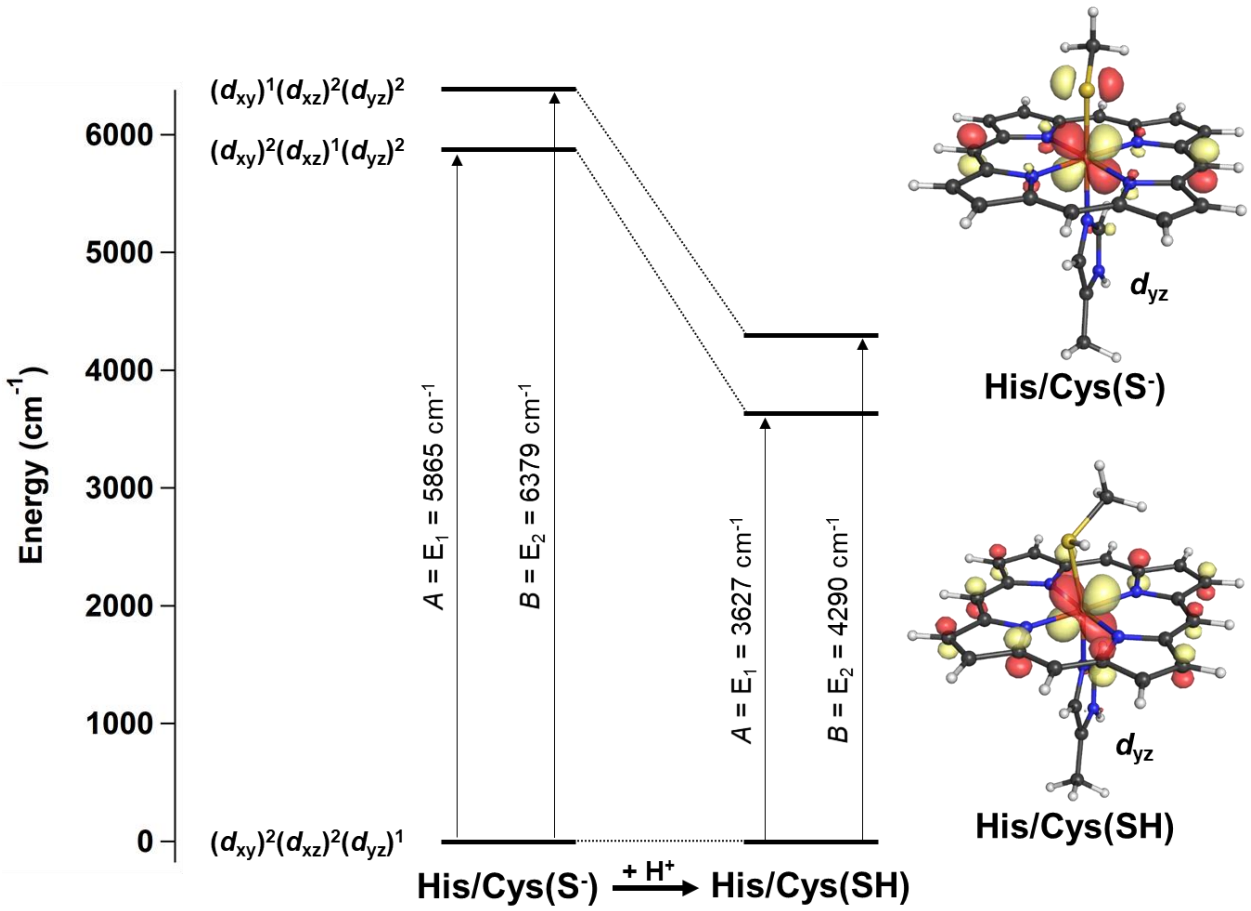
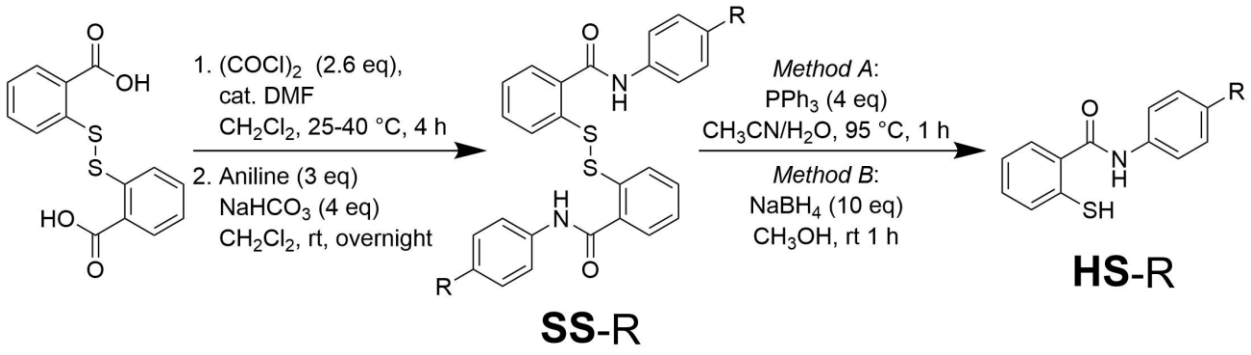


Figure 5.3 Effects of thiolate protonation on porphyrin electronic structure. *Left*: Comparison of energy levels between the ground state and two lowest-lying excited states in low-spin, Fe(III), d^5 porphyrin models, computed using TD-DFT. *Right*: Geometry-optimized structures of His/Cys(S^-) and His/Cys(SH) models with an overlay of the Fe $3d_{yz}$-based, spin-down MO. Orbital plots were generated with an isosurface value of 0.05, and structures were visualized using Pymol (v1.3).

effects of *para* substituent electron donating/withdrawing properties on the pK_a of a heteroatom proton are well-characterized and concisely encompassed in the Hammett parameter, σ.[67] We anticipate *para* substituents with a more positive σ value will lower the pK_a of the amide proton and result in a stronger H-bond and *vice versa*. Second, by making substitutions at the amide phenyl ring, we avoid substantially influencing the thiolate donor strength directly, and instead isolate the effects of changes in H-bond strength. Finally, synthesis of these compounds relies on a facile two-step syntheses from an acyl chloride and commercially-available, *para*-substituted aniline derivatives.[68,69]

We successfully synthesized five 2-mercapto-*N*-phenylbenzamide derivatives (**HS**-R, R = CH_3, H, Cl, CF_3, NO_2), which serve as intramolecularly H-bonded thiolate ligand precursors, using a two-step synthetic method (Scheme 5.1). We reacted 2,2'-disulfanediyldibenzoyl chloride with five different commercially-available, *para*-substituted anilines under basic conditions to generate 2,2'-disulfanediylbis(*N*-phenylbenzamide) derivatives with varying *N*-phenyl *para* substitution (**SS**-R, R = CH_3, H, Cl, CF_3, NO_2, Figures 5.S5-5.S14). Strategically, we chose to carry out the amide formation reaction using the these disulfides in order to eliminate side-reactions between the thiol moiety and acyl chloride intermediate. Reduction of the 2,2'-disulfanediylbis(*N*-phenylbenzamide) disulfide bond using either triphenylphosphine or sodium borohydride, followed by protonation of the charged thiolate, resulted in formation of our desired thiol ligand precursors (Figures 5.S15-5.S25). In ^{1}H NMR spectra of the thiols, we observed a direct correlation between the chemical shift observed for the amide proton and electron withdrawing strength of the *N*-phenyl *para* substituent, as represented by the two extremes: $\delta_{N\text{-}H}$ (**HS-**CH_3) = 7.65 ppm and $\delta_{N\text{-}H}$ (**HS-**NO_2) = 8.18 ppm. No such correlation was observed for protons in the aryl thiol ring. Taken together, these observations serve as preliminary evidence of a *localized*,



Scheme 5.1 Synthesis of 2-mercapto-N-phenylbenzamides derivatives from 2,2'-dithiodibenzoylchloride where R = CH_3, H, Cl, CF_3, and NO_2. Disulfides are abbreviated as '**SS**-R' and aryl thiols are abbreviated as '**HS**-R'.

systematic change in the electronic properties of the amide proton. The nature of this systematic change in electronics is enumerated below.

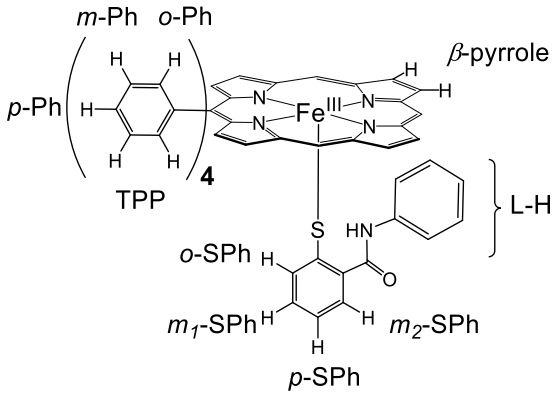
3.3 Synthesis of Five-Coordinate, Fe(III) Thiolate-Ligated Porphyrins

We prepared a series of five complexes of the form [Fe(TPP)(**S**-R)] using the intramolecularly H-bonded thiol ligand precursors, **HS**-R (R = CH_3, H, Cl, CF_3, NO_2). We assembled each complex using the procedure recently reported by Hunt and Lehnert for the synthesis of 5-c, Fe(III) tetraphenylporphyrinate complexes bearing an axial aryl thiolate ligand.[10] This method involves mixing [Fe(TPP)(OCH_3)] with a slight excess of thiol ligand precursor in dry CH_2Cl_2. The strongly basic methoxide ligand deprotonates the aryl thiol, resulting in coordination of the aryl thiolate and formation of one equivalent of methanol. The reaction progresses quickly (less than one minute) and can be monitored by a color change from dark green to dark red/brown. Characteristic features of 5-c aryl thiolate complexes, observed in electronic absorption spectra, include a split Soret peak with a shoulder at 380 nm and maximum at 407-408 nm, in addition to other peaks at 515, 576-578, and 717-722 nm (Figures 5.S27, S30, S33, S36, and S40).

Each recrystallized [Fe(TPP)(**S**-R)] compound was characterized using EPR and ^{1}H NMR spectroscopies and exhibited distinct features indicative of a high-spin, Fe(III) porphyrin complex. Effective *g*-values were observed at 7.1, 4.8, and 1.9 in the EPR spectrum of each compound, suggesting the presence of a high-spin (S = 5/2), rhombic paramagnetic species (Figures 5.S28, S31, S34, S37, and S41). An additional sharp signal around g = 6.0 was observed in all EPR spectra, derived from a very small amount of [Fe(TPP)Cl] impurity (Figure 5.S26). Assignment of ^{1}H NMR peaks is summarized in Table 5.6 and Scheme 5.2. Peaks for β-pyrrole protons were observed between 68.4 and 72.5 ppm, a range that is characteristic of paramagnetic shifts due to the presence of a high-spin, Fe(III) porphyrin species (Figures 5.S29, S32, S35, S39, and S42).

Table 5.6 Comparison of the NMR chemical shifts (ppm) of the 5-c aryl thiolate porphyrin complexes, [Fe(TPP)(**S**-R)], where R refers to the identity of the *N*-phenyl *para* substituent, recorded in CD_2Cl_2 at room temperature (between 20 and 22 °C).

R =	**β-pyrrole**	***m*-Ph, TPP**	***o*-Ph, TPP**	***p*-Ph, TPP**	***m*-Ph, SPh's**	***o*-Ph, SPh**	***p*-Ph, SPh**	***o*/*m*-Ph L-H**	***p*-Ph L-H**
CH_3	72.5	12.8	9.6	6.9	64.3, 56.6	-93.0	-85.6	-110.7, 8.4, 4.8	4.1
H	71.6	12.8	9.8	6.9	63.7, 56.4	-92.8	-85.3	-111.3, 8.3, 4.9	6.5
Cl	71.5	12.8	9.6	6.8	62.5, 55.6	-91.9	-84.5	-111.6, 8.3, 4.8	--
CF_3	68.4	12.8	9.9	6.9	61.2, 54.8	-90.4	-83.2	-110.5, 8.0, 5.1	--
NO_2	70.5	12.9	10.5	6.7	59.9, 54.3	-90.2	-82.7	-112.4, 8.2, 4.9	--



Scheme 5.2 Labeling scheme for the observed signals in the ^{1}H-NMR spectra reported in Table 5.6.

The crystal structure of [Fe(TPP)(**S**-NO_2)] revealed an intramolecular H-bond between the ligand amide and Fe-coordinated thiolate (Figure 5.4). The complex exhibits an Fe-S bond distance of 2.308 Å and an Fe-S-C bond angle of 102.8°, consistent with coordination of an aryl thiolate ligand. The average Fe-N bond distance (2.064 Å) and Fe atom displacement from the mean plane of the coordinating N atoms (0.418 Å) are comparable to those reported for [Fe(TPP)(SPh)] by Byrn and Strouse (Fe-N_{av} = 2.063 Å, Fe atom displacement = 0.470 Å).[70] The observation of a smaller Fe atom displacement in [Fe(TPP)(**S**-NO_2)] compared to the benzenethiolate complex suggests a somewhat diminished donor strength in the 2-mercapto-N-phenylbenzamide-derived ligands. The conformation of the amide relative to the coordinating S atom is consistent with the presence of an intramolecular interaction in which the amide acts as an H-bond donor to the coordinating S atom. This observation is supported by the fact that the amide H atom was identified in the Fourier difference map when refining the structure, suggesting that the atom was locked in a rigid position imposed by the H-bond. A donor-acceptor distance of 3.009 Å and a D-H···A angle of 149°, indicative of a H-bond of moderate strength. That this intramolecular H-bond contributes to the attenuation of the thiolate ligand strength is supported by spectroscopic and computational results detailed below.

5.3.4 Assembly and EPR Spectroscopic Characterization of Six-Coordinate, Fe(III) Thiolate-Ligated Porphyrins

By mixing each of the five-coordinate [Fe(TPP)(**S**-R)] complexes with a neutral donor ligand, we successfully assembled six-coordinate model complexes that exhibit EPR spectra characteristic of low-spin, Fe(III) thiolate-ligated porphyrins. We utilized two biologically-relevant neutral donor ligands to mimic heme-thiolate protein environments: 1-methylimidazole (1-MeIm), which gives rise to axial ligation similar to the Cys/His environment observed in Fe(III) hCBS and RcoM,

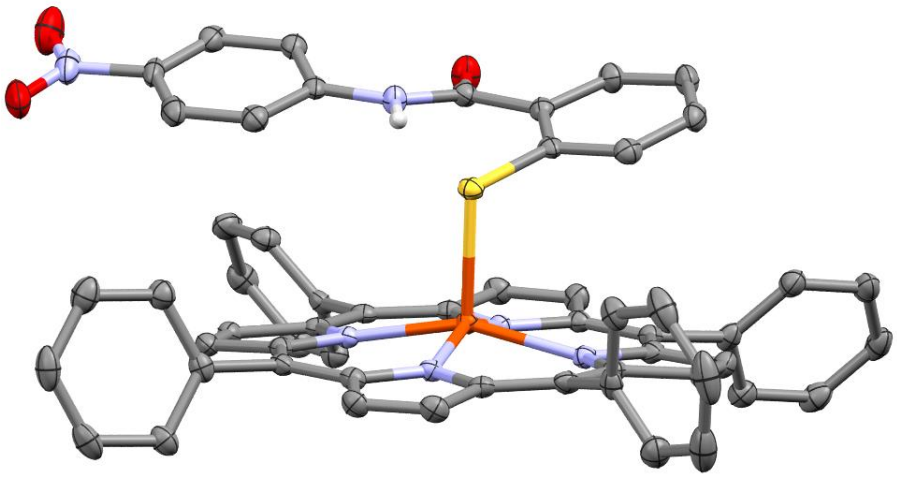


Figure 5.4 Molecular structure of [Fe(TPP)(**S**-NO_2)] shown with 50% probability ellipsoids. The amide H atom is displayed, and all other H atoms are omitted for clarity.

and water, which gives rise to axial ligation similar to the Cys/H_2O environment observed in Fe(III), substrate-free cytochromes P450. Upon addition of 1-$MeIm_{(l)}$ to a solution of each five-coordinate complex (in 50:50 CH_2Cl_2:toluene), we observe an EPR signal characteristic of a low-spin, Fe(III) thiolate-ligated porphyrin complex (g_z = 2.479-2.491, g_y = 2.280-2.283, and g_x = 1.868-1.877), which we attribute to the six-coordinate [Fe(TPP)(**S**-R)(1-MeIm)] complexes (Figure 5.S43). We also observe a very minor high-spin (S = 5/2) signal (less than 1% total signal intensity) with $g_\perp$ = 5.9 and $g_\parallel$ = 2.0 for each complex, which we attribute to a small [Fe(TPP)Cl] contaminant (Figure 5.S43). We were unable to observe formation of [Fe(TPP)(**S**-R)(OH_2)] complexes when adding exogenous water to a solution of five-coordinate complex due to the immiscibility between water and the 50:50 CH_2Cl_2:toluene solvent. To overcome this limitation, we prepared EPR samples using solvent that was not rigorously dried by distillation over CaH_2. Potentiometric titration revealed that this "wet" 50:50 CH_2Cl_2:toluene mixture contained a concentration of water greater than 2 mM. Upon dissolution of each five-coordinate porphyrin model complex to a concentration of ~500 μM in wet solvent, we observe an EPR signal characteristic of a low-spin, Fe(III) thiolate-ligated porphyrin complex which is distinct from that observed in the presence of 1-MeIm. This new signal, which we attribute to a six-coordinate [Fe(TPP)(**S**-R)(OH_2)] complex (g_z = 2.416-2.429, g_y = 2.267-2.272, and g_x = 1.912-1.917), is narrower than the signal observed for the 1-MeIm-bound complexes (Figure 5.S43). The minor [Fe(TPP)Cl] contaminant is also observed for the water-bound complexes. A representative overlay of low-spin, Fe(III) EPR spectra for both six-coordinate [Fe(TPP)(**S**-CH_3)(L')] complexes is displayed in Figure 5.5.

A positive correlation exists between EPR signal breadth and intramolecular H-bond strength for six-coordinate [Fe(TPP)(**S**-R)(L')] complexes. Figure 5.6 compares the low-spin,

Fe(III) EPR signals observed for six-coordinate complexes with either 1-MeIm or water bound as the sixth axial ligand *trans* to the aryl thiolate. A clear trend is apparent in both sets of complexes: the breadth of the low-spin EPR signal increases as the electron withdrawing character of the amide *N*-phenyl *para* substituent increases. As described above, the electron donating/withdrawing character of this substituent tunes the acidity of the amide proton and thereby modulates the strength of the N-H···S hydrogen bond *without* significantly altering the electronics of the thiophenolate ring directly. We therefore conclude that the observed trend is indicative of a direct correlation between intramolecular N-H···S hydrogen bond strength and low-spin, Fe(III) EPR signal breadth (Table 5.7). The connection of these trend to the porphyrin electronic structure are discussed in detail below.

3.5 Computationally-Observed Changes in Thiolate H-Bonding

We generated geometry-optimized models of six-coordinate [Fe(TPP)(**S**-R)(L')] (L' = 1-MeIm, H_2O) complexes using DFT calculations (BP86, TZVP). To generate initial coordinates for each model, we used the crystal structure of five-coordinate [Fe(TPP)(**S**-NO_2)]. We modified these atomic coordinates to reflect the appropriate *para* substitution and added a second axial ligand, either 1-MeIm or H_2O. Additionally, we employed porphinate (P^{2-}) as a simplified porphyrin structure by replacing each of the TPP^{2-} *meso* phenyl rings with a hydrogen atom. Frequency calculations ensured that each geometry-optimized model represents an energy minimum, as reflected by the lack of any computed negative vibrational frequencies. Representative geometry-optimized structures for both [Fe(P)(**S**-NO_2)(L')] models are displayed in Figure 5.7. Energy-minimized structures exhibited geometric parameters consistent with those observed experimentally.

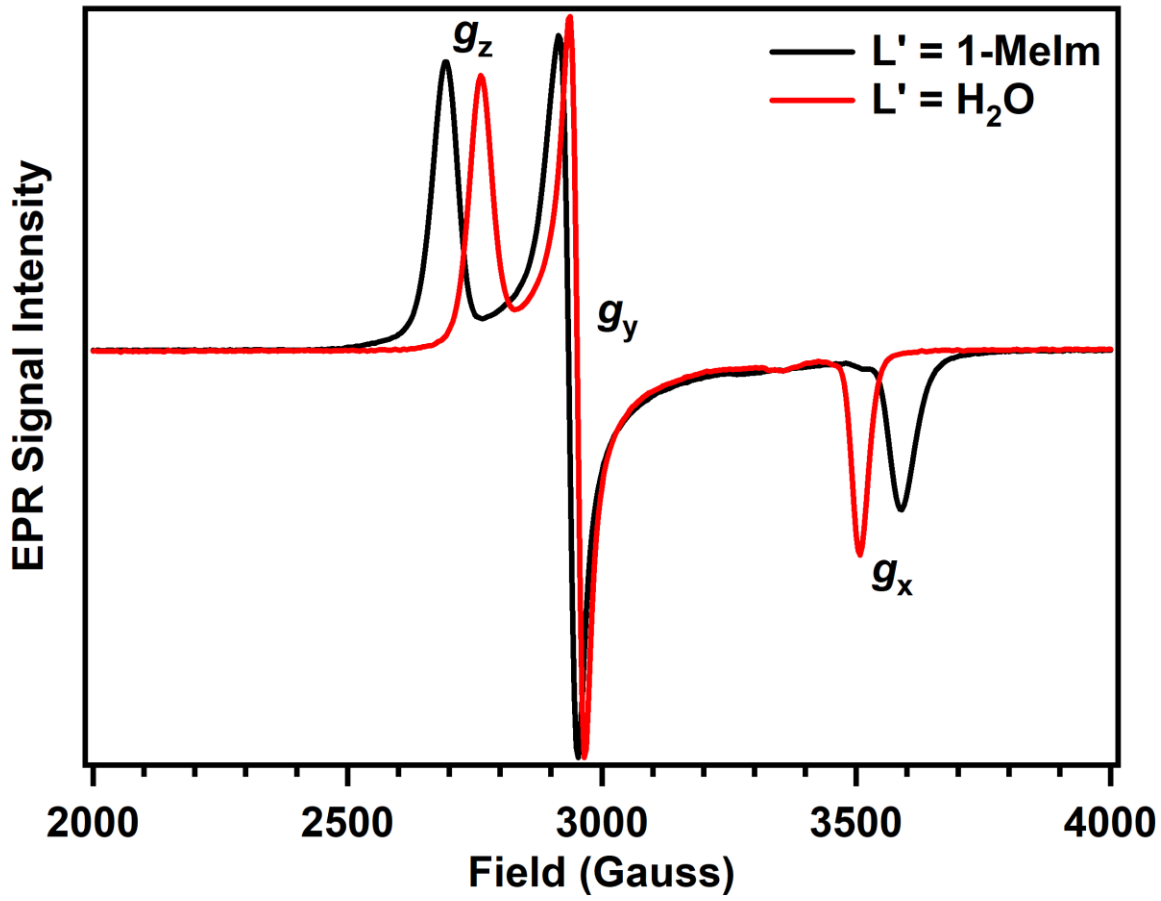


Figure 5.5 Overlay of low-spin, Fe(III) EPR signals observed for six-coordinate [Fe(TPP)(**S**-CH_3)(L')] complexes where L' = 1-MeIm (black) or H_2O (red). Samples were dissolved in 50:50 CH_2Cl_2:toluene, and spectra were recorded at 10 K. Signal intensities are normalized to the most intense feature at g_{mid}.

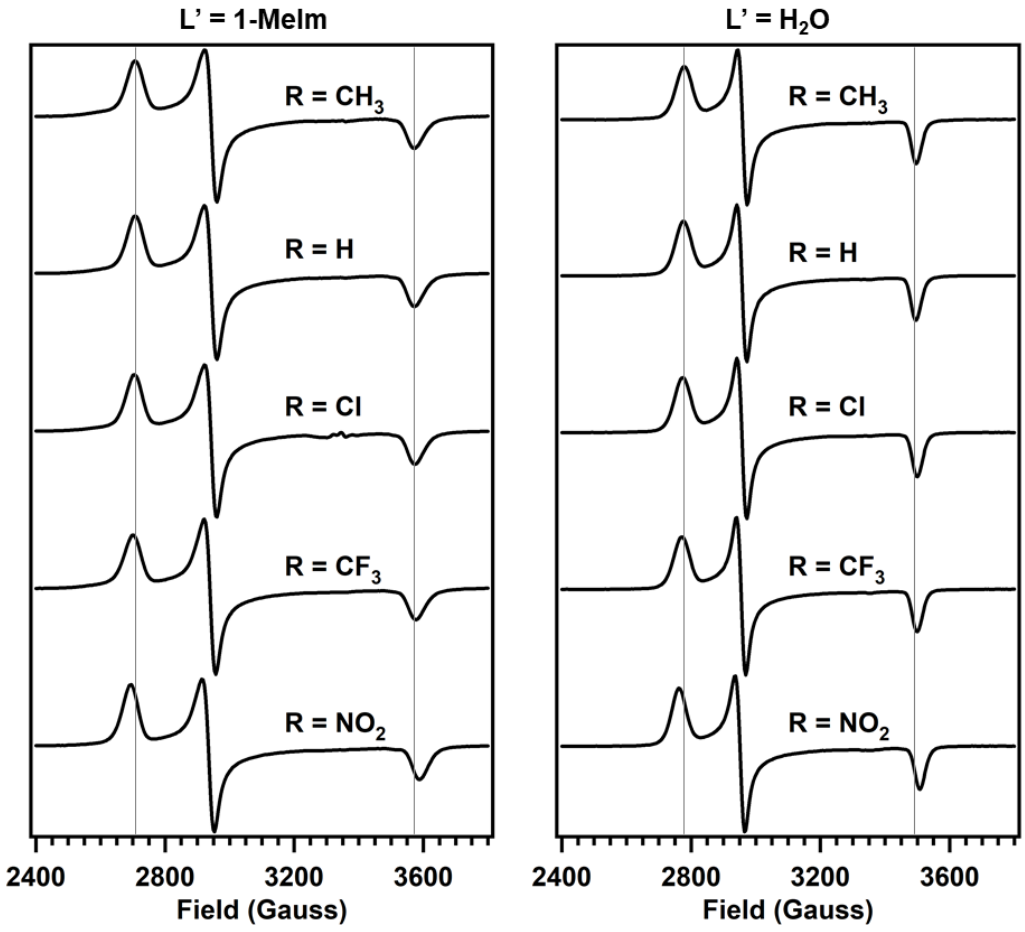


Figure 5.6 Comparison of rhombic, low-spin spectra observed for six-coordinate [Fe(TPP)(**S**-R)(L')] complexes where L' = 1-MeIm (left) or H_2O (right). Samples were prepared in 50:50 CH_2Cl_2:toluene, and spectra were recorded at 10 K. Signal intensities are normalized to the most intense feature at g_y. The vertical lines, centered on g_{max} and g_{mid} for [Fe(TPP)(**S**-CH_3)(L')], help depict how the low-spin signals broaden as the electron-withdrawing character of the amide *para* substituent (R) increases.

Table 5.7 Comparison of EPR parameters for [Fe(TPP)(**S**-R)(L')] complexes. Note: g_z, g_y, and g_x are given the historical assignments of g_{max}, g_{mid}, and g_{min} in this table to facilitate comparison of *g*-values to those of heme-thiolate proteins reported in the literature.

R =	**L' = 1-MeIm**			**L' = H_2O**		
	g_z	g_y	g_x	g_z	g_y	g_x
CH_3	2.479	2.280	1.877	2.416	2.267	1.917
H	2.478	2.279	1.877	2.417	2.268	1.919
Cl	2.481	2.280	1.875	2.419	2.268	1.916
CF_3	2.484	2.281	1.874	2.421	2.270	1.916
NO_2	2.491	2.283	1.868	2.429	2.272	1.912

Computed geometric parameters and vibrational frequencies for [Fe(P)(**S**-R)(L')] models support our hypothesis that the electron donating or withdrawing character of each *N*-phenyl *para* substituent directly correlates with the strength of the intramolecular N-H···S hydrogen bond. Increasing the electron withdrawing strength of the *para* substituent resulted in a marked decrease in calculated amide N-H vibrational frequency, as well as a slight lengthening of the N-H bond (Table 5.8). These observations suggest that changing the identity of the *para* substituent modulates the amide N-H bond strength, and by extension, the acidity of the amide proton. For example, in [Fe(P)(**S**-R)(1-MeIm)] models, the electron-donating methyl substituent gave rise to a *stronger* N-H bond ($\nu_{N\text{-}H}$ = 3258 cm^{-1}) and a *less acidic* amide proton, as compared to the strongly electron-withdrawing nitro substituent, which gave rise to a significantly *weaker* N-H bond ($\nu_{N\text{-}H}$ = 3176 cm^{-1}) and a *more acidic* amide proton. Amide proton acidity, approximated by N-H bond strength, directly influences the strength of the intramolecular N-H···S hydrogen bond: as the amide N-H bond is weakened, the H-bond donor-acceptor distance decreases, as does the distance between the amide H atom and thiolate S atom (Table 5.8). For example, the [Fe(P)(**S**-NO_2)(1-MeIm)] model, which exhibited the weakest amide N-H bond, possessed the strongest H-bond, represented by the shortest donor-acceptor distance (3.050 Å). The same trends are observed in to a similar degree in [Fe(P)(**S**-R)(H_2O)] models.

Computations predict that strengthening the intramolecular N-H···S hydrogen bond results in a weakening of the Fe-S bond. We observed a small but consistent increase in Fe-S bond distance as the electron withdrawing character of the *para* substituent (and H-bond strength) increased, ranging from 2.234 Å in [Fe(P)(**S**-CH_3)(1-MeIm)] to 2.245 Å in [Fe(P)(**S**-NO_2)(1-MeIm)] (Table 5.8). We observed this trend to a lesser extent, in the water-bound series, where the

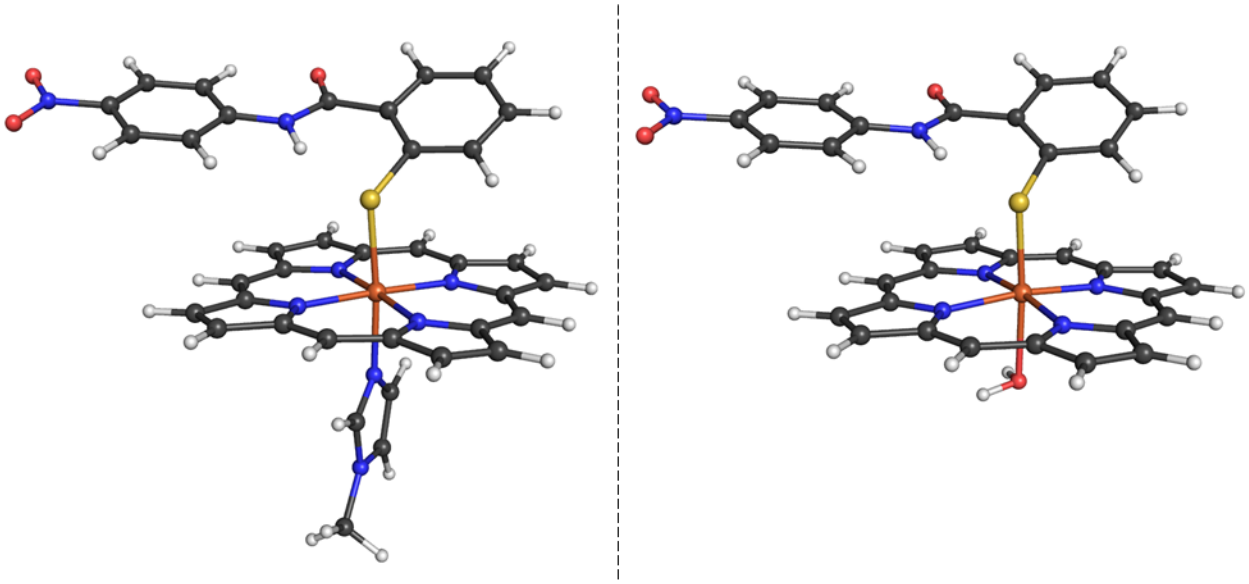


Figure 5.7 DFT-optimized structures of [Fe(P)(**S**-NO_2)(L')] complexes with L' = 1-MeIm (left) and L' = H_2O (right). Structures were visualized using Pymol (v1.3).

Table 5.8 Summary of geometric parameters and amide N-H vibrational frequencies for [Fe(TPP)(**S**-R)(L')] complexes computed using DFT.

R =	L' = 1-MeIm						L' = H_2O					
	ν_{N-H} (cm^{-1})	d_{N-H} (Å)	d_{D-A} (Å)	$d_{S-H(amide)}$ (Å)	d_{Fe-S} (Å)	$d_{Fe-N(ax)}$ (Å)	ν_{N-H} (cm^{-1})	d_{N-H} (Å)	d_{D-A} (Å)	$d_{S-H(amide)}$ (Å)	d_{Fe-S} (Å)	d_{Fe-O} (Å)
CH_3	3258	1.032	3.070	2.162	2.234	2.076	3306	1.029	3.089	2.198	2.192	2.215
H	3252	1.032	3.068	2.159	2.237	2.073	3301	1.030	3.087	2.193	2.194	2.212
Cl	3230	1.034	3.061	2.145	2.239	2.071	3279	1.031	3.079	2.178	2.195	2.211
CF_3	3214	1.034	3.059	2.142	2.243	2.071	3266	1.032	3.075	2.171	2.197	2.207
NO_2	3176	1.036	3.050	2.124	2.245	2.070	3234	1.034	3.067	2.155	2.198	2.207

Fe-S bond lengths ranged from 2.192 Å in [Fe(P)(**S**-CH_3)(H_2O)] to 2.198 Å in [Fe(P)(**S**-NO_2)(H_2O)]. Weakening of the Fe-S bond results in a concomitant strengthening of the bond between iron and the sixth axial ligand: the Fe-N(1-MeIm) bond shortens from 2.076 Å in [Fe(P)(**S**-CH_3)(1-MeIm)] to 2.070 Å in [Fe(P)(**S**-NO2)(1-MeIm)], while the Fe-O bond shortens from 2.215 in [Fe(P)(**S**-CH_3)(H_2O)] to 2.207 Å in [Fe(P)(**S**-NO_2)(H_2O)]. Using DFT, Paulat et al. predicted a similar direct correlation between intramolecular H-bond strength and Fe-S bond distance for a series of porphyrin-thiolate model compounds with zero, one, or two intramolecular H-bonds.[71] The electronic origins of the observed changes in bonding of axial porphyrin ligands will be discussed below.

5.3.6 Insight into H-Bond-Dependent Changes in Model Porphyrin-Thiolate Electronic Structure

As predicted in Section 5.3.1, the magnitude of low-spin, rhombic *g*-shifts increased as intramolecular H-bond strength increased in our 6-c porphyrin-thiolate models. Using the tetragonally-distorted, strong field d^5 model described above, we estimated ligand field parameters for these compounds from experimentally-observed *g*-values. (Table 5.9). The *trans* influence observed in heme-thiolate proteins was also observed in our porphyrin-thiolate models: rhombic and axial splittings were larger for the water-bound complexes than for 1-MeIm-bound complexes. For both sets of [Fe(TPP)(**S**-R)(L')] compounds, we observe a small *decrease* in rhombic and axial splittings as H-bond strength *increases* (for L' = 1-MeIm, $\Delta|V|' = 115$ cm^{-1}, $\Delta|\Delta'| = 82$ cm^{-1}; for L' = H_2O, $\Delta|V|' = 77$ cm^{-1}, $\Delta|\Delta'| = 61$ cm^{-1}). This observation demonstrates that 1) EPR spectroscopy is sensitive to the thiolate H-bonding environment and 2) that the intramolecular H-bond attenuates the thiolate ligand strength, ultimately giving rise to smaller *g*-shifts in the rhombic, low-spin EPR signal.

Computations recapitulate the H-bond-dependent trends observed in EPR spectra of [Fe(TPP)(**S**-R)(H_2O)] complexes. We successfully carried out TD-DFT calculations on the geometry-optimized structure of each [Fe(P)(**S**-R)(H_2O)], although we were unable to attain reasonable TD-DFT results for [Fe(P)(**S**-R)(1-MeIm)] complexes, likely due to the large number of atoms present in this second set of complexes. For [Fe(P)(**S**-R)(H_2O)] complexes, we observe a decrease in ΔE_1, the energy of the $(d_{xy})^2(d_{xz})^2(d_{yz})^1 \rightarrow (d_{xy})^2(d_{xz})^1(d_{yz})^2$ transition and ΔE_2, the energy of the $(d_{xy})^2(d_{xz})^2(d_{yz})^1 \rightarrow (d_{xy})^1(d_{xz})^2(d_{yz})^2$ transition (Table 5.10). Values for both ΔE_1 and ΔE_2 fell between those computed for the His/Cys(S^-) and His/Cys(SH) models, consistent with the fact that H-bonding represents an intermediate protonation state. We observed a decrease in values for ΔE_1 and ΔE_2 of approximately 100 cm^{-1} when comparing values between [Fe(P)(**S**-CH_3)(H_2O)], which contains the weakest H-bond, and [Fe(P)(**S**-NO_2)(H_2O)], which contains the strongest H-bond. These differences in energy are comparable to those computed for rhombic and axial splittings above.

As was the case for hemoprotein models with variable axial ligands, we can rationalize the difference in energy of the low-lying states for our H-bonding models in terms of Fe-S bonding interactions. H-bond donation to the coordinating thiolate (much like protonation of the thiolate) alters Fe-S bonding interactions in two ways. First, H-bonding effectively reduces the negative charge on sulfur and thereby reduces the ligand donor strength. Second, the H-bond diminishes the Fe-S π-bonding interaction, as evidenced by a decrease in sulfur 3*p* character observed in the Fe-based $3d_{yz}$ MOs of H-bonding models (Figure 5.S44). Increasing the strength of the H-bond enhances these two effects and leads to a weakening of the Fe-S bond, as evidenced by the correlation between H-bond strength and Fe-S bond distance observed in computational models of our complexes (Table 5.8). Weakening of the Fe-S bond causes a reduction in the π*-antibonding

Table 5.9 Summary of ligand field parameters computed from experimental *g*-values for [Fe(TPP)(**S**-R)(L')] compounds.

L' =	R =	g_x	g_y	g_z	V'/ξ [a]	Δ'/ξ [b]	k [c]	V'/Δ' [d]
1-MeIm	CH_3	-2.479	-2.280	1.877	-4.380	-4.665	1.068	0.939
	H	-2.478	-2.279	1.877	-4.377	-4.668	1.065	0.938
	Cl	-2.481	-2.280	1.875	-4.339	-4.637	1.063	0.936
	CF_3	-2.484	-2.281	1.874	-4.320	-4.631	1.065	0.933
	NO_2	-2.491	-2.283	1.868	-4.213	-4.535	1.056	0.929
H_2O	CH_3	-2.416	-2.267	1.917	-5.486	-5.354	1.147	1.025
	H	-2.417	-2.268	1.919	-5.564	-5.436	1.165	1.024
	Cl	-2.419	-2.268	1.916	-5.449	-5.340	1.148	1.020
	CF_3	-2.421	-2.270	1.916	-5.459	-5.335	1.155	1.023
	NO_2	-2.429	-2.272	1.912	-5.314	-5.257	1.147	1.011

a. V'/ξ = rhombic splitting, $\xi = 460$ cm^{-1}
b. Δ'/ξ = axial splitting
c. k = orbital reduction factor
d. V'/Δ' = rhombicity

Table 5.10 Electronic excitation energies for [Fe(P)(**S**-R)(H_2O)] computed using TD-DFT.

R =	E_1 (cm^{-1})	E_2 (cm^{-1})
CH_3	5442	5812
H	5428	5791
Cl	5388	5756
CF_3	5374	5724
NO_2	5337	5687

character and subsequent stabilization of the singly-occupied Fe $3d_{yz}$-based MO. Stabilization of d_{yz} brings the lowest-lying Kramer's doublets closer together in energy, giving rise to enhanced spin-orbit coupling and larger magnitude *g*-shifts in the presence of a stronger H-bond.

5.4 Discussion

We have uncovered the electronic origins of the uniquely narrow rhombic EPR signal observed in low-spin, Fe(III) heme-thiolate proteins. Our DFT computations revealed a π-bonding interaction between Fe and S that is unique to porphyrins bearing an axial thiolate ligand. This p_π-d_π interaction destabilizes the singly-occupied, Fe $3d_{yz}$-based MO and gives rise to larger energy differences between the $(d_{xy})^2(d_{xz})^2(d_{yz})^1$ ground state and the $(d_{xy})^2(d_{xz})^1(d_{yz})^2$ and $(d_{xy})^1(d_{xz})^2(d_{yz})^2$ excited states. These large ground/excited state energy differences, predicted using TD-DFT, reflect diminished spin-orbit coupling between the ground state Kramer's doublet and two low-lying excited states. This diminished spin-orbit coupling ultimately gives rise to the reduced magnitude of *g*-shifts observed for heme-thiolate proteins. Ligand field parameters for low-spin, Fe(III) porphyrins, estimated from experimentally-observed *g*-values, corroborate our computational results: we observed significantly larger rhombic and axial splittings in hemoproteins bearing an axial thiolate ligand compared to hemoproteins with no axial thiolate.

Through synthesis and characterization of porphyrin-thiolate model compounds bearing a tunable, intramolecular H-bond, we have established a direct correlation between the magnitude of *g*-shifts observed in the rhombic EPR signal of low-spin, Fe(III) heme-thiolates and thiolate H-bond strength. To model H-bonding in heme-thiolates, we designed a thiolate ligand with an intramolecular N-H···S H-bond between an amide and aryl thiolate S atom, and we systematically tuned the strength of this H-bond by varying the electronics of the amide. Spectroscopic

investigation of low-spin, Fe(III) porphyrin complexes bearing our model thiolates revealed a direct correlation between H-bond strength and *g*-shift in the rhombic EPR signal.

We employed DFT computations to better understand this correlation in terms of heme-thiolate electronic structure. From DFT, we observe that H-bonding attenuates the Fe-S π-bonding interaction and thereby stabilizes the Fe $3d_{yz}$-based MO. An important consequence of this stabilization is to decrease ground/excited state energy differences, and TD-DFT predicts a *systematic* decrease in these energy differences as H-bond strength increases in [Fe(P)(**S**-R)(H_2O)] compounds. These computational observations suggest that H-bonding gives rise to greater spin-orbit coupling through reduction in ground/excited state energy differences, as well as through decreased Fe-S covalency. Spin-orbit coupling directly contributes to the magnitude of the EPR *g*-shift, and thus an *increase* in heme-thiolate H-bond strength leads to an *increase* in the magnitude of *g*-shifts.

Given the well-defined relationship between heme-thiolate electronic structure and rhombic EPR signal, it is possible to compare thiolate donor strengths amongst heme-thiolate proteins through analysis of low-spin, Fe(III) EPR spectra. As an example, we compare the EPR signals (and resulting ligand field parameters) observed in imidazole-bound Fe(III) Cyt $P450_{cam}$, a type-1 heme-thiolate protein, with that of *Px*RcoM-1, a type-2 heme-thiolate protein (Table 5.5). These hemoproteins bear axial ligands that are nearly identical; however, rhombic splitting is slightly larger in *Px*RcoM-1 ($|V'| = 2019$ cm^{-1}) than in Cyt $P450_{cam}$ ($|V'| = 1868$ cm^{-1}), while axial splitting is significantly larger in Cyt $P450_{cam}$ ($|\Delta'| = 2512$ cm^{-1}) than in *Px*RcoM-1 ($|\Delta'| = 1960$ cm^{-1}). Slightly larger rhombic splitting in *Px*RcoM-1 suggests that the heme in this protein possesses slightly greater Fe-S p_π-d_π bonding character than in Cyt $P450_{cam}$. In contrast, the significantly *smaller* axial splitting in *Px*RcoM-1 suggests that the thiolate in this protein is a

weaker donor than the thiolate in Cyt $P450_{cam}$. These differences in heme-thiolate electronic structure are consistent with each protein's function. Cyt $P450_{cam}$ is a monooxygenase enzyme that requires a strong thiolate donor ligand to facilitate O-O bond cleavage and increase the basicity of the protonated ferryl species in compound II.[2,3,10,19-23] In *Px*RcoM-1, the thiolate ligand is replaced by a protein-derived methionine ligand upon heme reduction,[24,25] which may explain why the heme-coordinating thiolate is a weaker donor than that in Cyt $P450_{cam}$.

It is likely that second coordination sphere interactions, such as H-bonding, impart differential thiolate donor strengths in heme-thiolate proteins. The crystal structure of Cyt $P450_{cam}$ reveals a well-defined thiolate H-bonding network that includes three backbone amide H-bond donors.[15] Long donor-acceptor distances (3.08-3.56 Å) and non-linear D-H-A angles (88-126°) demonstrate that H-bonding to thiolate is relatively weak in this protein. Our EPR analysis, which points to a relatively strong thiolate donor in Cyt $P450_{cam}$, is consistent with a weak thiolate H-bonding network. While no structural data exist for *Px*RcoM-1, it is likely that the second coordination sphere differs significantly from that of Cyt $P450_{cam}$. Analysis of the *Px*RcoM-1 rhombic EPR signal suggests that the overall thiolate donor strength is weaker than that of Cyt $P450_{cam}$. We speculate that a strong H-bond may exist between the coordinating thiolate and a protein-derived H-bond donor in *Px*RcoM-1. This strong H-bond would facilitate redox-mediated ligand switching by weakening the Fe-S bond.

The intramolecularly H-bonded porphyrin complexes characterized in this study offer a new way to model the thiolate H-bonding network and its role in tuning the reactivity of heme-thiolate proteins. Hunt and Lehnert recently shed light on the electronic nature of the "thiolate push" effect, which gives rise to O-O bond cleavage in Cyt P450s, in terms of an admixture of Fe-S σ-bonding and Fe-O-O σ*-antibonding in a doubly-protonated model of the Compound 0 ferric

peroxo species.[10] This study demonstrated that modulating the thiolate donor strength influences this critical bonding interaction, and the authors hypothesize that H-bonding provides a means to attenuate thiolate donor strength. Our current study lends support to this hypothesis. Through computational and spectroscopic analysis of a model system in which we systematically tune the strength of a heme-thiolate H-bonding interaction, we have identified a clear connection between H-bonding and thiolate ligand strength. Based on changes in the magnitude of *g*-shifts in low-spin, rhombic EPR spectra, we demonstrate that increasing the strength of H-bonding to thiolate leads to a decrease in thiolate donor strength. The functional implications of these changes in H-bonding in our model system have yet to be explored.

In conclusion, we have developed a synthetic model of H-bonding in heme-thiolate proteins in which the strength of an intramolecular H-bond can be finely tuned. We utilized this model system to demonstrate that H-bond donation to thiolate influences the Fe-S bonding interaction in heme-thiolates using a combination of computational methods and EPR spectroscopy. We defined a direct correlation between *g*-shifts in low-spin, rhombic EPR spectra of Fe(III) heme-thiolates and the strength of H-bond donation to thiolate. Ultimately, we hope to use this relationship to better understand how H-bonding controls function in the diverse family of proteins bearing an axial thiolate ligand.

5.5 References

1. Smith, A. T.; Pazicni, S.; Marvin, K. A.; Stevens, D. J.; Paulsen, K. M.; Burstyn, J. N. Functional Divergence of Heme-Thiolate Proteins: A Classification Based on Spectroscopic Attributes. *Chem. Rev.* **2015**, *115*, 2532.

2. Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. Heme-containing oxygenases. *Chem. Rev.* **1996**, *96*, 2841.

3. Yosca, T. H.; Rittle, J.; Krest, C. M.; Onderko, E. L.; Silakov, A.; Calixto, J. C.; Behan, R. K.; Green, M. T. Iron(IV)hydroxide pK(a) and the Role of Thiolate Ligation in C-H Bond Activation by Cytochrome P450. *Science* **2013**, *342*, 825.

4. Yoshioka, S.; Tosha, T.; Takahashi, S.; Ishimori, K.; Hori, H.; Morishima, I. Roles of the proximal hydrogen bonding network in cytochrome P450cam-catalyzed oxygenation. *J. Am. Chem. Soc.* **2002**, *124*, 14571.

5. Galinato, M. G. I.; Spolitak, T.; Ballou, D. P.; Lehnert, N. Elucidating the role of the proximal cysteine hydrogen-bonding network in ferric cytochrome P450cam and corresponding mutants using magnetic circular dichroism spectroscopy. *Biochemistry* **2011**, *50*, 1053.

6. Lehnert, N. Elucidating second coordination sphere effects in heme proteins using low-temperature magnetic circular dichroism spectroscopy. *J. Inorg. Biochem.* **2012**, *110*, 83.

7. Mak, P. J.; Yang, Y.; Im, S.; Waskell, L. A.; Kincaid, J. R. Experimental documentation of the structural consequences of hydrogen-bonding interactions to the proximal cysteine of a cytochrome P450. *Angew. Chem. Int. Ed.* **2012**, *51*, 10403.

8. Krest, C. M.; Silakov, A.; Rittle, J.; Yosca, T. H.; Onderko, E. L.; Calixto, J. C.; Green, M. T. Significantly shorter Fe–S bond in cytochrome P450-I is consistent with greater reactivity relative to chloroperoxidase. *Nat. Chem.* **2015**, *7*, 696.

9. McQuarters, A. B.; Speelman, A. L.; Chen, L.; Elmore, B. O.; Fan, W.; Feng, C.; Lehnert, N. Exploring second coordination sphere effects in nitric oxide synthase. *J. Biol. Inorg. Chem.* **2016**, 1.

10. Hunt, A. P.; Lehnert, N. The Thiolate Trans Effect in Heme {FeNO}6 Complexes and Beyond: Insight into the Nature of the Push Effect. *Inorg. Chem.* **2019**.

11. Aono, S.; Ohkubo, K.; Matsuo, T.; Nakajima, H. Redox-controlled ligand exchange of the heme in the CO-sensing transcriptional activator CooA. *J. Biol. Chem.* **1998**, *273*, 25757.

12. Shelver, D.; Thorsteinsson, M. V.; Kerby, R. L.; Chung, S.-Y.; Roberts, G. P.; Reynolds, M. F.; Parks, R. B.; Burstyn, J. N. Identification of two important heme site residues (Cysteine 75 and Histidine 77) in CooA, the CO-sensing transcription factor of *Rhodospirillum rubrum*. *Biochemistry* **1999**, *38*, 2669.

13. Dhawan, I. K.; Shelver, D.; Thorsteinsson, M. V.; Roberts, G. P.; Johnson, M. K. Probing the heme axial ligation in the CO-sensing CooA protein with magnetic circular dichroism spectroscopy. *Biochemistry* **1999**, *38*, 12805.

14. Vogel, K. M.; Spiro, T. G.; Shelver, D.; Thorsteinsson, M. V.; Roberts, G. P. Resonance Raman evidence for a novel charge relay activation mechanism of the CO-dependent heme protein transcription factor CooA. *Biochemistry* **1999**, *38*, 2679.

15. Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **1987**, *195*, 687.

16. Cupp-Vickery, J. R.; Poulos, T. L. Structure of cytochrome P450 eryF: an enzyme involved in erythromycin biosynthesis. *Nat. Struct. Biol.* **1995**, *2*, 144.

17. Ravichandran, K. G.; Boddupalli, S. S.; Hasermann, C. A.; Peterson, J. A.; Deisenhofer, J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **1993**, *261*, 731.

18. Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **1995**, *3*, 41.

19. Groves, J. T. Key elements of the chemistry of cytochrome P-450: The oxygen rebound mechanism. *J. Chem. Ed.* **1985**, *62*, 928.

20. Dawson, J. H.; Sono, M. Cytochrome P-450 and chloroperoxidase: thiolate-ligated heme enzymes. Spectroscopic determination of their active-site structures and mechanistic implications of thiolate ligation. *Chem. Rev.* **1987**, *87*, 1255.

21. Behan, R. K.; Green, M. T. On the status of ferryl protonation. *J. Inorg. Biochem.* **2006**, *100*, 448.

22. Davydov, R.; Im, S.; Shanmugam, M.; Gunderson, W. A.; Pearl, N. M.; Hoffman, B. M.; Waskell, L. Role of the Proximal Cysteine Hydrogen Bonding Interaction in Cytochrome P450 2B4 Studied by Cryoreduction, Electron Paramagnetic Resonance, and Electron–Nuclear Double Resonance Spectroscopy. *Biochemistry* **2016**, *55*, 869.

23. Yosca, T. H.; Ledray, A. P.; Ngo, J.; Green, M. T. A new look at the role of thiolate ligation in cytochrome P450. *J. Biol. Inorg. Chem.* **2017**, *22*, 209.

24. Kerby, R. L.; Youn, H.; Roberts, G. P. RcoM: a new single-component transcriptional regulator of CO metabolism in bacteria. *J. Bacteriol.* **2008**, *190*, 3336.

25. Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met104 is the CO-replaceable ligand at Fe(II) heme in the CO-sensing transcription factor BxRcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21*, 559.

26. Roberts, G. P.; Youn, H.; Kerby, R. L. CO-sensing mechanisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 453.

27. Roberts, G. P.; Kerby, R. L.; Youn, H.; Conrad, M. CooA, a paradigm for gas sensing regulatory proteins. *J. Inorg. Biochem.* **2005**, *99*, 280.

28. Kerby, R. L.; Roberts, G. P. Burkholderia xenovorans RcoMBx-1, a Transcriptional Regulator System for Sensing Low and Persistent Levels of Carbon Monoxide. *J. Bacteriol.* **2012**, *194*, 5803.

29. Hines, J. P.; Dent, M. R.; Stevens, D. J.; Burstyn, J. N. Site-directed spin label electron paramagnetic resonance spectroscopy as a probe of conformational dynamics in the Fe(III) "locked-off" state of the CO-sensing transcription factor CooA. *Prot. Sci.* **2018**, *27*, 1670.

30. Deng, B.; Parthasarathy, S.; Wang, W.; Gibney, B. R.; Battaile, K. P.; Lovell, S.; Benson, D. R.; Zhu, H. Study of the Individual Cytochrome b5 and Cytochrome b5 Reductase Domains of Ncb5or Reveals a Unique Heme Pocket and a Possible Role of the CS Domain. *J. Biol. Chem.* **2010**, *285*, 30181.

31. Silkstone, G. G.; Cooper, C. E.; Svistunenko, D.; Wilson, M. T. EPR and Optical Spectroscopic Studies of Met80X Mutants of Yeast Ferricytochrome c. Models for Intermediates in the Alkaline Transition. *J. Am. Chem. Soc.* **2005**, *127*, 92.

32. Smith, A. T.; Su, Y.; Stevens, D. J.; Majtan, T.; Kraus, J. P.; Burstyn, J. N. Effect of the disease-causing R266K mutation on the heme and PLP environments of the human enzyme cystathionine ß-synthase. *Biochemistry* **2012**, *51*, 6360.

33. Bohan, T. L. Analysis of low-spin ESR spectra of ferric heme proteins: A reexamination. *J. Magn. Res.* **1977**, *26*, 109.

34. Peisach, J.; Blumberg, W. E. Electron paramagnetic resonance study of the high- and low-spin forms of cytochrome P-450 in liver and in liver microsomes from a methylcholanthrene-treated rabbit. *Proc. Nat. Acad. Sci. U.S.A.* **1970**, *67*, 172.

35. Blumberg, W. E.; Peisach, J. *Probes of Structure and Function of Macromolecules and Membranes*; Academic Press: New York, 1971; Vol. 2.

36. Palmer, G. In *The Porphyrins*; Dolphin, D., Ed.; Academic Press, Inc.: New York, New York, 1979; Vol. IV, p 313.

37. Barr, I.; Smith, A. T.; Senturia, R.; Chen, Y.; Scheidemantle, B. D.; Burstyn, J. N.; Guo, F. DiGeorge critical region 8 (DGCR8) is a double-cysteine-ligated heme protein. *J. Biol. Chem.* **2011**, *286*, 16716.

38. Stoll, S.; Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Res.* **2006**, *178*, 42.

39. Bruker-AXS; 2016.5-0 ed. Madison, Wisconsin, USA, 2016.

40. Krause, L.; Herbst-Irmer, R.; Sheldrick, G. M.; Stalke, D. Comparison of silver and molybdenum microfocus X-ray sources for single-crystal structure determination. *J. Appl. Crystallogr.* **2015**, *48*, 3.

41. Sheldrick, G. M. Georg-August-Universität Göttingen, Göttingen, Germany, 2013; Vol. 2013/1.

42. Sheldrick, G. M. 2013; Vol. 2018.

43. Sheldrick, G. M. SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr. Sect. A* **2015**, *71*, 3.

44. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. Sect. C* **2015**, *71*, 3.

45. Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.* **2009**, *42*, 339.

46. Guzei, I. A. University of Wisconsin-Madison, Madison, Wisconsin, USA., 2007-2013.

47. McGarvey, B. R. Survey of ligand field parameters of strong field d5 complexes obtained from the g matrix. *Coord. Chem. Rev.* **1998**, *170*, 75.

48. Corliss, C.; Sugar, J. *J. Phys. Chem. Ref. Data* **1982**, *11*, 135.

49. Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1327.

50. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalman, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; X. Li, H.; Hratchian, P.; Izmaylo, A. F.; Bloino, J.; Zhen, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpar, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gompert, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Wallingford, CT, USA, 2009.

51. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

52. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822.

53. Schafer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571.

54. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.

55. Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.

56. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

57. Durley, R. C.; Mathews, F. S. Refinement and structural analysis of bovine cytochrome b5 at 1.5 A resolution. *Acta Crystallogr. Sect. D* **1996**, *52*, 65.

58. Ochi, H.; Hata, Y.; Tanaka, N.; Kakudo, M.; Sakurai, T.; Aihara, S.; Morita, Y.; Huber, R. Structure of rice ferricytochrome c at 2·0 Å resolution. *J. Mol. Biol.* **1983**, *166*, 407.

59. Meier, M.; Janosik, M.; Kery, V.; Kraus, J. P.; Burkhard, P. Structure of human cystathionine β-synthase: a unique pyridoxal 5′-phosphate-dependent heme protein. *EMBO J.* **2001**, *20*, 3910.

60. Tang, S. C.; Koch, S.; Papaefthymiou, G. C.; Foner, S.; Frankel, R. B.; Ibers, J. A.; Holm, R. H. Axial ligation modes in iron(III) porphyrins. Models for the oxidized reaction states of cytochrome P-450 enzymes and the molecular structure of iron(III) protoporphyrin IX dimethyl ester p-nitrobenzenethiolate. *J. Am. Chem. Soc.* **1976**, *98*, 2414.

61. Verras, A.; Alian, A.; de Montellano, P. R. Cytochrome P450 active site plasticity: Attenuation of imidazole binding in cytochrome P450(cam) by an L244A mutation. *Protein Eng. Des. Sel.* **2006**, *19*, 491.

62. Schmidt, J. R.; Polik, W. F.; WebMO LLC: Holland, MI, USA, 2013; Vol. 2019.

63. Griffith, J. S. Binding in Hæmoglobin Azide as Determined by Electron Resonance: Theory of Electron Resonance in Ferrihæmoglobin Azide. *Nature* **1957**, *180*, 30.

64. Taylor, C. P. S. The EPR of low spin heme complexes Relation of the τ2g hole model to the directional properties of the g tensor, and a new method for calculating the ligand field parameters. *Biochim. Biophys. Acta Prot. Struct.***1977**, *491*, 137.

65. Palmer, G. The electron paramagnetic resonance of metalloproteins. *Biochem. Soc. Trans.* **1985**, *13*, 548.

66. Ueyama, N.; Nishikawa, N.; Yamada, Y.; Okamura, T.-a.; Nakamura, A. Cytochrome P-450 Model (Porphinato)(thiolato)iron(III) Complexes with Single and Double NH···S Hydrogen Bonds at the Thiolate Site. *J. Am. Chem. Soc.* **1996**, *118*, 12826.

67. McDaniel, D. H.; Brown, H. C. An Extended Table of Hammett Substitutent Constants Based on the Ionization of Substituted Benzoic Acids. *J. Org. Chem.* **1958**, *23*, 420.

68. Shinkai, H.; Maeda, K.; Yamasaki, T.; Okamoto, H.; Uchida, I. Bis(2-(Acylamino)phenyl) Disulfides, 2-(Acylamino)benzenethiols, and S-(2-(Acylamino)phenyl) Alkanethioates as Novel Inhibitors of Cholesteryl Ester Transfer Protein. *J. Med. Chem.* **2000**, *43*, 3566.

69. Sarma, B. K.; Mugesh, G. Redox Regulation of Protein Tyrosine Phosphatase 1B (PTP1B): A Biomimetic Study on the Unexpected Formation of a Sulfenyl Amide Intermediate. *J. Am. Chem. Soc.* **2007**, *129*, 8872.

70. Byrn, M. P.; Strouse, C. E. Porphyrin sponges. Inversion disorder and inversion twinning in lattice clathrates based on five-coordinate metallotetraarylporphyrin complexes. *J. Am. Chem. Soc.* **1991**, *113*, 2501.

71. Paulat, F.; Lehnert, N. Electronic structure of ferric heme nitrosyl complexes with thiolate coordination. *Inorg. Chem.* **2007**, *46*, 1547.

5.6 Supplementary Information

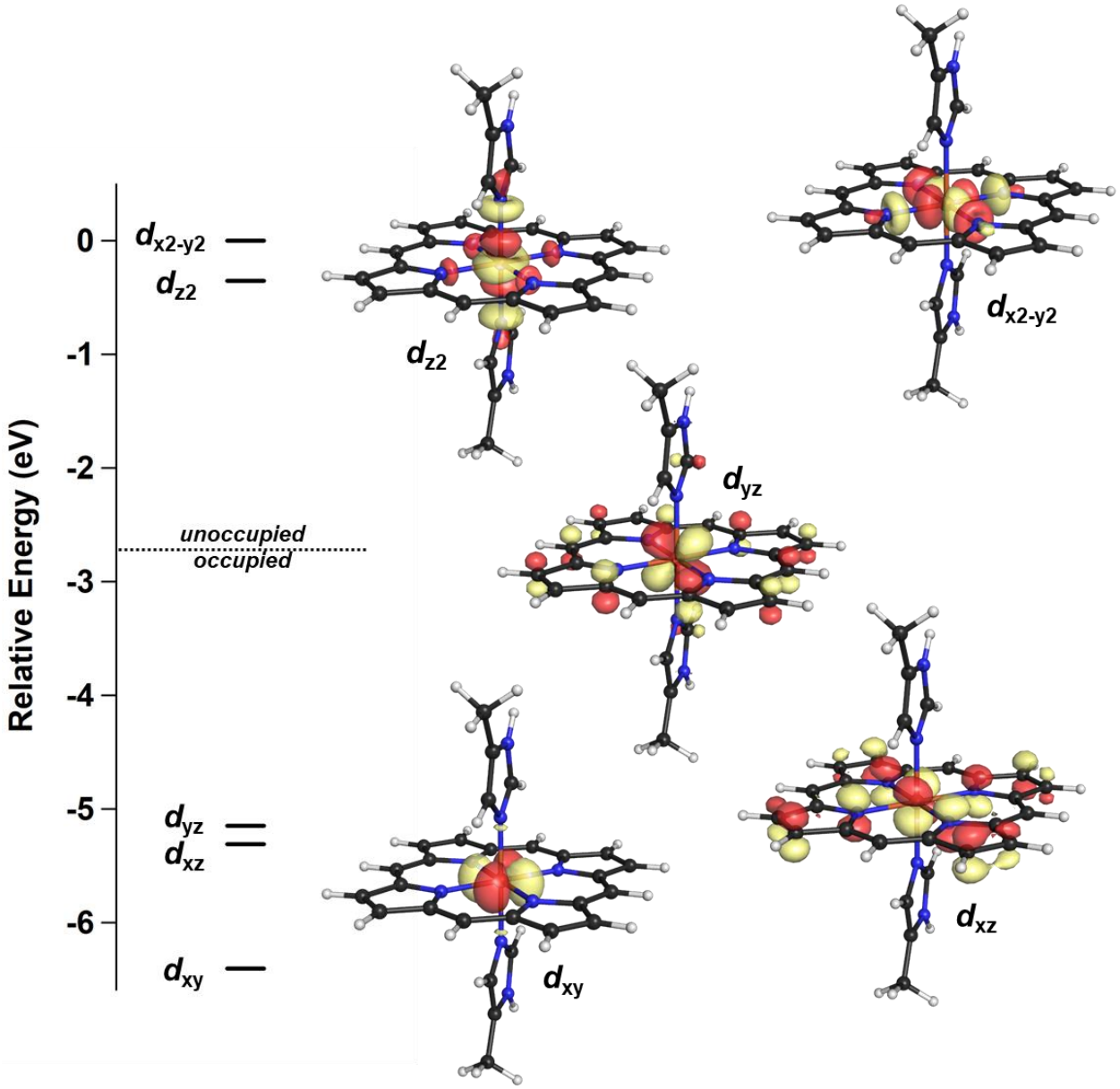


Figure S1. DFT-computed molecular orbital diagram for the Fe-based *d*-orbitals in $[Fe(P)(1\text{-MeIm})_2]^+$. Energies displayed relative to that of d_{x2-y2} for single-electron orbitals (spin-up) and the corresponding orbital wavefunctions (spin-down) are overlaid with the geometry-optimized structure. Contour plots were generated with an isosurface value of 0.05 and overlaid with the geometry-optimized structure using Pymol (v1.3).

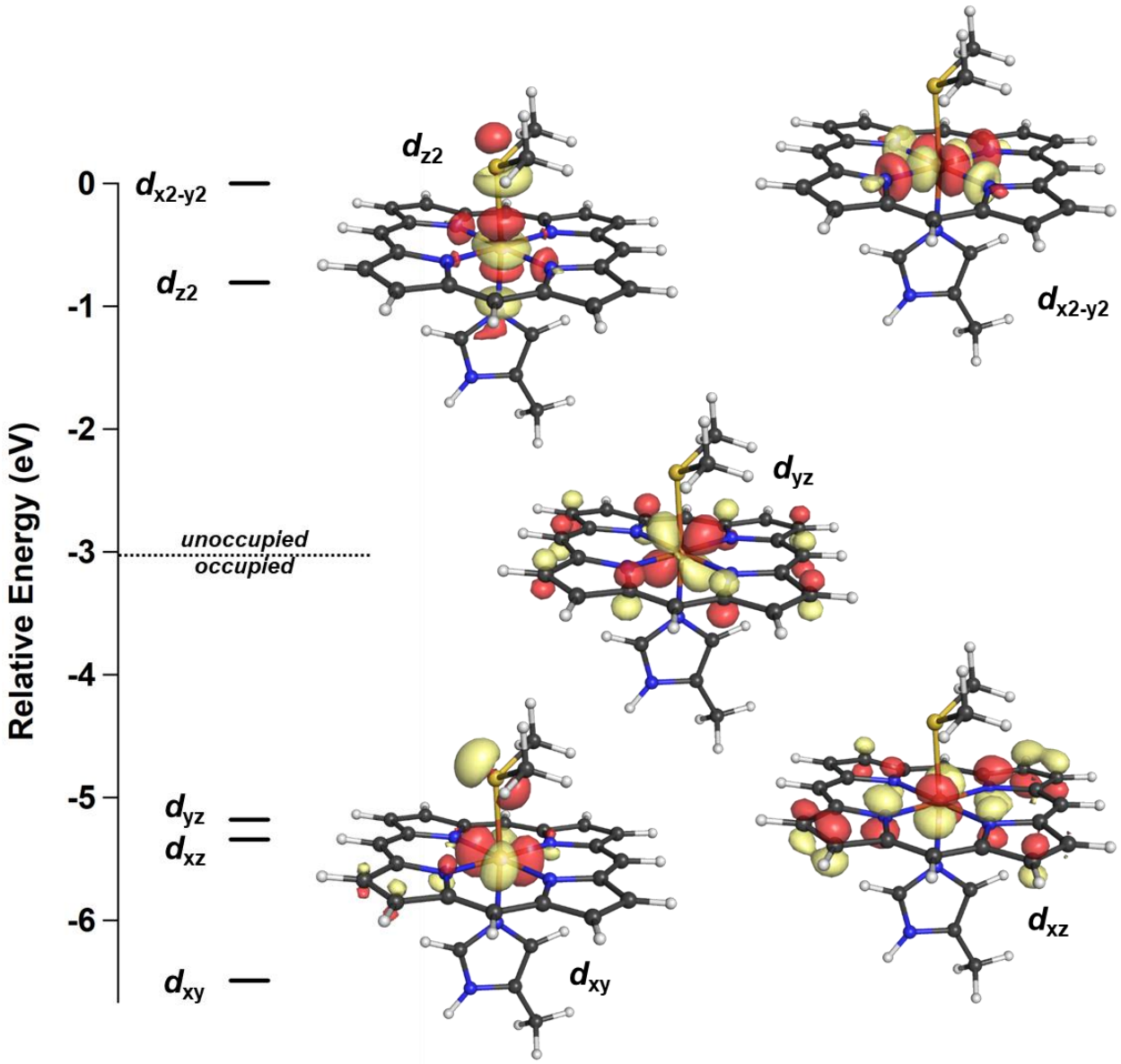


Figure S2. DFT-computed molecular orbital diagram for the Fe-based *d*-orbitals in $[Fe(P)(1\text{-MeIm})(S(CH_3)_2)]^+$. Energies displayed relative to that of d_{x2-y2} for single-electron orbitals (spin-up) and the corresponding orbital wavefunctions (spin-down) are overlaid with the geometry-optimized structure. Contour plots were generated with an isosurface value of 0.05 and overlaid with the geometry-optimized structure using Pymol (v1.3).

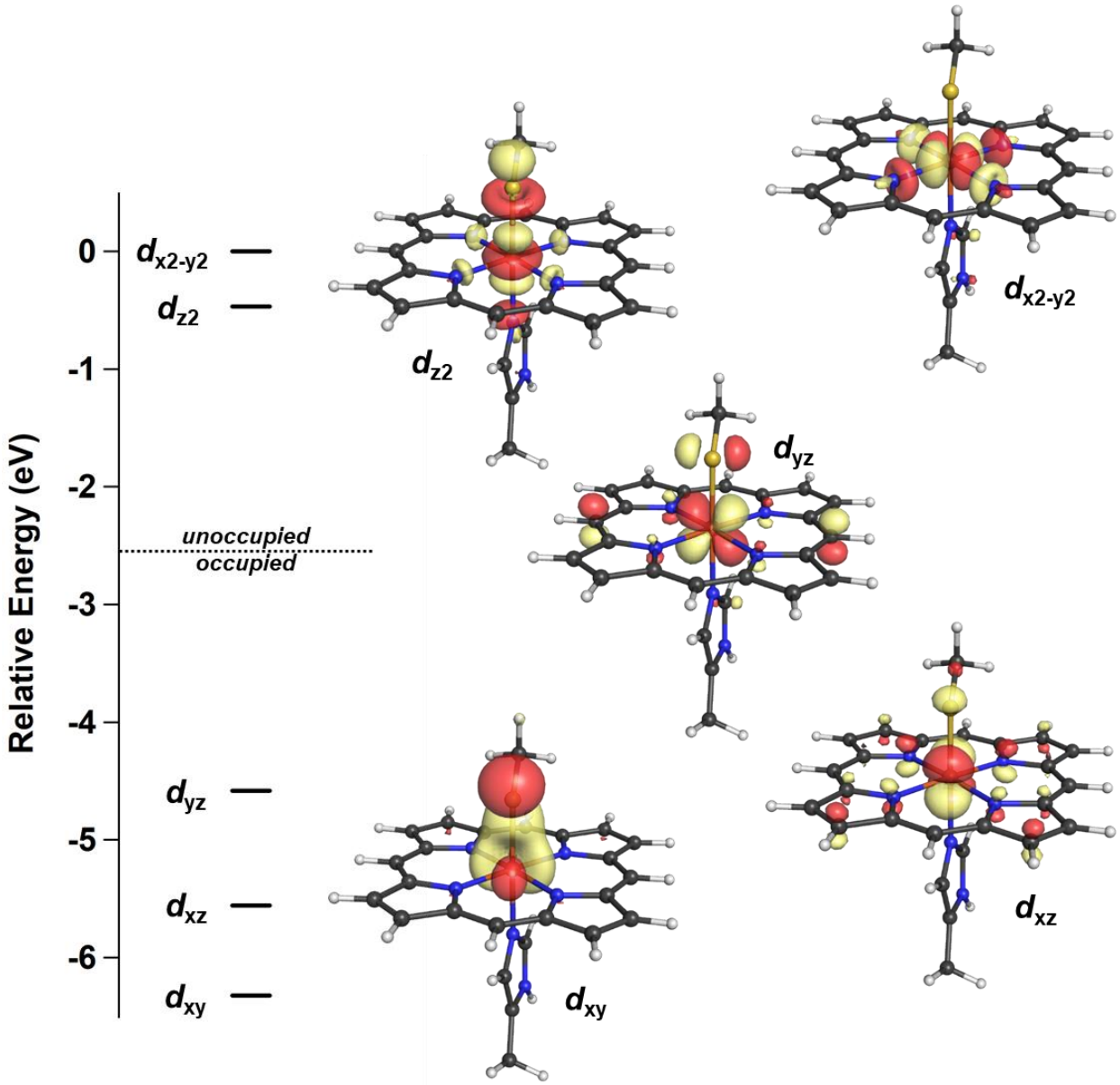


Figure S3. DFT-computed molecular orbital diagram for the Fe-based *d*-orbitals in [Fe(P)(1-MeIm)(SCH_3)]. Energies displayed relative to that of d_{x2-y2} for single-electron orbitals (spin-up) and the corresponding orbital wavefunctions (spin-down) are overlaid with the geometry-optimized structure. Contour plots were generated with an isosurface value of 0.05 and overlaid with the geometry-optimized structure using Pymol (v1.3).

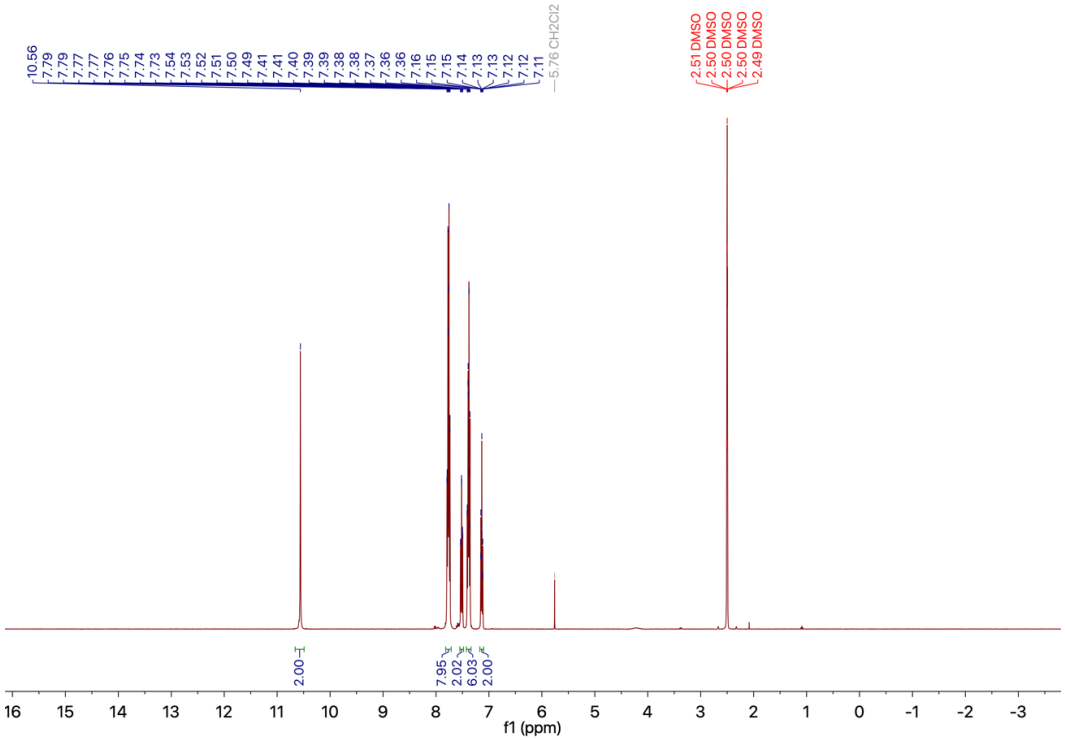


Figure S4. ^{1}H NMR of **SS**-H.

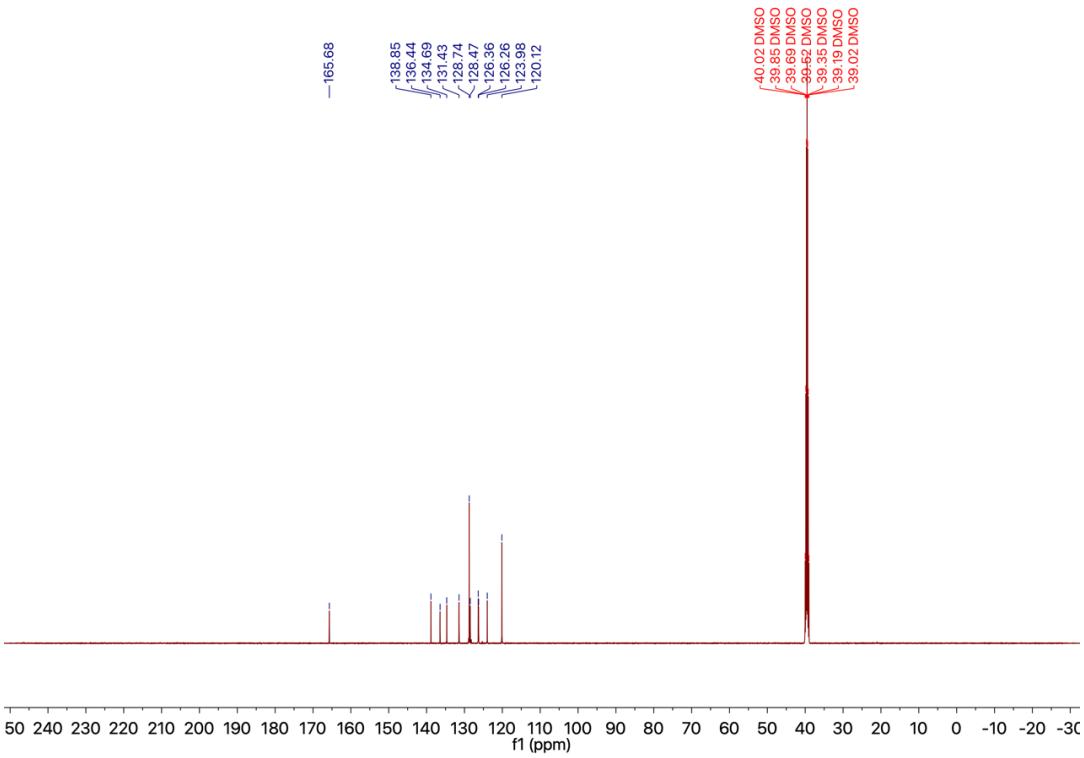


Figure S5. ^{13}C NMR of **SS**-H.

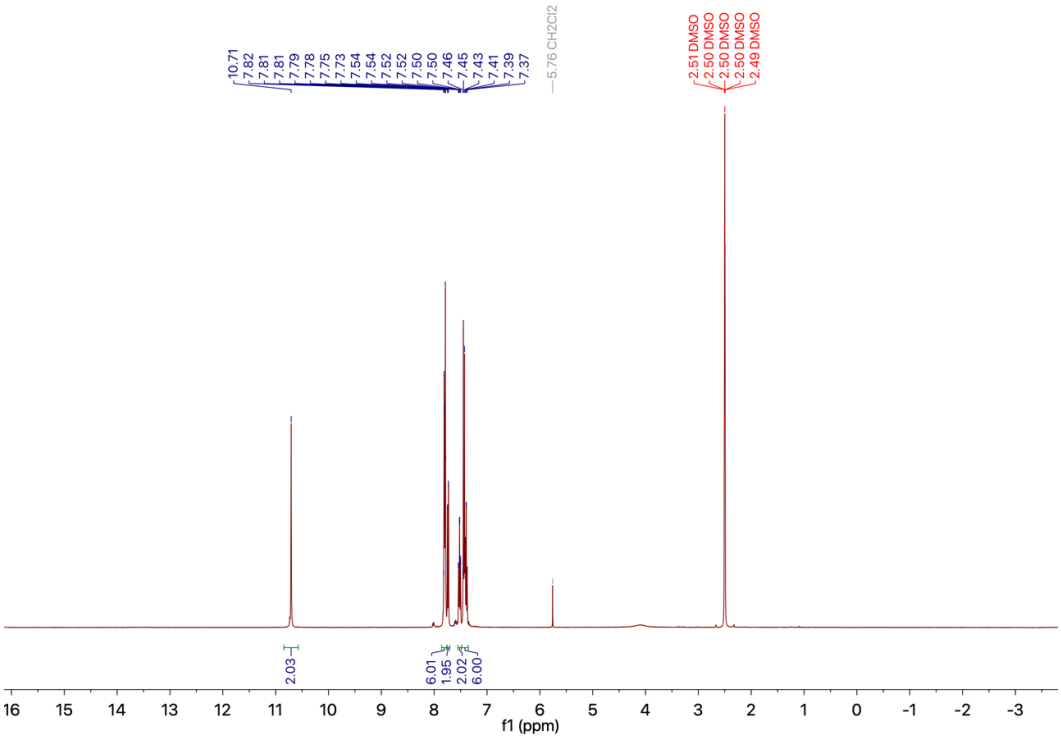


Figure S6. ^{1}H NMR of **SS**-Cl.

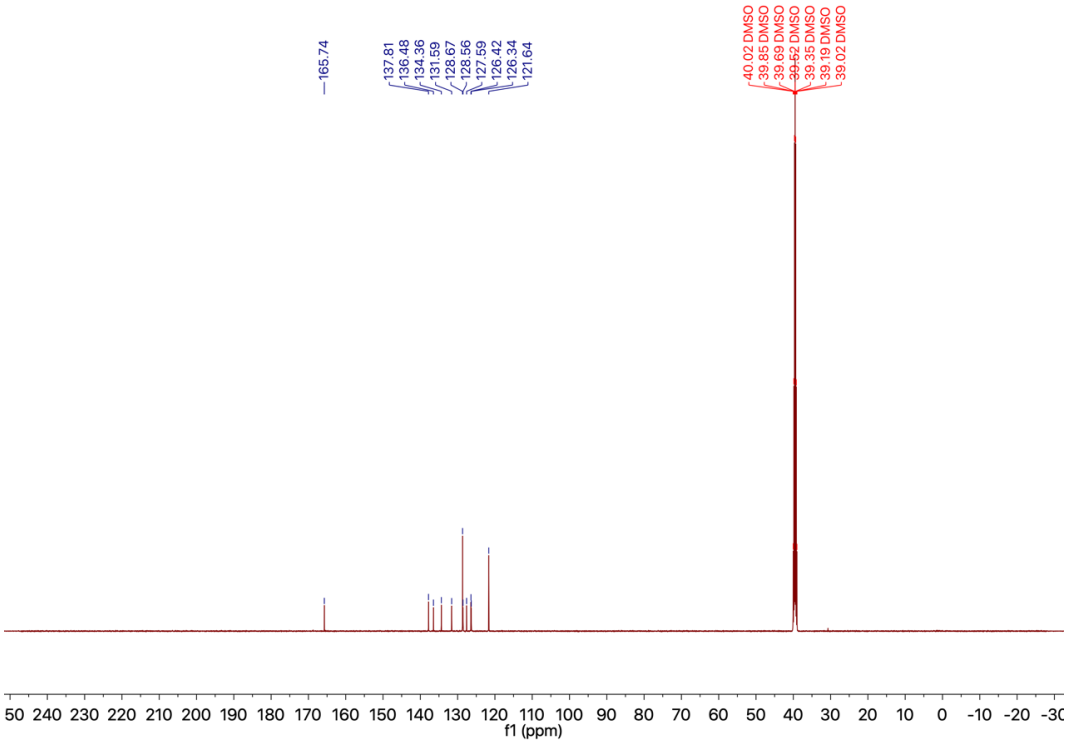


Figure S7. ^{13}C NMR of **SS**-Cl.

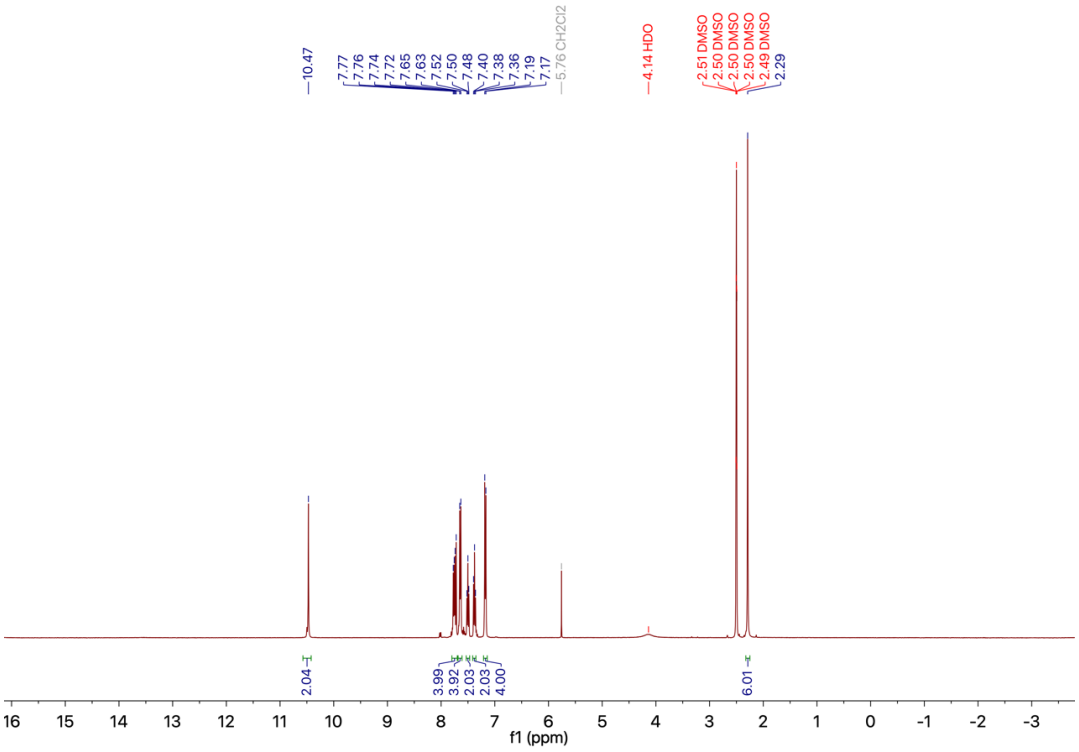


Figure S8. ^{1}H NMR of **SS**-CH_3.

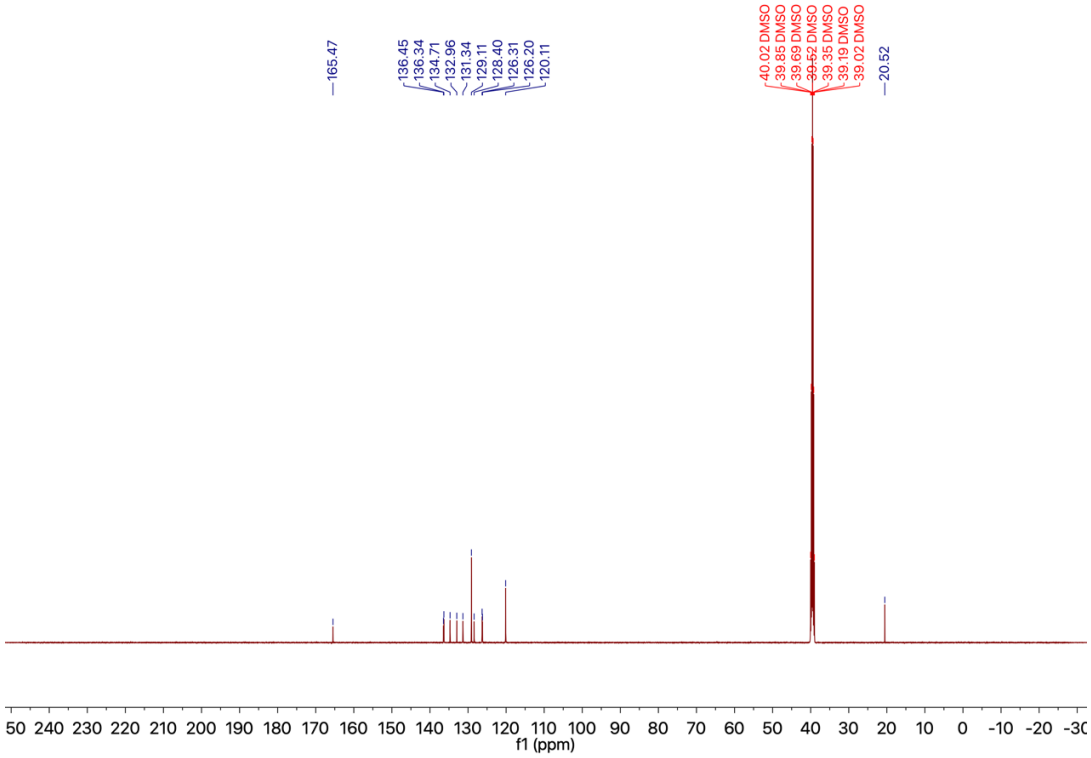


Figure S9. ^{13}C NMR of **SS**-CH_3.

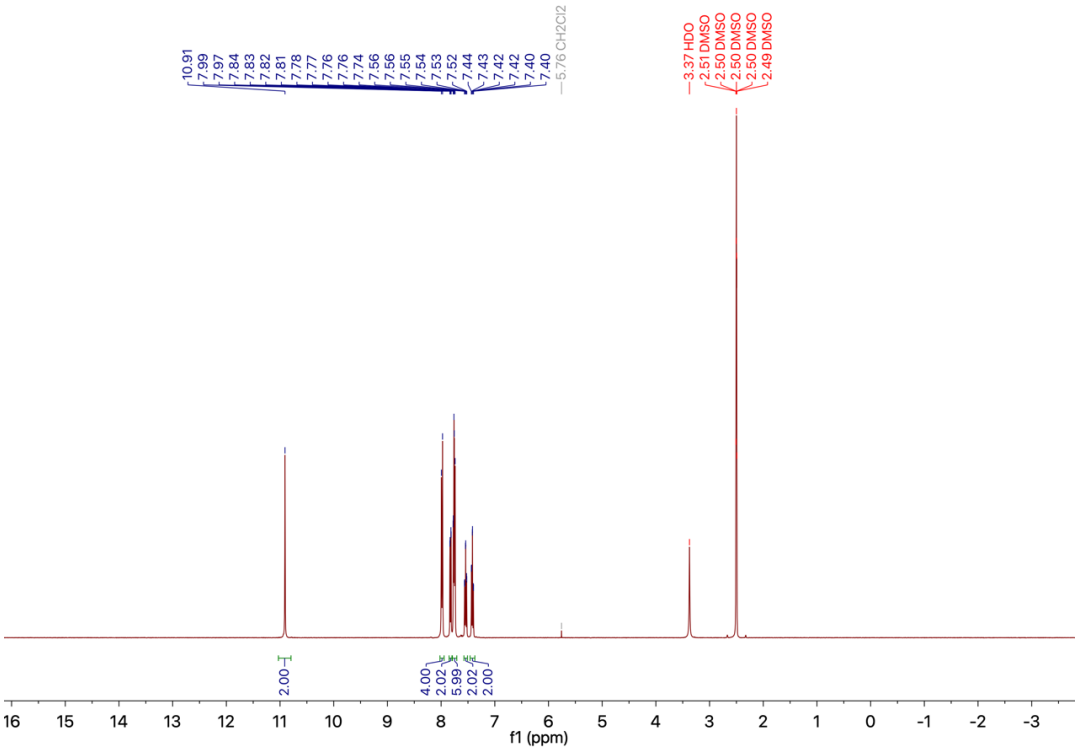


Figure S10. ^{1}H NMR of **SS**-CF_3.

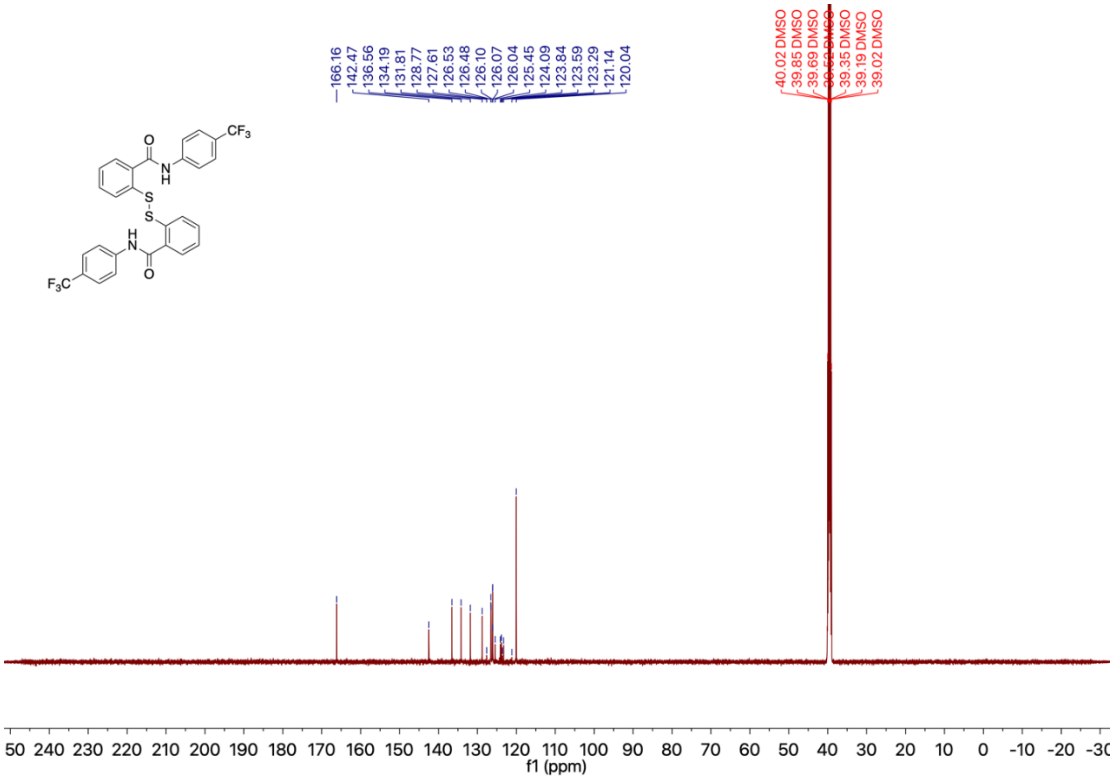


Figure S11. ^{13}C NMR of **SS**-CF_3.

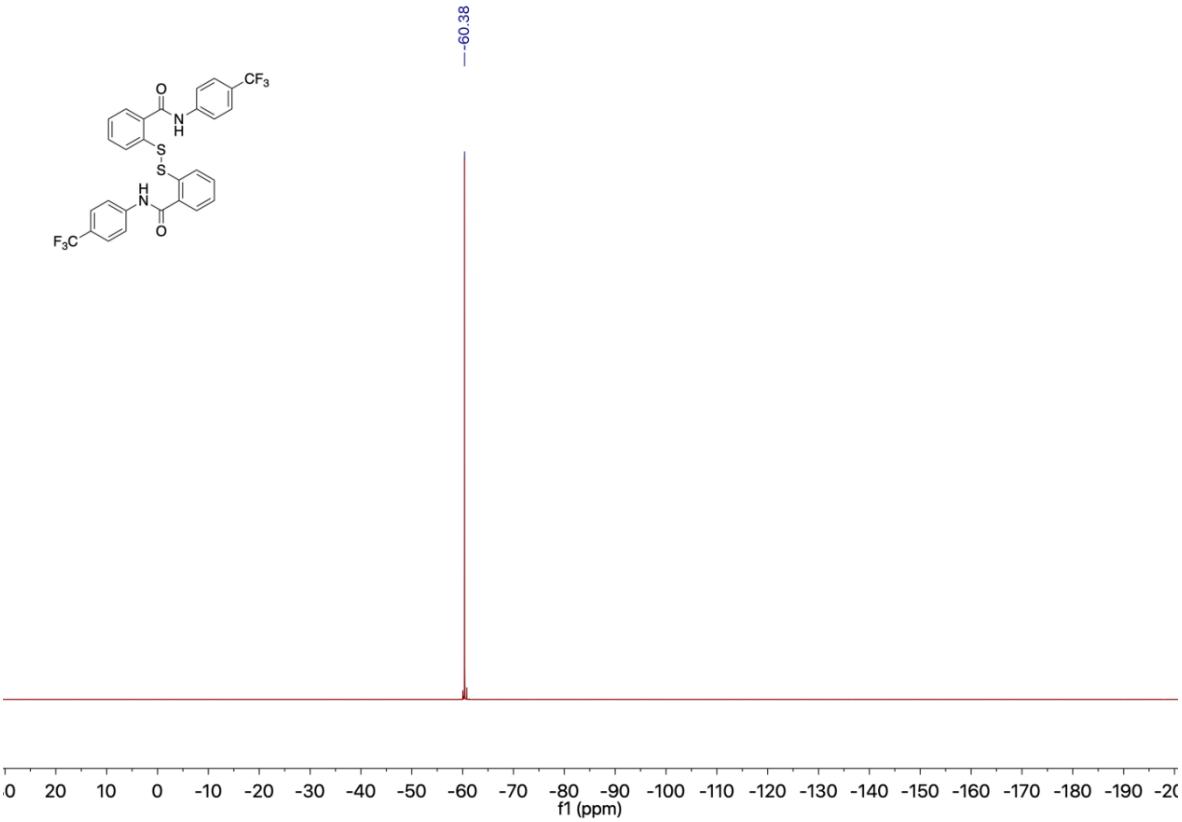


Figure S12. ^{19}F NMR of **SS**-CF_3.

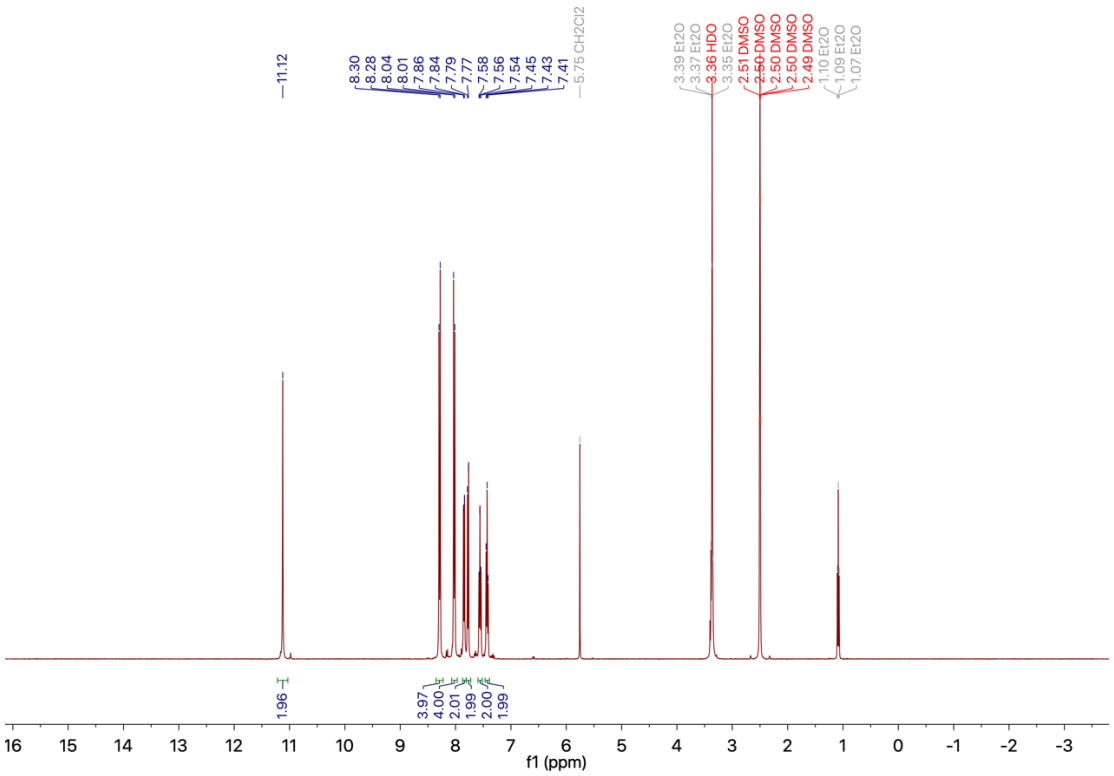


Figure S13. ^{1}H NMR of **SS-**NO_2.

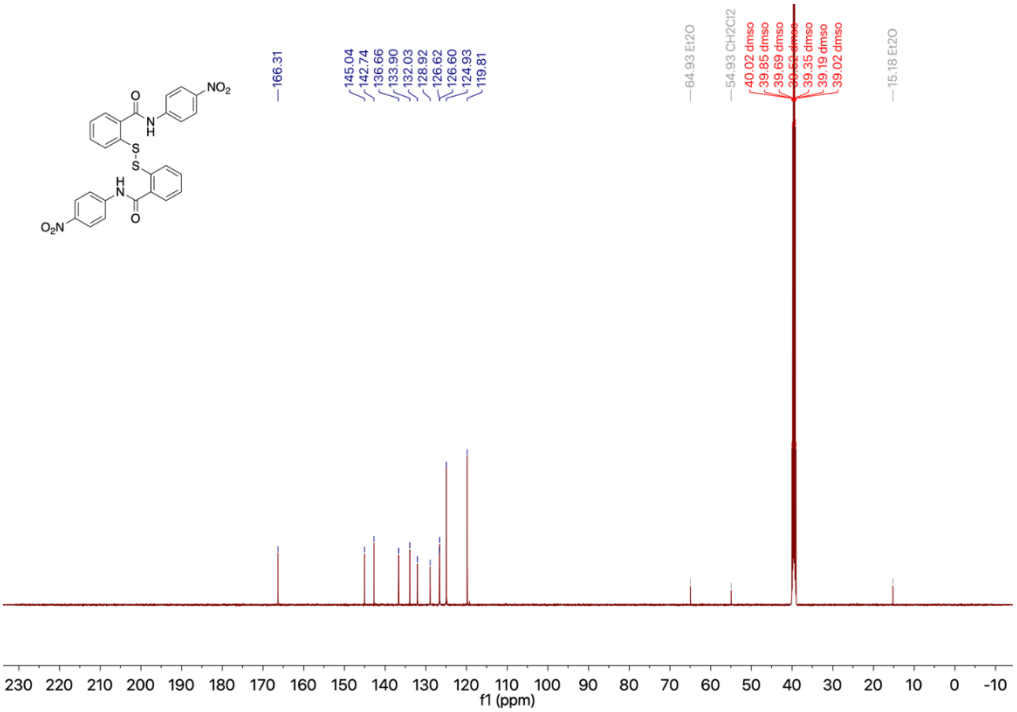


Figure S14. ^{13}C NMR of **SS-**NO_2.

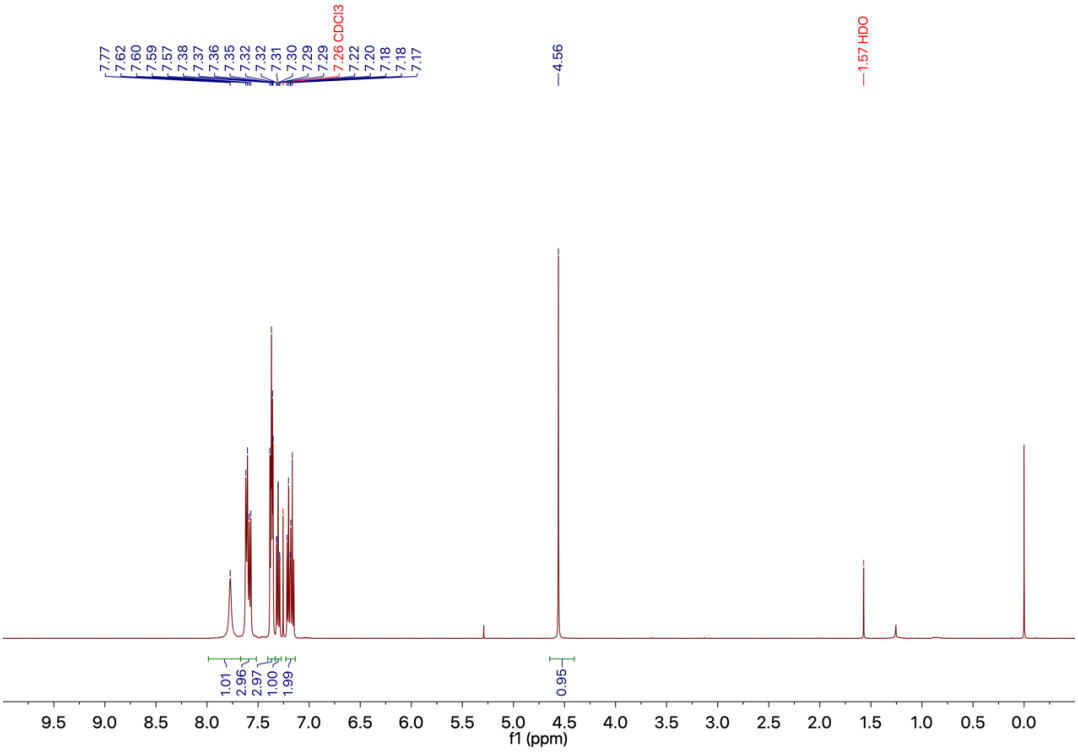


Figure S15. ^{1}H NMR of **HS**-H.

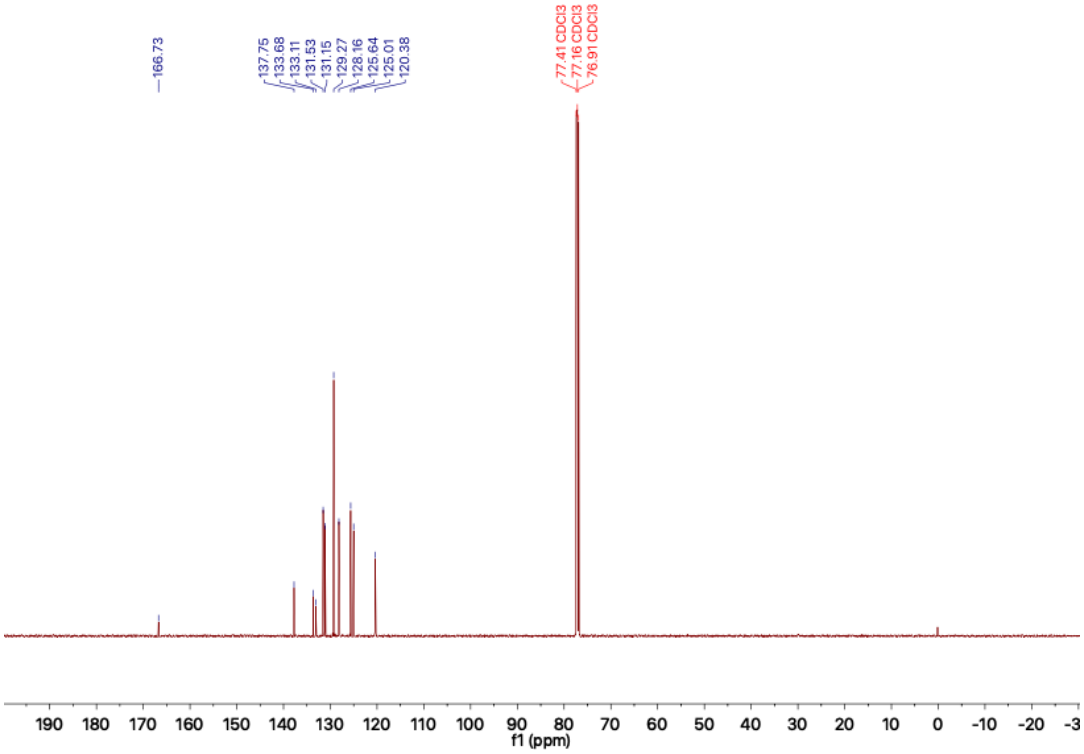


Figure S16. ^{13}C NMR of **HS**-H.

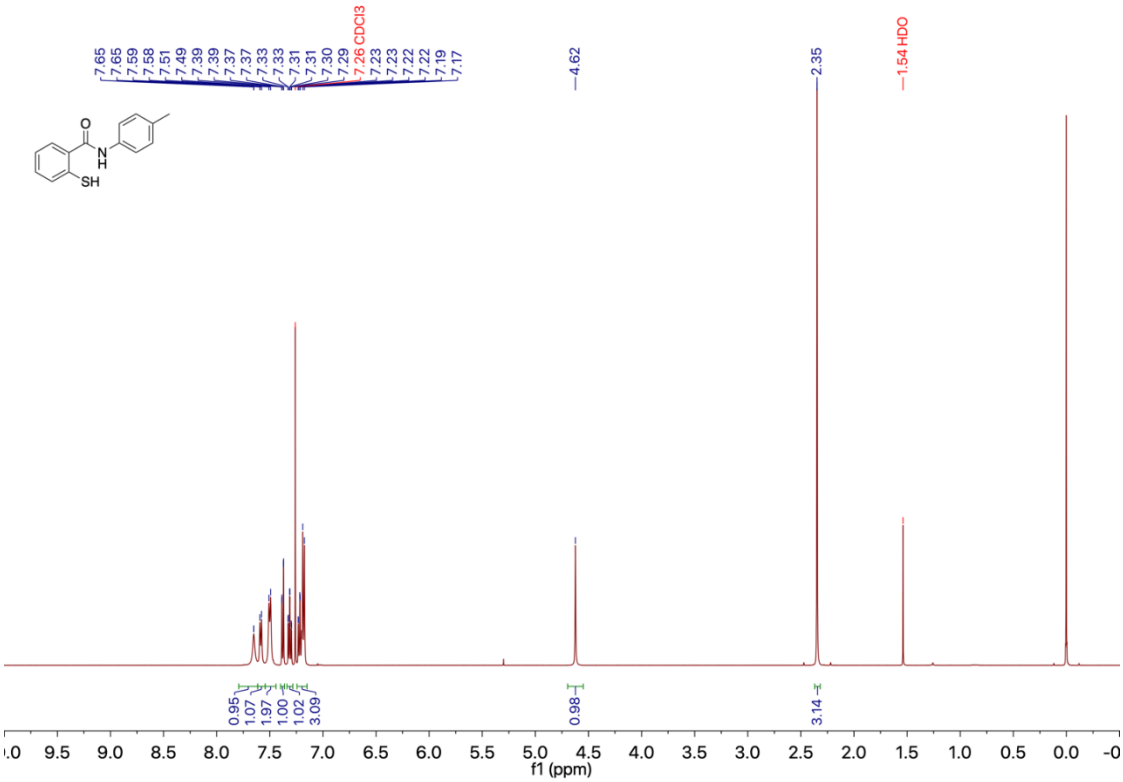


Figure S17. ^{1}H NMR of **HS**-CH_3.

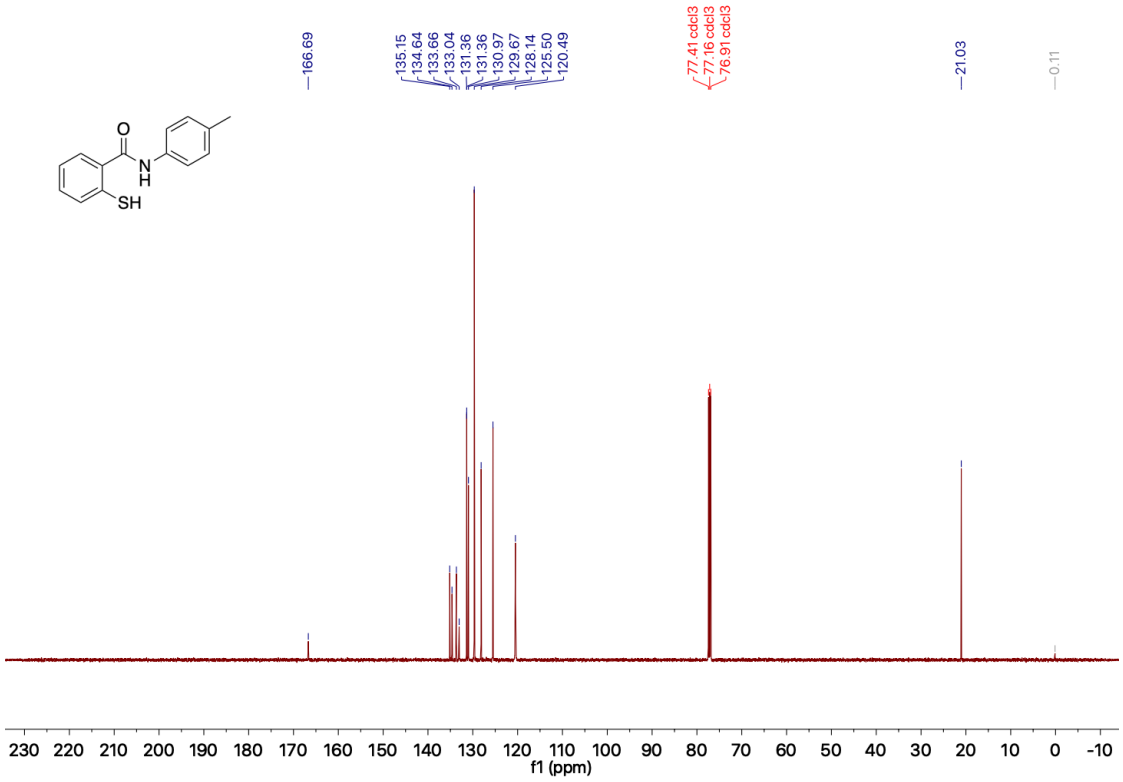


Figure S18. ^{13}C NMR of **HS**-CH_3.

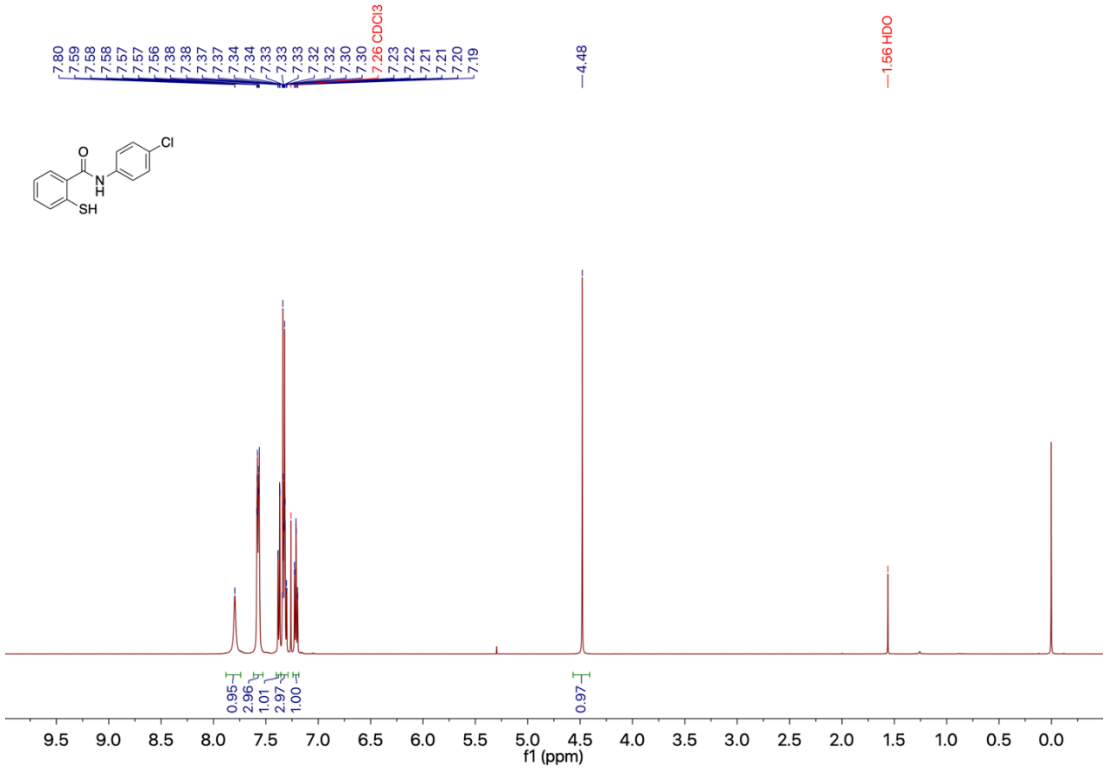


Figure S19. ^{1}H NMR of **HS**-Cl.

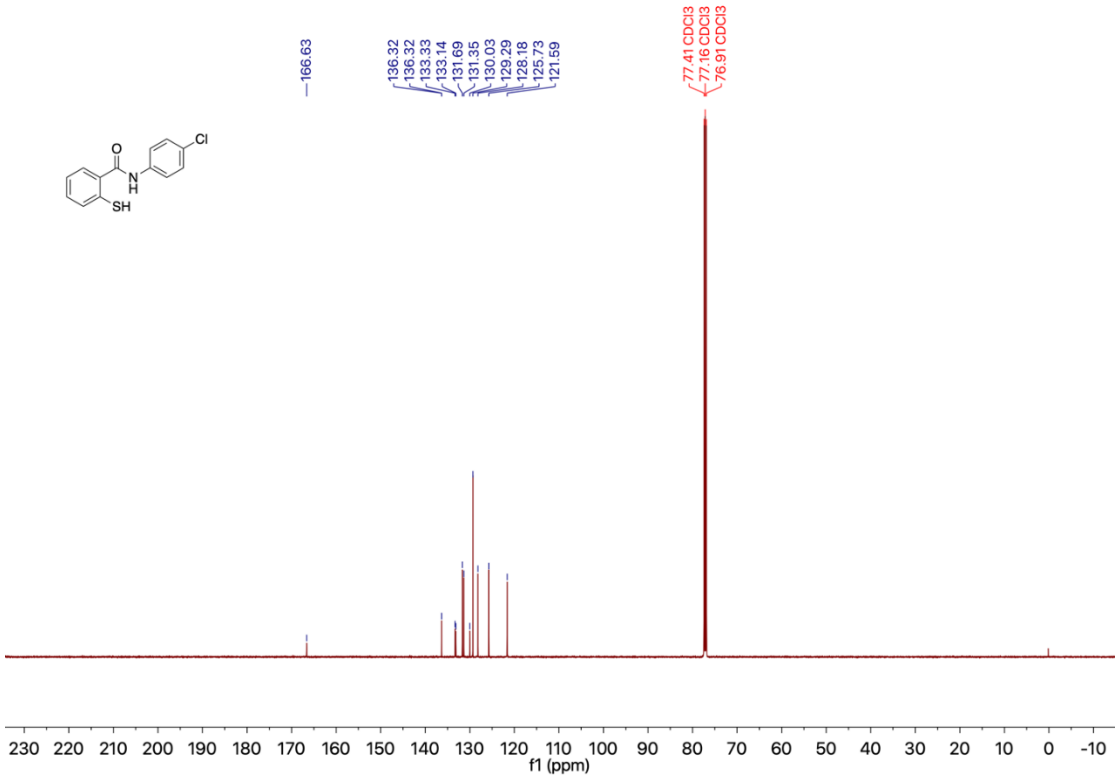


Figure S20. ^{13}C NMR of **HS**-Cl.

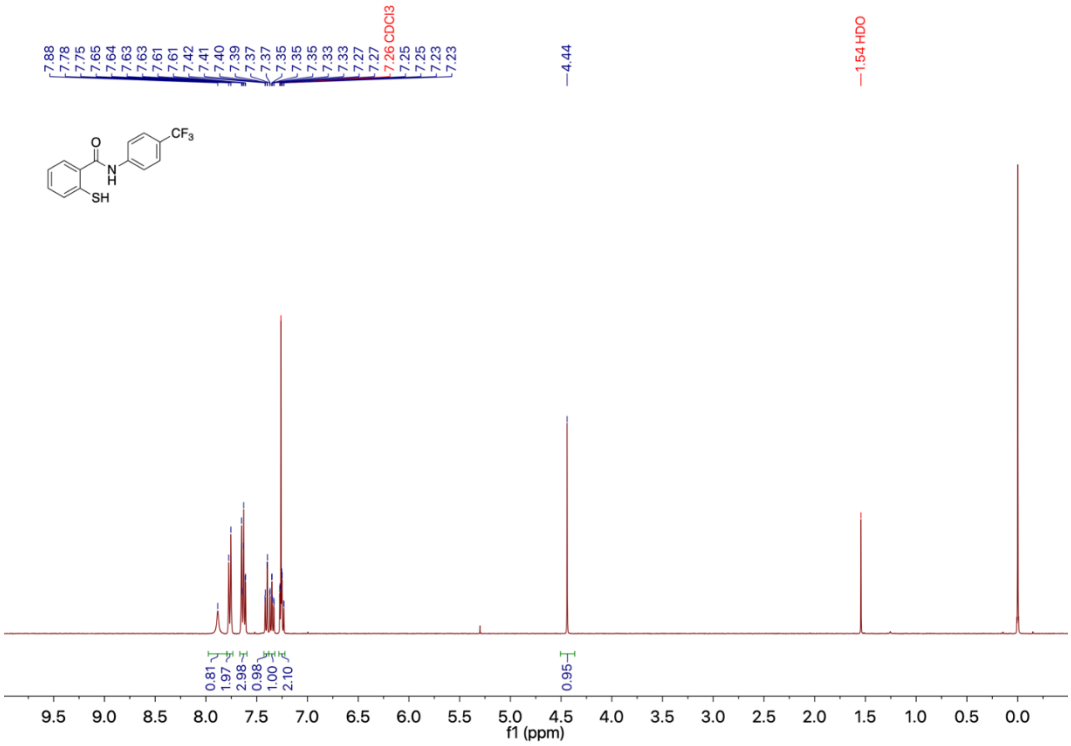


Figure S21. ^{1}H NMR of **HS**-CF_3.

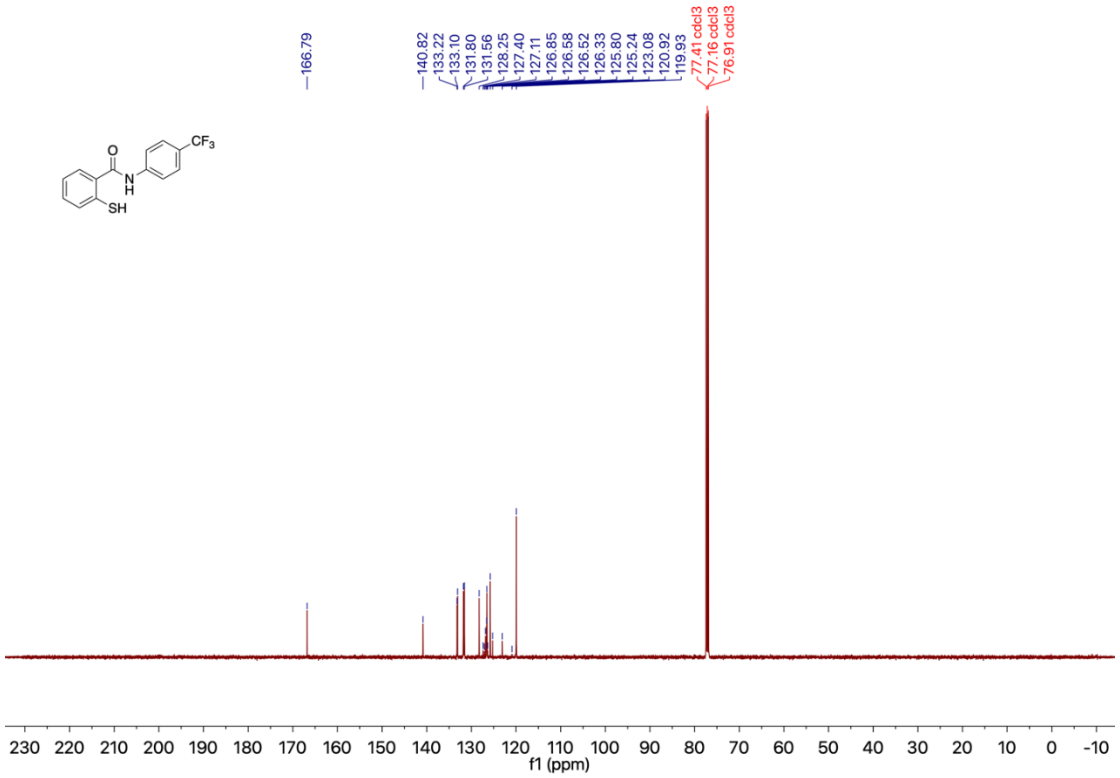


Figure S22. ^{13}C NMR of **HS**-CF_3.

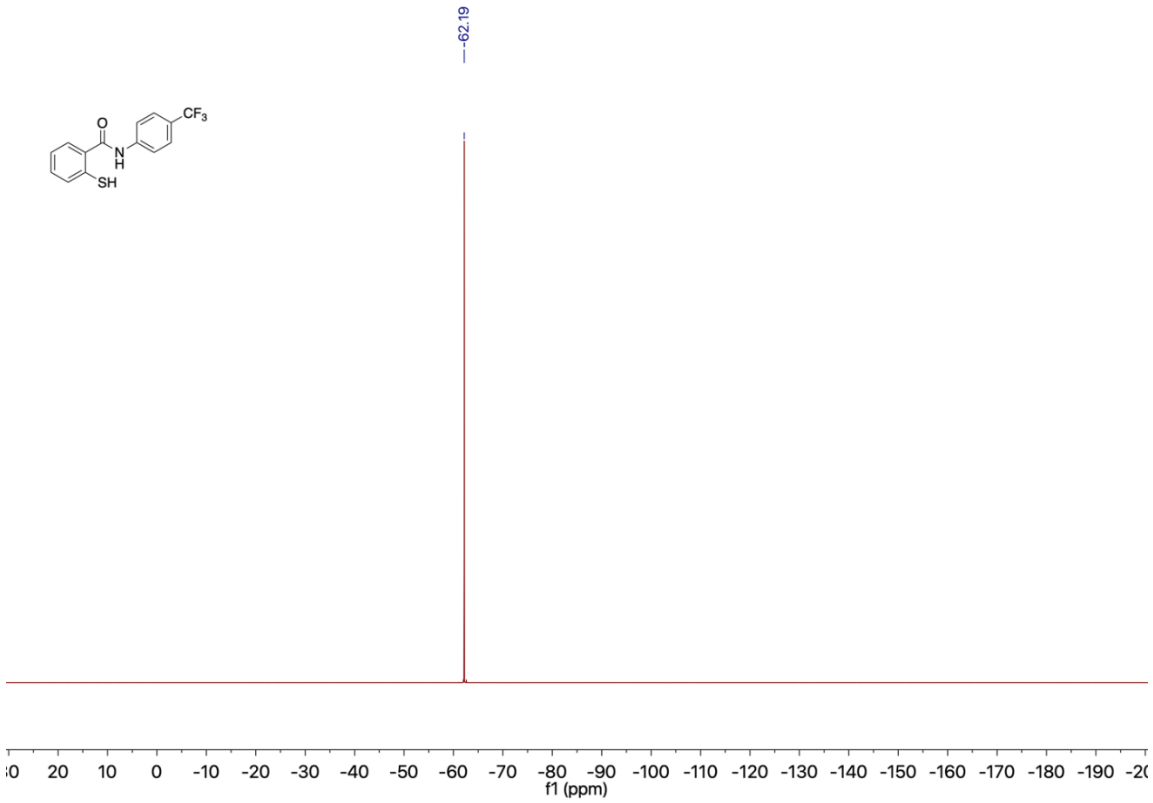


Figure S23. ^{19}F NMR of **HS**-CF_3.

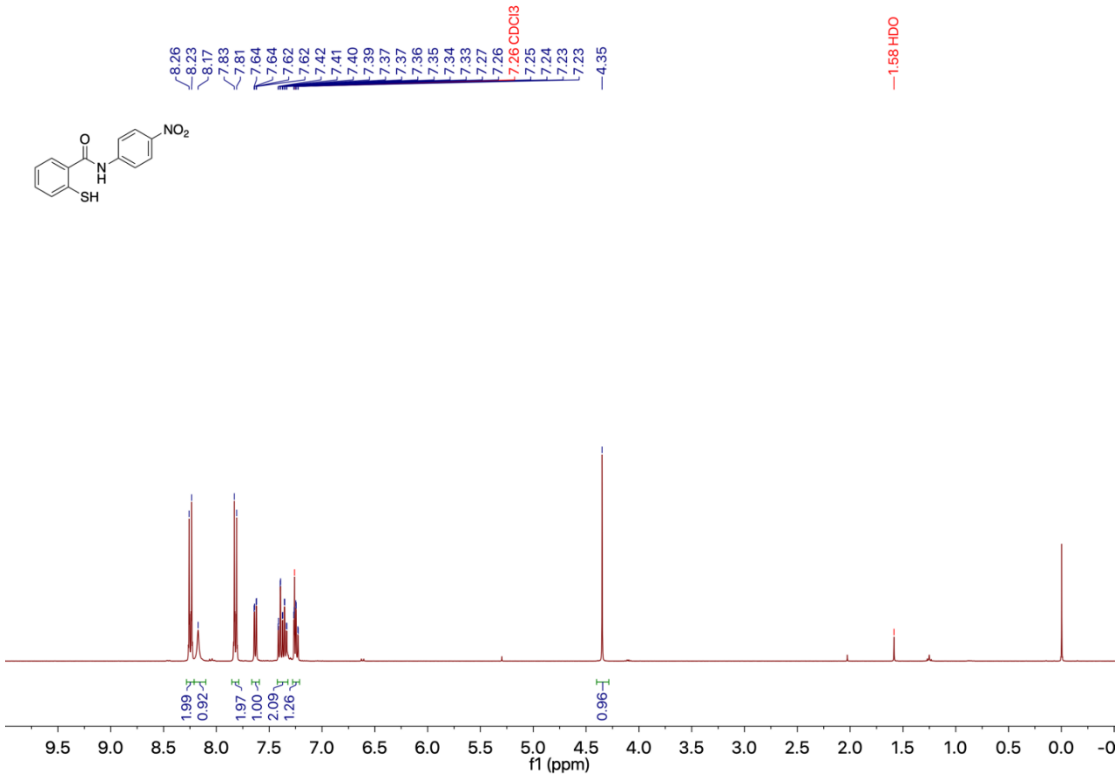


Figure S24. ^{1}H NMR of **HS**-NO_2.

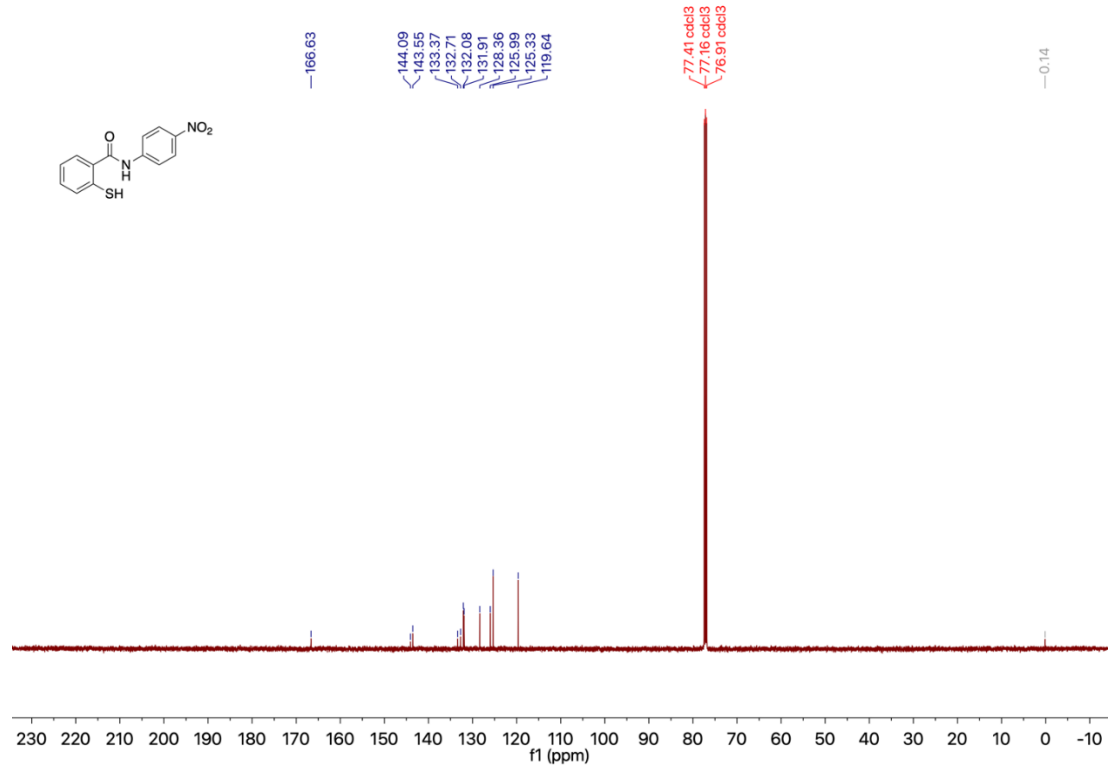


Figure S25. ^{13}C NMR of **HS**-NO_2.

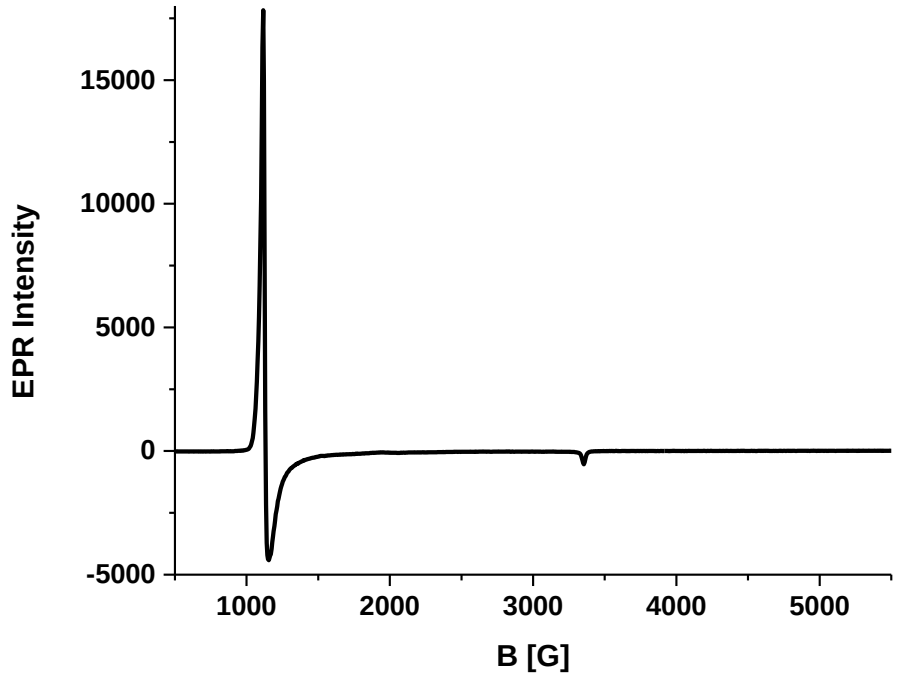


Figure S26. EPR spectrum of 2 mM [Fe(TPP)Cl] in 1:1 CH_2Cl_2:toluene, recorded at 4 K.

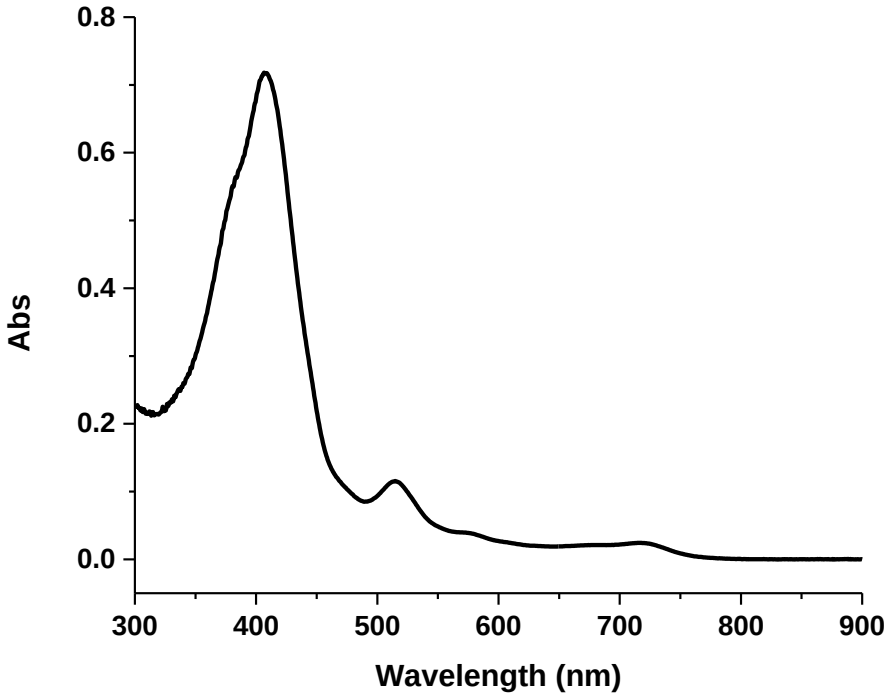


Figure S27. UV-Vis spectrum of [Fe(TPP)(**S**-CH_3)] in CH_2Cl_2.

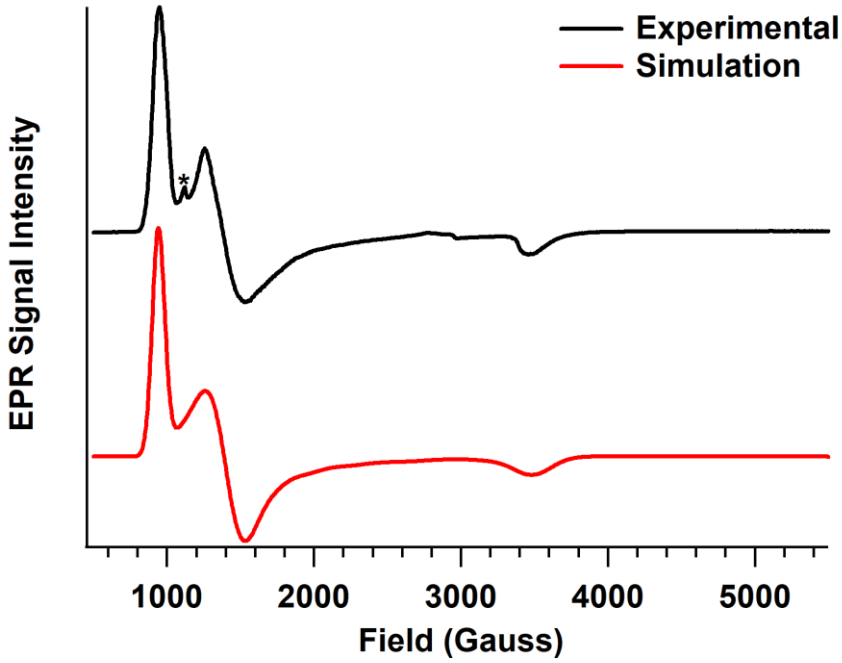


Figure S28. Experimental and simulated EPR spectrum of [Fe(TPP)(**S**-CH_3)] in 1:1 CH_2Cl_2:toluene, collected at 4 K. The asterisk indicates a small impurity of [Fe(TPP)Cl] (see Figure S26).

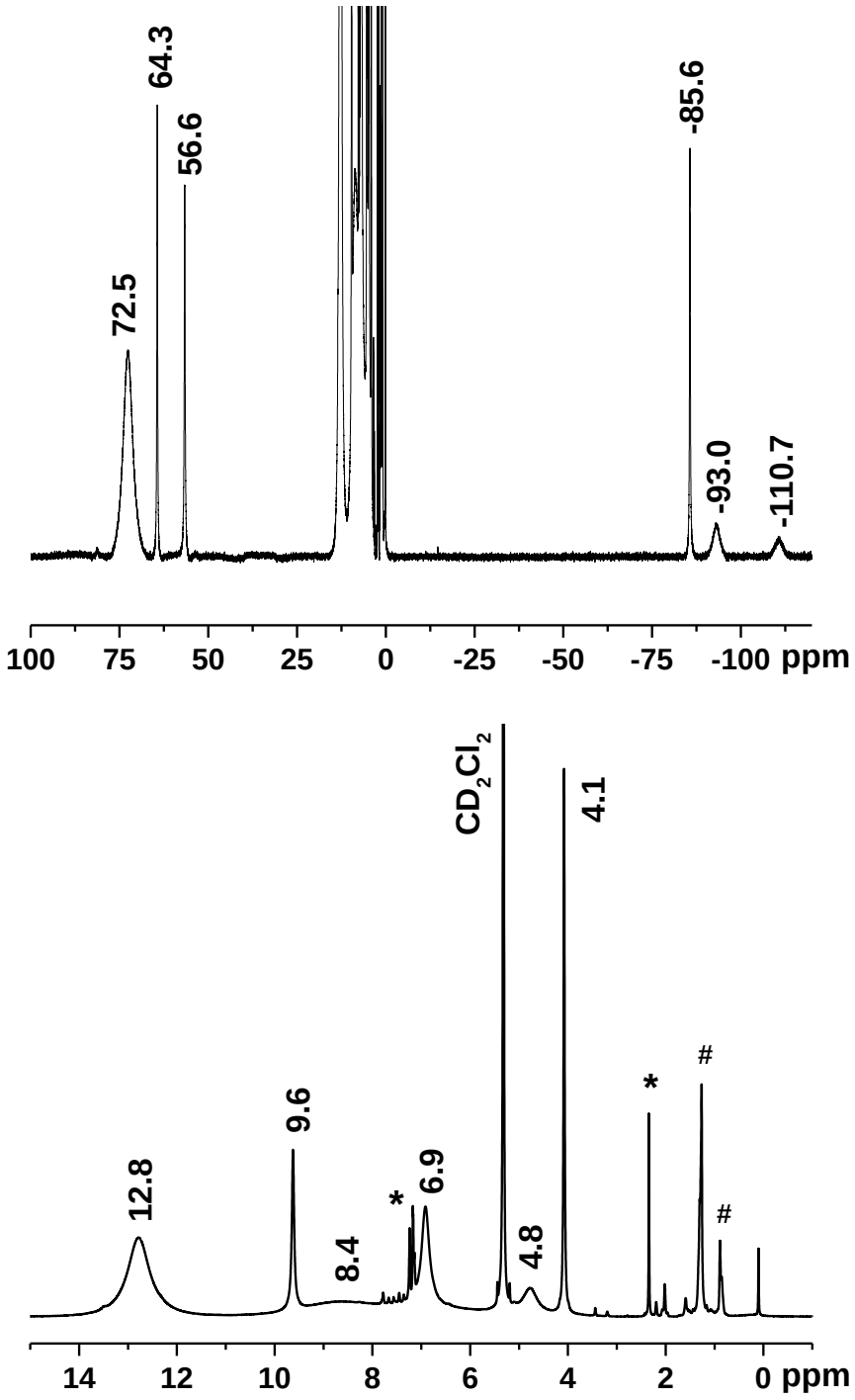


Figure S29. ^{1}H-NMR spectra of [Fe(TPP)(**S**-CH_3)] in CD_2Cl_2 at 20°C. The asterisk (*) and pound (#) symbols denote residual toluene (2.43, 7.15, 7.24 ppm) and hexanes (0.89 and 1.27 ppm), respectively, in the solid from recrystallization of the compound.

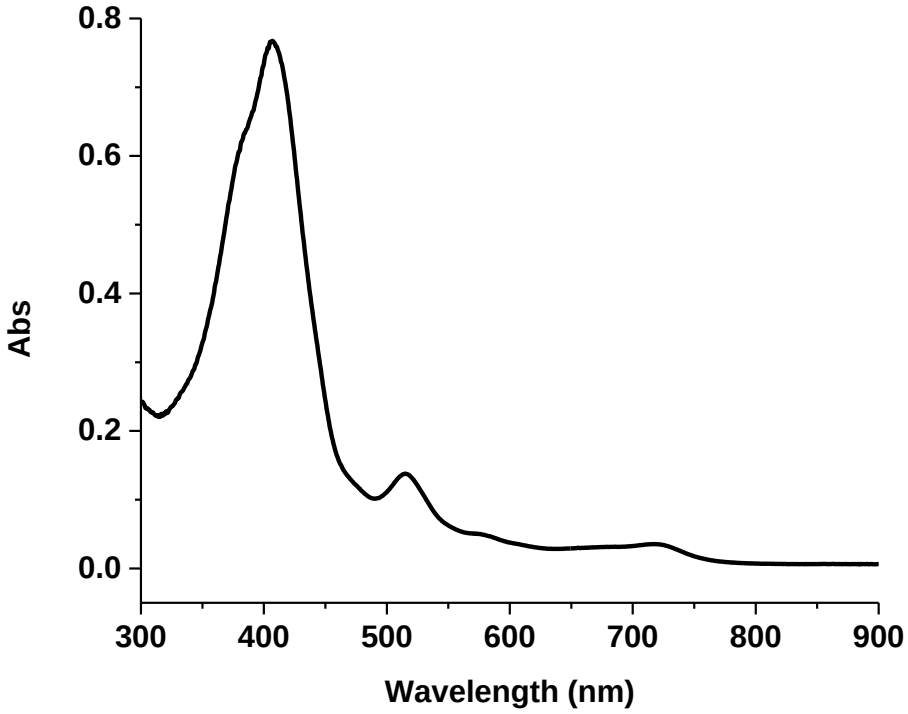


Figure S30. UV-Vis spectrum of [Fe(TPP)(**S**-H)] in CH_2Cl_2.

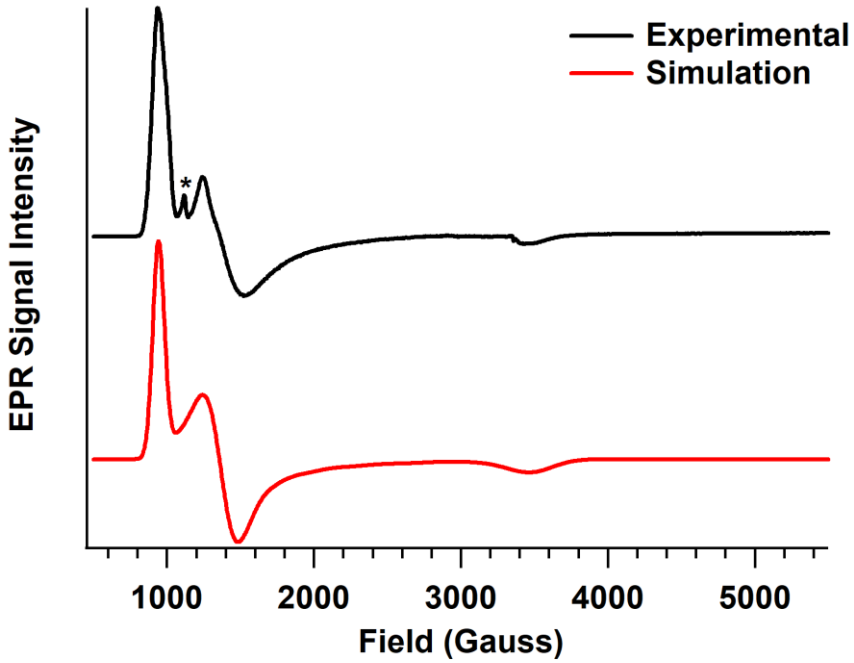


Figure S31. Experimental and simulated EPR spectrum of [Fe(TPP)(**S**-H)] in 1:1 CH_2Cl_2:toluene, collected at 4 K. The asterisk indicates a small impurity of [Fe(TPP)Cl] (see Figure S26).

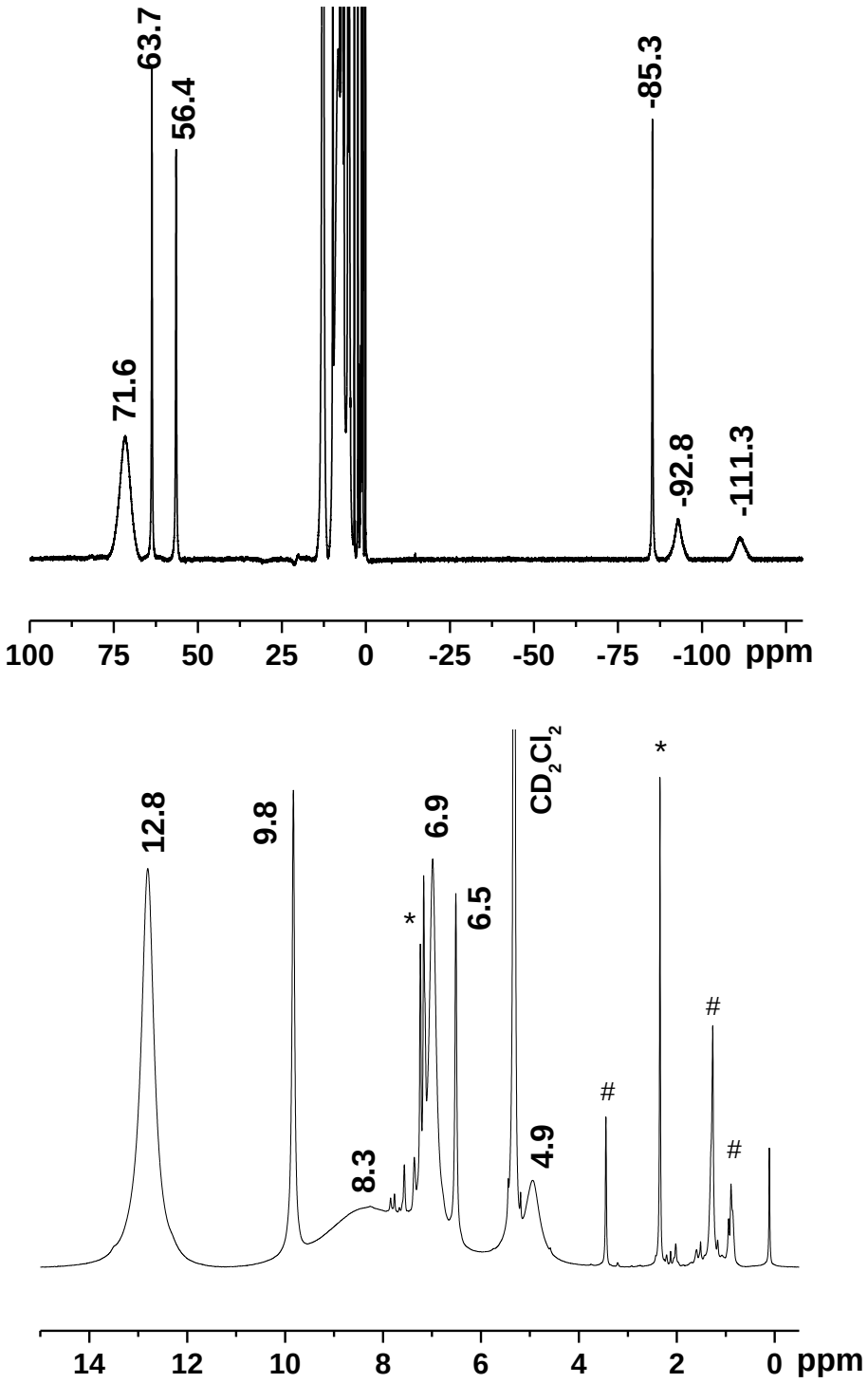


Figure S32. ^{1}H-NMR spectra of [Fe(TPP)(**S**-H)] in CD_2Cl_2 at 20°C. The asterisk (*) and pound (#) symbols denote residual toluene (2.43, 7.15, 7.24 ppm), methanol (1.09 and 3.42 ppm) and hexanes (0.89 and 1.27 ppm), respectively, in the solid from recrystallization of the compound.

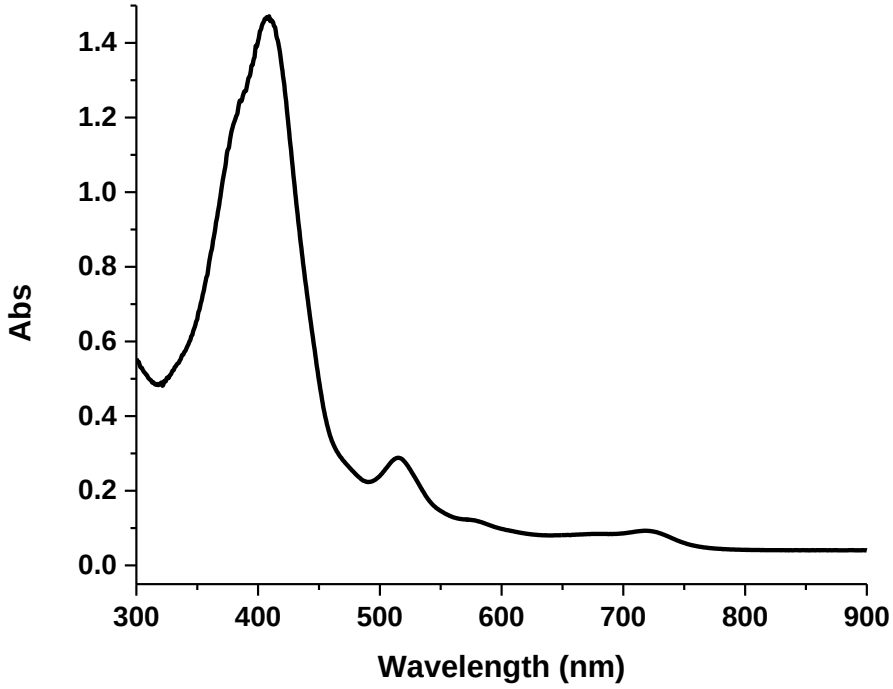


Figure S33. UV-Vis spectrum of [Fe(TPP)(**S**-Cl)] in CH_2Cl_2.

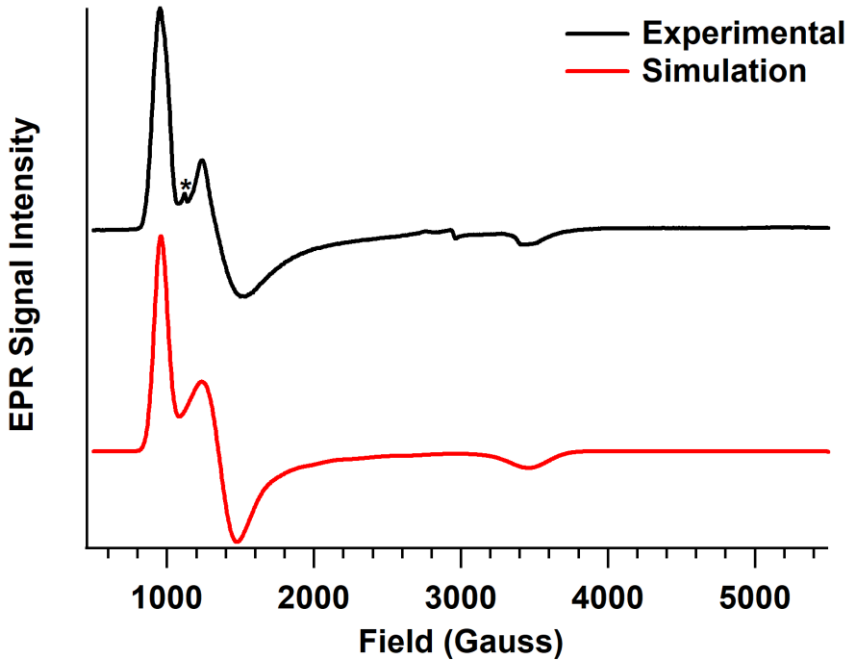


Figure S34. Experimental and simulated EPR spectrum of [Fe(TPP)(**S**-Cl)] in 1:1 CH_2Cl_2:toluene, collected at 4 K. The asterisk indicates a small impurity of [Fe(TPP)Cl] (see Figure S26).

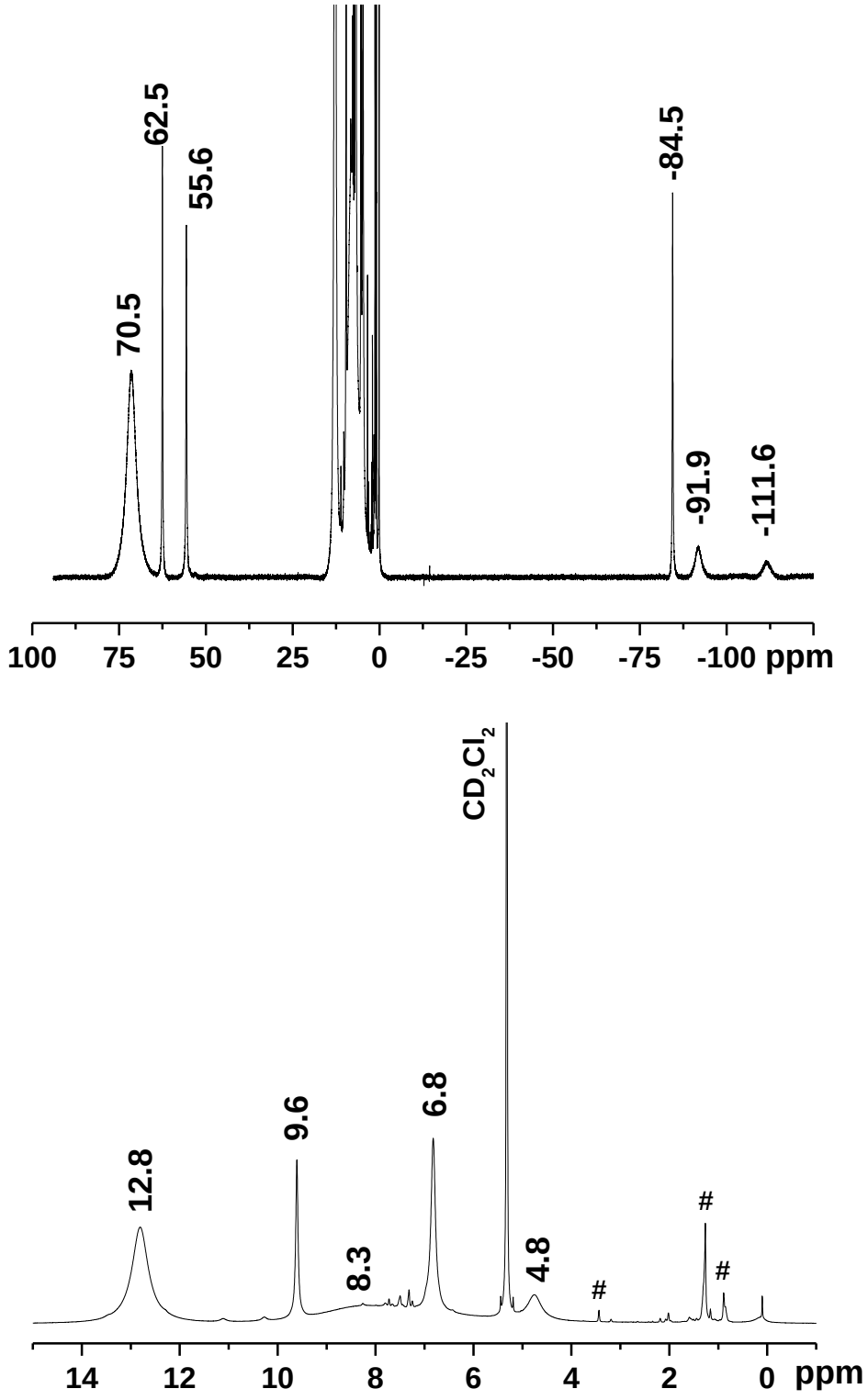


Figure S35. ^{1}H-NMR spectra of [Fe(TPP)(**S**-Cl)] in CD_2Cl_2. The pound (#) symbols denote residual methanol (1.09 and 3.42 ppm) and hexanes (0.89 and 1.27 ppm) in the solid from recrystallization of the compound.

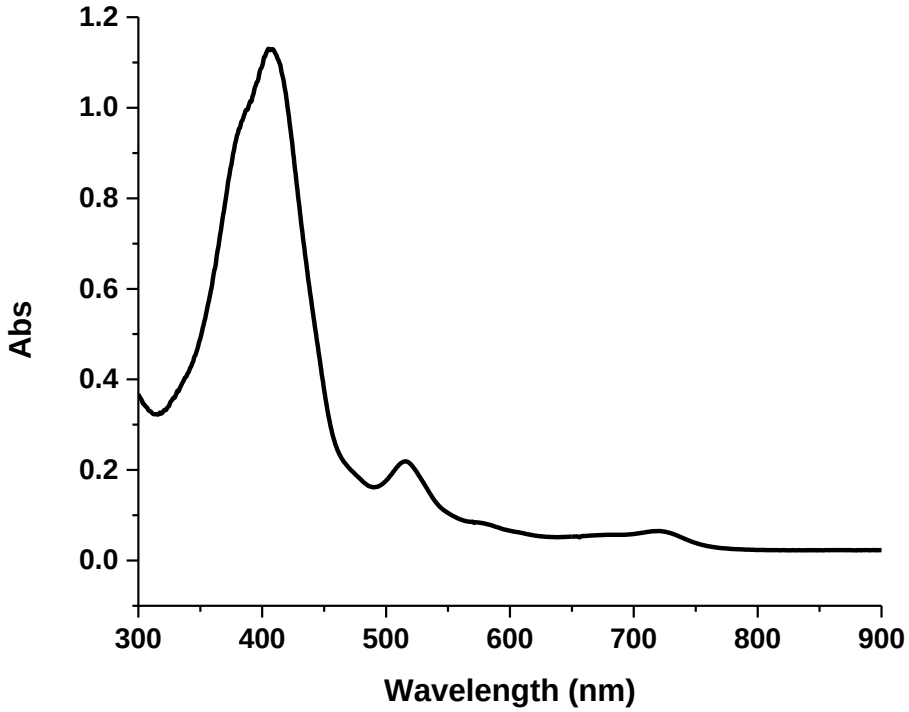


Figure S36. UV-Vis spectrum of [Fe(TPP)(**S**-CF_3)] in CH_2Cl_2.

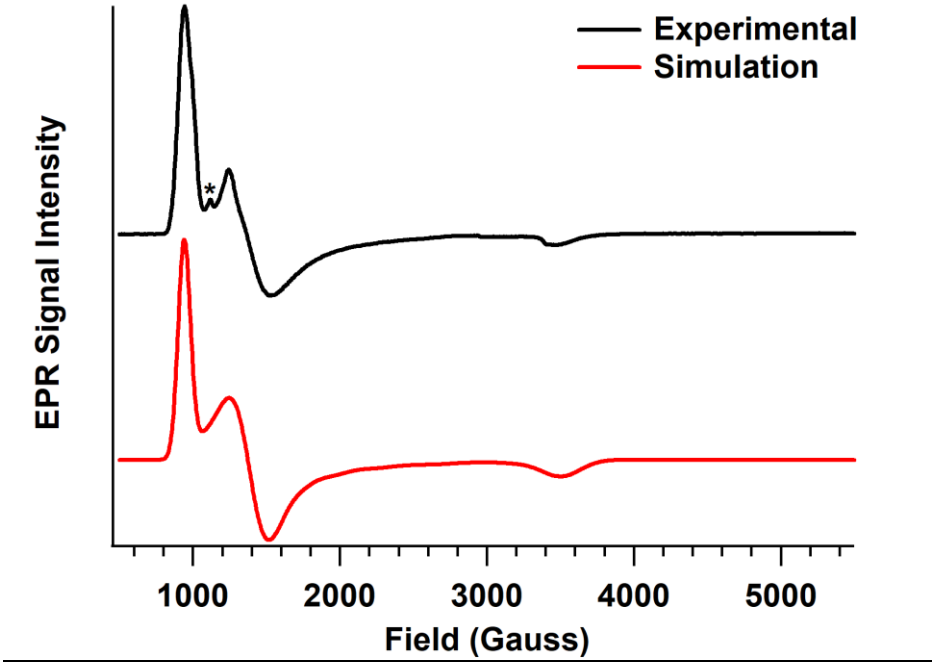


Figure S37. Experimental and simulated EPR spectrum of [Fe(TPP)(**S**-CF_3)] in 1:1 CH_2Cl_2:toluene, collected at 4 K. The asterisk indicates a small impurity of [Fe(TPP)Cl] (see Figure S26).

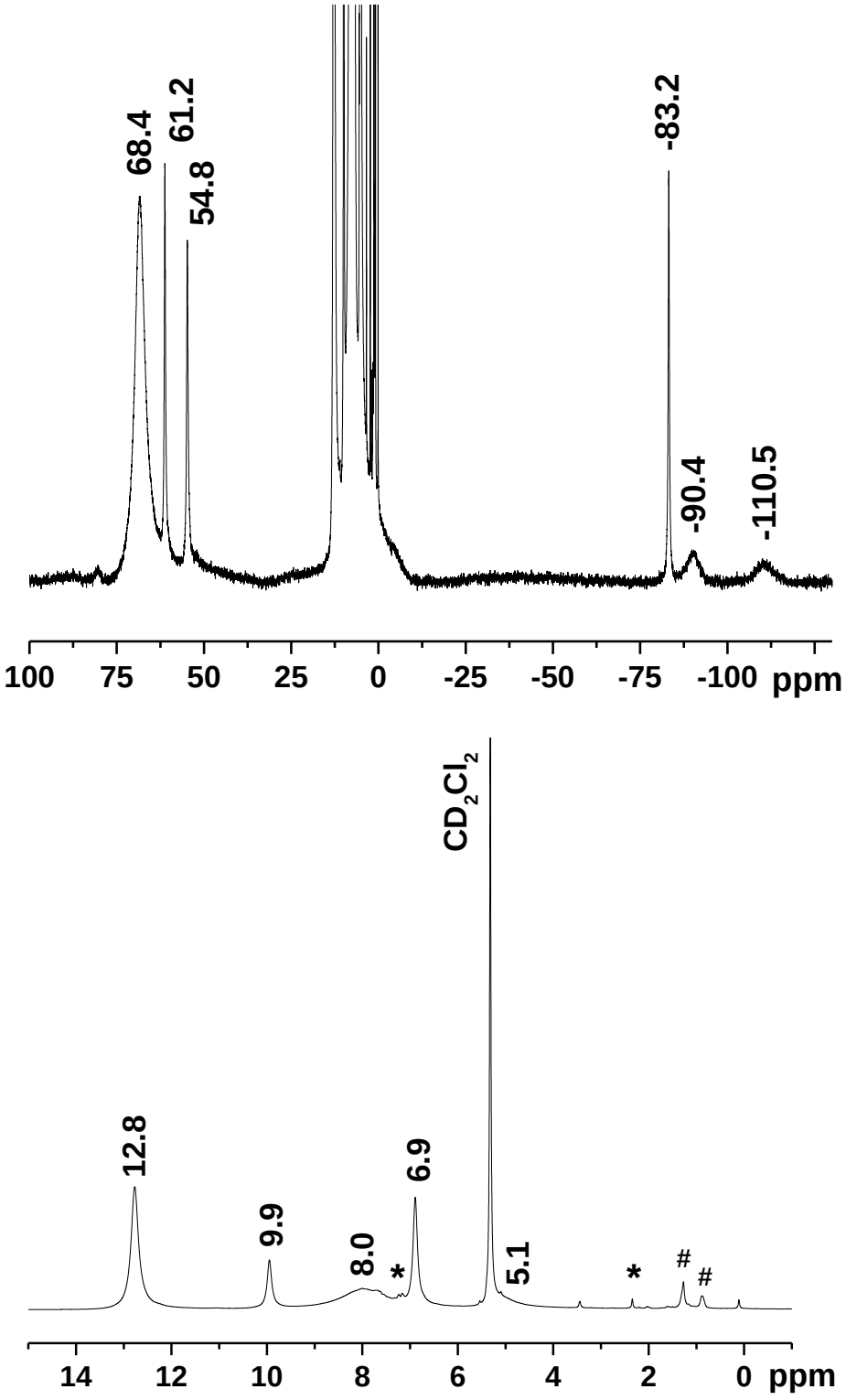


Figure S38. ^{1}H-NMR spectra of [Fe(TPP)(**S**-CF_3)] in CD_2Cl_2. The asterisk (*) and pound (#) symbols denote residual toluene (2.43, 7.15, 7.24 ppm) and hexanes (0.89 and 1.27 ppm), respectively, in the solid from recrystallization of the compound.

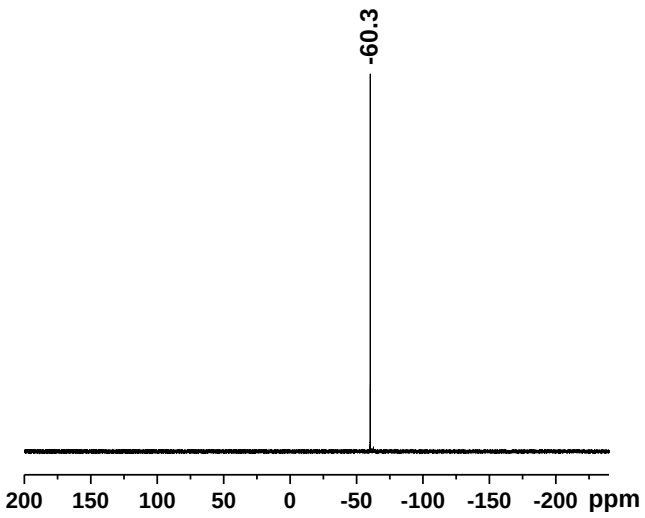


Figure S39. ^{19}F-NMR spectrum of [Fe(TPP)(**S**-CF_3)] in CD_2Cl_2. (Note: the ^{19}F signal of the free thiol, HSPhL$^{-CF_3}$, is observed at -62.47 ppm in CD_2Cl_2).

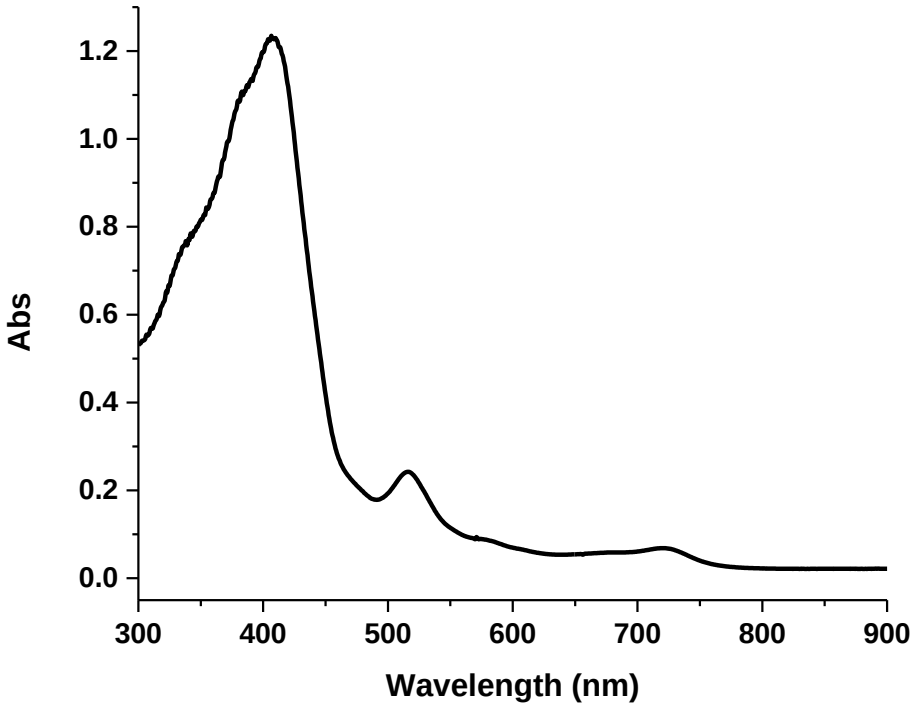


Figure S40. UV-Vis spectrum of [Fe(TPP)(**S**-NO_2)] in CH_2Cl_2.

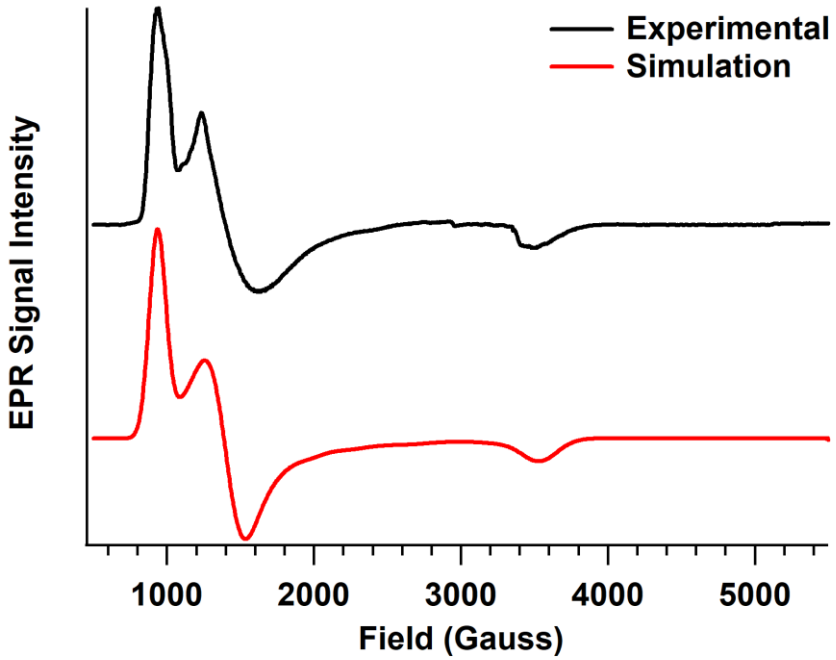


Figure S41. Experimental and simulated EPR spectrum of [Fe(TPP)(**S**-NO_2)] in 1:1 CH_2Cl_2:toluene, collected at 8 K.

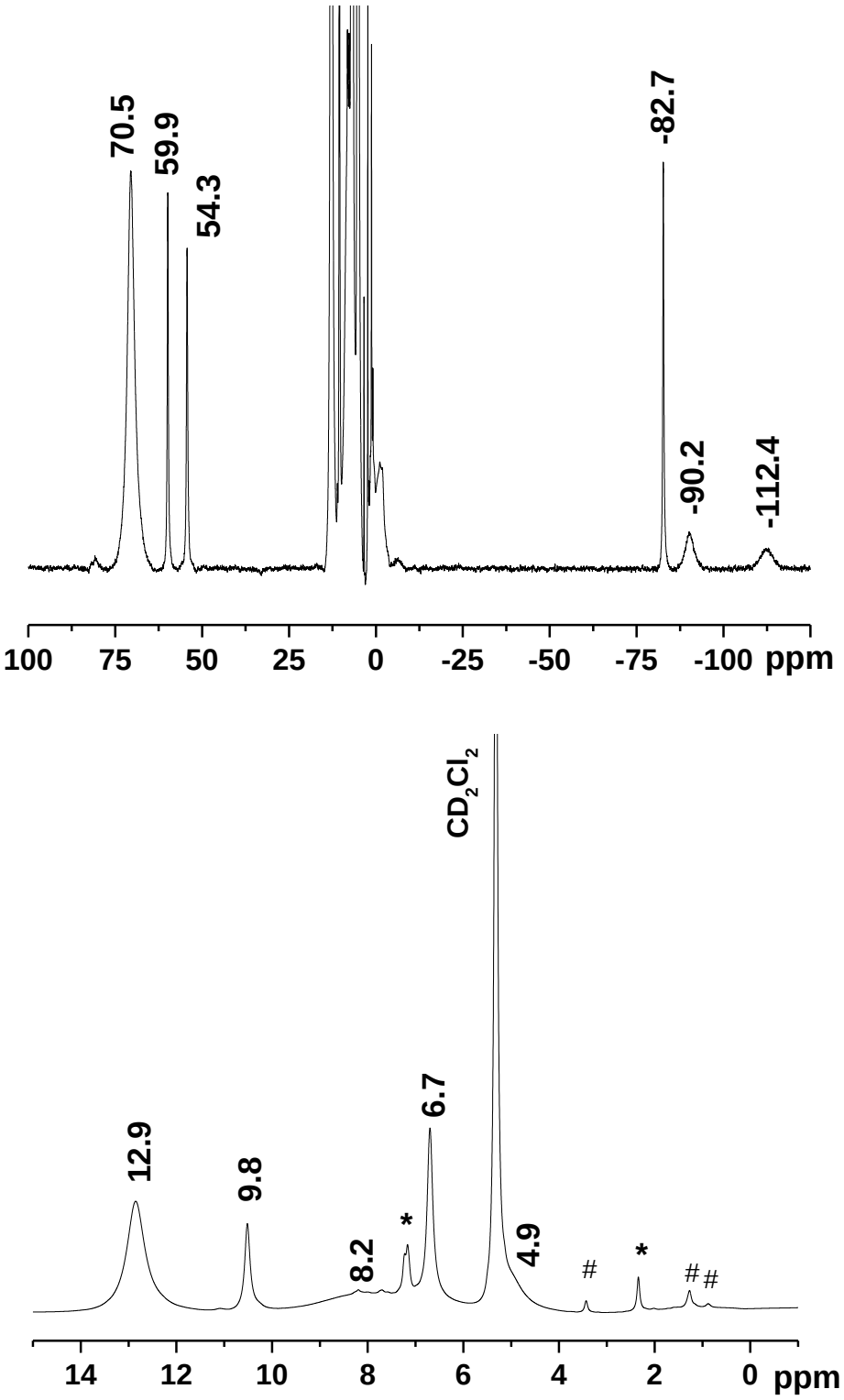


Figure S42. ^{1}H-NMR spectra of [Fe(TPP)(**S**-NO_2)] in CD_2Cl_2. The asterisk (*) and pound (#) symbols denote residual toluene (2.43, 7.15, 7.24 ppm), methanol (1.09 and 3.42 ppm) and hexanes (0.89 and 1.27 ppm), respectively, in the solid from recrystallization of the compound.

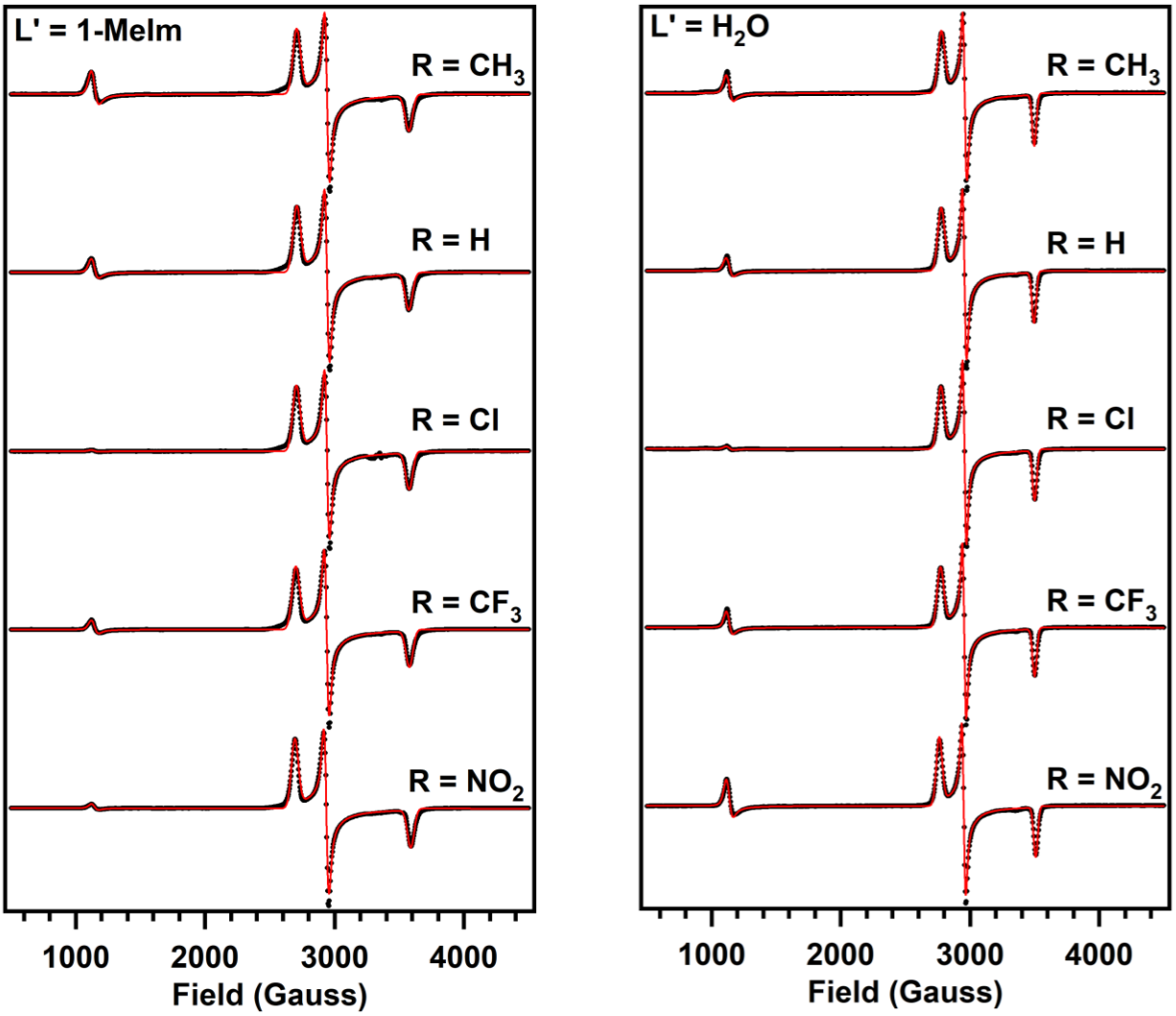


Figure S43. Experimental (black dots) and simulated (red lines) EPR spectra of six-coordinate [Fe(TPP)(**S**-R)(1-MeIm)] (left) and [Fe(TPP)(**S**-R)(H_2O)] (right) complexes recorded at 10 K. All spectra are normalized to maximum peak-to-peak height. The small signal observed at 1050 G in each spectrum is attributed to a [Fe(TPP)Cl] impurity. In all cases, this signal makes up less than one percent of the total signal intensity, as determined when fitting the relative ratios of the two species in each simulation.

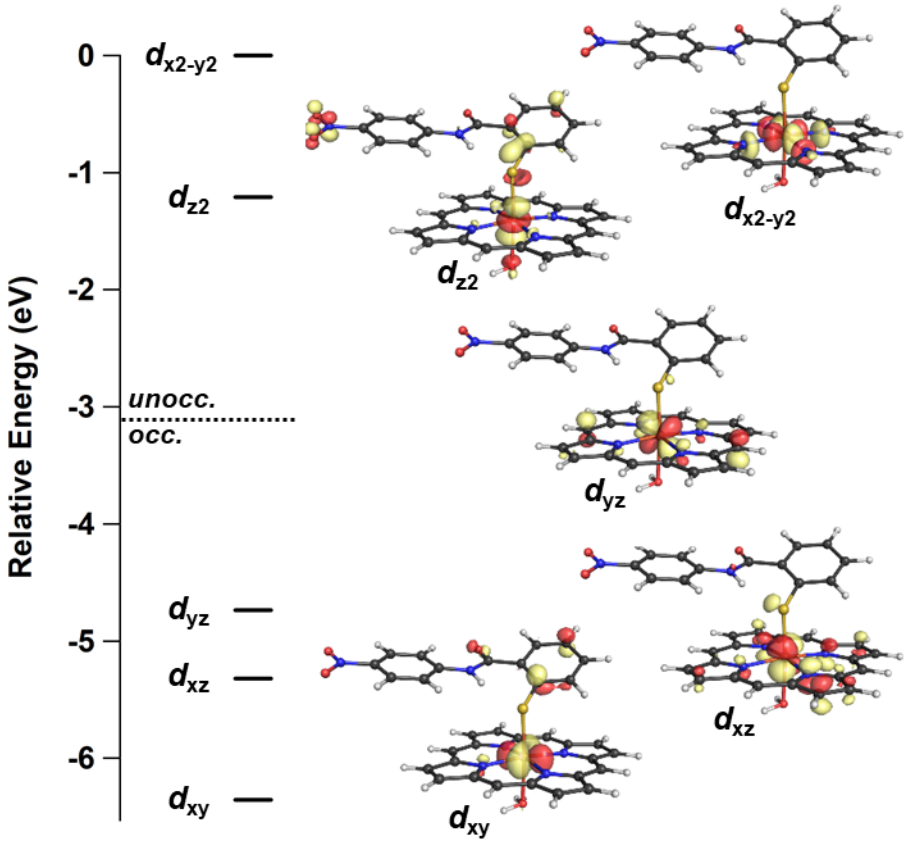


Figure S44. MO diagram for the Fe 3*d*-based MOs in six-coordinate, low-spin [Fe(P)(**S**-NO_2)(H_2O)] complexes. Spin up orbital energies are relative to the energy of the $3d_{x2-y2}$ orbital. Contour plots depicting spin-down single-electron orbitals were generated with an isosurface value of 0.05 and overlaid with the geometry-optimized structure of each complex using Pymol (v1.3). Orbital plots were generated with an isosurface value of 0.05, and structures were visualized using Pymol (v1.3).

Table S1. Fractional Atomic Coordinates ($\times 10^4$) and Equivalent Isotropic Displacement Parameters ($Å^2 \times 10^3$) for [Fe(TPP)(**S**-NO_2)]. U_{eq} is defined as 1/3 of of the trace of the orthogonalized U_{IJ} tensor.

Atom	***x***	***y***	***z***	**U(eq)**
Fe1	3606.9(2)	3075.6(2)	2923.2(2)	13.66(7)
S1	3531.0(3)	1281.2(4)	2845.6(2)	16.86(11)
O1	3761.5(10)	1126.1(13)	967.0(6)	27.8(3)
O2	8625.4(11)	667.8(15)	823.5(7)	42.5(4)
O3	8919.3(11)	964.9(14)	1682.0(8)	40.2(4)
N1	2218.8(11)	3384.5(12)	3121.4(6)	15.8(3)
N2	3945.7(11)	3114.4(12)	3746.3(6)	15.8(3)
N3	5017.5(11)	3395.3(12)	2775.8(6)	14.8(3)
N4	3301.6(11)	3695.7(12)	2161.4(6)	15.4(3)
N5	4513.1(12)	866.7(14)	1783.1(7)	19.3(4)
N6	8368.8(13)	806.2(15)	1297.3(9)	30.4(4)
C1	1824.4(14)	3186.7(15)	3623.8(8)	17.1(4)
C2	2328.5(14)	2991.4(15)	4110.2(8)	17.7(4)
C3	1757(3)	2646(4)	4591.2(14)	19.9(8)
C4	1226(2)	3347(4)	4894.2(12)	27.9(7)
C5	628(2)	2993(4)	5298.9(12)	38.3(12)
C6	558(2)	1937(3)	5400.1(12)	40.4(12)
C7	1086(2)	1232(3)	5105.1(12)	36.1(10)
C8	1691(2)	1588(3)	4706.1(12)	24.0(7)
C3A	1690(20)	2916(14)	4598(11)	17(8)
C4A	1239(14)	3774(15)	4818(7)	11(4)
C5A	601(11)	3649(15)	5237(6)	25(5)
C6A	464(14)	2659(17)	5448(8)	27(6)
C7A	895(12)	1789(15)	5225(7)	16(5)
C8A	1555(14)	1936(15)	4819(8)	26(6)
C9	3312.5(14)	3014.4(15)	4169.4(7)	16.9(4)
C10	3820.6(14)	2938.7(16)	4682.6(8)	20.2(4)
C11	4754.9(14)	2958.1(16)	4570.5(8)	20.2(4)
C12	4835.7(14)	3054.3(15)	3985.3(7)	17.0(4)
C13	5687.4(13)	3079.5(15)	3703.8(7)	16.6(4)
C14	6596.6(14)	2886.3(16)	4013.5(7)	18.4(4)
C15	7013.7(15)	3658.6(17)	4338.1(8)	23.7(5)
C16	7883.1(16)	3475(2)	4595.9(9)	31.1(5)
C17	8329.4(16)	2522(2)	4540.6(9)	32.8(6)
C18	7911.9(16)	1748(2)	4228.1(9)	29.9(5)
C19	7051.8(15)	1929.8(17)	3963.1(8)	23.4(4)
C20	5763.0(13)	3267.6(15)	3143.4(8)	16.1(4)

C21	6644.0(14)	3437.3(15)	2871.0(8)	17.7(4)
C22	6435.0(14)	3674.1(15)	2343.7(8)	17.4(4)
C23	5422.2(13)	3631.9(14)	2279.2(7)	15.1(4)
C24	4921.5(14)	3810.7(14)	1788.4(7)	15.3(4)
C25	5479.0(13)	4040.4(15)	1286.0(7)	16.2(4)
C26	6126.4(14)	3317.8(16)	1088.8(8)	20.8(4)
C27	6621.9(15)	3505.8(17)	612.8(8)	24.8(5)
C28	6484.4(15)	4435.0(17)	331.3(8)	24.7(5)
C29	5853.0(16)	5167.6(17)	526.0(8)	25.7(5)
C30	5357.2(15)	4972.6(16)	999.9(8)	20.7(4)
C31	3933.2(14)	3799.7(15)	1737.4(7)	16.3(4)
C32	3414.8(14)	3908.9(16)	1228.0(8)	20.4(4)
C33	2485.3(14)	3909.4(16)	1347.5(8)	19.7(4)
C34	2405.0(14)	3792.6(15)	1927.2(8)	16.2(4)
C35	1546.6(14)	3779.2(15)	2207.0(8)	16.9(4)
C36	655.3(14)	4020.4(16)	1889.0(8)	19.1(4)
C37	542.6(15)	4990.7(17)	1640.4(8)	22.4(4)
C38	-283.0(16)	5246.7(18)	1354.5(8)	27.0(5)
C39	-1021.2(16)	4532.8(18)	1325.5(8)	27.9(5)
C40	-926.2(15)	3567.8(18)	1574.3(9)	26.0(5)
C41	-89.7(14)	3307.9(17)	1850.4(8)	23.0(4)
C42	1470.0(14)	3575.6(15)	2766.6(8)	17.2(4)
C43	594.9(14)	3520.9(16)	3057.5(8)	19.8(4)
C44	809.5(14)	3260.0(16)	3580.8(8)	19.9(4)
C45	2648.2(14)	1097.7(14)	2324.2(8)	17.2(4)
C46	1725.7(14)	998.2(16)	2516.9(8)	20.5(4)
C47	974.5(14)	780.5(16)	2167.6(9)	22.9(4)
C48	1117.2(15)	694.7(17)	1610.8(9)	24.5(5)
C49	2018.4(14)	822.7(16)	1411.0(8)	22.0(4)
C50	2799.0(14)	1003.5(15)	1755.6(8)	17.5(4)
C51	3727.4(14)	1024.5(15)	1463.2(8)	19.1(4)
C52	5456.3(14)	819.3(15)	1630.9(8)	18.6(4)
C53	5749.5(14)	656.5(16)	1093.6(8)	21.0(4)
C54	6709.9(14)	641.4(16)	985.1(9)	23.3(4)
C55	7355.8(14)	787.2(16)	1410.4(9)	23.4(5)
C56	7079.9(15)	918.9(17)	1945.8(9)	24.4(5)
C57	6123.1(15)	928.5(16)	2055.6(9)	22.3(4)

Table S2. Fractional Atomic Coordinates ($\times 10^4$) and Equivalent Isotropic Displacement Parameters ($Å^2 \times 10^3$) for [Fe(TPP)(**S**-NO_2)]. U_{eq} is defined as 1/3 of of the trace of the orthogonalized U_{IJ} tensor.

Atom	***x***	***y***	***z***	**U(eq)**
Fe1	3606.9(2)	3075.6(2)	2923.2(2)	13.66(7)
S1	3531.0(3)	1281.2(4)	2845.6(2)	16.86(11)
O1	3761.5(10)	1126.1(13)	967.0(6)	27.8(3)
O2	8625.4(11)	667.8(15)	823.5(7)	42.5(4)
O3	8919.3(11)	964.9(14)	1682.0(8)	40.2(4)
N1	2218.8(11)	3384.5(12)	3121.4(6)	15.8(3)
N2	3945.7(11)	3114.4(12)	3746.3(6)	15.8(3)
N3	5017.5(11)	3395.3(12)	2775.8(6)	14.8(3)
N4	3301.6(11)	3695.7(12)	2161.4(6)	15.4(3)
N5	4513.1(12)	866.7(14)	1783.1(7)	19.3(4)
N6	8368.8(13)	806.2(15)	1297.3(9)	30.4(4)
C1	1824.4(14)	3186.7(15)	3623.8(8)	17.1(4)
C2	2328.5(14)	2991.4(15)	4110.2(8)	17.7(4)
C3	1757(3)	2646(4)	4591.2(14)	19.9(8)
C4	1226(2)	3347(4)	4894.2(12)	27.9(7)
C5	628(2)	2993(4)	5298.9(12)	38.3(12)
C6	558(2)	1937(3)	5400.1(12)	40.4(12)
C7	1086(2)	1232(3)	5105.1(12)	36.1(10)
C8	1691(2)	1588(3)	4706.1(12)	24.0(7)
C3A	1690(20)	2916(14)	4598(11)	17(8)
C4A	1239(14)	3774(15)	4818(7)	11(4)
C5A	601(11)	3649(15)	5237(6)	25(5)
C6A	464(14)	2659(17)	5448(8)	27(6)
C7A	895(12)	1789(15)	5225(7)	16(5)
C8A	1555(14)	1936(15)	4819(8)	26(6)
C9	3312.5(14)	3014.4(15)	4169.4(7)	16.9(4)
C10	3820.6(14)	2938.7(16)	4682.6(8)	20.2(4)
C11	4754.9(14)	2958.1(16)	4570.5(8)	20.2(4)
C12	4835.7(14)	3054.3(15)	3985.3(7)	17.0(4)
C13	5687.4(13)	3079.5(15)	3703.8(7)	16.6(4)
C14	6596.6(14)	2886.3(16)	4013.5(7)	18.4(4)
C15	7013.7(15)	3658.6(17)	4338.1(8)	23.7(5)
C16	7883.1(16)	3475(2)	4595.9(9)	31.1(5)
C17	8329.4(16)	2522(2)	4540.6(9)	32.8(6)
C18	7911.9(16)	1748(2)	4228.1(9)	29.9(5)
C19	7051.8(15)	1929.8(17)	3963.1(8)	23.4(4)
C20	5763.0(13)	3267.6(15)	3143.4(8)	16.1(4)

C21	6644.0(14)	3437.3(15)	2871.0(8)	17.7(4)
C22	6435.0(14)	3674.1(15)	2343.7(8)	17.4(4)
C23	5422.2(13)	3631.9(14)	2279.2(7)	15.1(4)
C24	4921.5(14)	3810.7(14)	1788.4(7)	15.3(4)
C25	5479.0(13)	4040.4(15)	1286.0(7)	16.2(4)
C26	6126.4(14)	3317.8(16)	1088.8(8)	20.8(4)
C27	6621.9(15)	3505.8(17)	612.8(8)	24.8(5)
C28	6484.4(15)	4435.0(17)	331.3(8)	24.7(5)
C29	5853.0(16)	5167.6(17)	526.0(8)	25.7(5)
C30	5357.2(15)	4972.6(16)	999.9(8)	20.7(4)
C31	3933.2(14)	3799.7(15)	1737.4(7)	16.3(4)
C32	3414.8(14)	3908.9(16)	1228.0(8)	20.4(4)
C33	2485.3(14)	3909.4(16)	1347.5(8)	19.7(4)
C34	2405.0(14)	3792.6(15)	1927.2(8)	16.2(4)
C35	1546.6(14)	3779.2(15)	2207.0(8)	16.9(4)
C36	655.3(14)	4020.4(16)	1889.0(8)	19.1(4)
C37	542.6(15)	4990.7(17)	1640.4(8)	22.4(4)
C38	-283.0(16)	5246.7(18)	1354.5(8)	27.0(5)
C39	-1021.2(16)	4532.8(18)	1325.5(8)	27.9(5)
C40	-926.2(15)	3567.8(18)	1574.3(9)	26.0(5)
C41	-89.7(14)	3307.9(17)	1850.4(8)	23.0(4)
C42	1470.0(14)	3575.6(15)	2766.6(8)	17.2(4)
C43	594.9(14)	3520.9(16)	3057.5(8)	19.8(4)
C44	809.5(14)	3260.0(16)	3580.8(8)	19.9(4)
C45	2648.2(14)	1097.7(14)	2324.2(8)	17.2(4)
C46	1725.7(14)	998.2(16)	2516.9(8)	20.5(4)
C47	974.5(14)	780.5(16)	2167.6(9)	22.9(4)
C48	1117.2(15)	694.7(17)	1610.8(9)	24.5(5)
C49	2018.4(14)	822.7(16)	1411.0(8)	22.0(4)
C50	2799.0(14)	1003.5(15)	1755.6(8)	17.5(4)
C51	3727.4(14)	1024.5(15)	1463.2(8)	19.1(4)
C52	5456.3(14)	819.3(15)	1630.9(8)	18.6(4)
C53	5749.5(14)	656.5(16)	1093.6(8)	21.0(4)
C54	6709.9(14)	641.4(16)	985.1(9)	23.3(4)
C55	7355.8(14)	787.2(16)	1410.4(9)	23.4(5)
C56	7079.9(15)	918.9(17)	1945.8(9)	24.4(5)
C57	6123.1(15)	928.5(16)	2055.6(9)	22.3(4)

Table S3. Bond Lengths for [Fe(TPP)(**S**-NO_2)].

Atom	Atom	Length/Å	Atom	Atom	Length/Å
Fe1	S1	2.3078(8)	C14	C19	1.390(3)
Fe1	N1	2.0653(17)	C15	C16	1.391(3)
Fe1	N2	2.0593(16)	C16	C17	1.380(4)
Fe1	N3	2.0711(17)	C17	C18	1.378(3)
Fe1	N4	2.0617(16)	C18	C19	1.387(3)
S1	C45	1.783(2)	C20	C21	1.437(3)
O1	C51	1.220(2)	C21	C22	1.351(3)
O2	N6	1.231(3)	C22	C23	1.438(3)
O3	N6	1.227(3)	C23	C24	1.401(3)
N1	C1	1.379(2)	C24	C25	1.497(3)
N1	C42	1.378(2)	C24	C31	1.400(3)
N2	C9	1.383(2)	C25	C26	1.392(3)
N2	C12	1.380(2)	C25	C30	1.392(3)
N3	C20	1.382(2)	C26	C27	1.387(3)
N3	C23	1.382(2)	C27	C28	1.386(3)
N4	C31	1.384(2)	C28	C29	1.384(3)
N4	C34	1.387(2)	C29	C30	1.385(3)
N5	C51	1.362(3)	C31	C32	1.440(3)
N5	C52	1.390(3)	C32	C33	1.349(3)
N6	C55	1.462(3)	C33	C34	1.430(3)
C1	C2	1.398(3)	C34	C35	1.401(3)
C1	C44	1.439(3)	C35	C36	1.501(3)
C2	C3	1.503(3)	C35	C42	1.397(3)
C2	C3A	1.505(15)	C36	C37	1.391(3)
C2	C9	1.395(3)	C36	C41	1.395(3)
C3	C4	1.390(5)	C37	C38	1.389(3)
C3	C8	1.387(5)	C38	C39	1.388(3)
C4	C5	1.386(5)	C39	C40	1.383(3)
C5	C6	1.378(5)	C40	C41	1.391(3)
C6	C7	1.381(5)	C42	C43	1.437(3)
C7	C8	1.383(4)	C43	C44	1.351(3)
C3A	C4A	1.386(10)	C45	C46	1.398(3)
C3A	C8A	1.382(10)	C45	C50	1.413(3)
C4A	C5A	1.383(10)	C46	C47	1.379(3)
C5A	C6A	1.384(11)	C47	C48	1.382(3)
C6A	C7A	1.385(11)	C48	C49	1.380(3)
C7A	C8A	1.384(10)	C49	C50	1.396(3)
C9	C10	1.437(3)	C50	C51	1.503(3)

C10	C11	1.352(3)	C52	C53	1.398(3)
C11	C12	1.441(3)	C52	C57	1.397(3)
C12	C13	1.395(3)	C53	C54	1.386(3)
C13	C14	1.500(3)	C54	C55	1.384(3)
C13	C20	1.395(3)	C55	C56	1.381(3)
C14	C15	1.392(3)	C56	C57	1.382(3)

Table S4. Bond Angles for [Fe(TPP)(**S**-NO_2)].

Atom	Atom	Atom	Angle/°	Atom	Atom	Atom	Angle/°
N1	Fe1	S1	99.61(5)	C17	C18	C19	120.1(2)
N1	Fe1	N3	157.27(6)	C18	C19	C14	120.6(2)
N2	Fe1	S1	96.54(5)	N3	C20	C13	126.00(18)
N2	Fe1	N1	88.47(6)	N3	C20	C21	109.74(16)
N2	Fe1	N3	87.46(6)	C13	C20	C21	124.10(18)
N2	Fe1	N4	155.96(7)	C22	C21	C20	107.31(17)
N3	Fe1	S1	103.07(5)	C21	C22	C23	107.32(17)
N4	Fe1	S1	107.49(5)	N3	C23	C22	109.69(16)
N4	Fe1	N1	87.25(6)	N3	C23	C24	125.19(17)
N4	Fe1	N3	87.42(6)	C24	C23	C22	125.12(17)
C45	S1	Fe1	102.76(6)	C23	C24	C25	117.92(17)
C1	N1	Fe1	124.89(13)	C31	C24	C23	124.30(17)
C42	N1	Fe1	127.48(13)	C31	C24	C25	117.77(16)
C42	N1	C1	105.96(16)	C26	C25	C24	120.65(17)
C9	N2	Fe1	125.73(13)	C26	C25	C30	118.21(18)
C12	N2	Fe1	127.41(12)	C30	C25	C24	121.14(18)
C12	N2	C9	105.93(15)	C27	C26	C25	121.16(19)
C20	N3	Fe1	126.06(13)	C28	C27	C26	119.7(2)
C20	N3	C23	105.91(16)	C29	C28	C27	119.84(19)
C23	N3	Fe1	127.46(12)	C28	C29	C30	120.1(2)
C31	N4	Fe1	125.74(13)	C29	C30	C25	120.92(19)
C31	N4	C34	106.14(15)	N4	C31	C24	126.04(17)
C34	N4	Fe1	125.67(12)	N4	C31	C32	109.31(17)
C51	N5	C52	128.95(18)	C24	C31	C32	124.64(17)
O2	N6	C55	118.63(19)	C33	C32	C31	107.23(17)
O3	N6	O2	123.44(19)	C32	C33	C34	107.87(17)
O3	N6	C55	117.9(2)	N4	C34	C33	109.35(16)
N1	C1	C2	125.54(18)	N4	C34	C35	126.10(17)
N1	C1	C44	109.84(16)	C35	C34	C33	124.54(18)
C2	C1	C44	124.56(17)	C34	C35	C36	118.06(17)
C1	C2	C3	116.4(2)	C42	C35	C34	124.08(18)
C1	C2	C3A	112.4(14)	C42	C35	C36	117.84(17)
C9	C2	C1	125.27(17)	C37	C36	C35	119.93(18)
C9	C2	C3	118.2(2)	C37	C36	C41	118.28(19)
C9	C2	C3A	121.6(15)	C41	C36	C35	121.75(18)
C4	C3	C2	121.8(4)	C38	C37	C36	121.3(2)
C8	C3	C2	119.0(3)	C39	C38	C37	119.6(2)
C8	C3	C4	119.0(3)	C40	C39	C38	119.9(2)

C5	C4	C3	120.6(4)	C39	C40	C41	120.2(2)
C6	C5	C4	119.6(3)	C40	C41	C36	120.7(2)
C5	C6	C7	120.5(3)	N1	C42	C35	125.26(18)
C6	C7	C8	119.8(3)	N1	C42	C43	109.76(16)
C7	C8	C3	120.5(3)	C35	C42	C43	124.98(18)
C4A	C3A	C2	122.9(12)	C44	C43	C42	107.33(17)
C8A	C3A	C2	117.2(11)	C43	C44	C1	107.07(17)
C8A	C3A	C4A	119.9(12)	C46	C45	S1	114.61(15)
C5A	C4A	C3A	120.5(12)	C46	C45	C50	118.60(18)
C4A	C5A	C6A	118.5(11)	C50	C45	S1	126.73(15)
C5A	C6A	C7A	121.8(11)	C47	C46	C45	121.48(19)
C8A	C7A	C6A	118.5(11)	C46	C47	C48	120.10(19)
C3A	C8A	C7A	120.3(11)	C49	C48	C47	119.23(19)
N2	C9	C2	125.41(17)	C48	C49	C50	122.05(19)
N2	C9	C10	109.73(17)	C45	C50	C51	127.69(18)
C2	C9	C10	124.86(17)	C49	C50	C45	118.46(18)
C11	C10	C9	107.32(17)	C49	C50	C51	113.78(17)
C10	C11	C12	107.18(17)	O1	C51	N5	122.72(19)
N2	C12	C11	109.77(17)	O1	C51	C50	121.47(18)
N2	C12	C13	125.26(17)	N5	C51	C50	115.71(17)
C13	C12	C11	124.97(18)	N5	C52	C53	123.86(18)
C12	C13	C14	119.06(16)	N5	C52	C57	115.77(18)
C20	C13	C12	124.56(18)	C57	C52	C53	120.36(19)
C20	C13	C14	116.38(17)	C54	C53	C52	119.13(19)
C15	C14	C13	121.38(19)	C55	C54	C53	119.4(2)
C19	C14	C13	119.62(18)	C54	C55	N6	119.7(2)
C19	C14	C15	118.96(19)	C56	C55	N6	117.96(19)
C16	C15	C14	120.0(2)	C56	C55	C54	122.30(19)
C17	C16	C15	120.4(2)	C55	C56	C57	118.5(2)
C18	C17	C16	119.8(2)	C56	C57	C52	120.3(2)

Table S5. Torsion Angles for [Fe(TPP)(**S**-NO_2)].

A	B	C	D	Angle/°	A	B	C	D	Angle/°
Fe1	S1	C45	C46	-90.87(14)	C14	C13	C20	N3	175.47(17)
Fe1	S1	C45	C50	91.91(17)	C14	C13	C20	C21	-9.6(3)
Fe1	N1	C1	C2	16.4(3)	C14	C15	C16	C17	-1.2(3)
Fe1	N1	C1	C44	-166.35(13)	C15	C14	C19	C18	-0.5(3)
Fe1	N1	C42	C35	-12.1(3)	C15	C16	C17	C18	0.0(3)
Fe1	N1	C42	C43	167.12(13)	C16	C17	C18	C19	0.9(3)
Fe1	N2	C9	C2	-7.8(3)	C17	C18	C19	C14	-0.7(3)
Fe1	N2	C9	C10	172.24(13)	C19	C14	C15	C16	1.4(3)
Fe1	N2	C12	C11	-171.74(13)	C20	N3	C23	C22	-1.3(2)
Fe1	N2	C12	C13	8.1(3)	C20	N3	C23	C24	179.11(18)
Fe1	N3	C20	C13	-12.1(3)	C20	C13	C14	C15	102.5(2)
Fe1	N3	C20	C21	172.35(12)	C20	C13	C14	C19	-75.3(2)
Fe1	N3	C23	C22	-172.97(12)	C20	C21	C22	C23	-1.2(2)
Fe1	N3	C23	C24	7.4(3)	C21	C22	C23	N3	1.6(2)
Fe1	N4	C31	C24	-20.7(3)	C21	C22	C23	C24	-178.78(18)
Fe1	N4	C31	C32	159.96(13)	C22	C23	C24	C25	1.5(3)
Fe1	N4	C34	C33	-160.28(13)	C22	C23	C24	C31	-177.82(18)
Fe1	N4	C34	C35	19.7(3)	C23	N3	C20	C13	176.10(18)
S1	C45	C46	C47	-175.46(16)	C23	N3	C20	C21	0.5(2)
S1	C45	C50	C49	177.58(15)	C23	C24	C25	C26	60.3(2)
S1	C45	C50	C51	0.7(3)	C23	C24	C25	C30	-120.5(2)
O2	N6	C55	C54	-1.6(3)	C23	C24	C31	N4	5.4(3)
O2	N6	C55	C56	178.5(2)	C23	C24	C31	C32	-175.37(19)
O3	N6	C55	C54	178.09(19)	C24	C25	C26	C27	177.66(18)
O3	N6	C55	C56	-1.8(3)	C24	C25	C30	C29	-177.98(19)
N1	C1	C2	C3	-171.6(3)	C24	C31	C32	C33	-177.14(19)
N1	C1	C2	C3A	174.0(9)	C25	C24	C31	N4	-173.90(17)
N1	C1	C2	C9	3.2(3)	C25	C24	C31	C32	5.3(3)
N1	C1	C44	C43	-1.2(2)	C25	C26	C27	C28	1.0(3)
N1	C42	C43	C44	-2.2(2)	C26	C25	C30	C29	1.2(3)
N2	C9	C10	C11	-2.0(2)	C26	C27	C28	C29	-0.1(3)
N2	C12	C13	C14	-173.22(18)	C27	C28	C29	C30	-0.2(3)
N2	C12	C13	C20	6.3(3)	C28	C29	C30	C25	-0.4(3)
N3	C20	C21	C22	0.5(2)	C30	C25	C26	C27	-1.6(3)
N3	C23	C24	C25	-178.98(17)	C31	N4	C34	C33	2.7(2)
N3	C23	C24	C31	1.7(3)	C31	N4	C34	C35	-177.36(18)
N4	C31	C32	C33	2.2(2)	C31	C24	C25	C26	-120.4(2)
N4	C34	C35	C36	173.00(18)	C31	C24	C25	C30	58.8(2)

N4	C34	C35	C42	-5.7(3)	C31	C32	C33	C34	-0.5(2)
N5	C52	C53	C54	178.85(19)	C32	C33	C34	N4	-1.4(2)
N5	C52	C57	C56	-178.39(18)	C32	C33	C34	C35	178.68(19)
N6	C55	C56	C57	178.40(18)	C33	C34	C35	C36	-7.1(3)
C1	N1	C42	C35	-177.81(19)	C33	C34	C35	C42	174.28(19)
C1	N1	C42	C43	1.5(2)	C34	N4	C31	C24	176.34(18)
C1	C2	C3	C4	-77.2(4)	C34	N4	C31	C32	-3.0(2)
C1	C2	C3	C8	97.1(4)	C34	C35	C36	C37	-61.5(3)
C1	C2	C3A	C4A	-71(3)	C34	C35	C36	C41	120.9(2)
C1	C2	C3A	C8A	107(3)	C34	C35	C42	N1	1.5(3)
C1	C2	C9	N2	-7.9(3)	C34	C35	C42	C43	-177.63(19)
C1	C2	C9	C10	172.02(19)	C35	C36	C37	C38	-178.43(18)
C2	C1	C44	C43	176.10(19)	C35	C36	C41	C40	177.03(18)
C2	C3	C4	C5	173.2(3)	C35	C42	C43	C44	177.04(19)
C2	C3	C8	C7	-172.5(3)	C36	C35	C42	N1	-177.14(18)
C2	C3A	C4A	C5A	175(2)	C36	C35	C42	C43	3.7(3)
C2	C3A	C8A	C7A	-172(2)	C36	C37	C38	C39	1.5(3)
C2	C9	C10	C11	178.14(19)	C37	C36	C41	C40	-0.6(3)
C3	C2	C9	N2	166.8(3)	C37	C38	C39	C40	-0.9(3)
C3	C2	C9	C10	-13.3(4)	C38	C39	C40	C41	-0.5(3)
C3	C4	C5	C6	-0.2(5)	C39	C40	C41	C36	1.2(3)
C4	C3	C8	C7	1.9(5)	C41	C36	C37	C38	-0.8(3)
C4	C5	C6	C7	0.6(4)	C42	N1	C1	C2	-177.45(19)
C5	C6	C7	C8	0.2(4)	C42	N1	C1	C44	-0.2(2)
C6	C7	C8	C3	-1.5(4)	C42	C35	C36	C37	117.2(2)
C8	C3	C4	C5	-1.0(5)	C42	C35	C36	C41	-60.4(3)
C3A	C2	C9	N2	-177.9(8)	C42	C43	C44	C1	2.0(2)
C3A	C2	C9	C10	2.0(9)	C44	C1	C2	C3	11.6(3)
C3A	C4A	C5A	C6A	3(4)	C44	C1	C2	C3A	-2.8(9)
C4A	C3A	C8A	C7A	6(5)	C44	C1	C2	C9	-173.65(19)
C4A	C5A	C6A	C7A	-5(3)	C45	C46	C47	C48	-2.5(3)
C5A	C6A	C7A	C8A	7(3)	C45	C50	C51	O1	-166.5(2)
C6A	C7A	C8A	C3A	-8(4)	C45	C50	C51	N5	17.0(3)
C8A	C3A	C4A	C5A	-4(5)	C46	C45	C50	C49	0.5(3)
C9	N2	C12	C11	-2.4(2)	C46	C45	C50	C51	-176.43(19)
C9	N2	C12	C13	177.43(19)	C46	C47	C48	C49	0.4(3)
C9	C2	C3	C4	107.7(4)	C47	C48	C49	C50	2.1(3)
C9	C2	C3	C8	-78.1(4)	C48	C49	C50	C45	-2.5(3)
C9	C2	C3A	C4A	100(3)	C48	C49	C50	C51	174.78(19)
C9	C2	C3A	C8A	-81(3)	C49	C50	C51	O1	16.5(3)
C9	C10	C11	C12	0.4(2)	C49	C50	C51	N5	-160.03(18)

C10	C11	C12	N2	1.2(2)	C50	C45	C46	C47	2.0(3)
C10	C11	C12	C13	-178.59(19)	C51	N5	C52	C53	-17.6(3)
C11	C12	C13	C14	6.6(3)	C51	N5	C52	C57	163.4(2)
C11	C12	C13	C20	-173.89(19)	C52	N5	C51	O1	1.9(3)
C12	N2	C9	C2	-177.44(19)	C52	N5	C51	C50	178.40(18)
C12	N2	C9	C10	2.6(2)	C52	C53	C54	C55	0.0(3)
C12	C13	C14	C15	-77.9(2)	C53	C52	C57	C56	2.6(3)
C12	C13	C14	C19	104.2(2)	C53	C54	C55	N6	-178.04(18)
C12	C13	C20	N3	-4.1(3)	C53	C54	C55	C56	1.9(3)
C12	C13	C20	C21	170.88(19)	C54	C55	C56	C57	-1.5(3)
C13	C14	C15	C16	-176.45(19)	C55	C56	C57	C52	-0.7(3)
C13	C14	C19	C18	177.42(19)	C57	C52	C53	C54	-2.2(3)
C13	C20	C21	C22	-175.20(18)					

Table S6. Hydrogen Atom Coordinates (Å×104) and Isotropic Displacement Parameters (Å2×103) for [Fe(TPP)(**S**-NO_2)].

Atom	***x***	***y***	***z***	**U(eq)**
H5	4439(17)	883(19)	2125(10)	31(7)
H5A	268.91	3475.84	5505.45	46
H6	144.09	1693.02	5674.81	49
H7	1032.87	505	5175.78	43
H8	2064.6	1102.99	4509.29	29
H4	1240(20)	4080(20)	4805(13)	30(10)
H4A	1366.99	4452.98	4681.13	13
H5B	264.86	4230.94	5376.99	30
H6A	62.84	2572.76	5752.98	33
H7A	741.25	1106.61	5348.05	19
H8A	1913.61	1359.69	4691.66	31
H10	3550.18	2885.02	5035.19	24
H11	5262.91	2916.02	4829.29	24
H15	6704.53	4311.08	4383.59	28
H16	8171.59	4008	4811.45	37
H17	8923.13	2401.01	4717.53	39
H18	8213.68	1088.98	4194.12	36
H19	6771.25	1395.11	3745.04	28
H21	7259.42	3391.48	3031.91	21
H22	6875.73	3838.18	2066.67	21
H26	6230.61	2684.82	1283.5	25
H27	7053.68	2999.94	480.3	30
H28	6822.77	4568.47	5.59	30
H29	5759.39	5805.69	334.1	31
H30	4928.13	5481.98	1131.71	25
H32	3678.06	3969.4	873.85	24
H33	1974.87	3975.56	1091.79	24
H37	1040.21	5488.2	1666.62	27
H38	-342.09	5906.89	1179.81	32
H39	-1591.15	4706.96	1134.85	33
H40	-1433.28	3080.96	1556.65	31
H41	-25.29	2638.25	2014.25	28
H43	-20.62	3645.18	2909.12	24
H44	372.73	3146.42	3866.58	24
H46	1613.93	1082.52	2896.9	25
H47	357.97	689.45	2310.01	27
H48	601.08	549.29	1368.51	29

H49	2110.15	786.59	1026.87	26
H53	5296.65	557.34	806.43	25
H54	6922.46	532.17	622.48	28
H56	7537.13	1001.02	2232.1	29
H57	5917.95	1009.83	2421.73	27

Table S7. Atomic Occupancy for [Fe(TPP)(**S**-NO_2)].

Atom	***Occupancy***	**Atom**	***Occupancy***	**Atom**	***Occupancy***
C3	0.867(8)	C4	0.867(8)	C5	0.867(8)
H5A	0.867(8)	C6	0.867(8)	H6	0.867(8)
C7	0.867(8)	H7	0.867(8)	C8	0.867(8)
H8	0.867(8)	H4	0.867(8)	C3A	0.133(8)
C4A	0.133(8)	H4A	0.133(8)	C5A	0.133(8)
H5B	0.133(8)	C6A	0.133(8)	H6A	0.133(8)
C7A	0.133(8)	H7A	0.133(8)	C8A	0.133(8)
H8A	0.133(8)				

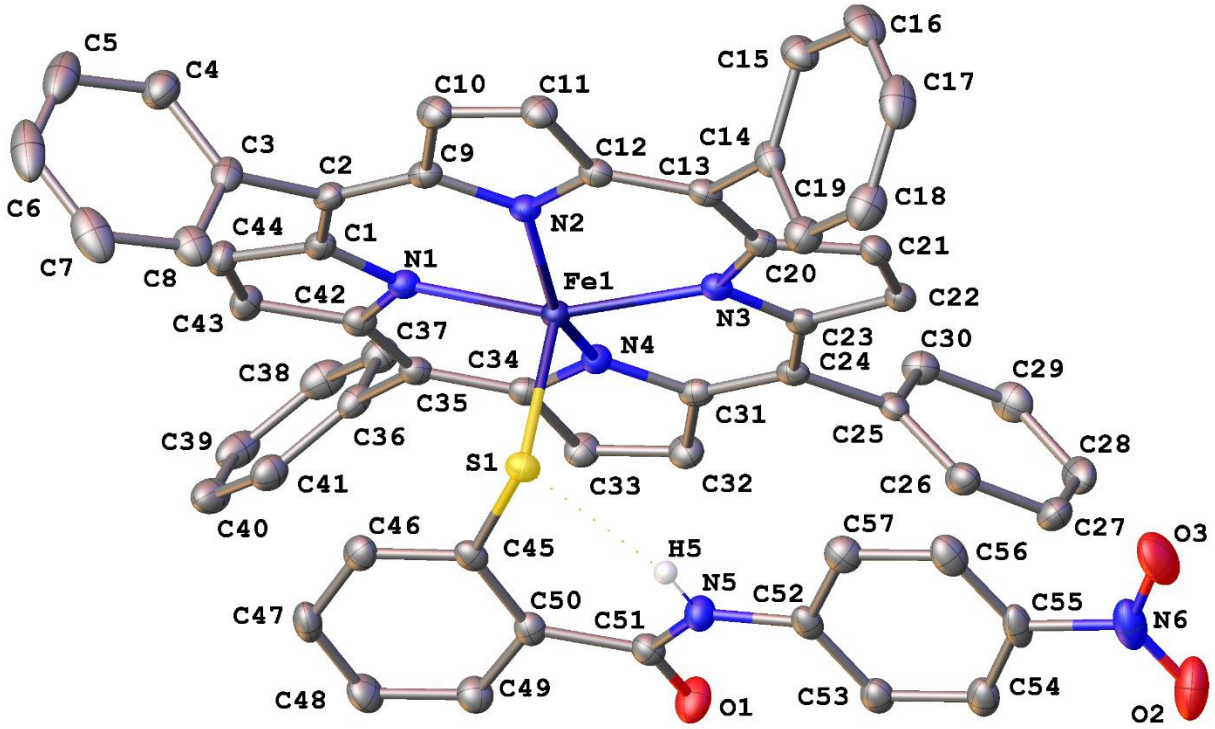


Figure S45. Molecular structure of the major component of [Fe(TPP)(**S**-NO_2)] shown with 50% probability ellipsoids and atom IDs. The amide H atom is displayed, and all other H atoms are omitted for clarity.

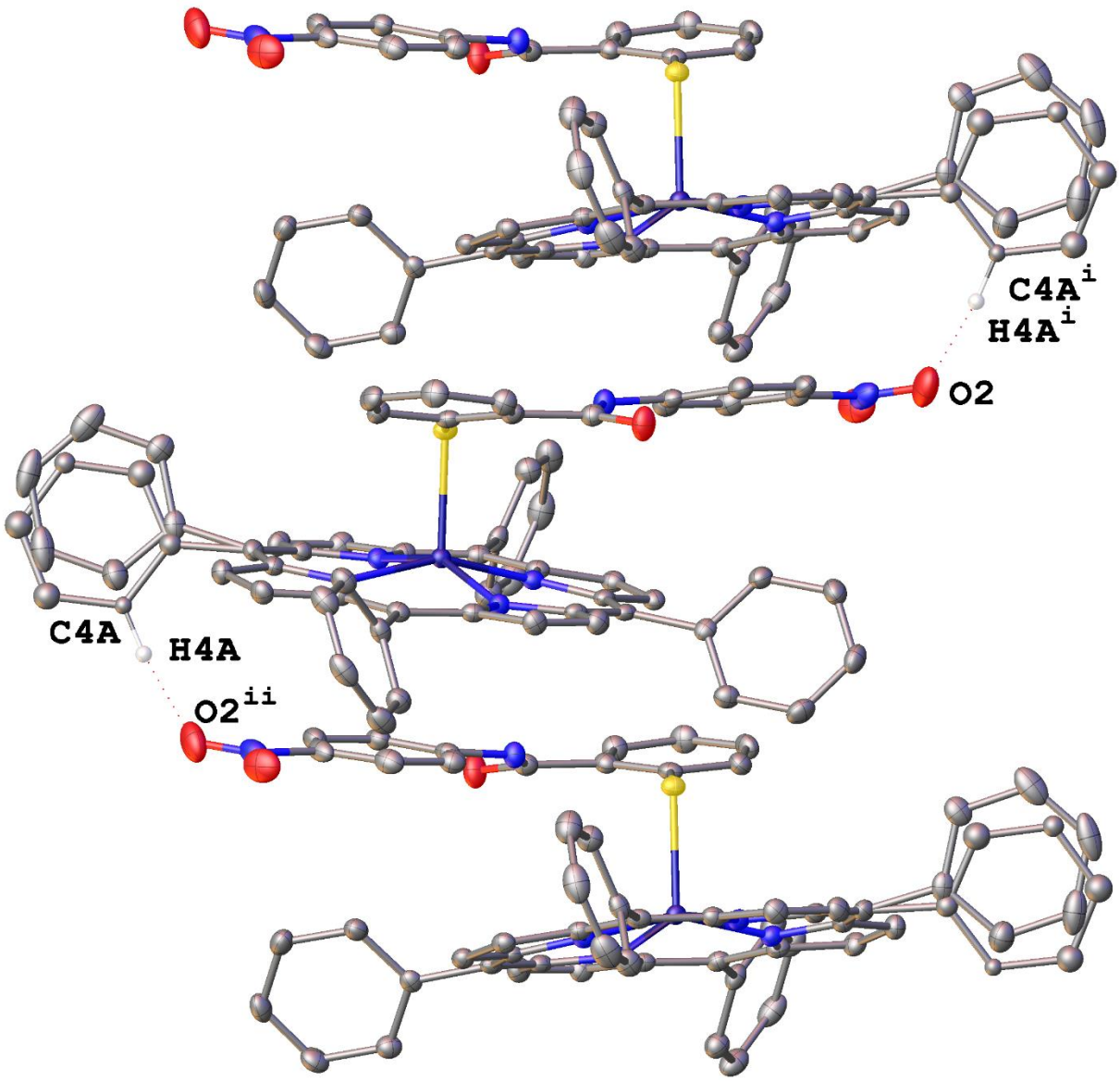


Figure S46. Crystallographic packing of [Fe(TPP)(**S**-NO_2)] shown with 50% probability ellipsoids. All H atoms, apart from those participating in the intermolecular H-bonding network, are omitted. This intermolecular H-bond between the phenyl ring of one porphyrin ligand and the *p*-NO_2 O atom of a different complex ($D_{D\text{-}A}$ = 2.895 Å) gives rise to the observed minor component. Symmetry codes: (i) 1-X,-1/2+Y,1/2-Z; (ii) 1-X,1/2+Y,1/2-Z.

5.7 Appendix 3: Computational Structural Coordinates

Table A3.1 Atomic coordinates for geometry-optimized for $[Fe(P)(His)_2]^+$ model.

	x	**y**	**z**
Fe	0	0	0
N	2.012	0	0
N	-0.033	2.006	0
N	0.037	-2.011	0
N	-2.005	-0.003	0
N	0.497	-0.478	-4.076
N	-0.006	0.005	-1.997
N	0.502	-0.485	4.073
N	-0.003	0.004	1.996
C	-0.64	-4.225	-0.006
C	0.725	-4.226	-0.005
C	0.644	4.22	0.002
C	-0.721	4.22	0.003
C	-4.207	-0.72	-0.002
C	-4.232	0.645	0
C	-2.86	1.085	0
C	-2.46	2.416	0.002
C	-1.136	2.84	0
C	1.065	2.84	0
C	2.392	2.432	0.002
C	2.826	1.113	0
C	4.214	0.717	0.002
C	4.238	-0.648	0.002
C	2.867	-1.089	0
C	2.465	-2.42	0
C	1.141	-2.846	-0.002
C	-1.061	-2.847	-0.004
C	-2.388	-2.437	-0.005
C	-2.82	-1.116	-0.003
C	0.77	-0.743	-2.777
C	-1.044	1.007	-5.43
C	-0.503	0.484	-4.145
C	-0.8	0.771	-2.835
C	0.771	-0.749	2.774
C	-1.029	1.005	5.434
C	-0.493	0.483	4.146
C	-0.792	0.773	2.837
H	-1.315	-5.077	-0.009
H	1.403	-5.076	-0.005

H	1.318	5.072	0.003
H	-1.4	5.07	0.003
H	-5.046	-1.412	-0.005
H	-5.094	1.305	0
H	-3.239	3.179	0.003
H	3.156	3.21	0.002
H	5.052	1.408	0.002
H	5.1	-1.309	0.002
H	3.243	-3.185	0
H	-3.152	-3.214	-0.007
H	1.513	-1.459	-2.444
H	0.952	-0.918	-4.87
H	-1.824	1.752	-5.23
H	-1.493	0.207	-6.039
H	-0.263	1.492	-6.035
H	-1.527	1.472	-2.447
H	1.51	-1.466	2.436
H	0.958	-0.927	4.865
H	-0.242	1.477	6.041
H	-1.488	0.208	6.037
H	-1.798	1.761	5.237
H	-1.518	1.478	2.453

Table A3.2 Atomic coordinates for geometry-optimized for $[Fe(P)(His)(Me_2S]^+$ model.

	x	**y**	**z**
Fe	0	0	0
S	0.191	-0.165	-2.324
N	-0.025	-0.008	2.009
N	0.275	-0.647	4.086
N	1.998	0	0
N	-0.016	-2.012	-0.026
N	0.016	2.007	0
N	-1.992	0	-0.018
C	-0.505	0.98	2.853
C	-0.331	0.598	4.161
C	-0.67	1.266	5.449
C	0.441	-0.985	2.783
C	-1.407	-0.605	-3.06
C	0.45	1.465	-3.075
C	2.825	-1.105	-0.09
C	2.404	-2.427	-0.178
C	1.081	-2.847	-0.131
C	-1.115	-2.85	0.056
C	-2.433	-2.424	0.135
C	-2.837	-1.095	0.084
C	-4.208	-0.667	0.081
C	-4.2	0.696	-0.036
C	-2.821	1.105	-0.083
C	-2.4	2.43	-0.147
C	-1.078	2.845	-0.086
C	1.114	2.841	0.104
C	2.435	2.419	0.179
C	2.843	1.093	0.11
C	4.214	0.664	0.096
C	4.203	-0.697	-0.039
C	0.7	4.223	0.083
C	-0.657	4.226	-0.047
C	-0.7	-4.229	0.006
C	0.659	-4.228	-0.121
H	-0.948	1.891	2.47
H	-1.131	2.241	5.251
H	-1.381	0.671	6.042
H	0.225	1.437	6.067
H	0.553	-1.22	4.876
H	0.896	-1.907	2.446

H	-2.176	0.134	-2.806
H	-1.273	-0.671	-4.147
H	-1.687	-1.589	-2.666
H	-0.368	2.152	-2.83
H	1.397	1.855	-2.686
H	0.53	1.329	-4.162
H	3.174	-3.194	-0.257
H	-3.21	-3.184	0.206
H	-5.064	-1.332	0.153
H	-5.048	1.374	-0.077
H	-3.171	3.2	-0.206
H	3.208	3.182	0.262
H	5.071	1.328	0.171
H	5.05	-1.375	-0.092
H	1.376	5.071	0.152
H	-1.328	5.079	-0.104
H	-1.372	-5.08	0.052
H	1.332	-5.079	-0.195

Table A3.3 Atomic coordinates for geometry-optimized for [Fe(P)(His)(Cys)] model.

	x	**y**	**z**
C	-0.03195	-1.48304	-2.97573
S	-0.00654	0.15744	-2.18016
Fe	0.000000	0.000000	0
C	0.003564	-3.44637	0.069127
C	-3.40675	0.000331	0.032813
C	0.003506	3.442657	0.079373
C	3.412529	0.000649	0.02556
N	-1.40639	-1.4314	0.049453
C	-1.22725	-2.80056	0.073191
C	-2.50493	-3.47713	0.088241
C	-3.4615	-2.50472	0.068134
C	-2.77034	-1.23516	0.046284
N	-1.41103	1.431324	0.064156
C	-2.77314	1.235362	0.038795
C	-3.46422	2.505333	0.027269
C	-2.50808	3.477247	0.042234
C	-1.22981	2.798823	0.060958
N	1.413581	1.433583	0.059414
C	1.234024	2.797399	0.082896
C	2.510325	3.474766	0.106133
C	3.469637	2.506307	0.089815
C	2.777995	1.236296	0.056261
N	1.413198	-1.43297	0.022571
C	2.777501	-1.23519	0.005398
C	3.467494	-2.50188	-0.0067
C	2.511567	-3.47642	0.011444
C	1.235007	-2.8022	0.030894
H	0.854007	-2.06885	-2.69382
H	-0.02747	-1.3155	-4.06128
H	-0.93845	-2.03815	-2.6972
H	0.002547	-4.53555	0.081552
H	-4.49874	-0.00109	0.019964
H	0.002547	4.530578	0.081552
H	4.503675	0.00145	0.019964
H	-2.63476	-4.55505	0.107136
H	-4.54195	-2.62104	0.068576
H	-4.54499	2.62117	0.007648
H	-2.63603	4.554181	0.037108
H	2.639217	4.554991	0.129718
H	4.549106	2.621188	0.097773

H	4.549305	-2.61797	-0.02546
H	2.643612	-4.55566	0.010528
C	0.192585	1.625984	5.476844
C	0.155207	0.808007	4.227258
N	0.391198	-0.5615	4.197887
C	-0.05169	1.151018	2.912796
C	0.320509	-0.98691	2.905994
N	0.056495	0.033506	2.102754
H	-0.41849	1.183929	6.279949
H	1.220395	1.743965	5.859716
H	-0.20166	2.629318	5.271282
H	0.576571	-1.15395	5.000363
H	-0.26647	2.12999	2.500474
H	0.464298	-2.0165	2.601035

Table A3.4 Atomic coordinates for geometry-optimized for $[Fe(P)(His)(CysSH)]^{+}$ model.

	x	**y**	**z**
Fe	0	0	0
S	-0.01518	-0.39877	-2.29586
N	1.437243	-1.41379	0.000161
N	-1.43875	-1.41047	-0.02061
N	-1.4235	1.395995	-0.05666
N	1.422307	1.400106	-0.05213
N	-0.03552	0.008166	1.983614
C	1.29176	0.466591	-3.24082
C	3.449911	-0.00619	-0.03566
C	-0.00203	-3.41431	-0.01668
C	-3.45266	-0.00692	-0.04344
C	-0.0015	3.407094	-0.0413
C	2.808491	-1.23881	-0.01326
C	3.475853	-2.5151	-0.01628
C	2.499733	-3.47007	-0.01379
C	1.234179	-2.78008	-0.00588
C	-1.23623	-2.77688	-0.02186
C	-2.50149	-3.4672	-0.03693
C	-3.47743	-2.51317	-0.04493
C	-2.81046	-1.23744	-0.03855
C	-2.80049	1.217626	-0.04257
C	-3.47057	2.490922	-0.01204
C	-2.49884	3.448899	-0.00955
C	-1.2313	2.764368	-0.03554
C	1.230418	2.767846	-0.05356
C	2.498155	3.452279	-0.06749
C	3.470061	2.49404	-0.06755
C	2.799263	1.219918	-0.0558
C	-1.54445	0.117278	5.391076
C	-0.76821	0.092557	4.11814
N	0.619357	0.176868	4.068558
C	-1.15505	-0.01307	2.805047
C	1.02704	0.127208	2.779394
H	1.311943	1.536244	-3.01319
H	2.237487	-0.00469	-2.95083
H	1.106723	0.286385	-4.30552
H	-1.07412	0.32721	-2.74625
H	4.539009	-0.00124	-0.0418
H	-0.0042	-4.50338	-0.03184
H	-4.54161	-0.00165	-0.0418

H	-0.0042	4.496299	-0.03184
H	4.553061	-2.65175	-0.02764
H	2.6143	-4.5502	-0.02178
H	-2.61609	-4.54744	-0.04221
H	-4.55517	-2.64809	-0.05773
H	-4.54803	2.621285	0.008222
H	-2.61691	4.52829	0.012035
H	2.616575	4.53213	-0.07494
H	4.547908	2.625163	-0.07491
H	-1.23455	0.953755	6.036706
H	-1.42022	-0.8161	5.962954
H	-2.61293	0.240762	5.175531
H	1.23902	0.273338	4.869022
H	-2.15756	-0.10479	2.404769
H	2.05998	0.1844	2.464365

Table A3.5 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-CH_3)(1-MeIm)].

	x	**y**	**z**
Fe	0	0	0
S	0.234	0.018	-2.222
O	-0.88	4.075	-4.172
N	0.565	-0.482	4.158
N	0.859	2.716	-3.547
N	0	0	2.076
N	-2.004	-0.015	0.034
N	-0.038	2.006	0.086
N	2.014	0.004	0.047
N	0.038	-2.017	0.059
C	1.269	-1.074	5.287
C	0.832	-0.695	2.845
C	-0.498	0.404	4.224
C	-0.837	0.692	2.926
C	5.536	6.048	-3.934
C	3.233	3.038	-3.406
C	4.376	3.83	-3.493
C	4.3	5.191	-3.824
C	3.022	5.723	-4.06
C	1.864	4.95	-3.976
C	1.959	3.584	-3.645
C	-0.462	2.968	-3.82
C	-2.25	-0.444	-3.214
C	-3.425	-0.2	-3.919
C	-3.612	1.038	-4.538
C	-2.616	2.004	-4.447
C	-1.413	1.782	-3.75
C	-1.234	0.527	-3.105
C	2.464	-2.408	-0.068
C	2.866	-1.078	-0.025
C	4.242	-0.636	-0.002
C	4.217	0.724	0.097
C	2.826	1.118	0.12
C	2.384	2.433	0.189
C	1.055	2.841	0.157
C	0.631	4.219	0.147
C	-0.732	4.214	0.053
C	-1.141	2.832	0.02
C	-2.461	2.398	-0.04
C	-2.858	1.069	-0.01

C	-4.234	0.629	0.03
C	-4.209	-0.73	0.118
C	-2.818	-1.125	0.115
C	-2.383	-2.443	0.159
C	-1.054	-2.854	0.111
C	-0.629	-4.232	0.079
C	0.735	-4.223	-0.007
C	1.142	-2.84	-0.016
H	2.061	-1.728	4.907
H	0.579	-1.671	5.9
H	1.721	-0.292	5.911
H	1.621	-1.351	2.491
H	-0.905	0.742	5.17
H	-1.617	1.344	2.55
H	5.69	6.4	-4.966
H	5.466	6.943	-3.297
H	6.434	5.492	-3.635
H	3.323	1.98	-3.151
H	5.352	3.376	-3.305
H	2.924	6.78	-4.32
H	0.884	5.381	-4.164
H	-2.095	-1.408	-2.729
H	-4.189	-0.977	-3.985
H	-4.525	1.248	-5.097
H	-2.735	2.981	-4.918
H	3.242	-3.17	-0.13
H	5.104	-1.297	-0.051
H	5.054	1.416	0.143
H	3.144	3.213	0.234
H	1.303	5.072	0.181
H	-1.411	5.06	0
H	-3.24	3.159	-0.092
H	-5.095	1.293	0
H	-5.046	-1.421	0.174
H	-3.145	-3.221	0.21
H	-1.3	-5.086	0.107
H	1.415	-5.069	-0.062
H	1.059	1.774	-3.175

Table A3.6 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-H)(1-MeIm)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	0.231	0.016	-2.225
O	-0.844	4.087	-4.168
N	0.078	-2.017	0.058
N	2.014	0.043	0.048
N	-0.077	2.004	0.082
N	-2.004	-0.055	0.034
N	0.000	0.000	2.073
N	0.574	-0.469	4.157
N	0.881	2.708	-3.546
C	1.198	-2.819	-0.014
C	0.818	-4.209	-0.007
C	-0.546	-4.244	0.075
C	-0.998	-2.875	0.107
C	-2.334	-2.490	0.153
C	-2.796	-1.181	0.112
C	-4.194	-0.813	0.115
C	-4.245	0.546	0.031
C	-2.878	1.012	-0.010
C	-2.507	2.350	-0.039
C	-1.195	2.808	0.018
C	-0.813	4.198	0.050
C	0.549	4.230	0.142
C	1.000	2.860	0.152
C	2.337	2.477	0.186
C	2.804	1.171	0.121
C	4.200	0.805	0.100
C	4.254	-0.554	0.005
C	2.887	-1.023	-0.021
C	2.511	-2.361	-0.064
C	-1.232	0.539	-3.104
C	-1.400	1.800	-3.747
C	-2.604	2.035	-4.439
C	-3.609	1.078	-4.532
C	-3.433	-0.164	-3.916
C	-2.259	-0.420	-3.214
C	-0.438	2.974	-3.817
C	1.991	3.562	-3.648
C	1.908	4.927	-3.983
C	3.077	5.687	-4.070

C	4.333	5.122	-3.829
C	4.412	3.767	-3.495
C	3.258	2.993	-3.406
C	-0.849	0.678	2.923
C	-0.504	0.397	4.223
C	0.845	-0.679	2.843
C	1.288	-1.047	5.285
H	1.514	-5.041	-0.061
H	-1.200	-5.111	0.100
H	-3.082	-3.283	0.200
H	-5.018	-1.520	0.170
H	-5.119	1.193	0.003
H	-3.300	3.095	-0.090
H	-1.508	5.031	-0.003
H	1.205	5.096	0.176
H	3.081	3.273	0.230
H	5.025	1.513	0.145
H	5.129	-1.198	-0.042
H	3.304	-3.108	-0.123
H	-2.714	3.013	-4.909
H	-4.521	1.298	-5.088
H	-4.204	-0.932	-3.981
H	-2.113	-1.386	-2.730
H	1.071	1.764	-3.174
H	0.933	5.367	-4.173
H	2.998	6.744	-4.335
H	5.236	5.729	-3.905
H	5.382	3.302	-3.309
H	3.328	1.934	-3.150
H	-1.640	1.316	2.547
H	-0.919	0.728	5.167
H	1.644	-1.321	2.490
H	1.728	-0.256	5.908
H	0.610	-1.655	5.900
H	2.092	-1.688	4.905

Table A3.7 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-Cl))(1-MeIm)].

	x	y	z
Fe	0.000	0.000	0.000
S	0.233	0.011	-2.227
O	-0.713	4.088	-4.227
N	0.961	2.671	-3.556
N	0.563	-0.479	4.155
N	0.000	0.000	2.071
N	-2.004	-0.015	0.033
N	-0.038	2.007	0.084
N	2.016	0.004	0.049
N	0.040	-2.016	0.054
Cl	5.935	5.925	-3.937
C	1.267	-1.070	5.285
C	0.832	-0.695	2.841
C	-0.500	0.407	4.220
C	-0.837	0.693	2.920
C	3.341	2.902	-3.387
C	4.517	3.640	-3.470
C	4.457	4.989	-3.830
C	3.230	5.592	-4.105
C	2.048	4.853	-4.021
C	2.090	3.494	-3.658
C	-0.348	2.971	-3.848
C	-2.272	-0.358	-3.208
C	-3.438	-0.071	-3.909
C	-3.579	1.170	-4.536
C	-2.544	2.095	-4.456
C	-1.348	1.829	-3.762
C	-1.215	0.571	-3.109
C	2.465	-2.408	-0.069
C	2.867	-1.077	-0.022
C	4.244	-0.636	0.008
C	4.217	0.723	0.111
C	2.826	1.117	0.130
C	2.385	2.431	0.200
C	1.056	2.841	0.160
C	0.631	4.219	0.154
C	-0.731	4.214	0.058
C	-1.139	2.832	0.019
C	-2.460	2.400	-0.040
C	-2.858	1.069	-0.010

C	-4.234	0.630	0.031
C	-4.209	-0.730	0.116
C	-2.818	-1.125	0.111
C	-2.381	-2.443	0.151
C	-1.053	-2.853	0.103
C	-0.628	-4.231	0.069
C	0.735	-4.223	-0.015
C	1.143	-2.840	-0.020
H	2.063	-1.721	4.906
H	0.577	-1.669	5.895
H	1.716	-0.287	5.909
H	1.620	-1.352	2.490
H	-0.907	0.746	5.165
H	-1.616	1.344	2.544
H	3.388	1.848	-3.110
H	5.479	3.171	-3.263
H	3.194	6.644	-4.388
H	1.088	5.315	-4.236
H	1.116	1.728	-3.162
H	-2.153	-1.324	-2.717
H	-4.232	-0.818	-3.966
H	-4.486	1.412	-5.092
H	-2.625	3.070	-4.936
H	3.242	-3.170	-0.131
H	5.105	-1.296	-0.039
H	5.054	1.415	0.163
H	3.144	3.213	0.252
H	1.303	5.072	0.196
H	-1.409	5.062	0.007
H	-3.239	3.160	-0.090
H	-5.094	1.293	0.003
H	-5.046	-1.421	0.170
H	-3.144	-3.222	0.200
H	-1.300	-5.086	0.094
H	1.416	-5.069	-0.071

Table A3.8 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-CF_3)(1-MeIm)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	0.193	-0.040	-2.234
F	7.638	3.682	-3.740
F	6.872	5.637	-3.115
F	6.919	5.114	-5.234
O	0.014	4.174	-4.153
N	0.545	-0.496	4.155
N	1.396	2.442	-3.556
N	-0.046	2.005	0.068
N	2.017	0.007	0.063
N	0.000	0.000	2.071
N	0.051	-2.019	0.046
N	-2.006	-0.020	0.042
C	1.158	-2.841	-0.015
C	0.752	-4.225	-0.028
C	-0.612	-4.237	0.029
C	-1.041	-2.860	0.067
C	-2.370	-2.453	0.109
C	-2.812	-1.137	0.098
C	-4.204	-0.748	0.116
C	-4.236	0.614	0.063
C	-2.864	1.061	0.022
C	-2.471	2.393	-0.008
C	-1.151	2.828	0.021
C	-0.746	4.212	0.048
C	0.618	4.220	0.123
C	1.046	2.842	0.131
C	2.375	2.439	0.177
C	2.822	1.124	0.136
C	4.214	0.736	0.142
C	4.247	-0.625	0.061
C	2.874	-1.073	0.016
C	2.478	-2.404	-0.035
C	-1.132	0.794	-3.093
C	-1.034	2.067	-3.722
C	-2.166	2.567	-4.395
C	-3.354	1.851	-4.482
C	-3.440	0.590	-3.884
C	-2.342	0.079	-3.200
C	0.162	3.000	-3.808

C	2.660	3.028	-3.672
C	2.878	4.377	-4.023
C	4.178	4.862	-4.122
C	5.283	4.035	-3.879
C	5.069	2.694	-3.534
C	3.775	2.198	-3.433
C	6.667	4.603	-3.990
C	-0.805	0.732	2.919
C	-0.478	0.435	4.218
C	0.800	-0.729	2.843
C	1.224	-1.111	5.287
H	1.436	-5.069	-0.076
H	-1.282	-5.093	0.035
H	-3.130	-3.236	0.138
H	-5.038	-1.444	0.161
H	-5.100	1.273	0.055
H	-3.254	3.151	-0.041
H	-1.426	5.058	0.009
H	1.288	5.075	0.154
H	3.132	3.221	0.223
H	5.048	1.432	0.193
H	5.113	-1.282	0.036
H	3.260	-3.163	-0.081
H	-2.069	3.550	-4.856
H	-4.205	2.268	-5.023
H	-4.358	0.004	-3.949
H	-2.400	-0.902	-2.729
H	2.023	5.019	-4.212
H	4.335	5.907	-4.394
H	5.917	2.037	-3.348
H	3.616	1.152	-3.166
H	-1.555	1.416	2.540
H	-0.870	0.794	5.163
H	1.562	-1.419	2.492
H	1.712	-0.346	5.904
H	0.510	-1.673	5.903
H	1.987	-1.803	4.909
H	1.370	1.481	-3.176

Table A3.9 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-NO_2)(1-MeIm)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	0.221	-0.008	-2.234
O	6.048	6.500	-4.313
O	7.062	4.684	-3.629
O	-0.459	4.109	-4.261
N	6.044	5.319	-3.939
N	1.112	2.588	-3.564
N	0.566	-0.474	4.153
N	0.000	0.000	2.070
N	-2.004	-0.018	0.035
N	-0.043	2.005	0.077
N	2.016	0.005	0.053
N	0.045	-2.016	0.049
C	1.273	-1.061	5.284
C	0.834	-0.693	2.841
C	-0.496	0.410	4.218
C	-0.838	0.693	2.919
C	3.493	2.644	-3.363
C	4.716	3.290	-3.450
C	4.753	4.628	-3.855
C	3.579	5.314	-4.177
C	2.350	4.670	-4.090
C	2.290	3.320	-3.677
C	-0.180	2.976	-3.867
C	-2.310	-0.213	-3.195
C	-3.460	0.148	-3.889
C	-3.525	1.393	-4.519
C	-2.432	2.249	-4.452
C	-1.250	1.905	-3.766
C	-1.194	0.644	-3.107
C	2.471	-2.406	-0.071
C	2.871	-1.075	-0.017
C	4.245	-0.631	0.022
C	4.215	0.730	0.129
C	2.824	1.119	0.140
C	2.378	2.434	0.206
C	1.049	2.840	0.158
C	0.623	4.219	0.151
C	-0.738	4.212	0.052
C	-1.146	2.830	0.014

C	-2.466	2.395	-0.038
C	-2.861	1.065	-0.003
C	-4.235	0.621	0.042
C	-4.206	-0.739	0.123
C	-2.815	-1.130	0.112
C	-2.375	-2.448	0.145
C	-1.046	-2.856	0.093
C	-0.619	-4.232	0.053
C	0.744	-4.223	-0.030
C	1.151	-2.839	-0.027
H	2.066	-1.716	4.905
H	0.584	-1.656	5.900
H	1.726	-0.276	5.902
H	1.624	-1.348	2.489
H	-0.904	0.748	5.163
H	-1.618	1.343	2.543
H	3.452	1.600	-3.049
H	5.647	2.777	-3.213
H	3.643	6.353	-4.493
H	1.427	5.189	-4.335
H	1.195	1.642	-3.151
H	-2.251	-1.183	-2.703
H	-4.302	-0.546	-3.938
H	-4.418	1.692	-5.070
H	-2.453	3.225	-4.937
H	3.251	-3.166	-0.132
H	5.110	-1.287	-0.022
H	5.051	1.422	0.188
H	3.136	3.215	0.267
H	1.293	5.072	0.198
H	-1.418	5.059	0.004
H	-3.247	3.154	-0.085
H	-5.097	1.282	0.020
H	-5.041	-1.432	0.180
H	-3.137	-3.229	0.190
H	-1.288	-5.089	0.075
H	1.425	-5.068	-0.087

Table A3.10 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-CH_3)(H_2O)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	-0.273	-0.212	2.165
O	-0.861	3.950	4.095
O	0.084	0.177	-2.206
N	-1.941	2.007	3.521
N	0.002	-1.994	-0.238
N	1.997	0.031	0.000
N	0.000	2.009	0.000
N	-2.004	-0.016	-0.268
C	0.850	0.800	3.121
C	0.510	2.020	3.766
C	1.502	2.664	4.528
C	2.778	2.134	4.686
C	3.103	0.925	4.067
C	2.147	0.274	3.293
C	-0.821	2.756	3.783
C	-3.291	2.391	3.588
C	-3.729	3.695	3.884
C	-5.096	3.968	3.945
C	-6.069	2.982	3.717
C	-5.615	1.689	3.420
C	-4.255	1.393	3.356
C	-7.541	3.302	3.800
C	-1.085	-2.838	-0.358
C	-0.651	-4.214	-0.409
C	0.710	-4.200	-0.317
C	1.107	-2.817	-0.211
C	2.425	-2.389	-0.117
C	2.834	-1.066	-0.035
C	2.831	1.129	0.084
C	4.213	0.715	0.095
C	4.214	-0.648	0.020
C	2.412	2.450	0.153
C	1.086	2.859	0.121
C	-1.112	2.833	-0.009
C	-0.714	4.213	0.106
C	0.647	4.229	0.187
C	-2.429	2.407	-0.128
C	-2.841	1.086	-0.251
C	-2.837	-1.113	-0.389

C	-4.213	-0.692	-0.467
C	-4.215	0.671	-0.381
C	-2.411	-2.434	-0.428
H	1.224	3.608	4.998
H	3.515	2.660	5.296
H	4.095	0.485	4.185
H	2.391	-0.669	2.802
H	-1.761	1.057	3.170
H	-2.992	4.473	4.068
H	-5.413	4.987	4.181
H	-6.341	0.893	3.239
H	-3.931	0.375	3.128
H	-8.154	2.423	3.560
H	-7.820	4.108	3.100
H	-7.822	3.642	4.809
H	-1.316	-5.068	-0.508
H	1.397	-5.041	-0.326
H	3.200	-3.156	-0.122
H	5.060	1.393	0.147
H	5.064	-1.324	-0.002
H	3.177	3.221	0.241
H	-1.403	5.053	0.134
H	1.309	5.085	0.291
H	-3.205	3.172	-0.115
H	-5.060	-1.366	-0.574
H	-5.064	1.348	-0.394
H	-3.172	-3.209	-0.522
H	0.237	1.139	-2.285
H	-0.867	0.058	-2.397

Table A3.11 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-H)(H_2O)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	-0.282	-0.206	2.166
O	-0.598	3.976	4.117
O	0.090	0.180	-2.203
N	-1.800	2.112	3.528
N	-0.030	-1.992	-0.244
N	1.998	0.000	0.000
N	0.032	2.010	0.003
N	-2.003	0.017	-0.261
C	0.905	0.727	3.123
C	0.647	1.965	3.773
C	1.682	2.544	4.531
C	2.923	1.934	4.682
C	3.166	0.706	4.060
C	2.166	0.121	3.288
C	-0.634	2.786	3.797
C	-3.122	2.580	3.593
C	-3.474	3.909	3.896
C	-4.822	4.269	3.951
C	-5.835	3.336	3.709
C	-5.484	2.016	3.407
C	-4.144	1.640	3.350
C	-1.130	-2.820	-0.359
C	-0.719	-4.200	-0.415
C	0.643	-4.208	-0.334
C	1.063	-2.833	-0.225
C	2.388	-2.427	-0.135
C	2.817	-1.109	-0.046
C	2.850	1.084	0.090
C	4.225	0.648	0.096
C	4.205	-0.713	0.011
C	2.452	2.411	0.164
C	1.133	2.842	0.128
C	-1.066	2.851	-0.008
C	-0.646	4.224	0.109
C	0.716	4.219	0.194
C	-2.390	2.446	-0.126
C	-2.823	1.131	-0.244
C	-2.855	-1.067	-0.378
C	-4.224	-0.624	-0.451

C	-4.204	0.738	-0.369
C	-2.450	-2.394	-0.419
H	1.467	3.503	5.006
H	3.694	2.410	5.290
H	4.128	0.205	4.172
H	2.347	-0.834	2.794
H	-1.680	1.154	3.170
H	-2.688	4.634	4.089
H	-5.081	5.302	4.191
H	-6.884	3.631	3.759
H	-6.258	1.270	3.219
H	-3.876	0.606	3.118
H	-1.397	-5.045	-0.512
H	1.318	-5.060	-0.348
H	3.152	-3.203	-0.148
H	5.083	1.312	0.152
H	5.044	-1.403	-0.018
H	3.229	3.170	0.255
H	-1.321	5.075	0.135
H	1.391	5.064	0.300
H	-3.154	3.225	-0.114
H	-5.083	-1.284	-0.553
H	-5.042	1.431	-0.382
H	-3.224	-3.157	-0.510
H	0.297	1.132	-2.284
H	-0.865	0.114	-2.400

Table A3.12 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-Cl)(H_2O)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	-0.282	-0.210	2.167
Cl	-7.484	3.826	3.795
O	-0.592	3.946	4.186
O	0.095	0.179	-2.200
N	-1.790	2.100	3.537
N	-0.030	-1.993	-0.245
N	1.998	0.000	0.000
N	0.032	2.009	0.005
N	-2.003	0.016	-0.261
C	0.910	0.713	3.128
C	0.656	1.946	3.792
C	1.695	2.516	4.551
C	2.936	1.902	4.690
C	3.174	0.681	4.055
C	2.171	0.103	3.284
C	-0.624	2.765	3.831
C	-3.108	2.573	3.607
C	-3.456	3.895	3.946
C	-4.800	4.271	4.003
C	-5.802	3.342	3.723
C	-5.474	2.027	3.387
C	-4.136	1.650	3.332
C	-1.129	-2.821	-0.354
C	-0.718	-4.202	-0.412
C	0.644	-4.209	-0.338
C	1.064	-2.833	-0.231
C	2.389	-2.426	-0.145
C	2.818	-1.109	-0.051
C	2.850	1.084	0.096
C	4.225	0.648	0.100
C	4.205	-0.712	0.007
C	2.451	2.410	0.177
C	1.131	2.841	0.138
C	-1.067	2.850	-0.013
C	-0.647	4.224	0.109
C	0.715	4.218	0.204
C	-2.390	2.446	-0.138
C	-2.823	1.131	-0.254
C	-2.856	-1.069	-0.375

C	-4.225	-0.626	-0.454
C	-4.203	0.738	-0.382
C	-2.450	-2.396	-0.412
H	1.483	3.468	5.037
H	3.710	2.371	5.300
H	4.136	0.177	4.158
H	2.347	-0.848	2.781
H	-1.669	1.152	3.152
H	-2.668	4.611	4.165
H	-5.065	5.294	4.267
H	-6.261	1.303	3.175
H	-3.879	0.621	3.072
H	-1.396	-5.046	-0.506
H	1.319	-5.061	-0.355
H	3.154	-3.203	-0.161
H	5.082	1.313	0.161
H	5.045	-1.400	-0.027
H	3.227	3.169	0.273
H	-1.322	5.075	0.130
H	1.389	5.062	0.315
H	-3.152	3.224	-0.135
H	-5.082	-1.285	-0.558
H	-5.042	1.430	-0.406
H	-3.224	-3.159	-0.500
H	0.302	1.130	-2.283
H	-0.860	0.112	-2.406

Table A3.13 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-CF_3)(H_2O)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	-0.278	-0.208	2.169
F	-7.520	4.791	4.472
F	-7.722	4.010	2.439
F	-8.096	2.702	4.149
O	-0.617	3.949	4.174
O	0.093	0.172	-2.200
N	-1.800	2.091	3.530
N	2.000	0.000	0.000
N	0.033	2.010	0.000
N	-2.002	0.016	-0.261
N	-0.030	-1.993	-0.239
C	0.909	0.725	3.127
C	2.174	0.125	3.284
C	3.173	0.712	4.054
C	2.927	1.933	4.687
C	1.683	2.537	4.546
C	0.648	1.958	3.787
C	-0.636	2.767	3.825
C	-3.120	2.549	3.597
C	-4.136	1.613	3.316
C	-5.474	1.980	3.366
C	-5.833	3.293	3.702
C	-7.280	3.697	3.700
C	-4.828	4.225	3.986
C	-3.481	3.868	3.937
C	2.851	1.084	0.093
C	2.451	2.410	0.168
C	1.132	2.841	0.130
C	-1.065	2.851	-0.022
C	-2.388	2.447	-0.147
C	-2.821	1.131	-0.259
C	-2.855	-1.069	-0.372
C	-2.449	-2.396	-0.406
C	-1.129	-2.822	-0.345
C	1.064	-2.834	-0.222
C	2.389	-2.426	-0.137
C	2.819	-1.109	-0.048
C	4.206	-0.713	0.009
C	4.226	0.648	0.098

C	0.645	-4.211	-0.326
C	-0.717	-4.204	-0.400
C	-4.224	-0.626	-0.454
C	-4.203	0.738	-0.387
C	-0.645	4.224	0.092
C	0.717	4.218	0.188
H	2.357	-0.825	2.783
H	4.139	0.215	4.159
H	3.698	2.408	5.295
H	1.464	3.490	5.030
H	-1.669	1.141	3.147
H	-3.866	0.588	3.060
H	-6.247	1.242	3.151
H	-5.100	5.246	4.254
H	-2.703	4.590	4.166
H	3.230	3.169	0.261
H	-3.151	3.225	-0.151
H	-3.223	-3.160	-0.489
H	3.154	-3.204	-0.153
H	5.045	-1.402	-0.023
H	5.083	1.312	0.156
H	1.319	-5.061	-0.341
H	-1.397	-5.047	-0.491
H	-5.081	-1.286	-0.556
H	-5.039	1.430	-0.415
H	-1.319	5.077	0.108
H	1.392	5.063	0.294
H	0.302	1.121	-2.295
H	-0.859	0.100	-2.407

Table A3.14 Atomic coordinates for geometry-optimized for [Fe(P)(**S**-NO_2)(H_2O)].

	x	**y**	**z**
Fe	0.000	0.000	0.000
S	-0.282	-0.216	2.169
O	-8.035	2.852	3.519
O	-7.443	4.846	4.204
O	-0.575	3.908	4.255
O	0.097	0.178	-2.198
N	-1.774	2.080	3.553
N	-7.182	3.692	3.838
N	-0.031	-1.991	-0.247
N	1.998	0.000	0.000
N	0.034	2.011	0.007
N	-2.003	0.019	-0.262
C	0.918	0.691	3.135
C	0.675	1.918	3.813
C	1.719	2.477	4.574
C	2.958	1.858	4.700
C	3.187	0.642	4.050
C	2.178	0.076	3.278
C	-0.600	2.738	3.872
C	-3.084	2.544	3.639
C	-3.437	3.850	4.043
C	-4.777	4.215	4.109
C	-5.770	3.294	3.775
C	-5.443	1.994	3.376
C	-4.108	1.625	3.311
C	-1.132	-2.819	-0.351
C	-0.721	-4.200	-0.410
C	0.641	-4.207	-0.343
C	1.062	-2.832	-0.238
C	2.387	-2.426	-0.155
C	2.818	-1.109	-0.059
C	2.851	1.083	0.102
C	4.225	0.646	0.100
C	4.205	-0.714	0.000
C	2.453	2.409	0.189
C	1.134	2.841	0.147
C	-1.064	2.853	-0.017
C	-0.643	4.226	0.107
C	0.718	4.218	0.211
C	-2.387	2.450	-0.151

C	-2.821	1.135	-0.264
C	-2.856	-1.065	-0.373
C	-4.224	-0.620	-0.458
C	-4.202	0.743	-0.396
C	-2.452	-2.394	-0.406
H	1.513	3.424	5.074
H	3.738	2.317	5.309
H	4.148	0.133	4.142
H	2.348	-0.872	2.766
H	-1.647	1.142	3.139
H	-2.651	4.556	4.300
H	-5.067	5.219	4.419
H	-6.238	1.294	3.127
H	-3.841	0.612	3.004
H	-1.400	-5.044	-0.500
H	1.316	-5.060	-0.364
H	3.151	-3.202	-0.177
H	5.084	1.310	0.163
H	5.045	-1.402	-0.040
H	3.230	3.167	0.288
H	-1.316	5.078	0.119
H	1.393	5.063	0.324
H	-3.148	3.230	-0.160
H	-5.083	-1.280	-0.561
H	-5.039	1.435	-0.430
H	-3.228	-3.156	-0.489
H	0.308	1.127	-2.287
H	-0.857	0.112	-2.407