Carbon Monoxide-Dependent DNA Binding of Heme-Containing Transcription Factors CooA and RcoM

By

Madeleine G. Roberts

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 08/25/2023

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

Judith N. Burstyn, Professor, Chemistry

Thomas C. Brunold, Professor, Chemistry

M. Thomas Record, Professor, Chemistry and Biochemistry

Helen Blackwell, Norman C. Craig Professor, Chemistry

Carbon Monoxide-Dependent DNA Binding of Heme-Containing Transcription Factors CooA and RcoM

By

Madeleine G. Roberts

Under the supervision of Professor Judith N. Burstyn

at the University of Wisconsin-Madison

Abstract

Carbon monoxide (CO), despite its reputation as a toxic gas, is an essential molecule to biological systems. For a group of microbes, CO serves as the sole source of energy and carbon. The genes that control this interesting biological strategy, the metabolism of CO, require strict transcriptional regulation by two biochemically characterized, nonhomologous proteins CooA (CO Oxidation Activator) and RcoM (Regulator of CO Metabolism). The first chapter of this work will highlight our current understanding of these two proteins, both of which use an iron-containing cofactor, heme, to sense and respond to environmental CO. Both CooA and RcoM undergo similar redox cycling at heme to bind CO, and both proteins rely on allostery to propagate the signal of CO-binding to distant parts of the protein controlling DNA binding. The rest of this dissertation will focus on gaining a more complete understanding of the CO-dependent mechanisms of regulation used by CooA and RcoM. The second chapter describes a biochemical characterization of RcoM-1 from *Paraburkholderia xenovorans*, defines the RcoM DNA binding site as a triplet set of direct repeats, and speculates as to the binding mode between RcoM and DNA given the unusual stoichiometry of an interaction between a dimeric protein and triplet direct repeat promoter. The third chapter provides further support for the hypothesized

RcoM:DNA complex using low-resolution small-angle X-ray scattering structural models, presenting the first experimentally-supported model of the RcoM dimer and the RcoM:DNA complex. Finally, the fourth chapter seeks to broaden our understanding of the CooA allosteric mechanism by including how protein dynamics are modulated upon effector binding and by binding DNA. These chapters all seek to connect how CO binding at heme leads to functional change in CO-sensing transcription factors, CooA and RcoM. The last chapter of this dissertation highlights work done outside the laboratory in preparation for a career in policy. This chapter discusses the potential for implementation of natural climate solutions in Wisconsin agriculture as well as the global cycling of nitrogen and carbon that controls the fundamentals of carbon storage.

Acknowledgements

I have a great number of people to thank for their contributions to getting me to the end of my PhD. All of the people listed contributed greatly to the quality of the work in this dissertation through facilitating my development as a scientist and researcher or by providing immense personal support. I am incredibly grateful to those listed in the following paragraphs for their support through these challenging years.

Some of my best graduate school memories take place in the first several months of joining the Burstyn group. I want to thank Matt and Brian for being the best possible friends, mentors, and now collaborators I could have asked for. When I joined the lab, I immediately felt welcomed into our tiny office, and it became so fun to learn by following you both around the building. We used to say that it was a good thing we got along so well because the office was too small for tension, but I now think that the tiny office facilitated our wild scientific speculation, funny conversation, and fast friendship. Matt—Thank you for remaining my mentor and friend even after moving on from UW. I have learned so much from our RcoM collaboration these many years. I am so excited to see where you and your future students can take our understanding of this protein; you are the best possible person to shepherd this research after the Burstyn Lab. Brian—I benefited from your expertise for my entire graduate career; thank you for teaching me so much and always pushing me to be better.

I would like to thank my advisor, Judith, for encouraging my non-traditional goals. Your belief that my graduate school experience should be made up of what I need made all the difference in my ability to pursue training outside the laboratory in science policy. The ability to use our time outside the building for skill-building and professional development meant I was

well-prepared to apply for the jobs and fellowship opportunities I wanted. Thank you for supporting my goals, helping me build my professional network, and sharing my interests.

I would like to thank our collaborators, on- and off-campus, that contributed to this work. From James Madison University, Dr. Chris Berndsen and his undergraduate students Mikayla Riney and Abigail Sholes made it possible for us to use SAXS techniques to understand RcoM structure and drove the complicated data analysis required to interpret this data. SIBYLS Beamline at the Advanced Light Source and Lawrence Berkley National Laboratory collected the data. Here at UW-Madison, I worked with Dr. Craig Bingman in the Crystallography Core and Drs. Darrell McCaslin and Dan Stevens in the Biophysics Instrumentation Center throughout my time in graduate school.

I would next like to thank my roommate and friend, Gina, for sticking with me through all five years of graduate school, through thick and thin. I was so lucky in choosing you to live with our first year. I could not have imagined then all that we would go through together in five years. I will forever miss our pandemic Friday night cheeseburgers, porch beers, and movie nights with Julia. Thank you for your immense support and positivity; I do not know what I would have done without you.

To Tom—thank you for being my friend and teaching me so much about myself. You work harder than anyone, and I know that your persistence and work ethic will carry you wherever you choose. And thank you for letting me teach you about Taylor Swift.

I would like to thank Haley, Hanna, Emily, and Kelly, my rocks for almost ten years. I cannot believe that I met you all within the first few days of starting college, and that we would

turn out to be close friends for life. I love you all so much and appreciate your support-from-afar as I moved to Wisconsin and started the next phases of my life.

Lastly, I would like to thank my family. I would not be here without your support, and I appreciate all you have done for our family to allow Kate and me to pursue higher education. I love you all so much.

Table of Contents

Abstract i
Acknowledgements iii
List of Tables viii
List of Figures ix
List of Abbreviations xiii
Chapter 1: Carbon Monoxide Sensing Transcription Factors: Regulators of Microbial Carbon Monoxide Oxidation Pathway Gene Expression 1
1.1 Introduction 2
1.2 Overview of microbial CO oxidation 3
1.3 CooA, the archetypical CO-sensing transcription factor 8
1.4 RcoM, a high affinity CO sensor that regulates aerobic CO oxidation 14
1.5 Functional Contexts of CooA and RcoM 22
1.6 Putative Novel Regulators of CO Oxidation 25
1.7 Concluding Remarks 27
1.8 Appendix 29
1.9 References 30
Chapter 2: Quaternary Structure and DNA Binding Properties of the Heme-Dependent, CO-Sensing Transcriptional Regulator *Px*RcoM 50
2.1 Introduction 51
2.2 Materials and Methods 56
2.3 Results 63
2.4 Discussion 76
2.5 Supplementary Information 81
2.6 References 98
Chapter 3: Insight Into the Protein:DNA Complex of Heme-Containing Transcription Factor, *Px*RcoM-1 108
3.1 Introduction 109
3.2 Materials and Methods 113
3.3 Results 120
3.4 Discussion 132
3.5 Supplementary Information 137
3.6 References 141

Chapter 4: Probing Conformational Dynamics of DNA Binding by CO-Sensing Trascription Factor, CooA, with Site-Directed Spin Label Electron Paramagnetic Resonance Spectroscopy 149

4.1 Introduction 150

4.2 Materials and Methods 155

4.3 Results 158

4.4 Discussion 165

4.5 References 168

Chapter 5: Opportunities for Implementing Natural Claimate Solutions in Wisconsin Agriculture 175

5.1 Introduction 176

5.2 Review of the Scientific Literature on Natural Climate Solutions in Agriculture 181

5.3 Limitations and Continued Debate 187

5.4 Policy Pathways 189

5.5 Summary and Conclusions 190

5.6 References 191

List of Tables

Chapter 1

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

Chapter 2

Table 2.1 Oligonucleotide sequences employed in fluorescence polarization DNA binding assays.

Table 2.S1 Heme quantitation of WT and HBD *Px*RcoM-1.

Chapter 3

Table 3.1 Oligonucleotide sequences used in fluorescence anisotropy DNA binding assays and for SAXS structural analysis.

Table 3.2 Characteristic parameters of SAXS experimental curves and theoretical models

Chapter 4

Table 4.1 *In vitro* fluorescence anisotropy single-point DNA binding assays of cysteine-substitution variants on a Δ_4Cys CooA background, before and after MAL-6 labeling.

Chapter 5

Table 5.1 Summary of six on-field agricultural natural climate solutions practices.

List of Figures

Chapter 1

Figure 1.1 Exemplary operons encoding Ni,Fe-CODH and Cu,Mo-CODH.

Figure 1.2 Structures and coordination states of CooA.

Figure 1.3 The CooA protein sequence similarity network colored according to genomic context.

Figure 1.4 Domain organization and coordination states of RcoM.

Figure 1.5 Genomic contexts of *rcoM* genes mapped onto the RcoM protein sequence similarity network.

Chapter 2

Figure 2.1 Organization of the $rcoM_1$/$coxM_1$ intergenic region in *P. xenovorans*

Figure 2.2 Size exclusion chromatograms of full-length WT and heme-binding domain (HBD) truncate *Px*RcoM-1.

Figure 2.3 (A) Electronic absorption spectra of WT, C94S, and H74A *Px*RcoM-1 and (B) Size exclusion chromatograms of full-length WT, C94S, and H74A *Px*RcoM-1.

Figure 2.4 RcoM sequence similarity network images at an alignment score cutoff of (A) 33 and (B) 45.

Figure 2.5 Logo plot generated from multiple sequence alignment of 40 putative *cox*-associated RcoM promoter regions.

Figure 2.6 Fluorescence polarization data for inactive Fe(III) and active Fe(II)-CO *Px*RcoM-1 incubated with two-site 8 bp direct repeats, “d + e” (top), “e + f” (middle), and a single 8 bp site “d” (bottom) Texas-Red labeled dsDNA.

Figure 2.7 Fluorescence polarization data for active Fe(II)-CO *Px*RcoM-1 and triplet direct repeat site dsDNA oligonucleotide.

Figure 2.8 A proposed model of RcoM DNA binding based on the structural model of CarH tetrameric DNA binding.

Figure 2.S1 Representative SDS-PAGE for a typical preparation of full-length, WT *Px*RcoM-1 (expected molecular weight of 31.5 kDa).

Figure 2.S2 Sedimentation equilibrium data plotted as ln(absorbance) vs. radial position squared.

Figure 2.S3 X-band EPR spectra of WT *Px*RcoM-1 as isolated after purification and reconstituted with hemin chloride.

Figure 2.S4 A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography profiles (lower panel) for as-isolated and reconstituted full-length WT *Px*RcoM-1.

Figure 2.S5 A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography profiles (lower panel) for the as-isolated and reconstituted *Px*RcoM-1 heme binding domain truncate.

Figure 2.S6 RcoM sequence similarity network shown as a function of alignment score cutoff.

Figure 2.S7 Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cox*-associated RcoM proteins.

Figure 2.S8 Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cowN*-associated RcoM proteins.

Figure 2.S9 Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cox*- and *cowN*-associated RcoM proteins.

Figure 2.S10 Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the “d+e” Texas Red labeled oligonucleotide.

Figure 2.S11 Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the “e+f” Texas Red labeled oligonucleotide.

Figure 2.S12 Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the “a” site Texas Red-labeled dsDNA oligonucleotide.

Figure 2.S13 Fluorescence anisotropy data for active, Fe(II)-CO *Px*RcoM-2 DNA binding to Texas Red-labeled dsDNA oligonucleotides bearing $coxM_1$ promoter elements.

Figure 2.S14 Complete fluorescence anisotropy data for WT *Px*RcoM-1 DNA binding to the “d + e + f” Texas Red labeled oligonucleotide.

Figure 2.S15 Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the “d + e + f” Texas Red labeled oligonucleotide fit using a simple bimolecular

interaction model ($\chi^2 = 1.0 \times 10^{-3}$, dashed line) and cooperative binding model ($\chi^2 = 3.8 \times 10^{-4}$, solid line).

Chapter 3

Figure 3.1 Heme coordination states of *Px*RcoM-1 and logo plot from the alignment of 40 *coxM* promoters showing the conservation the *coxM* promoter.

Figure 3.2 ColabFold model of *Px*RcoM-1 dimer and the refinement of the predicted structure with SEC-SAXS data.

Figure 3.3 SEC and SEC-SAXS data of Fe(III) and Fe(II)-CO RcoM indicates no effector-dependent change in oligomeric state.

Figure 3.4 *In vitro* fluorescence polarization data for RcoM with a higher-affinity oligonucleotide and two oligonucleotides sequentially dropping out binding sites.

Figure 3.5 SEC-SAXS data for crosslinked RcoM with native and higher-affinity promoter sequence oligonucleotides.

Figure 3.6 Box plot displaying the fit of HADDOCK-generated models of RcoM:DNA complex to SAXS data by number of RcoM dimers in the complex and the best HADDOCK model of RcoM:DNA complex involving two RcoM dimers.

Figure 3.S1 PDDT and PAE plots for the ColabFold *Px*RcoM-1 predicted dimer structure.

Figure 3.S2 Size exclusion chromatogram and fluorescence anisotropy based DNA binding assay of constitutively active *Px*RcoM-1 variant, H74Y.

Figure 3.S3 HT-SAXS data showing formation of the RcoM:DNA complex and SDS-PAGE gel of product of glutaraldehyde crosslinking of RcoM and promoter DNA sequence.

Chapter 4

Figure 4.1 Structure and heme coordination of the three functional states of CooA.

Figure 4.2 Location of MAL-6 labeling sites on a crystal structure of CooA and the chemical reaction between MAL-6 and an introduced cysteine to form redox-stable bond with protein.

Figure 4.3 Characterization of a new Δ_4Cys CooA variant, I173C, with absorbance spectra of Fe(III), Fe(II), and Fe(II)-CO protein and *in vitro* fluorescence anisotropy data of MAL-6 labeled and unlabeled I173C CooA with pCooF.

Figure 4.4 EPR spectra of Fe(III) and Fe(II)-CO MAL-6 labeled Δ_4Cys CooA variants.

Figure 4.5 EPR spectra of Fe(III), Fe(II)-CO, Fe(II)-CO+DNA of selected MAL-6 labeled Δ_4Cys CooA variants.

Chapter 5

Figure 5.1 Simplified global carbon and nitrogen cycles.

List of Abbreviations

A

Acetyl-CoA	acetyl-coenzyme A
ACS	acetyl CoA synthase
Amp	ampicillin
AMP	adaptive multipaddock grazing
AUC-SE	analytical ultracentrifugation-sedimentation equilibrium

B

BCA	bicinchoninic acid
BLAST	basic local alignment search tool
bp	base pair

C

cAMP	cyclic adenosine monophosphate
Ch	*Carboxydothermus hydrogenoformans*
CO	carbon monoxide
CO_2	carbon dioxide
CoA	coenzyme A
CODH	carbon monoxide dehydrogenase
COG	cluster of orthologous groups
CooA	Carbon monoxide oxidation Activator
CRP	cyclic AMP receptor protein
CW	continuous wave

D

Da	Dalton
DENSS	DENsity from Solution Scattering
DNA	deoxyribonucleic acid
DTT	dithiothreitol

E

EFI-EST	Enzyme Function Initiative-Enzyme Similarity Tool
EFI-GNT	Enzyme Function Initiative-Genome Neighborhood Tool
EPR	electron paramagnetic resonance
ExPASy	Expert Protein Analysis System

F

FAD	flavin adenine dinucleotide
FNR	fumarate and nitrate reduction regulator
FoXS	Fast X-ray Scattering
FPLC	fast protein liquid chromatography

G

GHG	greenhouse gas

H

HBD	heme binding domain
HPLC	high-performance liquid chromatography
HTH	helix-turn-helix
HT-SAXS	high throughput small-angle x-ray scattering

I

IDT	Integrated DNA Technologies
INSDC	International Nucleotide Sequence Database Collaboration
IPTG	isopropyl β-D-1-thiogalactopyranoside

M

MALS	multi-angle light scattering
MAL-6	N-(2,-2,6,6-tetramethylpiperidin-4-yl-1-oxyl)maleimide
MarR	multiple antibiotic resistance regulator
MOPS	3-(*N*-morpholino)propanesulfonic acid
M_s	sequence molecular weight
M_w	average molecular weight

N

NCS	natural climate solutions
NHE	normal hydrogen electrode

P

PAE	predicted alignment error
PAGE	polyacrylamide gel electrophoresis
PAS	Per-Arnt-Sim
pCooF	CooF promoter
PDB	Protein Data Bank
PDDF	pair-distance distribution function
pLDDT	predicted local distance difference test
pTM	predicted template modeling
Px	*Paraburkholderia xenovorans*

R

RcoM	Regulator of CO Metabolism
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAP	RNA polymerase
Rr	*Rhodospirillum rubrum*

S

SAXS	small angle X-ray scattering
SDSL-EPR	site directed spin label electron paramagnetic resonance spectroscopy
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEC-SAXS	size-exclusion chromatography coupled small angle X-ray scattering
SHE	standard hydrogen electrode
SOC	soil organic carbon
SSN	sequence similarity network

T

TCEP	tris(2-carboxyethyl)phosphine
Tris	tris(hydroxymethyl)aminomethane
TRX	Texas Red-X

W

WICCI	Wisconsin Initiative on Climate Change Impacts
WT	wild type

δ

δALA	δ-aminolevulinic acid

Chapter 1

Carbon Monoxide-Sensing Transcription Factors: Regulators of Microbial Carbon Monoxide Oxidation Pathway Gene Expression

A version of this chapter was published as:

Dent, M. R.; Weaver, B. R.; Roberts, M. G.; Burstyn, J. N. Carbon Monoxide-Sensing Transcription Factors: Regulators of Microbial Carbon Monoxide Oxidation Pathway Gene Expression. *Journal of Bacteriology* **2023**, *61* (8), 678–688. https://doi.org/10.1128/jb.00332-22

M.R.D. and B.R.W. wrote the initial draft of this manuscript. M.G.R. rewrote a significant portion of the section concerning RcoM. M.R.D., B.R.W., M.G.R. and J.N.B. heavily edited and incorporated reviewer comments into the manuscript.

1.1 Introduction

Microbes that use carbon monoxide (CO) as an energy and/or carbon source play important roles in a variety of biological contexts. Computational models estimate that CO-oxidizing bacteria in soils remove between 145 and 163 Tg of CO from the atmosphere each year.[1–4] Recent metagenomic, proteomic, and transcriptomic analyses have revealed that CO oxidation is widespread and aids survival in aerobic heterotrophic bacteria under conditions of carbon starvation.[5,6] Microbes that metabolize dissolved CO, largely derived from photodegradation of photochromic dissolved organic matter,[7,8] are also found in freshwater and marine environments under both anoxic and oxygen-replete conditions.[9,10] Important symbiotic relationships are believed to exist between plants and CO metabolizers in the rhizosphere,[11,12] and CO-metabolizing bacteria likely provide nutrients to a gutless marine worm, *Olavius algarvensis*, in nutrient-poor coastal sediments.[13,14] Numerous organisms that populate the microbiomes of humans and other higher organisms also support CO metabolism, and there is some evidence that CO metabolizers in the gut may help maintain CO balance.[15] The natural diversity of CO metabolizing organisms underscores the importance of CO as a biological resource, despite its toxicity in most organisms at high concentrations.

To metabolize CO, microbes use the enzyme CO dehydrogenase (CODH), which catalyzes the reversible oxidation of CO to CO_2 via Equation 1.

$$CO + H_2O \rightleftharpoons CO_2 + 2\,H^+ + 2\,e^- \quad (1)$$

CODH enzymes belong to one of two broad classes, Ni,Fe-CODHs or Cu,Mo-CODHs. Enzymes in each of these classes employ unique metallocofactors, exhibit distinct structural features, and operate under different oxygen tensions and CO concentrations.[10,16–19] Both enzyme classes

require myriad accessory proteins to assemble active site metallocofactors, and as a result, production of CO oxidation systems incurs a high energetic cost.[20–27]

Herein, we examine two nonhomologous, CO-dependent transcriptional activators known to regulate oxidative CO metabolism in microbes, CO oxidation activator (CooA) and regulator of CO metabolism (RcoM). We compare the biochemical sensory mechanisms employed by these proteins to regulate CODH expression and connect functional sensing properties to the distinct physiological contexts in which these proteins operate. In light of recent studies uncovering tremendous breadth and diversity of CO metabolizers present in the environment,[5,6] we highlight evidence for the existence of additional uncharacterized regulators of microbial CO metabolism.

1.2 Overview of microbial CO oxidation

A variety of bacteria and archaea, found in both anoxic and oxygen-replete 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.[28] Many microorganisms that utilize anaerobic CO metabolism have been identified and characterized, including carboxydotrophic extremophiles found near undersea hydrothermal vents or volcanic hot springs,[29–31] as well as photosynthetic, acetogenic, and methanogenic microorganisms found in a variety of anoxic environments.[32] Aerobic CO metabolizers occupy an equally diverse range of niches and are found in soils, sediments, and aquatic environments.[5,6,10] Many of these aerobic organisms are carboxydovores, meaning CO oxidation supports the organism's metabolic needs, but CO cannot act as its sole energy source. Aerobic carboxydovores

utilize CO at ambient concentrations (0.1–25 nM),[1] which are much lower than those typically found in anoxic environments.[5,33,34]

Both anaerobic and aerobic microorganisms harness energy from oxidative conversion of CO to CO_2, as shown in Equation 1.[9] In CO oxidation, CODH generates electrons that drive a variety of reductive processes. In anaerobic CO-oxidizing organisms, these electrons may be used in sulfate reduction to sulfide (desulfurication), proton reduction to molecular hydrogen (hydrogenogenesis), and carbon dioxide reduction to acetate (acetogenesis) or methane (methanogenesis).[17] In aerobic CO-oxidizing bacteria, the terminal oxidant is most commonly molecular oxygen (aerobic respiration), although CO oxidation may also drive dissimilatory nitrate reduction or denitrification.[10, 35]

Anaerobic and aerobic microorganisms employ divergent pathways to fix carbon derived from CO. In anaerobic bacteria and archaea, the CO_2 produced by CO oxidation is converted to acetyl-coenzyme A (acetyl-CoA) via the Wood-Ljungdahl pathway.[36,37] In this pathway, CODH operates in reverse and catalyzes the reduction of CO_2 to produce CO (the reverse reaction as shown in Equation 1). The CO is subsequently shuttled to acetyl-CoA synthase (ACS), which condenses CO, CoA, and a methyl group (delivered by a corrinoid iron-sulfur protein) to produce acetyl-CoA. The methyl group on the corrinoid iron-sulfur protein is regenerated through the six-electron reduction of CO_2 in a tetrahydrofolate-dependent process.[38,39] Alternatively, in aerobic and purple nonsulfur bacteria, CO_2 produced by the oxidation of CO is fixed via the Calvin-Benson-Bassham cycle.[40] It is important to note that CODH has distinct roles in CO oxidation and

[1]Range of CO concentrations in aerobic environments estimated from CO mixing levels quantified in the atmosphere and soils.[6,33,34] To approximate aqueous CO concentrations, CO mixing levels were converted to molar solubility using Henry's Law (K_H = 0.00099 M/bar for CO in water at 25 °C).[146]

CO fixation; the nature of the specific CODH employed and the relative concentrations of CO and CO_2 determine whether CO is used as an energy and/or carbon source.

Anaerobic CO metabolism

Ni,Fe-CODHs, found in anaerobic carboxydotrophs, are oxygen sensitive enzymes that catalyze the interconversion of CO and CO_2, as shown in Equation 1, using a [Ni-4Fe-4S] cofactor.[16,41–43] Ni,Fe-CODHs are encoded in genes found on *acs*, *cdh*, and *coo* operons depending on their functional context (Figure 1.1). Proteins encoded by *acs* and *cdh* operons are associated with CO_2 fixation via the Wood-Ljungdahl pathway.[36,37] Conversely, genes found on *coo* operons encode CODH along with accessory proteins needed to liberate energy from CO oxidation when coupled with proton reduction (hydrogen formation) or another reductive process described above.[44–47] These operons exhibit broad genetic diversity and may bear genes that encode accessory proteins associated with CODH [Ni-4Fe-4S] assembly (CooC, CooT, and CooJ),[20,21,24–26] electron transport (CooF),[23] and hydrogen production (CooMKLXUH).[22,27]

CO-dependent transcriptional regulation is primarily observed for *coo* operons, but not *cdh* or *acs* operons. Acetogens and methanogens require an operative Wood-Ljungdahl pathway under all growth conditions. As a result, expression of *cdh* and *acs* operons is not affected by the presence or absence of CO in these organisms.[48] In direct contrast, CO-dependent transcriptional activation of *coo* operons is well established and was first identified by Bonam *et al.* in *Rhodospirillum rubrum*, a purple, nonsulfur bacterium capable of facultative CO oxidation.[49] Follow-up experiments demonstrated that this regulation was dependent on the CO-sensing transcription factor, CooA.[50] In *Carboxydothermus hydrogenoformans*, a thermophilic carboxydotroph, it has been proposed that two CooA paralogs coordinately control CO metabolism under different CO concentration regimes by differential regulation of *coo*- and *acs*-like operons.[51]

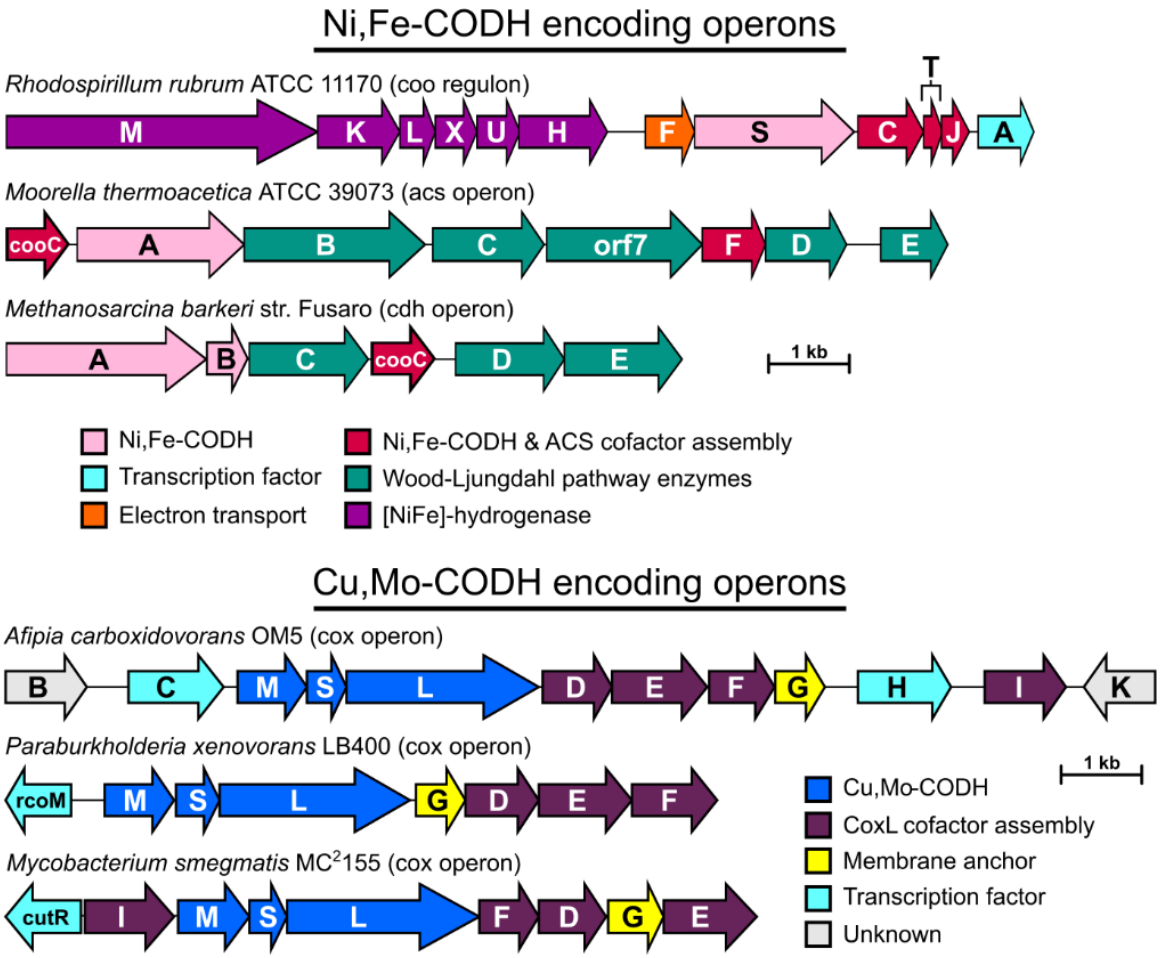


Figure 1.1. Exemplary operons encoding Ni,Fe-CODH and Cu,Mo-CODH. UniProt accession codes for Ni,Fe-CODH genes depicted are as follows: *Rhodospirillium rubrum*, Q2RUG7; *Moorella thermoacetica*, P27988; *Methanosarcina barkeri*, Q46G04. UniProt accession codes for Cu,Mo-CODH genes depicted are as follows: *Afipia carboxidovorans*, P19919; *Paraburkholderia xenovorans*, Q13R29; *Mycobacterium smegmatis*, I7F6J6.

Aerobic CO metabolism

CO oxidation in aerobic bacteria is catalyzed by O_2-tolerant Cu,Mo-CODHs.[52–54] These enzymes are heterotrimeric complexes of CoxS, CoxM, and CoxL.[18] The large subunit (CoxL) harbors a unique molybdenum- and copper-containing active site that catalyzes CO oxidation, and resulting electrons are shuttled by [2Fe-2S] clusters and a flavin adenine dinucleotide (FAD) cofactor, found in the CoxS and CoxM subunits, respectively. Genes encoding Cu,Mo-CODH are found within carbon monoxide oxidase, or *cox*, operons. These operons contain *coxS*, *coxM*, and *coxL* in addition to a variable set of accessory genes (Figure 1.1).10 From studies of the *Afipia carboxidovorans* (formerly *Oligotropha carboxidovorans*) CODH, *coxDEF* and *coxI* accessory genes were determined to facilitate posttranslational assembly of the CoxL bimetallic active site, and *coxG* anchors the Cu,Mo-CODH complex to the cytoplasmic membrane.[55–58]

Transcription of *cox* operons is regulated by both CO and organic carbon availability. Addition of CO to cell cultures has been shown to increase the abundance or apparent activity of CODH in several bacteria species.[59–61] In select microbes, this behavior is mediated by activity of the CO-sensing transcription factor, RcoM.[62] The mechanism of CO-dependent activation in other systems may be mediated by putative CO-sensing transcription factors discussed at the end of this review (see below). In addition to CO-dependent transcriptional activation, Cu,Mo-CODH activity was enhanced when aerobic soil bacteria were challenged by carbon limitation and diminished when cultures were supplemented with an organic carbon source.[5,59,63] The mechanism behind this regulatory behavior is poorly understood, but likely involves regulators associated with carbon catabolite repression, organic carbon-sensing, and CO sensing.

Given the energy input required to generate CO-oxidizing machinery, microorganisms using these proteins must tightly regulate expression of genes involved in CO oxidation. In these

microorganisms, rapid response to environmental stimuli is often accomplished through transcriptional regulation, and many CO-metabolizing microorganisms employ CO-sensing transcription factors to regulate expression of proteins associated with CO oxidation. Despite the genetic diversity exhibited by CO-metabolizing machinery in microorganisms, only two CO-sensing transcription factors have been characterized: CooA and RcoM.

1.3 CooA, the archetypical CO-sensing transcription factor

CooA is a member of the cyclic AMP receptor protein/fumarate and nitrate reductase (CRP/FNR) structural superfamily (COG0664), a well-studied and diverse group of bacterial transcription factors.[64] CRP/FNR proteins have an N-terminal regulatory domain that facilitates homodimerization via a leucine zipper motif and a C-terminal, DNA-binding domain bearing a winged helix-turn-helix motif.[65] In CooA, each regulatory domain in the homodimeric structure binds a single molecule of heme. CO binding to Fe(II) heme in the regulatory domain allosterically activates CooA to bind promoter sites upstream of the *coo* operon and recruit RNA polymerase (RNAP) for transcription of downstream genes.[66,67] CooA binds to a palindromic sequence of DNA composed of a set of 5 base pair (bp) inverted repeats spaced 6 bp apart, an archetypical motif of CRP/FNR-targeted promoters.[50,68,69] Random mutagenesis approaches have identified CooA variants that exhibit CO-dependent DNA binding with little to no transcriptional activation activity *in vivo*.[66,70] These results led to the identification of "activating regions" in the effector binding and DNA binding domains which directly bind and recruit RNAP to facilitate transcription.

CO binding to heme was first characterized in CooA from *R. rubrum* (*Rr*CooA, UniProt ID Q2RUG3). This homolog bears a six-coordinate, ferric heme axially coordinated by protein-derived cysteine-thiolate (Cys75) and N-terminal proline (Pro2).[71,72] The negatively charged

thiolate stabilizes the ferric oxidation state, and as a result, *Rr*CooA exhibits a low reduction midpoint potential (average -290 mV vs. Normal Hydrogen Electrode [NHE]),[73] which corresponds to the midpoint potential for *Rr*CODH activation (-316 mV vs Standard Hydrogen Electrode [SHE]).[74] Given that *R. rubrum* is a facultative anaerobe, it is crucial that *Rr*CooA does not activate expression of oxygen-sensitive Ni,Fe-CODH under aerobic conditions. Thus, the low reduction potential of the *Rr*CooA heme enables this protein to act as a redox (and by proxy, oxygen) sensor as well as a CO sensor. Upon heme iron reduction to the ferrous oxidation state, the axial Cys94 is replaced by a neutral histidine (His77) ligand (Figure 1.2A). CO subsequently binds to the Fe(II) heme, replacing the weakly-bound Pro2 ligand and enhancing protein affinity for its target promoter upstream of the *coo* operon.[72,75–79] Importantly, these *in vitro* observations corroborate early *in vivo* data, which showed loss of CODH protein synthesis within 5 min of oxygen exposure.[49] At present, the source of electrons for *Rr*CooA reduction is not known, but reduction could be achieved by soluble, low-potential ferredoxins.

Interestingly, the redox-mediated ligand switch observed in *Rr*CooA does not occur in another well-studied CooA homolog from the thermophilic carboxydotroph *C. hydrogenoformans* (*Ch*CooA, UniProt ID Q3AB29). Although this homolog bears a cysteine residue at a position analogous to that observed in *Rr*CooA, spectroscopic data are not consistent with the presence of a thiolate-bound ferric heme.[80] Instead, a histidine ligand (His85) remains bound to the *Ch*CooA heme in the ferric, ferrous, and ferrous-CO-bound states. As in *Rr*CooA, the heme in *Ch*CooA is constitutively six-coordinate with the N-terminal amine group likely bound to heme in the ferric and ferrous states, but replaced by CO. Importantly, the lack of an axial cysteine(thiolate) ligand increases the heme reduction midpoint potential (average +190 mV vs. NHE) and thereby abrogates a redox/oxygen sensing function.[80] The lack of redox sensing in *Ch*CooA is consistent

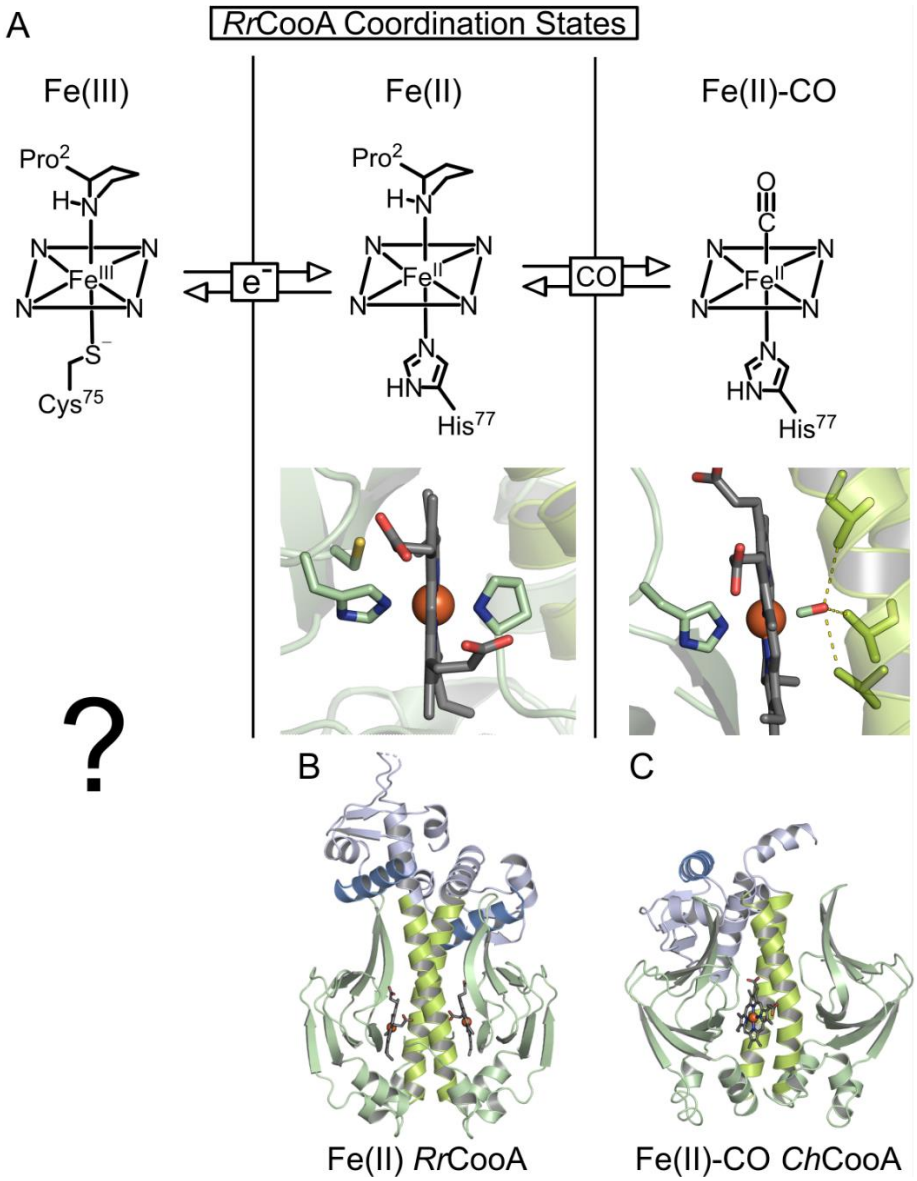


Figure 1.2. Structures and coordination states of CooA. A) Coordination structures of *Rr*CooA in the Fe^{III}, Fe^{II}, and Fe^{II}-CO states. The Fe^{II} and Fe^{II}-CO images are accompanied with the corresponding structure from 1FT9 and 2HKX, respectively. B) Structure of inactive, Fe^{II} CooA from *R. rubrum*, PDB 1FT9. C) Structure of partially active CooA from *C. hydrogenoformans* where only one monomer has retained its heme, PDB 2HKX. The question mark denotes that no high resolution structure of CooA exists in the Fe^{III} oxidation state. For each structure, the effector binding domain is shown in pale green, DNA binding domain in pale blue, and heme cofactors as dark grey. The F-helices, which make close contacts with DNA, are dark blue. The C-helices, which form the dimer interface, are yellow-green. CooA UniProt accession codes are: *Rr*CooA, Q2RUG3 and *Ch*CooA, Q3AB29.

with the biological niche of *C. hydrogenoformans*, which is an obligate anaerobe and only survives under anoxic conditions.

Despite differing heme environments, *Rr*CooA and *Ch*CooA may share an allosteric activation model that invokes a structural change upon CO binding to Fe(II) heme (Figure 1.2). Unique conformations are associated with inactive Fe(II) *Rr*CooA (Figure 2A) and a constitutively active N127L/S128L *Ch*CooA (Figure 2B).[65,81] The active structure of N127L/S128L *Ch*CooA closely matches that of cAMP-bound CRP.[82] 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 binding.[83] This rigid body rotation may be facilitated by CO binding through loss of the Pro2 heme ligand and subsequent rotation of the C-helices at the dimer interface.[84–88] Citing constitutive activity of a Q4C/M77C *Ch*CooA variant containing a disulfide linkage, Tripathi and Poulos suggested that a structural transition to the active conformation may be facilitated by stabilizing interactions between the N-terminal loop and DNA-binding domain.[89]

While a well-enumerated structural model for CO-dependent allosteric activation in CooA exists, CO-induced changes in protein dynamics represent a heretofore unrecognized factor in CooA allosteric activation. Biophysical investigations of a growing number of bacterial transcriptional regulators, including CRP, show that effector-induced changes in protein dynamics play a key role in allosteric activation.[90–93] Specifically, effector binding can lead to changes in fast (ps-ns timescale) protein motions, giving rise to changes in conformational entropy. CooA exhibits conformational heterogeneity that is modulated upon ligand binding, suggesting that dynamics could play an important role in CooA allosteric regulation. CO recombination following

flash photolysis exhibited nonexponential behavior when probed by sub-picosecond mid-IR spectroscopy.[94] Additionally, biphasic kinetics for CO association and dissociation were observed using a combination of rapid mixing electronic absorption, flash photolysis, and time-resolved resonance Raman spectroscopy.[77] These data provide spectroscopic evidence that Fe(II)–CO CooA samples multiple conformational states on the fast (ns-ps) timescales. Recent experiments using site-directed spin label electron paramagnetic resonance spectroscopy demonstrated that Fe(III) *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.[96]

The CO affinity of CooA is relatively weak; however, CO binding to heme is cooperative, and active CO-bound CooA exhibits a high affinity for its target promoter DNA. Using the Hill equation to fit CO titration data, a $P_{1/2}$ value of 2.2 μM with a Hill coefficient (*n)* of 1.4 was observed for *Rr*CooA, indicative of relatively weak, yet cooperative CO binding.[77] These data suggest that high (μM) concentrations of CO are required to activate CooA, but full activation occurs rapidly once the appropriate CO levels are reached due to positive cooperativity. CO binding affinity in *Rr*CooA appears to be tuned to that of the *Rr*Ni,Fe-CODH, which exhibits K_M for CO of 32 μM and a K_d for CO of 4.2 μM.[96] Furthermore, Fe(II)–CO CooA exhibits a high affinity (K_d = 6-26 nM) for its target promoter site,[95,97] implying that upregulation of *coo* expression occurs promptly upon activation of CooA. While the kinetics of CO-dependent gene activation have not been studied *in vitro,* expression of CODH was observed within 5 min of CO exposure *in vivo* in *R. rubrum*.[49]

Taken together, these observations present a model for physiological response to CO in anaerobic CO metabolizers bearing monofunctional CODHs. When these microorganisms experience a spike in local CO concentration, CooA is quickly activated and binds upstream of the

coo operon. Transcriptional activation leads to rapid expression of CODH and other accessory proteins needed to capitalize on an abundant, but potentially fleeting source of energy. In facultative anaerobes, the presence of CO must also coincide with the absence of oxygen to prevent immediate inactivation of CODH. This dual sensing is accomplished by CooA homologs bearing heme with a low reduction midpoint potential, which is readily oxidized, and thereby inactivated, in the presence of oxygen.

Inspired by recent studies interrogating the structural and genomic diversity of Ni,Fe-CODHs,[17,28,38,41,98] we assessed the diversity of genomic contexts of CooA using genome neighborhood network analysis.[99] We assembled a curated list of 519 CooA homologs using BLAST with *Rr*CooA and *Ch*CooA as query sequences, ensuring that all identified sequences conserved both the heme-coordinating histidine residue and residues within the DNA recognition helix (see Appendix 1). Using these CooA sequences, we developed a sequence similarity network (SSN) using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST).[100] We then investigated the neighboring genes surrounding *cooA* and classified the genomic contexts into five categories: *cooS*, *cooMKLXUH* (Ni,Fe-hydrogenase), *cooS + cooMKLXUH*, *cowN*, and unknown (Table 1.1). Genes associated with aerobic CO oxidation were not observed in the genomic context of CooA homologs identified in our network.

According to our analysis, CooA is strongly associated with Ni,Fe-CODH, Ni,Fe-hydrogenase and CowN, a small nitrogenase accessory protein that has been shown to alleviate CO-dependent inhibition of Mo-nitrogenase.[101] Previous reports demonstrated that CowN is also associated with RcoM and that *cowN* is upregulated in response to CO in a CooA- or RcoM-dependent manner.[101–103] While the precise mechanism of CowN-dependent protection against CO is unknown, a recent report from Medina *et al.* showed that CowN directly interacts with Mo-

nitrogenase during turnover and reduces the kinetic inhibition constant (K_I) by five- to ten-fold without significantly altering substrate K_M.[102]

Mapping the genomic context of each CooA homolog onto its corresponding node in the SSN reveals a connection between CooA regulatory targets and protein sequence (Figure 1.3). At an alignment score cutoff of 35 (37% sequence identity), a supercluster of CooA nodes emerges with distinct sub-clustering based on genomic context, as well as one additional cluster associated with *cooMKLXUH*. This organization becomes clearer when the clustering stringency is increased by increasing the cutoff score. The *cowN*-associated proteins assemble into two large subclusters as they diverge from homologs associated with *cooS* and *cooMKLXUH*. At the high alignment score, most CooA proteins that are associated with anaerobic CO oxidation remain clustered together, including *Ch*CooA, with few outlying clusters. One of these smaller clusters includes *Rr*CooA, and further inspection of this cluster indicates that these proteins conserve the redox switching ligand, Cys75 in *Rr*CooA. This observation suggests that this cluster represents CooA proteins with high redox potentials, as has been shown with *Rr*CooA.[73] These homologs are unique because they act as dual redox and CO sensors that ensure CO metabolizing operons are only upregulated under both anaerobic and CO-replete conditions.

1.4 RcoM, a high affinity CO sensor that regulates aerobic CO oxidation

Genes encoding the heme-dependent transcription factor RcoM (UniProt ID: Q13YL3, *Px*RcoM-1; Q13IY4, *Px*RcoM-2) were originally identified upstream of *coxMSL* genes in the soil bacterium *Paraburkholderia xenovorans* in 2008, more than ten years after the discovery of CooA.[62] We note that the two RcoM paralogs identified in *P. xenovorans* are 88% identical, and no significant different in functional or biophysical characteristics have been identified between these two proteins. RcoM exhibits a unique domain architecture that includes an N-terminal, heme-

Table 1.1. Distribution of genomic contexts for CooA- and RcoM-encoding genes.[a]

CooA	
genomic context	**number of genes**
cooS	194
cooS + *cooMKLXUH*	69
cooMKUXLS	23
cowN	199
unknown	34
total	**519**

RcoM	
genomic context	**number of genes**
coxMSL	156
cowN	99
cooS	31
unknown	25
total	**311**

[a]The genomic contexts of cooA and rcoM were classified based on the presence of genes associated with CO oxidation (*cooS, coxMSL*), hydrogen production (*cooMKLXUH*), and CO protection of nitrogenase (*cowN*). CooA and RcoM homologs encoded by genes that do not appear nearby any of these genes were classified with unknown genomic context. Further details are provided in the supplemental material.

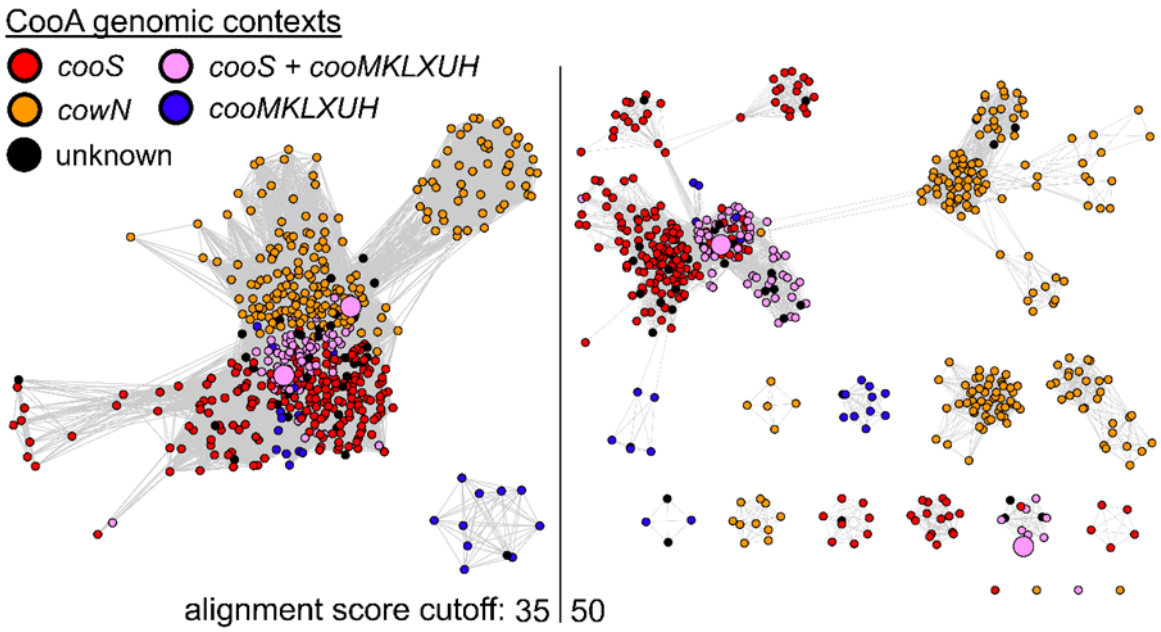


Figure 1.3. The CooA protein sequence similarity network colored according to genomic context. The network is displayed at alignment score cutoffs of 35 (left) and 50 (right). These cutoffs correspond to 37% and 44% average sequence identity, respectively. Node colors correspond to *cooA* genomic contexts consisting of genes within 10 genes upstream or downstream. Nodes corresponding to *Ch*CooA and *Rr*CooA are enlarged for context.

binding, regulatory domain with a PAS (Per-Arnt-Sim) fold and a C-terminal, DNA binding domain with a LytTR fold (Figure 1.4A).[104–107] We are beginning to understand the biophysical properties of RcoM. Size exclusion chromatography and analytical ultracentrifugation sedimentation equilibrium experiments indicate that *Px*RcoM-1 is primarily homodimeric, that dimerization occurs via the heme-binding PAS domain, and that dimerization does not require heme (Figure 1.4B).[108] These observations are consistent with the behavior of other prokaryotic PAS domain-containing proteins, which tend to form homooligomers in solution.[104] These observations represent an important first step in characterizing the structure of RcoM, which is the only known fusion of a PAS sensory domain and a LytTR DNA-binding domain.

In a manner similar to CooA, RcoM uses heme to sense CO. Reduction of Fe(III) heme results in a redox-mediated ligand switch in which a charged cysteine(thiolate) ligand (Cys94) is replaced by a neutral methionine ligand (Met104, Figure 1.4C).[109,110] In *Px*RcoM-1, one of two RcoM paralogs in *P. xenovorans*, Met104 is replaced by CO, resulting in activation of the protein to bind to its promoter upstream of *coxM*.[111,112] The redox-mediated ligand switch observed in RcoM is analogous to that seen in *Rr*CooA. While this Cys residue is completely conserved amongst RcoM homologs found upstream of *coxM*,[108] we note that in our hands, *Px*RcoM-1 bearing Fe(II) heme does not undergo rapid oxidation in the presence of oxygen. This empirical observation suggests that the RcoM heme may exhibit a relatively high redox potential (akin to that observed in *Ch*CooA), despite bearing an axial Cys ligand in the Fe(III) oxidation state (like *Rr*CooA). Currently, there is no reported redox potential measurement for any RcoM homolog, and the functional implications of redox-mediated ligand switching remains unresolved.

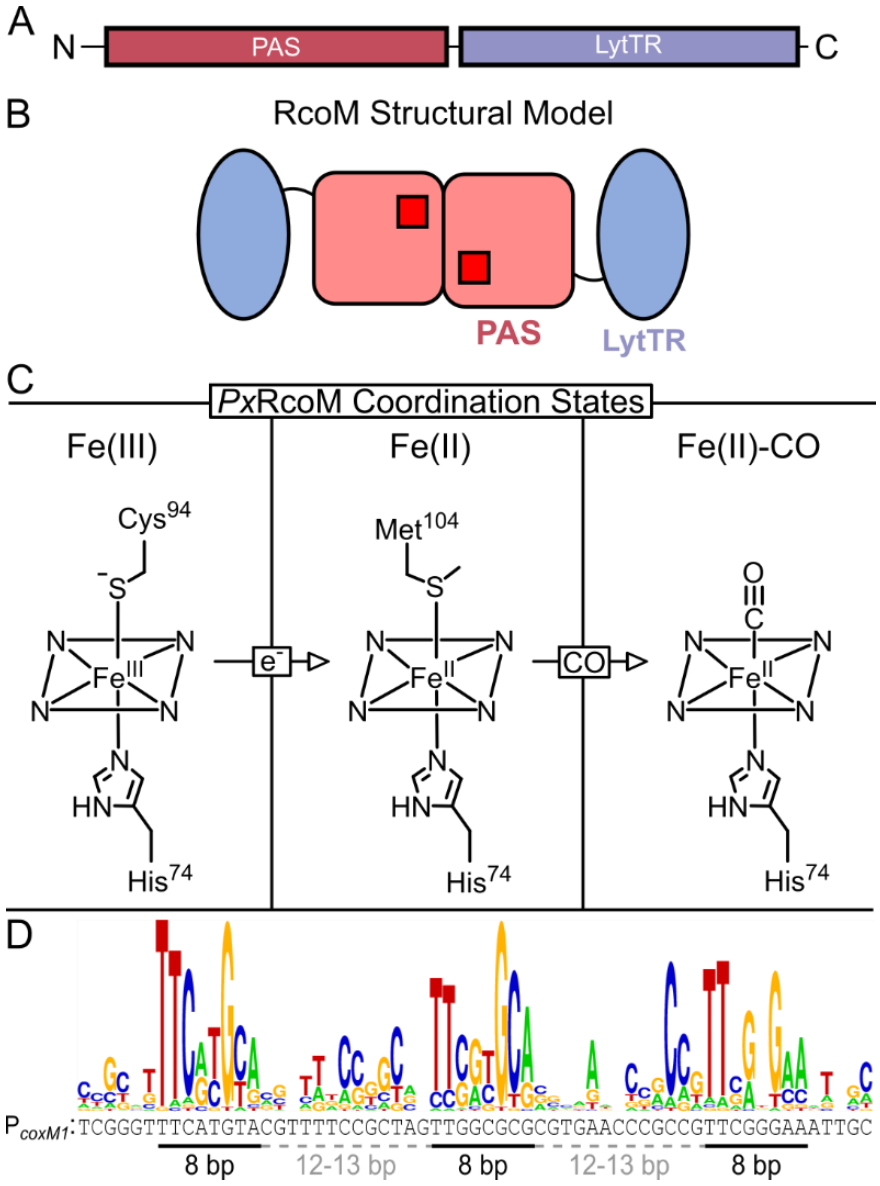


Figure 1.4. Domain organization and coordination states of RcoM. A) Domain architecture of RcoM. Amino acid numbering depicts the start and end of the PAS and LytTR domains as defined by InterPro (PAS: IPR035965, LytTR: IPR046947). B) Cartoon model of RcoM shown with the heme-binding, PAS domain in red along with the DNA-binding LytTR domain shown in blue. C) Coordination structures of *Px*RcoM-1 in the Fe^{III}, Fe^{II}, and Fe^{II}-CO states. D) Logo plot derived from alignment of *coxM* promoters associated with RcoM revealing a triplet direct repeat sequence motif. The native promoter sequence recognized by *Px*RcoM-1, P_{coxM1}, is shown below the plot. RcoM UniProt accession codes are: *Px*RcoM-1, Q13YL3 and *Px*RcoM-2, Q13IY4.

Unlike CooA, RcoM exhibits very high affinity for CO. Several independent research groups have observed Fe(II)-CO heme in recombinantly expressed RcoM, hinting at the high CO binding affinity of this protein.[62,112,113] Consistent with this empirical observation, spectroscopic analysis of CO binding and dissociation kinetics revealed a nanomolar binding affinity (K_d = 4 nM) for the *Px*RcoM-2 paralog.[114] This remarkably high CO binding affinity was largely attributed to a very slow CO dissociation rate constant (k_{off} = 6.4•10^{-5} s^{-1}). Using ultrafast pump-probe spectroscopy, the authors observe fast and complete geminate recombination and suggest that a very small escape probability for RcoM (0.5% on the nanosecond timescale) contributes to the slow CO dissociation rate, although we caution that thermal and photolytic ligand dissociation mechanisms may differ.[114] At present, there is no high-resolution structure available for any RcoM homolog, and the heme pocket residues that may contribute to this protein's remarkably high CO binding affinity have not been identified experimentally.

The six-coordinate ferrous heme of RcoM enables CO binding selectivity in aerobic environments where oxygen, another potential heme ligand, is abundant. A distinctive feature of these CO sensing transcription factors is the presence of a sixth, lower affinity protein-derived ligand. In *Rr*CooA, this functionally important, lower affinity ligand is Pro2.[87,115] By analogy, Met104 fulfills this role in RcoM.[115] Exogenous ligand binding is attenuated by this sixth, protein-derived ligand, which must first dissociate from heme for another molecule to bind.[77,113] Due to differences in electronic structure, CO exhibits significantly slower off rates and higher affinities for ferrous heme than oxygen. As a result, only CO, but not oxygen, replaces the endogenous sixth ligand and binds to ferrous heme.

The RcoM promoter binding site differs significantly from those of other LytTR-containing transcription factors. The RcoM DNA binding motif was originally identified

upstream of *coxM₁* adjacent to a putative σ^{70} RNA polymerase -35/-10 binding sequence.[111] Based on sequence alignment of putative *coxM* promoters, this region has since been refined to consist of three 8-bp imperfect direct repeats of the sequence "TTnnnGCA", with 12- to13-bp spacing regions (Figure 1.4D).[108] The repeat sequence and spacing are consistent with other LytTR-containing transcription factors; however, LytTR-containing proteins typically bind a pair of repeat sites.[116–121] Despite a report that claimed to observe tight RcoM binding to a single repeat site,[114] later experiments demonstrate that all three binding sites are necessary for modest affinity, cooperative RcoM binding ($[P]_{1/2}$ = 250 ± 10 nM).[108]

Experiments exploring the activity of *coxM*$_1$ promoter variants *in vivo* demonstrated that the three direct repeat DNA sequences are nonequivalent. By screening promoter variants that randomized all but the "TT" motif of the central repeat, Kerby *et al.* identified several sequences with higher CO-dependent activity, indicating that this repeat motif could be readily improved.[111] Substituting the second repeat with an "improved" motif resulted in 2- to 5-fold higher β-galactosidase activity in *E. coli* reporter strains compared to a strain bearing the native promoter. Interestingly, replacement of either the first or third repeat with the improved motif resulted in diminished activity, suggesting that the three sites are nonequivalent in promoting transcription.

Like CooA, RcoM protein sequences cluster according to genomic context (Figure 1.5). The genomic contexts of RcoM were recently explored using the same methodology applied to CooA, described above.[108] As with CooA, RcoM associates with anaerobic CO oxidation genes and CowN; however, *rcoM* genes are also found in the context of aerobic CO oxidation, leading to classification into three separate genomic contexts: *coxMSL*, *cooS*, and *cowN* (Table 1.1). At lower alignment cutoff scores, all RcoM proteins fall into the same cluster with clear organization based on genomic context; however, *cooS*-associated RcoM proteins clearly diverge from those

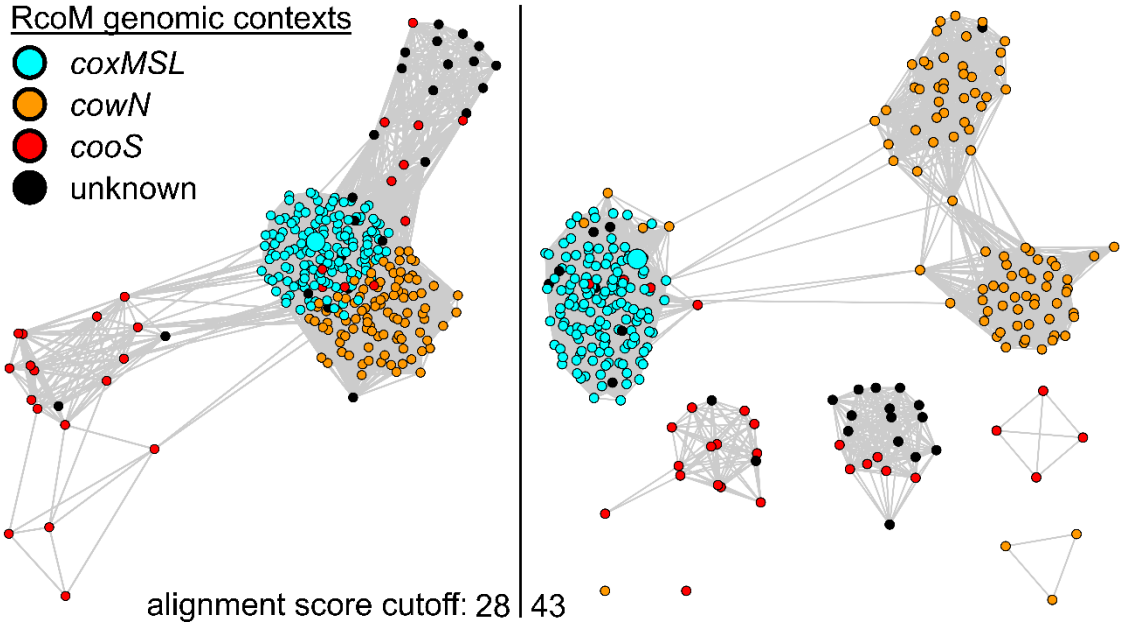


Figure 1.5. Genomic contexts of *rcoM* genes mapped onto the RcoM protein sequence similarity network. The same 95% identity representative node network is displayed at alignment score cutoffs of 28 and 43. These cutoffs correspond to 32% and 37% average sequence identity, respectively. Node colors correspond to *rcoM* genomic contexts, which are labeled according to the identity of genes neighboring *rcoM*, as identified through genome neighborhood analysis. The node corresponding to the experimentally characterized protein, *Px*RcoM-1, is enlarged for context. This network is adapted from Ref 108, and the color scheme is altered to match Figure 1.3.

associated with *cox* and *cowN*. At a higher alignment cutoff score, most *coo*-associated proteins separate into two clusters, and divergence between *cox*- and *cowN*-associated RcoM proteins becomes more apparent.

As with CooA, clustering among *coo*-, *cox*-, and *cowN*-associated RcoM homologs suggests that these RcoM proteins and/or promoters are tuned to activate transcription of downstream genes in disparate CO concentration regimes. Upon closer examination of separate sequence alignments of the *cox*-, *cowN*-, and *cooS*-associated RcoM proteins, we observed that the His heme ligand is conserved across all three genomic contexts, but the redox switching ligands are not. The cysteine(thiolate) ligand is highly conserved among *cox*-associated RcoMs, while it is not found in *cowN*- and *cooS*-associated RcoM sequences (see Appendix 1). The methionine ligand is highly conserved in *cox*-associated RcoMs, but less well conserved among *cowN*- and *cooS*-associated proteins. The functional significance of the conservation of the redox switching ligands is still unknown. In the DNA binding domain, variable conservation of the presumed residues that specifically contact DNA is observed. His218 and Arg254 of *Px*RcoM-1 are predicted to make specific contacts with DNA based on analogy to the structure of a DNA-bound LytTR domain of AgrA.[107] These DNA-contacting residues are highly conserved among *cox*- and *cowN*-associated RcoMs but are less well-conserved among *cooS*-associated proteins. Further biochemical investigation of RcoM homologs from other organisms and genomic contexts is warranted to understand how nature has adjusted the properties of this protein to regulate three distinct CO-related processes.

1.5 Functional Contexts of CooA and RcoM

While CooA and RcoM both use heme to sense CO, these proteins possess remarkably different biochemical properties. CooA exhibits a relatively low affinity for CO and a high affinity

for its corresponding promoter when bound to CO. In contrast, RcoM exhibits a very high affinity for CO and a relatively low affinity for its triplet repeat promoter when bound to CO. The disparate biochemical properties of CooA and RcoM coincide with distinct physiological roles.

On one hand, a majority of *cooA* genes are positioned upstream of *coo* operons, which encompass O_2-sensitive Ni,Fe-CODHs (Table 1.1). Studies of *Rr*CooA discussed here suggest that CooA homologs enable fast transcriptional activation of *coo* genes in response to micromolar levels of CO. While some Ni,Fe-CODH homologs exhibit catalytic efficiencies that approach the diffusion limit (k_{cat}/K_M = 1.7•10^9 M^{-1}s^{-1} for CODH II from *C. hydrogenoformans*, UniProt Q9F8A8),[47] the catalytic properties of Ni,Fe-CODH homologs are highly variable. Reported K_M^{CO} values for various Ni,Fe-CODH range from 20 nM to 30 µM.[96,122–125] We recognize that K_M^{CO} values are not equivalent to equilibrium binding constants (K_d^{CO}), and K_M^{CO} only provides a theoretical upper limit for this value. The identification of high affinity Ni,Fe-CODH proteins suggests that CooA homologs with higher CO affinity might exist in nature.

On the other hand, a majority of *rcoM* genes are positioned upstream of *cox* operons, which encompass O_2-tolerant Cu,Mo-CODHs (Table 1.1 and Figure 1.1). Kinetics studies of Cu,Mo-CODH from *A. carboxidovorans* reveal a k_{cat} for fully active enzyme is between 90-100 s^{-1}, and K_M values for CO range from 0.52 μM to 10.7 μM.[19,126] Taken together, these observations suggest that although Cu,Mo-CODHs may be slower enzymes, they exhibit comparable catalytic efficiencies as high as 1.9•10^8 M^{-1}s^{-1}. Recent experiments monitoring aerobic CO oxidation in liquid cultures of *Mycobacterium smegmatis* have measured an apparent K_M^{CO} of 350 nM and an apparent threshold of CO oxidation of 43 pM.[5] These data indicate that aerobic carboxydovores can oxidize ambient CO from the atmosphere, which is present at very low concentrations (0.1-2 ppm).[6,33,34] Characterized RcoM homologs exhibit low nanomolar binding affinity for CO and are

readily activated at these low concentrations.[114] *In vitro*, several studies have shown that RcoM exhibits a relatively weak affinity for its promoter.[108,111] This weak binding interaction would necessitate the buildup of a larger fraction of CO-bound RcoM before initiation of transcription, leading to speculation that aerobic CO oxidizing microbes have specifically evolved this protein to sense persistently low CO concentrations.[111] However, we recognize that *in vitro* assessment of RcoM promoter affinity does not account for potential interaction with other proteins or molecules *in vivo* that could strengthen RcoM DNA binding. Given that expression of *coxMSL* is also subject to carbon catabolite repression by unknown mechanisms, further studies of RcoM activity *in vivo* are required to further enumerate the regulatory behavior of this protein.

Interestingly, RcoM and CooA both regulate expression of CowN, a small accessory protein that mitigates CO-dependent inhibition of bacterial Mo-containing nitrogenase (Mo-nitrogenase) enzymes.[101–103] A recent study indicates that the Mo-nitrogenase from *Gluconacetobacter diazotrophicus* exhibits 50-60% activity in the presence of CO at 0.001 atm (roughly 1 μM in aqueous solution) relative to that observed in the absence of CO; however, this inhibition was completely abolished in the presence of exogenous CowN.[102] This low micromolar CO concentration threshold for Mo-nitrogenase inhibition is compatible with CO binding affinities of CooA and RcoM, although we note that neither CooA nor RcoM homologs associated with *cowN* genes have been investigated to date. Sequence similarity networks of CooA and RcoM proteins reveal significant sequence divergence between *cowN*-associated homologs and homologs that regulate CO metabolism (Figures 1.3 and 1.5). A comparative functional analysis of *cowN*-associated CooA and RcoM homologs is needed to reveal how these proteins optimally regulate CowN expression. Further, maximal CowN expression is observed under conditions of nitrogen starvation in the presence of CO, but even nitrogen starvation alone increased CowN

expression.[102] Detailed analysis of the regulatory networks that influence *cowN* transcription may reveal how expression of this accessory protein may be linked to CO concentrations and nitrogen availability.

1.6 Putative Novel Regulators of CO Oxidation

Putative Regulators of Anaerobic CO Oxidation

Given the abundance of CO-responsive operons in microbes, it is highly likely that other CO-dependent regulatory proteins exist. In a recent genomic analysis of Ni,Fe-CODHs, Inoue *et al*. identified several classes of transcriptional regulators associated with anaerobic CO oxidation.[99] These classes include proteins in the IscR/NsrR, MarR, AraC, and TetR/AcrR families. IscR and NsrR are members of the Rrf2 structural superfamily that use [2Fe-2S] or [4Fe-4S] clusters to regulate Fe-S cluster assembly[127,128] or sense nitric oxide,[129,130] respectively. Several members of the multiple antibiotic resistance regulator (MarR) family also bind Fe-S clusters,[131] including ChlR, which uses a [4Fe-4S] cluster to regulate pigment biosynthesis in response to oxygen.[132] Several AraC-like transcriptional regulators are heme sensor proteins that control expression of heme iron utilization pathways in bacteria.[133,134] Proteins within the AcrR/TetR family regulate lipid transport and antibiotic efflux in response to lipophilic ligands, although no characterized members of the AcrR/TetR family bind metallocofactors.[135,136] Given the diverse nature of Ni,Fe-CODH and CO metabolism, we emphasize that transcription factors associated with Ni,Fe-CODH may sense a variety of different molecules or cellular states (e.g., redox). Identification of putative transcription factors via genomic analysis serves as a foothold for further biochemical investigation of the regulators and the associated proteins.

Two-component transcriptional regulators also have been identified in the genomic context of anaerobic CO oxidation.[98] These regulators include EnzV/OmpR-like proteins, which are involved in bacterial osmoregulation,[138] and FixJ/NarR-like proteins, which are involved in heme-dependent oxygen sensing.[138] Two proteins which comprise a CO-responsive two-component transcriptional regulatory system, CorQ (UniProt A0A0A7DZ72) and CorR (UniProt B6YWP1) were recently shown to regulate expression of a *coo* operon in the carboxydotrophic hydrogenogenic archaeon *Thermococcus onnurineus*.[139] The sensory component, CorQ, contains a 4-vinyl reductase domain with several conserved Cys residues that may bind an Fe-S cluster.[140] Neither CorQ nor CorR have been biochemically characterized, and it has not been shown whether these proteins bind CO directly.

As summarized above, many of the putative transcriptional regulators associated with anaerobic CO oxidation 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. While RcoM and CooA demonstrate that heme is a well-suited cofactor for CO sensing, many 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.[129,141] Use of Fe-S clusters to directly sense CO would present a novel functional strategy. Alternatively given that Ni,Fe-CODH is oxygen sensitive, these putative Fe-S cluster-dependent regulators may modulate CO metabolism in response to O_2 tension. The potential to discover novel regulatory processes governing anaerobic CO oxidation underscores the need for thorough functional characterization of these putative sensors.

Putative Regulators of Aerobic CO Oxidation

Two putative CO-sensing transcription factors, CoxC (UniProt: Q9KX27) and CoxH (UniProt: F8C101), were identified in the *cox* operon of carboxidotrophic bacterium *A. carboxidovorans*.[59,142] Both CoxC and CoxH possess an N-terminal MHYT domain, an integral membrane domain with putative roles in signal transduction,[143] linked to a C-terminal, DNA-binding LytTR domain similar to that found in RcoM. It has been proposed that the MHYT domain binds metal ions, based on the conservation of potential metal-coordinating methionine and histidine residues.[143] To date, no biochemical investigation of CoxC or CoxH has been carried out, and no evidence has been reported demonstrating that the MHYT domain binds metals.

In *Mycobacterium sp.* str. JC1, Oh *et al.* showed that CutR (UniProt: D5G1X9), a LysR-type transcriptional regulator, was critical for transcriptional activation of genes encoding Cu,Mo-CODH in the presence of CO.[60] The authors demonstrated that CutR is conserved among mycobacteria, is generally encoded directly adjacent to the *cox* operon (Figure 1.1), and they identified a potential CutR binding site (TTAAG-N6-CTTAA) adjacent to a putative -35 region upstream of *coxM*. To date, CutR binding to this site has not been conclusively shown. Presently, it is unclear whether CutR senses CO directly or if CutR functions in coordination with another sensor protein. No characterized LysR-type regulator is known to bind or sense a transition metal ion.[144]

1.7 Concluding Remarks

Microorganisms have evolved elaborate biochemical systems that use CO as a source of carbon and/or energy. Large energy inputs are needed to assemble these CO-metabolizing systems, which require specialized cofactors and accessory proteins. To ensure that these metabolic pathways are only active when CO is a suitable energy source, microorganisms have developed CO-responsive transcriptional regulators. Two such regulators, CooA and RcoM, are non-

homologous transcription factors that each employ unique, heme-dependent allosteric mechanisms to sense and respond to the presence of CO at varied oxygen tensions and CO concentration regimes. While CooA serves as the paradigm CO sensor, we are only just beginning to uncover the mechanistic details of CO-dependent transcriptional regulation in RcoM. Based on genomic context analysis, it appears that CooA and RcoM each primarily regulate anaerobic and aerobic CODH expression, respectively. However, both transcription factors are found in the genomic context of CowN, an accessory protein that mitigates CO-dependent inhibition of Mo-nitrogenases. Future biochemical and biophysical analysis of CooA and RcoM homologues will connect differences in protein sequence and functional properties (ligand and promoter binding affinities) with overlapping and complementary biological contexts. Moreover, the biological, genetic, and biochemical approaches developed to study these CO-sensing archetypes will serve as guiding frameworks to explore novel regulators of microbial CO metabolism.

1.8 Appendix

Generation and analysis of CooA Sequence Similarity Network

To curate a list of CooA homologs, we performed two BLAST searches with *Rr*CooA (UniProt ID: Q2RUG3) and *Ch*CooA (UniProt ID: Q3AB29) as query sequences and obtained the 1,000 most related proteins for each. The resulting 1,598 sequences were aligned using MUSCLE5.1.[146] All sequences that did not contain both the iron-coordinating histidine residue (H77 in *Rr*CooA) and residues within the DNA recognition helix (Arg177, Gln178 in *Rr*CooA) were removed. The remaining 519 CooA sequences were used to generate an SSN using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST).[100] The 10 nearest genes upstream and downstream of *cooA* were accessed manually via the GenBank or via the EFI-Genome Neighborhood Tool (EFI-GNT).[99,147] Each genome neighborhood was classified based on the presence of anaerobic CODH (*cooS*), components of the CO-driven hydrogenase (*cooMKLXUH*), or *cowN*. These classifications were mapped onto the SSN in Cytoscape 3.8,[148] and network edges were filtered based on alignment score to demonstrate relationships between proteins within the SSN.

Generation and analysis of RcoM Sequence Alignments

We generated sequence alignments of RcoM homologs in each genomic context with the goal of comparing protein features between contexts. Starting from the list of RcoM homologs Accession IDs containing 31 *coo*-associated RcoMs, 99 *cowN*-associated RcoMs, 156 *cox*-associated RcoMs, and 25 RcoMs with unknown contexts, we generated sequence alignments of each genomic context separately and one alignment of all RcoM sequences using MUSCLE 5.1.[146]

1.9 References

1. Khalil MAK, Rasmussen RA. 1990. The Global Cycle of Carbon Monoxide - Trends and Mass Balance. Chemosphere 20:227–242.
2. Stein O, Schultz MG, Bouarar I, Clark H, Huijnen V, Gaudel A, George M, Clerbaux C. 2014. On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations. Atmos Chem Phys 14:9295–9316.
3. Liu LC, Zhuang QL, Zhu Q, Liu SQ, van Asperen H, Pihlatie M. 2018. Global soil consumption of atmospheric carbon monoxide: an analysis using a process-based biogeochemistry model. Atmos Chem Phys 18:7913–7931.
4. Ehhalt D, Prather M, Dentener F, Derwent R, Dlugokencky E, Holland E, Isaksen I, Katima J, Kirchhoff V, Matson P, Midgley P, Wang M, 2001. Atmospheric Chemistry and Greenhouse Gases, p. 239–288. *In* Houghton, JT, Ding, Y, Griggs, DJ, Noguer, M, van der Linden, PJ, Dai, X, Maskell, K, Johnson, CA (eds.), Climate Change, 2001: The Scientific Basis. Cambridge University Press, Cambridge, UK and New York, NY, USA.
5. Cordero PRF, Bayly K, Man Leung P, Huang C, Islam ZF, Schittenhelm RB, King GM, Greening C. 2019. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. ISME Journal 13:2868–2881.
6. Bay SK, Dong X, Bradley JA, Leung PM, Grinter R, Jirapanjawat T, Arndt SK, Cook PLM, LaRowe DE, Nauer PA, Chiri E, Greening C. 2021. Trace gas oxidizers are widespread and active members of soil microbial communities. Nat Microbiol 6:246–256.
7. Stubbins A, Uher G, Law CS, Mopper K, Robinson C, Upstill-Goddard RC. 2006. Open-ocean carbon monoxide photoproduction. Deep Sea Research Part II: Topical Studies in Oceanography 53:1695–1705.

8. Ossola R, Gruseck R, Houska J, Manfrin A, Vallieres M, McNeill K. 2022. Photochemical Production of Carbon Monoxide from Dissolved Organic Matter: Role of Lignin Methoxyarene Functional Groups. Environ Sci Technol 56:13449–13460.

9. Oelgeschlager E, Rother M. 2008. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. Arch Microbiol 190:257–269.

10. King GM, Weber CF. 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat Rev Microbiol 5:107.

11. King GM, Crosby H. 2002. Impacts of plant roots on soil CO cycling and soil-atmosphere CO exchange. Glob Chang Biol 8:1085–1093.

12. Del Moral R, Clampitt CA. 1985. Growth of Native Plant Species on Recent Volcanic Substrates from Mount St. Helens. Am Midl Nat 114:374–383.

13. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang Y-J, Shah M, VerBerkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier N. 2012. 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 109:E1173–E1182.

14. Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, Littmann S, Kuypers MMM, Dubilier N. 2015. Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. Environ Microbiol 17:5023–5035.

15. Hopper CP, de La Cruz LK, Lyles K v, Wareham LK, Gilbert JA, Eichenbaum Z, Magierowski M, Poole RK, Wollborn J, Wang B. 2020. Role of Carbon Monoxide in Host-Gut Microbiome Communication. Chem Rev 120:13273–13311.

16. Biester A, Marcano-Delgado AN, Drennan CL. 2022. Structural Insights into Microbial One-Carbon Metabolic Enzymes Ni–Fe–S-Dependent Carbon Monoxide Dehydrogenases and Acetyl-CoA Synthases. Biochemistry 61:2797–2805.

17. Can M, Armstrong FA, Ragsdale SW. 2014. Structure, Function, and Mechanism of the Nickel Metalloenzymes, CO Dehydrogenase, and Acetyl-CoA Synthase. Chem Rev 114:4149–4174.

18. Dobbek H, Gremer L, Meyer O, Huber R. 1999. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S -selanylcysteine. Proc Natl Acad Sci U S A 96:8884–8889.

19. Dobbek H, Gremer L, Kiefersauer R, Huber R, Meyer O. 2002. Catalysis at a dinuclear [CuSMo(O)OH] cluster in a CO dehydrogenase resolved at 1.1-Å resolution. Proc Natl Acad Sci U S A 99:15971–15976.

20. Watt RK, Ludden PW. 1998. 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 273:10019–10025.

21. Jeon WB, Cheng J, Ludden PW. 2001. 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 276:38602–38609.

22. Ragsdale SW. 2004. Life with Carbon Monoxide. Crit Rev Biochem Mol Biol 39:165–195.

23. Singer SW, Hirst MB, Ludden PW. 2006. CO-dependent H2 evolution by Rhodospirillum rubrum: Role of CODH:CooF complex. Biochim Biophys Acta Bioenerg 1757:1582–1591.

24. Jeoung J-H, Giese T, Grünwald M, Dobbek H. 2009. CooC1 from Carboxydothermus hydrogenoformans Is a Nickel-Binding ATPase. Biochemistry 48:11505–11513.

25. Jeoung J-H, Giese T, Grünwald M, Dobbek H. 2010. Crystal structure of the ATP-dependent maturation factor of Ni,Fe-containing carbon monoxide dehydrogenases. J Mol Biol 396:1165–1179.

26. Timm J, Brochier-Armanet C, Perard J, Zambelli B, Ollagnier-De-Choudens S, Ciurli S, Cavazza C. 2017. The CO dehydrogenase accessory protein CooT is a novel nickel-binding protein. Metallomics 9:575–583.

27. Schoelmerich MC, Müller V. 2020. Energy-converting hydrogenases: the link between H2 metabolism and energy conservation. Cell Mol Life Sci 77:1461–1481.

28. Adam PS, Borrel G, Gribaldo S. 2018. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. Proc Natl Acad Sci U S A 115:E1166–E1173.

29. Jannasch HW, Wirsen CO. 1981. Morphological Survey of Microbial Mats Near Deep-Sea Thermal Vents. Appl Environ Microbiol 41:528–538.

30. Karl DM, Wirsen CO, Jannasch HW. 1980. Deep-Sea Primary Production at the Galápagos Hydrothermal Vents. Science 207:1345–1347.

31. Lilley MD, de Angelis MA, Gordon LI. 1982. CH4, H2, CO and N2O in submarine hydrothermal vent waters. Nature 300:48–50.

32. Techtmann SM, Colman AS, Robb FT. 2009. ‘That which does not kill us only makes us stronger’: the role of carbon monoxide in thermophilic microbial consortia. Environ Microbiol 11:1027–1037.

33. Novelli PC, Masarie KA, Lang PM. 1998. Distributions and recent changes of carbon monoxide in the lower troposphere. J Geophys Res Atmos 103:19015–19033.

34. Novelli PC, Masarie KA, Tans PP, Lang PM. 1994. Recent Changes in Atmospheric Carbon Monoxide. Science 263:1587–1590.

35. King GM. 2006. Nitrate-dependent anaerobic carbon monoxide oxidation by aerobic CO-oxidizing bacteria. FEMS Microbiol Ecol 56:1–7.

36. Ragsdale SW. 2008. Enzymology of the Wood-Ljungdahl pathway of acetogenesis. Ann N Y Acad Sci 1125:129–136.

37. Ragsdale SW, Pierce E. 2008. Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. Biochim Biophys Acta Proteins Proteom 1784:1873–1898.

38. Techtmann SM, Lebedinsky A V., Colman AS, Sokolova TG, Woyke T, Goodwin L, Robb FT. 2012. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. Front Microbiol 3:132.

39. Jeoung J-H, Goetzl S, Hennig SE, Fesseler J, Wörmann C, Dendra J, Dobbek H. 2014. The extended reductive acetyl-CoA pathway:ATPases in metal cluster maturation and reductive activation. Biol Chem 395:545–558.

40. Bowien B, Schlegel HG. 1981. Physiology and Biochemistry of Aerobic Hydrogen-Oxidizing Bacteria. Annu Rev Microbiol 35:405–452.

41. Jeoung J-H, Fesseler J, Goetzl S, Dobbek H. 2014. Carbon Monoxide. Toxic Gas and Fuel for Anaerobes and Aerobes: Carbon Monoxide Dehydrogenases, p. 37–69. *In*

Kroneck, PMH, Torres, MES (eds.), Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment. Springer, Dordrecht.

42. Dobbek H. 2018. Mechanism of Ni,Fe-Containing Carbon Monoxide Dehydrogenases, p. 153–166. *In* Ribbe, MW (ed.), Metallocofactors that Activate Small Molecules. Springer, Cham.

43. Stripp ST, Duffus BR, Fourmond V, Léger C, Leimkühler S, Hirota S, Hu Y, Jasniewski A, Ogata H, Ribbe MW. 2022. Second and Outer Coordination Sphere Effects in Nitrogenase, Hydrogenase, Formate Dehydrogenase, and CO Dehydrogenase. Chem Rev 122:11900–11973.

44. Kerby RL, Ludden PW, Roberts GP. 1995. Carbon monoxide-dependent growth of Rhodospirillum rubrum. J Bacteriol 177:2241–2244.

45. Sekar B, Raj S, Seol E, Ainala SK, Lee J, Park S. 2014. Cloning and functional expression of *Citrobacter amalonaticus* Y19 carbon monoxide dehydrogenase in *Escherichia coli*. Int J Hydrogen Energy 39:15446–15454.

46. Lee J, Seol E, Kaur G, Oh Y-K, Park S. 2012. Hydrogen production from C1 compounds by a novel marine hyperthermophilic archaeon Thermococcus onnurineus NA1. Int J Hydrogen Energy 37:11113–11121.

47. Svetlitchnyi V, Peschel C, Acker G, Meyer O. 2001. Two Membrane-Associated NiFeS-Carbon Monoxide Dehydrogenases from the Anaerobic Carbon-Monoxide-Utilizing Eubacterium Carboxydothermus hydrogenoformans. J Bacteriol 183:5134–5144.

48. Arendsen AF, Soliman MQ, Ragsdale SW. 1999. Nitrate-dependent regulation of acetate biosynthesis and nitrate respiration by *Clostridium thermoaceticum*. J Bacteriol 181:1489–1495.

49. Bonam D, Lehman L, Roberts GP, Ludden PW. 1989. Regulation of carbon monoxide dehydrogenase and hydrogenase in *Rhodospirillum rubrum*: effects of CO and oxygen on synthesis and activity. J Bacteriol 171:3102–3107.

50. Fox JD, He Y, Shelver D, Roberts GP, Ludden PW. 1996. Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. J Bacteriol 178:6200–6208.

51. Techtmann SM, Colman AS, Murphy MB, Schackwitz WS, Goodwin LA, Robb FT. 2011. Regulation of Multiple Carbon Monoxide Consumption Pathways in Anaerobic Bacteria. Front Microbiol 2:12.

52. Meyer O, Rajagopalan K v. 1984. Molybdopterin in carbon monoxide oxidase from carboxydotrophic bacteria. J Bacteriol 157:643–648.

53. Bell JM, Colby J, Williams E. 1988. CO oxidoreductase from Streptomyces strain G26 is a molybdenum hydroxylase. Biochem J 250:605–612.

54. Stein BW, Kirk ML. 2015. Electronic structure contributions to reactivity in xanthine oxidase family enzymes. J Biol Inorg Chem 20:183–194.

55. Kaufmann P, Duffus BR, Teutloff C, Leimkühler S. 2018. Functional Studies on Oligotropha carboxidovorans Molybdenum−Copper CO Dehydrogenase Produced in Escherichia coli. Biochemistry 57:2889–2901.

56. Pelzmann A, Ferner M, Gnida M, Meyer-Klaucke W, Meyer O. 2009. 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 284:9578–9586.

57. Maisel T, Joseph S, Mielke T, Bürger J, Schwarzinger S, Meyer O. 2012. The CoxD Protein, a Novel AAA+ ATPase Involved in Metal Cluster Assembly: Hydrolysis of Nucleotide-Triphosphates and Oligomerization. PLoS One 7:e47424.

58. Pelzmann AM, Mickoleit F, Meyer O. 2014. 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 19:1399–1414.

59. Santiago B, Schübel U, Egelseer C, Meyer O. 1999. Sequence analysis, characterization and CO-specific transcription of the cox gene cluster on the megaplasmid pHCG3 of Oligotropha carboxidovorans. Gene 236:115–124.

60. Oh JI, Park SJ, Shin SJ, Ko IJ, Han SJ, Park SW, Song T, Kim YM. 2010. Identification of trans- and cis-Control Elements Involved in Regulation of the Carbon Monoxide Dehydrogenase Genes in Mycobacterium sp Strain JC1 DSM 3803. J Bacteriol 192:3925–3933.

61. Lee JH, Park SW, Kim YM, Oh J-I. 2017. Identification and characterization of the genes encoding carbon monoxide dehydrogenase in Terrabacter carboxydivorans. Res Microbiol 168:431–442.

62. Kerby RL, Youn H, Roberts GP. 2008. RcoM: a new single-component transcriptional regulator of CO metabolism in bacteria. J Bacteriol 190:3336–3343.

63. Weber CF, King GM. 2012. The phylogenetic distribution and ecological role of carbon monoxide oxidation in the genus Burkholderia. FEMS Microbiol Ecol 79:167–175.

64. Körner H, Sofia HJ, Zumft WG. 2003. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol Rev 27:559–592.

65. Lanzilotta WN, Schuller DJ, Thorsteinsson M v., Kerby RL, Roberts GP, Poulos TL. 2000. Structure of the CO sensing transcription activator CooA. Nat Struct Biol 7:876–80.

66. Leduc J, Thorsteinsson M V., Gaal T, Roberts GP. 2001. Mapping CooA·RNA Polymerase Interactions: Identification of Activating Regions 2 and 3 in CooA, the CO-sensing Transcriptional Activator. J Biol Chem 276:39968–39973.

67. Roberts GP, Youn H, Kerby RL. 2004. CO-Sensing Mechanisms. Microbiol Mol Biol Rev 68:453–473.

68. He Y, Shelver D, Kerby RL, Roberts GP. 1996. Characterization of a CO-responsive Transcriptional Activator from *Rhodospirillum rubrum*. J Biol Chem 271:120–123.

69. Kolb A, Busby S, Buc H, Garges S, Adhya S. 1993. Transcriptional Regulation by cAMP and its Receptor Protein. Annu Rev Biochem 62:749–797.

70. He Y, Gaal T, Karls R, Donohue TJ, Gourse RL, Roberts GP. 1999. Transcription activation by CooA, the CO-sensing factor from *Rhodospirillum rubrum*: The interaction between CooA and the C-terminal domain of the α subunit of RNA polymerase. J Biol Chem 274:10840–10845.

71. Thorsteinsson M v, Kerby RL, Conrad M, Youn H, Staples CR, Lanzilotta WN, Poulos TJ, Serate J, Roberts GP. 2000. Characterization of Variants Altered at the N-terminal Proline, a Novel Heme-Axial Ligand in CooA, the CO-sensing Transcriptional Activator. J Biol Chem 275:39332–39338.

72. Shelver D, Thorsteinsson M v., Kerby RL, Chung S-Y, Roberts GP, Reynolds MF, Parks RB, Burstyn JN. 1999. Identification of Two Important Heme Site Residues (Cysteine 75 and Histidine 77) in CooA, the CO-Sensing Transcription Factor of Rhodospirillum rubrum. Biochemistry 38:2669–2678.

73. Nakajima H, Honma Y, Tawara T, Kato T, Park S-Y, Miyatake H, Shiro Y, Aono S. 2001. Redox Properties and Coordination Structure of the Heme in the CO-sensing Transcriptional Activator CooA. J Biol Chem 276:7055–7061.

74. Heo J, Halbleib CM, Ludden PW. 2001. Redox-dependent activation of CO dehydrogenase from Rhodospirillum rubrum. Proc Natl Acad Sci U S A 98:7690–7693.

75. Youn H, Conrad M, Chung S-Y, Roberts GP. 2006. Roles of the heme and heme ligands in the activation of CooA, the CO-sensing transcriptional activator. Biochem Biophys Res Commun 348:345–350.

76. Nakajima H, Nakagawa E, Kobayashi K, Tagawa SI, Aono S. 2001. Ligand-switching Intermediates for the CO-sensing Transcriptional Activator CooA Measured by Pulse Radiolysis. J Biol Chem 276:37895–37899.

77. Puranik M, Nielsen SB, Youn H, Hvitved AN, Bourassa JL, Case MA, Tengroth C, Balakrishnan G, Thorsteinsson M v, Groves JT, McLendon GL, Roberts GP, Olson JS,

Spiro TG. 2004. Dynamics of Carbon Monoxide Binding to CooA. J Biol Chem 279:21096–21108.

78. Aono S, Nakajima H, Saito K, Okada M. 1996. A Novel Heme Protein That Acts as a Carbon Monoxide-Dependent Transcriptional Activator inRhodospirillum rubrum. Biochem Biophys Res Commun 228:752–756.

79. Youn H, Kerby RL, Thorsteinsson M v., Clark RW, Burstyn JN, Roberts GP. 2002. 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 277:33616–33623.

80. Inagaki S, Masuda C, Akaishi T, Nakajima H, Yoshioka S, Ohta T, Pal B, Kitagawa T, Aono S. 2005. Spectroscopic and redox properties of a CooA homologue from *Carboxydothermus hydrogenoformans*. J Biol Chem 280:3269–3274.

81. Borjigin M, Li H, Lanz ND, Kerby RL, Roberts GP, Poulos TL. 2007. Structure-based hypothesis on the activation of the CO-sensing transcription factor CooA. Acta Crystallogr D Biol Crystallogr 63:282–287.

82. Schultz S, Shields G, Steitz T. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 253:1001–1007.

83. Akiyama S, Fujisawa T, Ishimori K, Morishima I, Aono S. 2004. Activation mechanisms of transcriptional regulator CooA revealed by small-angle X-ray scattering. J Mol Biol 341:651–668.

84. Yamamoto K, Ishikawa H, Takahashi S, Ishimori K, Morishima I. 2001. Binding of CO at the Pro2 Side Is Crucial for the Activation of CO-sensing Transcriptional Activator CooA. J Biol Chem 276:11473–11476.

85. Kerby RL, Youn H, Thorsteinsson M v, Roberts GP. 2003. 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 325:809–823.

86. Youn H, Kerby RL, Roberts GP. 2003. The Role of the Hydrophobic Distal Heme Pocket of CooA in Ligand Sensing and Response. J Biol Chem 278:2333–2340.

87. Clark RW, Youn H, Parks RB, Cherney MM, Roberts GP, Burstyn JN. 2004. Investigation of the Role of the N-Terminal Proline, the Distal Heme Ligand in the CO Sensor CooA. Biochemistry 43:14149–14160.

88. Otomo A, Ishikawa H, Mizuno M, Kimura T, Kubo M, Shiro Y, Aono S, Mizutani Y. 2016. 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 120:7836–7843.

89. Tripathi S, Poulos TL. 2018. Testing the N-terminal velcro model of CooA carbon monoxide activation. Biochemistry 57:3059–3064.

90. Popovych N, Sun S, Ebright RH, Kalodimos CG. 2006. Dynamically driven protein allostery. Nat Struct Mol Biol 13:831–838.

91. Tzeng S-R, Kalodimos CG. 2015. The role of slow and fast protein motions in allosteric interactions. Biophys Rev 7:251–255.

92. Capdevila DA, Walsh BJC, Zhang Y, Dietrich C, Gonzalez-Gutierrez G, Giedroc DP. 2020. Structural basis for persulfide-sensing specificity in a transcriptional regulator. Nat Chem Biol 1–6.

93. Tzeng S-R, Kalodimos CG. 2013. Allosteric inhibition through suppression of transient conformational states. Nat Chem Biol 9:462–465.

94. Rubtsov I V, Zhang T, Nakajima H, Aono S, Rubtsov GI, Kumazaki S, Yoshihara K. 2001. Conformational dynamics of the transcriptional regulator CooA protein studied by subpicosecond mid-infrared vibrational spectroscopy. J Am Chem Soc 123:10056–10062.

95. Hines JP, Dent MR, Stevens DJ, Burstyn JN. 2018. 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. Protein Science 27:1670–1679.

96. Ensign SA, Hyman MR, Ludden PW. 1989. Nickel-specific, slow-binding inhibition of carbon monoxide dehydrogenase from Rhodospirillum rubrum by cyanide. Biochemistry 28:4973–4979.

97. Clark RW, Lanz ND, Lee AJ, Kerby RL, Roberts GP, Burstyn JN. 2006. Unexpected NO-dependent DNA binding by the CooA homolog from Carboxydothermus hydrogenoformans. Proc Natl Acad Sci U S A 103:891–896.

98. Inoue M, Nakamoto I, Omae K, Oguro T, Ogata H, Yoshida T, Sako Y. 2019. Structural and phylogenetic diversity of anaerobic carbon-monoxide dehydrogenases. Front Microbiol 10:3353.

99. Zallot R, Oberg N, Gerlt JA. 2019. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. Biochemistry 58:4169–4182.

100. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim Biophys Acta 1854:1019–1037.

101. Kerby RL, Roberts GP. 2011. Sustaining N 2 -Dependent Growth in the Presence of CO. J Bacteriol 193:774–777.

102. Medina MS, Bretzing KO, Aviles RA, Chong KM, Espinoza A, Garcia CNG, Katz BB, Kharwa RN, Hernandez A, Lee JL, Lee TM, lo Verde C, Strul MW, Wong EY, Owens CP. 2021. CowN sustains nitrogenase turnover in the presence of the inhibitor carbon monoxide. J Biol Chem 296:100501.

103. Hoffmann M-C, Pfänder Y, Fehringer M, Narberhaus F, Masepohl B. 2014. NifA- and CooA-Coordinated cowN Expression Sustains Nitrogen Fixation by Rhodobacter capsulatus in the Presence of Carbon Monoxide. J Bacteriol 196:3494–3502.

104. Möglich A, Ayers RA, Moffat K. 2009. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. Structure 17:1282–1294.

105. Galperin MY. 2008. Telling Bacteria: Do Not LytTR. Structure 16:657–659.

106. Zou Z, Qin H, Brenner AE, Raghavan R, Millar JA, Gu Q, Xie Z, Kreth J, Merritt J. 2018. LytTR Regulatory Systems: A potential new class of prokaryotic sensory system. PLoS Genet 14:e1007709.

107. Sidote DJ, Barbieri CM, Wu T, Stock AM. 2008. Structure of the Staphylococcus aureus AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with an Unusual Mode of Binding. Structure 16:727–735.

108. Dent MR, Roberts MG, Bowman HE, Weaver BR, McCaslin DR, Burstyn JN. 2022. Quaternary Structure and Deoxyribonucleic Acid-Binding Properties of the Heme-Dependent, CO-sensing Transcriptional Regulator PxRcoM. Biochemistry 61:678–688.

109. Marvin KA, Kerby RL, Youn H, Roberts GP, Burstyn JN. 2008. The transcription regulator RcoM-2 from Burkholderia xenovorans is a cysteine-ligated hemoprotein that undergoes a redox-mediated ligand switch. Biochemistry 47:9016–9028.

110. Smith AT, Marvin KA, Freeman KM, Kerby RL, Roberts GP, Burstyn JN. 2012. 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 17:1071–1082.

111. Kerby RL, Roberts GP. 2012. *Burkholderia xenovorans* $RcoM_{Bx}$-1, a Transcriptional Regulator System for Sensing Low and Persistent Levels of Carbon Monoxide. J Bacteriol 194:5803–5816.

112. Bowman HE, Dent MR, Burstyn JN. 2016. Met104is the CO-replaceable ligand at Fe(II) heme in the CO-sensing transcription factor *Bx*RcoM-1. J Biol Inorg Chem 21:559–569.

113. Bouzhir-Sima L, Motterlini R, Gross J, Vos MH, Liebl U. 2016. Unusual Dynamics of Ligand Binding to the Heme Domain of the Bacterial CO Sensor Protein RcoM-2. J Phys Chem B 120:10686–10694.

114. Salman M, Villamil Franco C, Ramodiharilafy R, Liebl U, Vos MH. 2019. Interaction of the Full-Length Heme-Based CO Sensor Protein RcoM-2 with Ligands. Biochemistry 58:4028–4034.

115. Pinkert JC, Clark RW, Burstyn JN. 2006. Modeling proline ligation in the heme-dependent CO sensor, CooA, using small-molecule analogs. J Biol Inorg Chem 11:642–650.

116. Nikolskaya AN, Galperin MY. 2002. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. Nucleic Acids Res 30:2453–2459.

117. Ween O, Gaustad P, Havarstein LS. 1999. Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. Mol Microbiol 33:817–827.

118. Reyes D, Andrey DO, Monod A, Kelley WL, Zhang G, Cheung AL. 2011. Coordinated Regulation by AgrA, SarA, and SarR To Control agr Expression in Staphylococcus aureus. J Bacteriol 193:6020–6031.

119. Cheung JK, Rood JI. 2000. The VirR Response Regulator from Clostridium perfringens Binds Independently to Two Imperfect Direct Repeats Located Upstream of the *pfoA* Promoter. J Bacteriol 182:57–66.

120. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, Keck W, Amrein KE, Lange R. 2000. Microarray-Based Identification of a Novel *Streptococcus pneumoniae* Regulon Controlled by an Autoinduced Peptide. J Bacteriol 182:4696–4703.

121. Diep DB, Håvarstein LS, Nes IF. 1996. Characterization of the locus responsible for the bacteriocin production in Lactobacillus plantarum C11. J Bacteriol 178:4472–4483.

122. Domnik L, Merrouch M, Goetzl S, Jeoung J-H, Léger C, Dementin S, Fourmond V, Dobbek H. 2017. CODH-IV: A High-Efficiency CO-Scavenging CO Dehydrogenase with Resistance to O2. Angew Chem Int Ed 56:15466–15469.

123. Benvenuti M, Meneghello M, Guendon C, Jacq-Bailly A, Jeoung J-H, Dobbek H, Léger C, Fourmond V, Dementin S. 2020. The two CO-dehydrogenases of Thermococcus sp. AM4. Biochim Biophys Acta Bioenerg 1861:148188.

124. Parkin A, Seravalli J, Vincent KA, Ragsdale SW, Armstrong FA. 2007. Rapid and efficient electrocatalytic CO2/CO interconversions by Carboxydothermus hydrogenoformans CO dehydrogenase I on an electrode. J Am Chem Soc 129:10328–10329.

125. Hadj-Saïd J, Pandelia ME, Léger C, Fourmond V, Dementin S. 2015. The Carbon Monoxide Dehydrogenase from Desulfovibrio vulgaris. Biochim Biophys Acta Bioenerg 1847:1574–1583.

126. Zhang B, Hemann CF, Hille R. 2010. Kinetic and Spectroscopic Studies of the Molybdenum-Copper CO Dehydrogenase from *Oligotropha carboxidovorans*. J Biol Chem 285:12571–12578.

127. Giel JL, Nesbit AD, Mettert EL, Fleischhacker AS, Wanta BT, Kiley PJ. 2013. Regulation of iron–sulphur cluster homeostasis through transcriptional control of the Isc pathway by [2Fe–2S]–IscR in Escherichia coli. Mol Microbiol 87:478–492.

128. Schwartz CJ, Giel JL, Patschkowski T, Luther C, Ruzicka FJ, Beinert H, Kiley PJ. 2001. IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins. Proc Natl Acad Sci U S A 98:14895–14900.

129. Kommineni S, Lama A, Popescu B, Nakano MM. 2012. Global Transcriptional Control by NsrR in Bacillus subtilis. J Bacteriol 194:1679–1688.

130. Volbeda A, Dodd EL, Darnault C, Crack JC, Renoux O, Hutchings MI, le Bru NE, Fontecilla-Camps JC. 2017. Crystal structures of the NO sensor NsrR reveal how its iron-sulfur cluster modulates DNA binding. Nat Commun 2017 8:1 8:1–10.

131. Perera IC, Grove A. 2010. Molecular Mechanisms of Ligand-Mediated Attenuation of DNA Binding by MarR Family Transcriptional Regulators. J Mol Cell Biol 2:243–254.

132. Ludwig M, Pandelia ME, Chew CY, Zhang B, Golbeck JH, Krebs C, Bryant DA. 2014. 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 289:16624–16639.

133. Amarelle V, Koziol U, Fabiano E. 2019. 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 32:273–291.

134. Vanderpool CK, Armstrong SK. 2003. Heme-Responsive Transcriptional Activation of Bordetella bhu Genes. J Bacteriol 185:909–917.

135. Kang SM, Kim DH, Jin C, Ahn HC, Lee BJ. 2019. The crystal structure of AcrR from Mycobacterium tuberculosis reveals a one-component transcriptional regulation mechanism. FEBS Open Bio 9:1713–1725.

136. Deng W, Li C, Xie J. 2013. The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. Cell Signal 25:1608–1613.

137. Cai SJ, Inouye M. 2002. EnvZ-OmpR Interaction and Osmoregulation in *Escherichia coli*. J Biol Chem 277:24155–24161.

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

139. Kim M-S, Jung H-C, Yang T-J, Youn H. 2015. A Novel CO-Responsive Transcriptional Regulator and Enhanced H 2 Production by an Engineered *Thermococcus onnurineus* NA1 Strain. Appl Environ Microbiol 81:1708–1714.

140. Anantharaman V, Koonin E V., Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307:1271–1292.

141. Crack JC, Green J, Thomson AJ, Brun NE le. 2014. Iron–Sulfur Clusters as Biological Sensors: The Chemistry of Reactions with Molecular Oxygen and Nitric Oxide. Acc Chem Res 47:3196–3205.

142. Fuhrmann S, Ferner M, Jeffke T, Henne A, Gottschalk G, Meyer O. 2003. Complete nucleotide sequence of the circular megaplasmid pHCG3 of Oligotropha carboxidovorans: function in the chemolithoautotrophic utilization of CO, H2 and CO2. Gene 322:67–75.

143. Galperin MY, Gaidenko TA, Mulkidjanian AY, Nakano M. 2001. MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett 205:17–23.

144. Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology 154:3609–3623.

145. Sander R. 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmos Chem Phys 15:4399–4981.

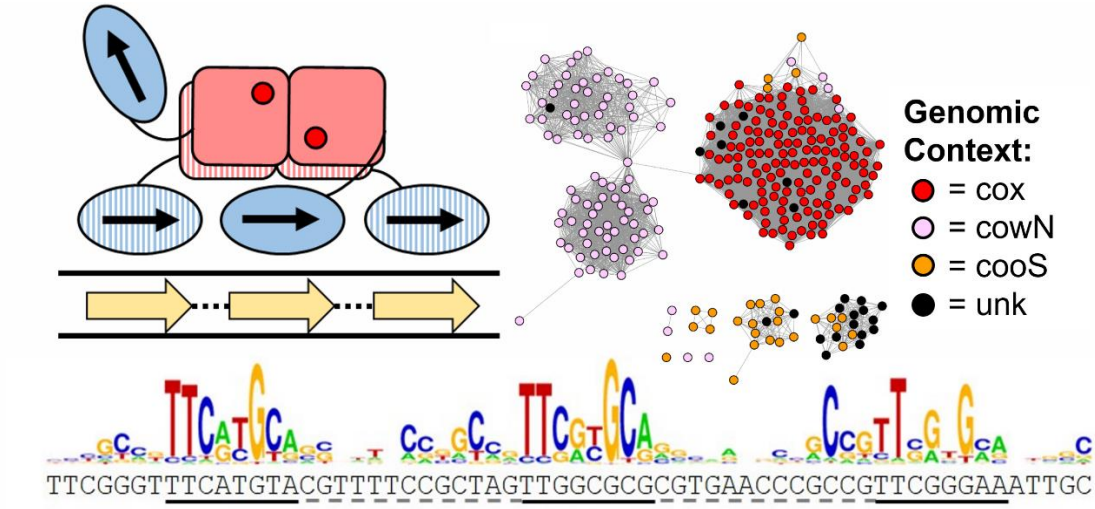
146. Edgar RC. 2021. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. bioRxiv 2021.06.20.449169.

147. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44:D67–D72.

148. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13:2498–2504.

Chapter 2

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



A version of this chapter was published as:

Dent, M. R.*; Roberts, M. G.*; Bowman, H. E.; Weaver, B. R.; McCaslin, D. R.; Burstyn, J. N. Quaternary Structure and Deoxyribonucleic Acid-Binding Properties of the Heme-Dependent, CO-Sensing Transcriptional Regulator *Px*RcoM. *Biochemistry* **2022**, *61* (8), 678–688. https://doi.org/10.1021/acs.biochem.2c00086.

*MRD and MGR contributed equally to this work.

M.R.D. wrote the majority of the manuscript, with help from M.G.R., B.R.W., H.E.B., and J.N.B. Experiments were designed and performed by M.R.D., M.G.R., and H.E.B. B.R.W. conducted bioinformatics analysis. D.R.M. conducted and analyzed the analytical ultracentrifugation experiments. J.N.B. supervised the project and advised experiments.

2.1 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] To regulate expression of the complex 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, is the paradigm in transcriptional regulation via heme-based CO-sensing.[3,7,8] CO binding to Fe(II) heme in the beta-rich effector binding domain allosterically activates CooA to bind DNA via a C-terminal helix-turn-helix (HTH) DNA-binding domain.[9–11] Active CooA binds to its consensus site upstream of the *coo* operon and enhances the transcription of genes linked to anaerobic, oxidative CO metabolism.[5,12–14] RcoM, like CooA, is a single-component transcriptional activator that utilizes heme to sense CO;[6,15–17] however, RcoM exhibits a much higher affinity for CO and bears no homology to CooA.[18–20] Given its unique structural topology as the only known single-component fusion of a sensory PAS (Per-Arnt-Sim) domain and a DNA-binding LytTR domain, RcoM is predicted to exhibit unique DNA binding properties and CO-driven allosteric regulation. By developing a more complete understanding of the DNA binding properties of RcoM, we aim to understand CO sensing and CO metabolism regulation in bacteria.

Originally identified in thirteen prokaryotic organisms by Kerby et al.,[6] the *rcoM* genes were found in three distinct genomic contexts. Homologs were identified through genetic analysis based on the sequence of *Px*RcoM-1, one of two RcoM paralogs from *P. xenovorans*.[6] As is the case for *cooA, rcoM* genes were found adjacent to *cooS* genes, which encode oxygen-sensitive, Fe-Ni CO dehydrogenase (CODH) enzymes.[2,13] All *cooS*-adjacent *rcoM* genes were identified in

anaerobes that do not contain a *cooA* gene, an observation that is consistent with RcoM functioning as a transcriptional regulator of oxygen-sensitive *coo* operons. Unlike *cooA*, *rcoM* genes were also identified adjacent to *cox* operons, which encode oxygen-tolerant, molybdo-iron-sulfur-flavoprotein CODHs.[6,21,22] Thus far, RcoM is the only experimentally characterized transcriptional regulator of *cox* genes. Two other putative CO-sensing transcriptional regulators, CoxC and CoxH, were identified in *O. carboxidovorans* and may regulate the *cox* operon; however, unlike RcoM, CoxC and CoxH appear to be transmembrane proteins based on the presence of an N-terminal MHYT domain.[21,23]. Both *rcoM* and *cooA* genes were also identified directly upstream of *cowN*, a gene that encodes a small (100 amino acid) accessory protein.[24] CowN protects Mo-containing nitrogenases against inhibition by CO binding; however, the molecular mechanism underlying this protection is unknown.[25]

Based on sequence homology, RcoM is predicted to have a unique domain architecture comprised of an N-terminal sensory domain that adopts a PAS fold and a C-terminal DNA-binding domain that adopts a LytTR fold. The PAS fold is a sensory domain 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 as the first component of two-component signal transduction systems.[26] Several such bacterial oxygen sensors possess a heme-containing PAS domain, including structurally-characterized FixL and *Ec*DOS.[4,27,28] In each of these proteins, oxygen binding to ferrous heme modulates activity of the C-terminal catalytic domain.[29,30] The LytTR fold exhibits a DNA binding mode distinct from that of the well-studied HTH motif, and LytTR domains are 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*.[31–34] A crystal structure of a single LytTR domain of AgrA bound to a 15 base pair

(bp) DNA oligomer revealed interactions between loops of the β-β-β fold and specific nucleobases, in addition to enforcement of significant bending of the oligonucleotide.[35] While RcoM is composed of two well-characterized domains, this CO-sensing transcription factor is unique among PAS- and LytTR-containing proteins as it is the only known single-component fusion of these two domains.

CO binds with high affinity to Fe(II) heme in the C-terminal PAS domain of RcoM via a ligand replacement mechanism. Previous spectroscopic investigations established that the RcoM heme is coordinatively saturated in the Fe(III) and Fe(II) oxidation states.[6,15–17] In the Fe(III) state, two protein-derived amino acid residues, His^{74} and Cys^{94} serve as axial heme ligands.[6,15,16,18] Heme reduction is accompanied by a ligand switch in which Met^{104} replaces Cys^{94} as the sixth axial ligand. In Fe(II) RcoM, His^{74} remains bound to heme, while Met^{104} serves as a weak axial ligand that may be replaced by CO.[17,18] Fe(II) RcoM binds CO with a very high affinity, estimated to be 10^{10} M^{-1} for *Px*RcoM-2, [19] and the CO binding rate constant is limited by dissociation of the Met^{104} ligand. Extremely fast and complete geminate recombination gives rise to very slow CO dissociation kinetics and subsequently large CO binding affinity, processes that are allosterically coupled to the DNA binding domain.[20] While the details of CO binding to RcoM are established, 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 bacterial, two-component transcription factors that contain LytTR domains. The canonical DNA binding site for such proteins is a pair of direct repeats upstream of the -35 region of the relevant operon.[31,36–39] 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.[40,41] 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*.[42] 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*.[43–45] For all other characterized LytTR DNA binding proteins, the data are consistent with a model in which a dimeric protein, bearing a single LytTR domain per monomer, binds to the pair of repeats.

The proposed DNA-binding motif for *Px*RcoM-1 differs significantly from all previously identified binding sites for LytTR-containing proteins. 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 between $rcoM_1$ and $coxM_1$ on the first chromosome of the *P. xenovorans* genome.[18] One of these two binding regions is located directly upstream of the -10/-35 extension preceding $coxM_1$, while the second binding region is ~100 bp further upstream (Figure 2.1). These two putative *Px*RcoM-1 binding sites contain three direct repeats of the sequence pattern TTNNNG, whose sequence modestly resembles the repeat motifs identified for other LytTR-containing transcription factors.[31,34,36,46] Not only does the proposed number of direct repeats differ, 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 *Px*RcoM-1 and those known for other LytTR-containing transcription factors, we sought to better characterize the quaternary structure of *Px*RcoM-1 and re-examine the proposed DNA binding site for this CO-sensing transcription factor. In this study, we demonstrate that *Px*RcoM-1 exists primarily as a homodimer through size-exclusion chromatography and analytical ultracentrifugation sedimentation equilibrium. Using bioinformatics, we establish a relationship between RcoM

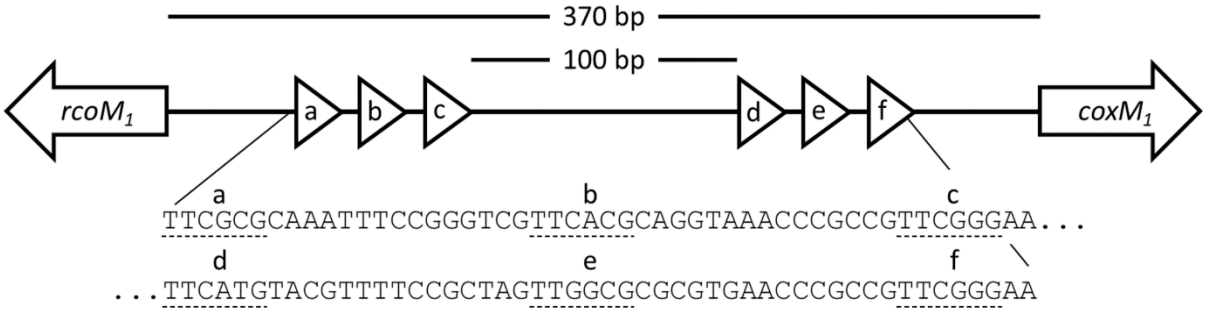


Figure 2.1. Organization of the *rcoM₁*/*coxM₁* intergenic region in *P. xenovorans* as described in Kerby et al.[18] Putative *Px*RcoM-1 binding sites that comprise the *coxM₁* promoter are highlighted. The sequences of two proposed triplet repeats "a + b + c" and "d + e + f" are shown with the "TTNNNG" motif underlined.

protein sequences and genomic context. For *cox*-associated RcoM proteins, we identify a consensus DNA binding sequence that encompasses *one* of the two originally proposed *Px*RcoM-1 binding regions directly upstream of the -10/-35 extension. *In vitro* DNA binding experiments corroborate bioinformatics analysis, demonstrating conclusively that a triplet repeat is required to achieve CO-dependent DNA binding activity in *Px*RcoM-1. Together these data highlight unique features of RcoM and its mechanism of DNA recognition: binding of a *pair of dimers* to a direct *triplet repeat* sequence.

2.2 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 were originally provided by Dr. Robert Kerby (University of Wisconsin-Madison, Department of Bacteriology).[6] Texas-Red-X labeled oligonucleotides for fluorescence anisotropy DNA binding assays were synthesized by Integrated DNA Technologies (IDT). Unlabeled complementary strands (IDT) underwent a standard desalting purification, whereas HPLC or PAGE purification was performed on labeled strands shorter or longer than 50 nucleotides, respectively.

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 H74A mutations

were created using Agilent's online tool (forward: 5'-cggcttttctccggggccagttgcaccacgt-3'; reverse: 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 VJS6737[47] *E. coli* 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, -2 and variants were isolated and purified in a manner similar to that previously described.[6,16–18] *E. coli* VJS6737 containing the pEXT20 expression vector[6,48] for each *Px*RcoM variant was grown in Lennox-LB medium with nutrient broth (3 g/L) or in Terrific Broth. All growth media were supplemented with ferric citrate (20 mg/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 50 μM 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 [50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.4, 500 mM NaCl, 1 mM dithiothreitol (DTT) or 1 mM TCEP], 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 (50 mM MOPS, 150 mM NaCl, 10 mM imidazole, pH 7.5). The cell supernatant was applied to the column and washed with binding buffer until the 280 nm absorbance reached baseline. Additional washing of bound *Px*RcoM-1 was achieved by adding four column volumes of 15:85, 25:75, and 35:65 (% v:v) mixtures of elution buffer (50 mM MOPS, 150 mM NaCl, 220 mM imidazole, pH 7.5) and binding buffer to the column. Elution of bound *Px*RcoM-1 was achieved by washing with four column volumes of elution buffer. The

red-colored fractions were pooled and incubated in 45%-saturated ammonium sulfate (v/v) 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 or TCEP). To ensure complete desalting, resuspended protein was passed through a Sephadex G-25 gravity column pre-equilibrated with storage buffer. 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; the protein purity was greater than 90% as verified by SDS-PAGE (Figure 2.S1). Heme content was determined using the pyridine hemochromagen assay.[49] 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 $M^{-1}cm^{-1}$).[50,51]

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 or TCEP. 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 $CO_{(g)}$ were added to the headspace, followed by gentle mixing to form the Fe(II)-CO heme adduct. When indicated in the main text and figure legend, we reconstituted the protein to ensure complete heme loading according to Method A in the Supplemental Information.

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. Eluted protein was detected by absorbance at 280 nm. For each experiment, the column was pre-equilibrated with the running buffer stated in the figure legends. Each 100 μL sample (41-280 μM protein) 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. When indicated in the main text, we reconstituted the protein to ensure complete heme loading according to Method A in the Supplemental Information.

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 418 nm without a cut-off filter was used to monitor the absorbance of the samples as a function of radial position. Two samples were prepared in 50 mM MOPS, pH 7.4, 150 mM NaCl, 1 mM DTT. Initial absorbances as measured in the centrifuge were 0.88 and 0.50. The samples were spun at 12.0k, 15.0k, 19.0k, and 26.0k rpm. Gradients were recorded every 3-4 hours until the consecutive gradients 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 showed a small non-sedimentable absorbance and those values were subtracted as a known fixed baseline during global fitting. A sequence molecular

weight (M_s) of 31,503 Da and a partial specific volume of 0.743 mL/g were computed 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.

Bioinformatics analyses

A sequence similarity network (SSN) containing putative RcoM proteins was generated using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) using the "FASTA" input option.[52] Potential RcoM protein sequences were gathered from three protein-protein BLAST searches using previously identified RcoM proteins from each of the three genomic contexts (*Rhodospirillum rubrum* RcoM: Q2RNI6, *Paraburkholderia xenovorans* RcoM-1: Q13YL3, *Geobacter sulfurreducens* RcoM: Q74BE2).[6] BLAST hits were included in the network if they contained a query cover of greater than 72%, included both a LytTR domain and a PAS domain, and conserved the heme-coordinating histidine residue. The RcoM SSN contained 361 protein sequences in total. All SSNs analyzed in this work are 100% representative node networks and were visualized using Cytoscape 3.8.2 using the organic layout.[53]

Inspection of genome neighborhoods allowed us to classify each RcoM based on its association with nearby *cox*, *cowN*, or *cooS* genes. EFI Genome Neighborhood Tool (EFI-GNT) was used to retrieve neighborhood diagrams (a 40 gene window centered on the gene of interest) for RcoM proteins associated with a UniProt ID in the SSN.[54,55] For proteins with RefSeq and INSDC accession IDs, genome neighborhoods were accessed and inspected manually for the presence of *cox*, *cowN*, or *cooS* genes using the associated nucleotide accession ID through the NCBI website. These three genomic contexts were based on the presence of genes showing significant homology to experimentally verified homologs of the aerobic CODH subunits (CoxL: P19919, CoxM: P19920, CoxS: P19921), CowN (Q2RNI5), or the anerobic CODH (CooS:

Q9F8A8). RcoM proteins that did not contain *cox*, *cowN*, or *cooS* genes within their genome neighborhood were classified as unknown or "unk". All protein accession IDs and genomic context classifications can be found in the supplementary "RcoM genomic context" spreadsheet file.

The nucleotide archive IDs, provided by the EFI-GNT, were used to procure the nucleotide sequence of 44 putative *cox*-associated RcoM promoter regions, 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 Pro-Coffee via the T-Coffee online web server and visualized using WebLogo.[56–58] 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. The final sequence alignment was visualized using BOXSHADE 7.0 through the ExPASy server.

Fluorescence polarization DNA binding assay

For each fluorescence polarization assay, oxidized (ferric) and reconstituted *Px*RcoM (Supplemental Information, Method B) was mixed with double-stranded DNA (dsDNA) oligonucleotide containing a 5' Texas Red-X fluorescent label and a putative *Px*RcoM binding site (Table 2.1). Final protein concentrations varied from 0.1-3 μM, and assay conditions in each 500 μL sample were as follows: 40 mM Tris·HCl pH 8.0, 20 mM KCl, 5% (v/v) glycerol, 5.0 mM DTT or TCEP, 48 μg sheared salmon sperm DNA, and 3.2 nM Texas Red-X-labeled dsDNA. Anisotropy values were recorded using an ISS PC1 fluorimeter. Each Fe(III) *Px*RcoM sample was

Table 2.1. Oligonucleotide sequences employed in fluorescence polarization DNA binding assays. The 5' → 3' sense strand is displayed, and the complementary strand is omitted for clarity.

Oligonucleotide Name	Length (bp)	Sequence (5' → 3')[a]
d + e	42	TRX/CGGGTTTCATGTACGTTTTCCGCTAGTTGGCGCGCGTGAACC
e + f	41	TRX/GCTAGTTGGCGCGCGTGAACCCGCCGTTCGGGAAATTGCGT
d + e + f	62	TRX/CGGGTTTCATGTACGTTTTCCGCTAGTTGGCGCGCGTGAACCCGCCGTTCGGGAAATTGCGT
d	18	TRX/TCGGGTTTCATGTACGTT
a	18	TRX/TCGATTTTCGCGCAAATT

[a]TRX = Texas Red-X, underlined residues highlight 8 bp imperfect repeat binding motifs

loaded into a septum-sealed glass culture tube (6×50 mm dimensions, VWR). The headspace of each tube was purged with $Ar_{(g)}$ for 10 minutes, and protein samples were reduced through anaerobic addition of 10 μL of a 100 mM sodium dithionite solution, prepared in 40 mM Tris·HCl pH 8.0, 20 mM KCl, 5% (v/v) glycerol). Fe(II)-CO adducts were generated by introducing 100 μL $CO_{(g)}$ to the headspace of each tube using a gas-tight syringe and gently vortexing. Anisotropy values were recorded immediately after introduction of CO and after a 40 min incubation period at room temperature to ensure complete conversion to the Fe(II)-CO species and equilibration of DNA binding. Each data point represents an average of three technical repeat measurements of anisotropy.

The Hill equation was used to fit a cooperative binding model to fluorescence anisotropy data taken of the Fe(II)-CO species after the 40 minute incubation period:

$$A = A_f + \frac{(A_b - A_f)}{1 + \left(\frac{[P]_{1/2}}{[P]}\right)^n}$$

where A is the observed fluorescence anisotropy, A_f is the baseline fluorescence anisotropy in the absence of binding, A_b is the maximum fluorescence anisotropy (i.e. when all protein is bound to DNA), $[P]_{1/2}$ is the protein concentration at half saturation, $[P]$ is the total protein concentration, and n is the Hill coefficient.

2.3 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), the protein as studied should exhibit a monomeric molecular weight of 30.8 kDa, including the C-terminal 6-His tag and a single heme b molecule, and a dimeric molecular weight of 61.6 kDa. The majority *Px*RcoM-1 species observed in solution by SEC has an apparent mass of 70 kDa (Figure

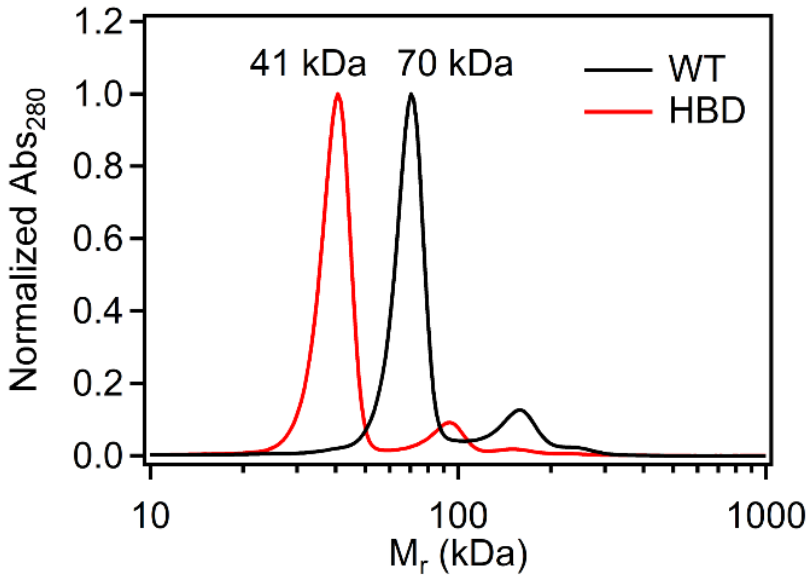


Figure 2.2. Size exclusion chromatograms of full-length WT and heme-binding domain (HBD) truncate *Px*RcoM-1. Calculated monomeric MW values are 30.8 kDa for the full-length *Px*RcoM-1 and 17.7 kDa for the HBD. Protein samples (5.58 mg/mL WT, 3.90 mg/mL HBD) in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT (100 μL) were injected via a 1 mL sample loop at a flow rate of 0.6 mL/min. Eluted protein was detected by absorbance at 280 nm.

2.2), which is 8.4 kDa greater than expected for a homodimeric species. Analysis by SEC assumes that the shape and other properties of a protein are similar to those of the globular calibration standards; thus, an apparent molecular weight larger than predicted suggests that the dimer is non-spherical. Several small peaks were observed with larger molecular weights than the dimeric fraction, suggesting that at least a portion of the protein exists as higher order oligomers in solution (e.g., tetrameric *Px*RcoM-1 likely gave rise to the distinct peak observed at an apparent molecular weight of 160 kDa). The observation of *Px*RcoM-1 oligomeric species is unsurprising given the tendency of prokaryotic PAS domains to form homo-oligomers.[59]

Analytical ultracentrifugation sedimentation equilibrium (AUC-SE), which is thermodynamically rigorous and does not require calibration standards, confirmed that *Px*RcoM-1 exists as a dimer (Figure 2.S2). At holoprotein concentrations yielding A_{418} of 0 to 1.75, samples behaved as a single species with minimal loss of material due to aggregation. Global fits of the absorbance gradients to single species models gave an average molecular weight (M_w) of 66,282 ± 160 Da. Comparing M_w to M_s (the sequence molecular weight) gave a M_w/M_s of 2.10 ± 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. Higher order oligomers are likely to form in a concentration-dependent manner, and protein concentrations were significantly lower in AUC-SE (0.1 mg/mL) experiments than in SEC (5 mg/mL) experiments. Furthermore, oligomers would likely be depleted at the AUC-SE speeds used for monomeric and dimeric species.

A truncate of *Px*RcoM-1 encompassing the heme-binding domain (HBD, residues 1-153) largely exists as a non-spherical homodimer in solution. The calculated monomeric molecular weight for the *Px*RcoM-1 HBD truncate is 17.7 kDa, while the majority truncate species runs with an apparent molecular weight of 44 kDa by SEC (Figure 2.2). The apparent molecular weight of

the homodimeric truncate is somewhat (8.6 kDa) larger than that predicted by sequence alone, suggesting that, like full-length *Px*RcoM-1, the truncated HBD homodimer is not spherical. Importantly, these data demonstrate that dimerization in *Px*RcoM-1 occurs via the heme-binding domain. As with full-length *Px*RcoM-1, several small peaks were observed at higher apparent molecular weights for the HBD truncate, including a peak at an apparent molecular weight of 94 kDa that may be attributable to a HBD tetramer. These minority species suggest that a fraction of the protein exists as higher order oligomers in solution.

*Px*RcoM-1 binds heme with a 1:1 heme-to-monomer ratio. While as-isolated WT *Px*RcoM-1 exhibited heme loading around 50% (assuming a 1:1 ratio of heme to monomer), heme loading improved maximally to just above 100% after reconstitution with excess hemin chloride (Table 2.S1). A small amount of adventitiously bound heme, detected by EPR spectroscopy, likely accounts for the slightly greater than 100% heme loading (Figure 2.S3). Consistent with an increase in heme loading relative to as-isolated protein, reconstituted RcoM-1 exhibited greater intensity of the Soret band at 423 nm relative to that of the protein peak at 280 nm (Figure 2.S4). Similarly, heme loading in the *Px*RcoM-1 HBD truncate approached 100% (assuming a 1:1 heme-to-monomer ratio) after reconstitution with excess hemin chloride (Table 2.S1 and Figure 2.S5). Taken together, these observations demonstrate that one heme molecule binds to each *Px*RcoM-1 monomer within the N-terminal PAS 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 spectroscopy and SEC. Previous studies of Fe(III) *Px*RcoM-2 demonstrated that the variant C94S binds heme in a five-coordinate, high-spin state, retaining the native His^{74} ligand in the absence of the sixth Cys94 ligand.[16] C94S *Px*RcoM-1 binds ferric heme in an

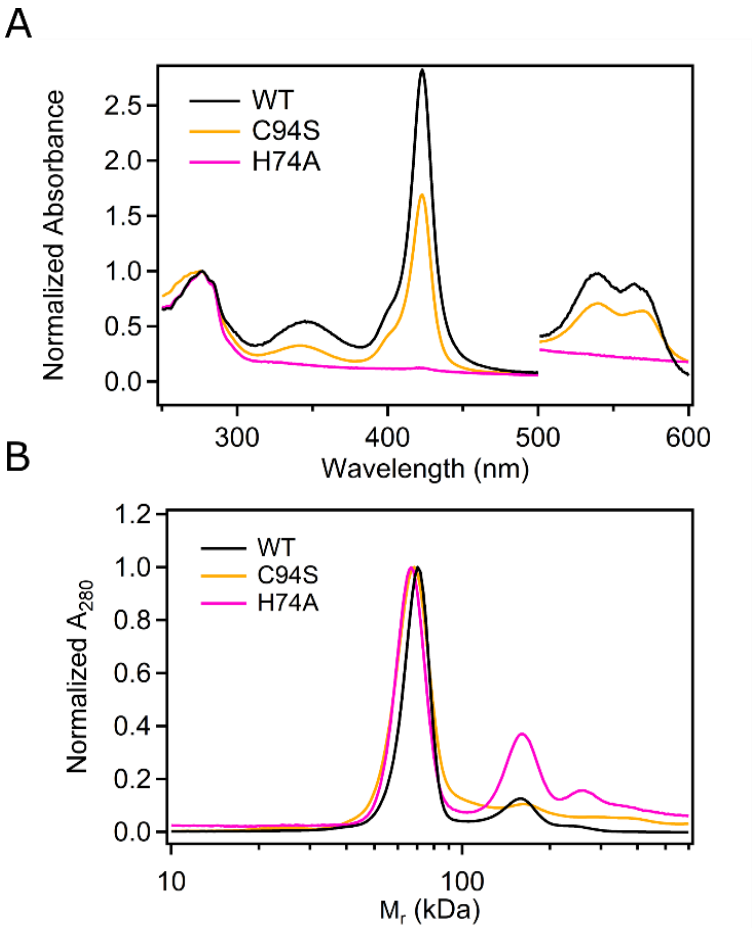


Figure 2.3 (A) Electronic absorption spectra of WT, C94S, and H74A *Px*RcoM-1 in 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT. Absorption maxima for C94S are: Soret, 423 nm and α/β 569 nm and 540 nm. Traces are normalized to absorbance at 280 nm. (B) Size exclusion chromatograms of full-length WT, C94S, and H74A *Px*RcoM-1. Protein samples of 100 µL (5.58 mg/mL for WT, 1.46 mg/mL for C94S, and 1.19 mg/mL 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 280-nm absorbance peak.

analogous fashion, as indicated by the visible spectral features (Figure 2.3A). It was previously shown that H74A *Px*RcoM-1 does not bind heme,[6] and the absence of heme was corroborated by the lack of heme-derived peaks in the visible spectrum (Figure 2.3A). By SEC, the majority fractions of C94S (68.9 kDa) and H74A (66.8 kDa) were dimeric, revealing that the heme-free protein is still a dimer. The dimeric variants C94S and H74A exhibited slightly smaller apparent molecular weights compared to WT (70 kDa), indicating that these proteins adopt a shape that is slightly more spherical or compact than that of the WT protein (Figure 2.3B). The difference in apparent molecular weight was larger for heme-free H74A *Px*RcoM-1, compared to the C94S variant bearing a high-spin heme, suggesting that loss of heme may result in a slight collapse of the PAS domain heme-binding pocket. This H74A variant also exhibited a larger relative fraction of tetrameric RcoM, suggesting that loss of heme may increase the propensity to form higher-order oligomers.

RcoM sequence similarity network reflects genomic context

Through generation of an RcoM sequence similarity network (SSN) and classification of the genomic context of these homologs, we observed protein clustering based on the identity of the adjacent, RcoM-regulated genes. We assembled a bank of 361 RcoM protein sequences using protein-protein BLAST searches and generated a RcoM SSN using the EFI Sequence Similarity Tool (EFI-EST). We successfully classified 311 of these RcoM proteins based on their genomic proximity to *cox*, *cowN*, or *cooS* genes. In total, we identified 156 proteins adjacent to a *cox* operon, 99 proteins adjacent to a *cowN* gene, and 31 proteins adjacent to a *cooS* gene. The observation of *rcoM* genes adjacent to *cox*, *cowN*, and *coo* operons is consistent with previous studies.[6,18] Upon increasing the stringency of the alignment cutoff score, we observed sub-clustering in the SSN that correlates with the RcoM genomic context (Figure 2.S6). Nodes associated with *cooS* separate

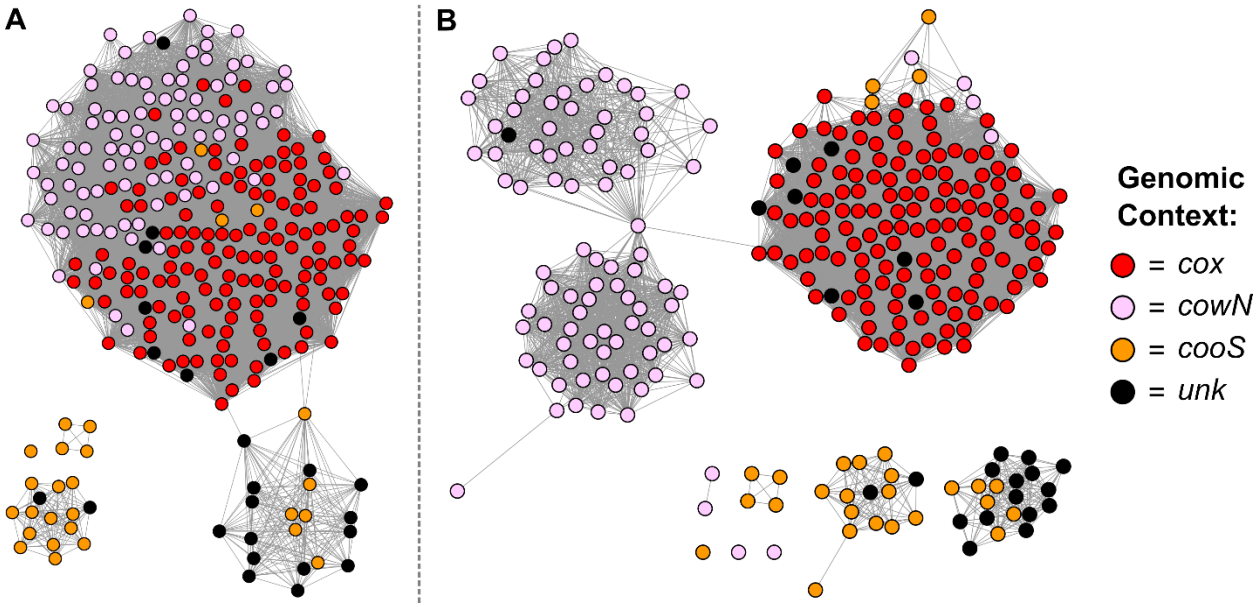


Figure 2.4. RcoM sequence similarity network images at an alignment score cutoff of (A) 33 and (B) 45. Nodes are colored based on their genomic context classification.

from the central cluster earlier, suggesting these RcoM proteins are more distantly related to the *cox*- and *cowN*-associated proteins (Figure 2.4A). At an alignment cutoff score of 45, we observe separation between *cox*-, *cowN*-, and *cooS*-associated nodes (Figure 2.4B). At this cutoff score, the *cowN*-associated proteins organize into a bimodal subcluster, the *cooS*-associated proteins organize into three subclusters, while the *cox*-associated protein cluster remains intact. 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 amino acid sequence of RcoM and its regulatory target.

Re-examining DNA binding in RcoM

Through alignment of 40 putative *cox* promoters, we identified a well-conserved pair of DNA repeats that may reflect an RcoM consensus binding sequence consistent with binding sites for other LytTR-containing transcription factors, with a third less well-conserved repeat site downstream. Importantly, the far upstream "a + b + c" triplet repeat identified by Kerby et al. was not conserved among these *cox* promoters.[18] 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 2.S7). This window encompassed the "d + e + f" triplet motif site originally identified as a *Px*RcoM-1 binding site.[18] Sequence alignment revealed a highly conserved 8 bp motif, TTC[AG][TC]GCA, which encompasses the 6 bp "TTnnnG" sequence originally proposed as a recognition sequence for *Px*RcoM-1 (Figure 2.5). This motif is strongly conserved at the "d" and "e" sites and modestly conserved at the "f" site, and each site is separated by a 12-13 bp linker. Sequence alignment of 27 *cowN*-associated promoters revealed the same 8 bp triplet repeat motif, as did a combined sequence alignment of

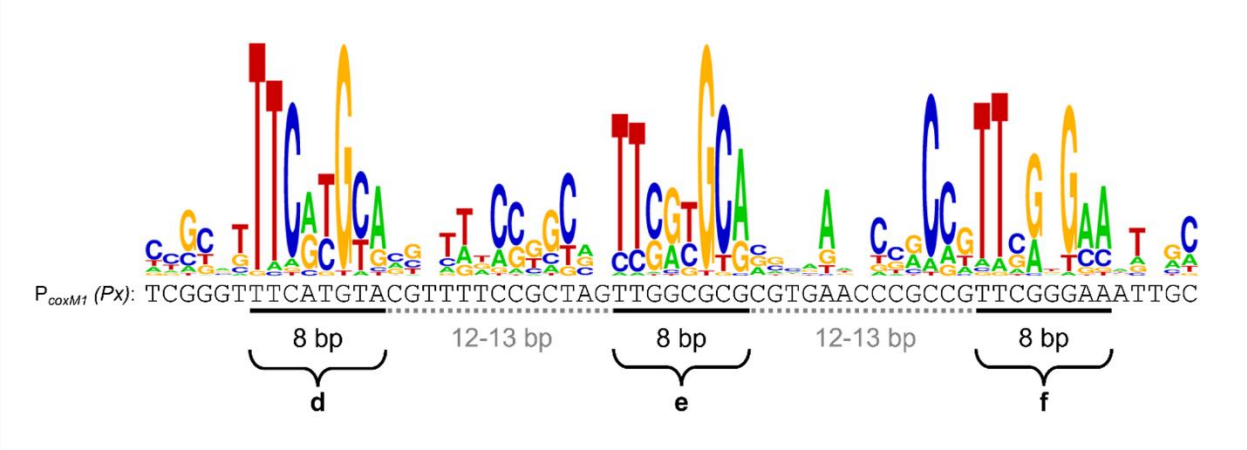


Figure 2.5. Logo plot generated from multiple sequence alignment of 40 putative *cox*-associated RcoM promoter regions. The three 8 bp repeats, identified through sequence alignment, are annotated below the logo plot, with the newly refined 8-bp "d", "e", and "f" sites labeled.

cox- and *cowN*-associated promoters (Figures 2.S8 and 2.S9). No strong trend emerged when aligning *coo*-associated promoters (data not shown).

The above bioinformatics results, which suggest a doublet repeat *Px*RcoM-1 promoter, are consistent with the dimeric nature of RcoM and previously described promoters of LytTR transcription factors; however, *Px*RcoM-1 does not bind to oligonucleotides containing the repeat pairs "d + e" or "e + f". We used fluorescence anisotropy to measure *Px*RcoM-1 binding *in vitro* to two-site 8 bp direct repeats ("d + e" and "e + f") by incubating active Fe(II)-CO *Px*RcoM-1 with the two-site dsDNA oligonucleotides. No binding was observed; there was minimal CO-dependent increase in fluorescence anisotropy (Figure 2.6). Active *Px*RcoM-1 exhibited a slight increase in fluorescence anisotropy at protein dimer concentrations greater than 200 μM in the presence of the "d + e" oligo; however, the binding affinity was too weak for measurement of a binding curve or for biological relevance.

*Px*RcoM-1 does not bind single 8-bp site promoters, "a" or "d", as has been suggested in a recent report focusing primarily on *Px*RcoM-2 CO binding kinetics. After the surprising result that *Px*RcoM-1 does not bind a doublet promoter, we moved to test single site promoters, as it was described in Salman et al. that *Px*RcoM-2 binds an 18 bp oligonucleotide encompassing the "a" site with high-affinity (K_d < 2 nM).[20] Upon incubation of *Px*RcoM-1 with single-site dsDNA oligonucleotide ("d"), we did not observe a CO-dependent increase in fluorescence anisotropy (Figure 2.6). We have repeated this experiment with *Px*RcoM-1 and *Px*RcoM-2 under experimental conditions matching those of the previous study using the identical 18-mer encompassing the "a" site but have been unable to reproduce their results (Figures 2.S12, 2.S13). Differences in experimental design, ill-suited for measuring specific DNA binding, may have led

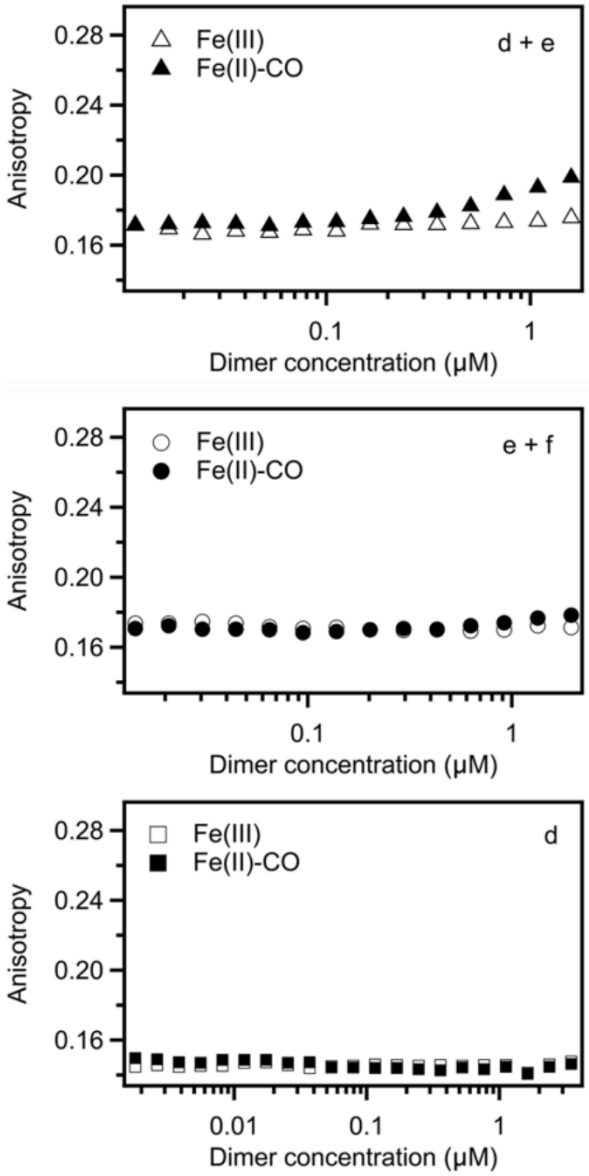


Figure 2.6 Fluorescence polarization data for inactive Fe(III) and active Fe(II)-CO *Px*RcoM-1 incubated with two-site 8 bp direct repeats, "d + e" (top), "e + f" (middle), and a single 8 bp site "d" (bottom) Texas-Red labeled dsDNA. Replicate experiments with "d + e" and "e + f" oligonucleotides are shown in Figures 2.S10 and 2.S11, respectively.

the authors to conclude that the slight increase in anisotropy seen with CO-bound RcoM-2 was a specific, strong interaction.

*Px*RcoM-1 exhibited a substantial CO-dependent increase in DNA binding affinity in the presence of the three-site "d + e + f" dsDNA oligonucleotide (Figure 2.7). This observation demonstrates that the *Px*RcoM-1 promoter site encompasses all three conserved repeats observed in the sequence alignment of putative *cox* promoter regions. These DNA binding experiments were repeated with the paralog RcoM-2 (88% sequence similarity). RcoM-2, which did not bind single repeat sites in our hands, binds to the triplet direct repeat sequence ("d + e + f") with a similar affinity as RcoM-1, demonstrating that both proteins significantly interact with the triplet repeat promoter and not to single site sequences ("a" or "d") (Figure 2.S13). These observations confirm that the triplet repeat is the requisite sequence for DNA binding.

*Px*RcoM-1 exhibits moderate affinity for the "d + e + f" sequence and binds cooperatively, suggesting that more than one RcoM dimer is required to bind the promoter. Using the Hill equation, and assuming that *Px*RcoM-1 binds as a homodimer, we observed an average $[P]_{1/2}$ value of 250 ± 10 nM and an average Hill coefficient, *n*, of 1.7 ± 0.1. Comparison of a simple bimolecular interaction model and the Hill model, which accounts for cooperativity and molecularity, demonstrates that the Hill equation provides a better fit of the experimental data ($\chi^2 = 1.0 \times 10^{-3}$ for simple bimolecular versus $\chi^2 = 3.8 \times 10^{-4}$ for cooperative, Figure 2.S15). An average *n* value of 1.7 suggests that *Px*RcoM-1 binding to "d + e + f" is cooperative, involving more than one (plausibly two) *Px*RcoM-1 homodimers. This result is unsurprising given that each *Px*RcoM-1 homodimer contains two LytTR domains, which most likely bind two of the three repeat sites present in the "d + e + f" sequence. A second *Px*RcoM-1 homodimer would be necessary to bind the third available repeat site.

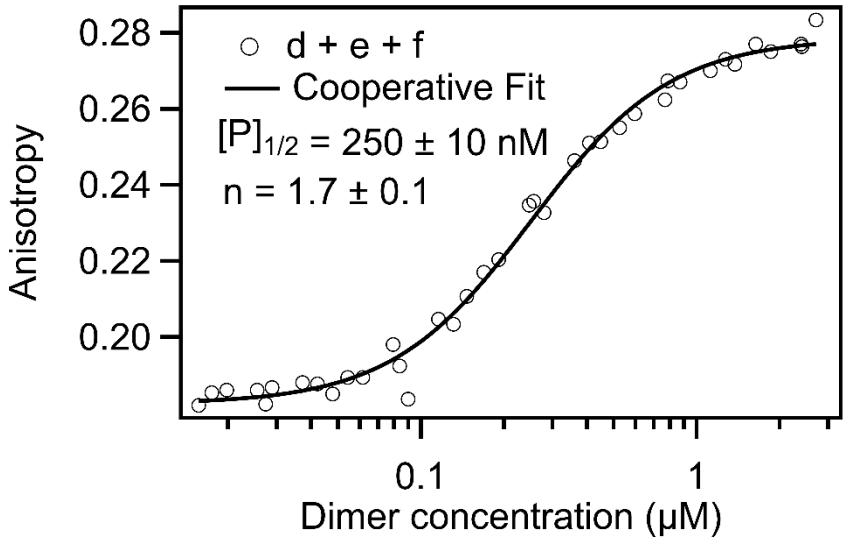


Figure 2.7. Fluorescence polarization data for active Fe(II)-CO *Px*RcoM-1 and triplet direct repeat site dsDNA oligonucleotide. All three data sets for the "d + e + f" oligonucleotide are shown with circles. The solid line represents the best fit of all "d + e + f" data to the Hill equation ($\chi^2 = 3.8 \times 10^{-4}$), as determined by nonlinear least squares analysis. The average $[P]_{1/2}$ and n values of the three trials are displayed with standard errors. Independent plots and fits for the three data sets are shown in Figure 2.S14.

2.4 Discussion

In this study, we provide insight into the quaternary structure and DNA-binding properties of the heme-dependent, single-component transcriptional regulator *Px*RcoM-1. Based on 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 consistent with other PAS domain-containing proteins.[60] Heme is not required for dimerization, and the 1:1 heme:monomer stoichiometry is consistent with that of other heme-dependent gas sensors.[7,61] Bioinformatics analysis of *cox*-associated promoter sequences implies that RcoM interacts with two or three conserved direct repeat DNA binding sites upstream of the *coxM1* core promoter region. *In vitro* DNA binding assays showed that RcoM binds only to the triplet repeat sequence and does not bind singlet or doublet repeats. DNA binding to a triplet direct repeat sequence is different than has been observed for other LytTR-type transcription factors, which bind a doublet pair of direct repeat sites.[31] Additionally, RcoM DNA binding requires more than one dimer to fully saturate the three-site promoter, and this binding process is cooperative.

Divergence among RcoM protein sequences is dependent upon genomic context. Genes encoding RcoM appear upstream of three types of operons: *cox*, *coo*, and *cowN* (Figure 2.4). These operons are associated with distinct physiological processes: aerobic CO metabolism, anaerobic CO metabolism, and protection of nitrogenase enzymes from inactivation by CO, respectively.[1,6,12,14,24,25] The CO binding affinities of CODH enzymes that carry out aerobic (K_M = 0.52 μM)[62] and anaerobic (K_M = 18 μM)[63] CO metabolism differ, suggesting that these pathways are operative and need to be regulated at different CO concentrations.[18] The differences in RcoM

protein sequences that correlate with genomic context may reflect differences in CO binding affinity, DNA binding affinity or both, and may in turn give rise to differential gene expression as a function of CO concentration. Further study of RcoM proteins encoded in the context of *cooS* and *cowN* genes is required to substantiate this hypothesis.

Using bioinformatics and *in vitro* DNA binding studies, we refined the identity of the cognate *Px*RcoM-1 binding site upstream of $coxM_1$. The binding sites originally proposed for *Px*RcoM-1 consisted of two sets of triplet repeats spaced ~100 bp apart.[18] A sequence alignment of 40 putative *cox*-associated promoters revealed no conservation of the far upstream triplet repeat; however, a triplet repeat consensus motif (3 x 8 bp, "d + e + f") emerged directly upstream of the -10/-35 $coxM_1$ core promoter region. Using fluorescence polarization to probe DNA binding *in vitro*, we demonstrated that all three 8 bp repeats are required to observe specific, CO-dependent *Px*RcoM-1 binding. Taken together, these data allow us to present a refined cognate DNA binding motif for *Px*RcoM-1 that consists of 8 bp repeats with the sequence TTC[AG][TC]GCA separated by 13 bp spacers (Figure 2.5). The paralog *Px*RcoM-2 binds equally well to this same site (Figure 2.S13). The repeat sequence, repeat length, and spacer length of this newly identified repeat motif are consistent with other LytTR promoters[31]; however, the presence of three repeat sites instead of two is unusual, especially given the dimeric nature of the transcription factor.

Our observations are inconsistent with an earlier report of high affinity binding to a single 8 bp site. Salman et al.[20] claimed high affinity binding (K_d less than 2 nM) of *Px*RcoM-2 to a Texas Red-labelled dsDNA encapsulating the previously identified "a" site alone. In our hands, neither *Px*RcoM-1 nor *Px*RcoM-2 bind to dsDNA containing a single "a" or "d" motif (Figures 2.6, 2.S12, and S13). Our data are inconsistent with RcoM binding to a single 8 bp site; instead, we observe that all three 8 bp direct repeats are necessary for high affinity, CO-dependent binding.

Analysis of the RcoM:DNA interaction suggests a unique DNA binding mode for this transcription factor. RcoM differs from many well-known transcription factors, which dimerize with C_2 symmetry to interact with a pair of palindromic repeat binding sites. Instead, the triplet direct repeat consensus site and the dimeric nature of RcoM requires three LytTR domains, from at least two RcoM dimers, to bind to the promoter. This model is supported by the cooperativity observed in our *in vitro* DNA binding data, suggesting that RcoM is unique among LytTR-containing proteins. Such proteins typically bind to a pair of direct repeats with tandem-oriented LytTR domains from one dimer.[31,64–66] A single reported exception to this model is AlgR, which binds to a promoter site consisting of a pair of perfect direct repeats far upstream from a third imperfect, inverted repeat.[43–45] This unexpected binding mode was explained by invoking DNA looping.

The RcoM:DNA binding interaction may resemble that of CarH, an adenosylcobalamin-dependent, HTH-containing transcription factor whose DNA binding site is composed of a triplet of 11 bp direct repeats.[67] CarH tetramerizes through its C-terminal adenosylcobalamin-binding sensory domain upon activation by red light, and subsequently binds to DNA using only three of the four available DNA binding domains. The three domains exhibit translational symmetry upon binding to the promoter, with each oriented in the same direction along the DNA.[68] A long unstructured region, ranging from 6-120 residues in length, connects the DNA binding domain to the C-terminal sensory domain of CarH proteins, allowing each individual DNA binding domain to adopt the different orientations necessary for interaction with direct repeats in the promoter. Similar to CarH, the RcoM DNA binding domain connects to the sensory domain through a long (~40 residues), unstructured linker that could be key to the activity of RcoM. Thus, CarH provides precedent and a potential recognition strategy for the binding stoichiometry observed for RcoM.

A proposed model for the RcoM:DNA complex consistent with our observation that two RcoM dimers are required to bind the trimeric promoter, based on the Drennan lab structural model of CarH DNA binding, is presented in Figure 2.8. Further structural investigation into the RcoM:DNA complex will allow us to uncover the unique mode of interaction with DNA and enhance our understanding of how LytTR-containing transcription factors regulate critical biological processes.

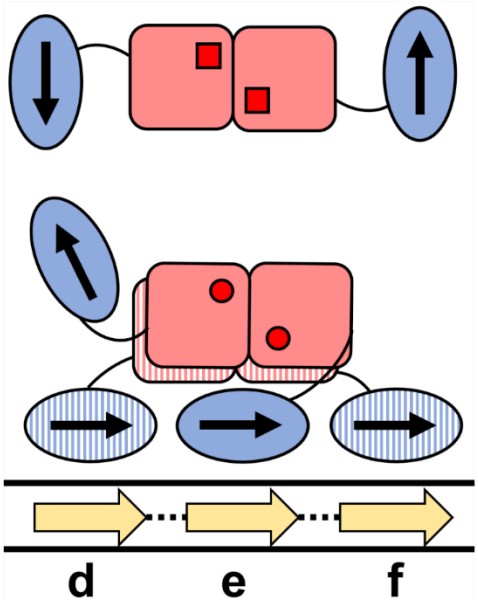


Figure 2.8. A proposed model of RcoM DNA binding based on the structural model of CarH tetrameric DNA binding. (Top) Two RcoM monomers form an elongated homodimer through the heme binding domains. DNA binding domains are connected to the sensory domains through a long flexible linker. (Bottom) Upon activation by CO binding at the heme, two RcoM dimers come together to interact with the full trimeric promoter.

2.5 Supplementary Information

Heme Reconstitution

When indicated in the main text, one of the following heme reconstitution protocols was performed to ensure full heme loading prior to experimentation.

Method A: 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 spin dialysis using a 600 μL 30K MWCO spin concentrator and dilution into 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT.

Method B: A combined heme reconstitution and re-oxidation procedure was performed by diluting concentrated protein stocks to 2.5 mL using buffer containing 50 mM MOPS, pH 7.4, 150 mM NaCl, 1 mM DTT, 3 mM potassium ferricyanide, and hemin chloride (from a 20 mM stock in DMSO) equimolar to the amount of protein monomer present in solution. DMSO never exceeded 0.2% (v:v). After incubating on ice for 30 min, DMSO, ferricyanide, and excess hemin chloride were removed by passing the protein through a PD-10 desalting column (GE Healthcare), followed by three successive rounds of spin dialysis using a 600 μL 30K MWCO spin concentrator and dilution into 50 mM MOPS pH 7.4, 150 mM NaCl, 1 mM DTT.

Table 2.S1. 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 hemochromogen 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	

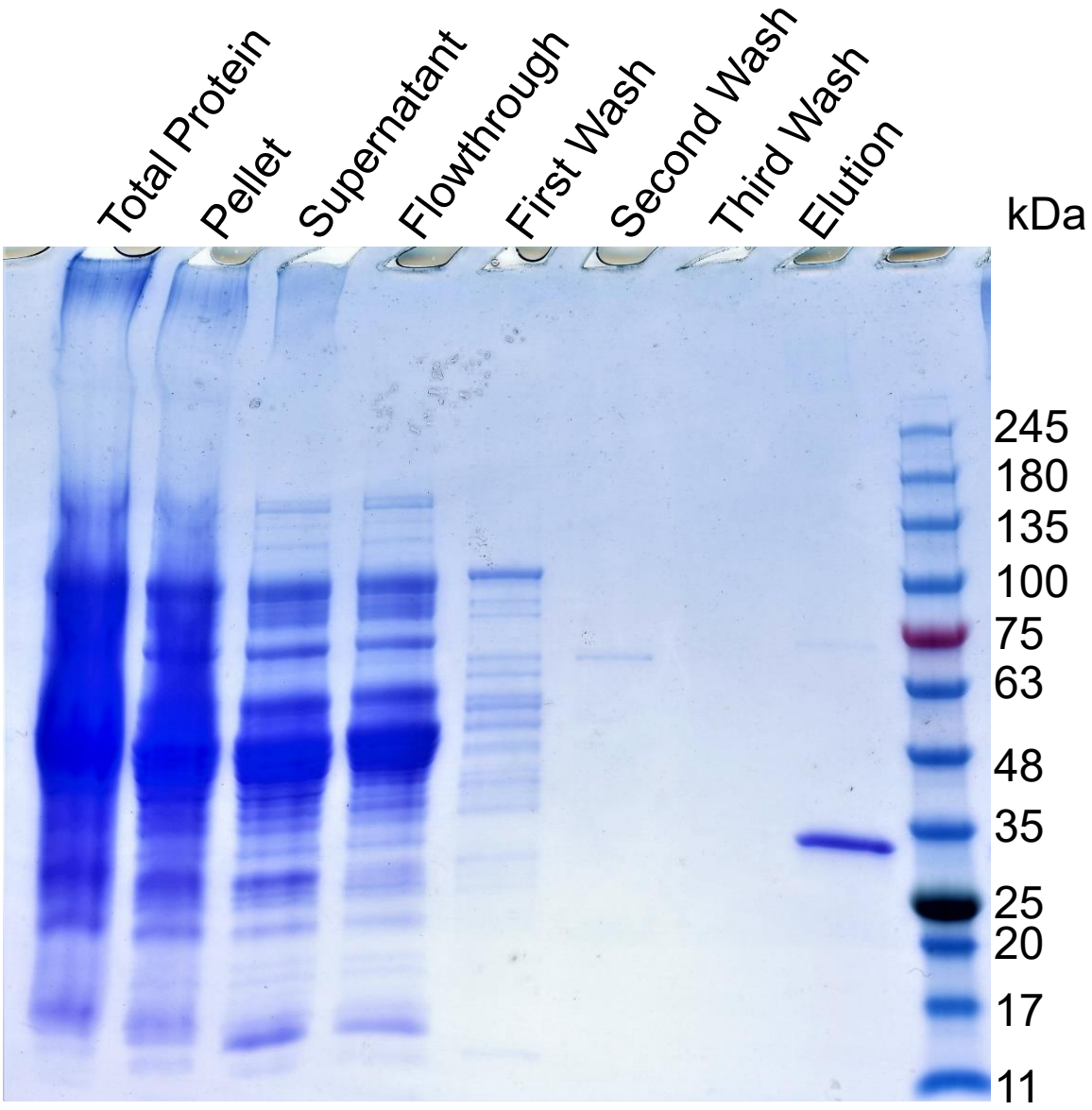


Figure 2.S1. Representative SDS-PAGE for a typical preparation of full-length, WT *Px*RcoM-1 (expected molecular weight of 31.5 kDa).

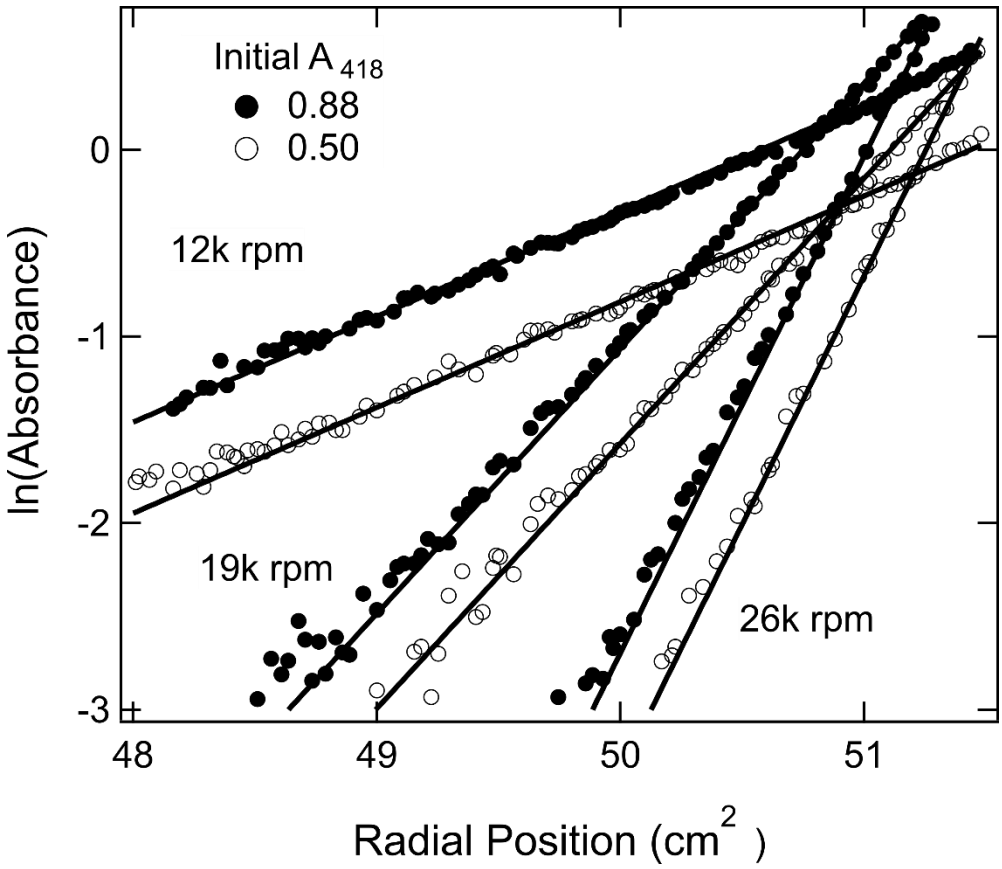


Figure 2.S2. Sedimentation equilibrium data plotted as ln(absorbance) vs. radial position squared. In this format, a single species presents as a line with slope proportional to M_w, speed, and temperature (4 °C). Data from samples with two different initial absorbances at 418 nm are shown as filled and open circles. Data from three of the four speeds included in the analysis are shown (15k rpm not shown). For each speed, the data from the different samples appear parallel, consistent with a single species. The solid lines are from a global fit to all the data as a single species. The analysis demonstrated the data to be well described by a single species with an M_w/M_s ratio of 2.10, a dimeric species. The data sets include absorbances at 418 nm ranging from zero to ~2.1.

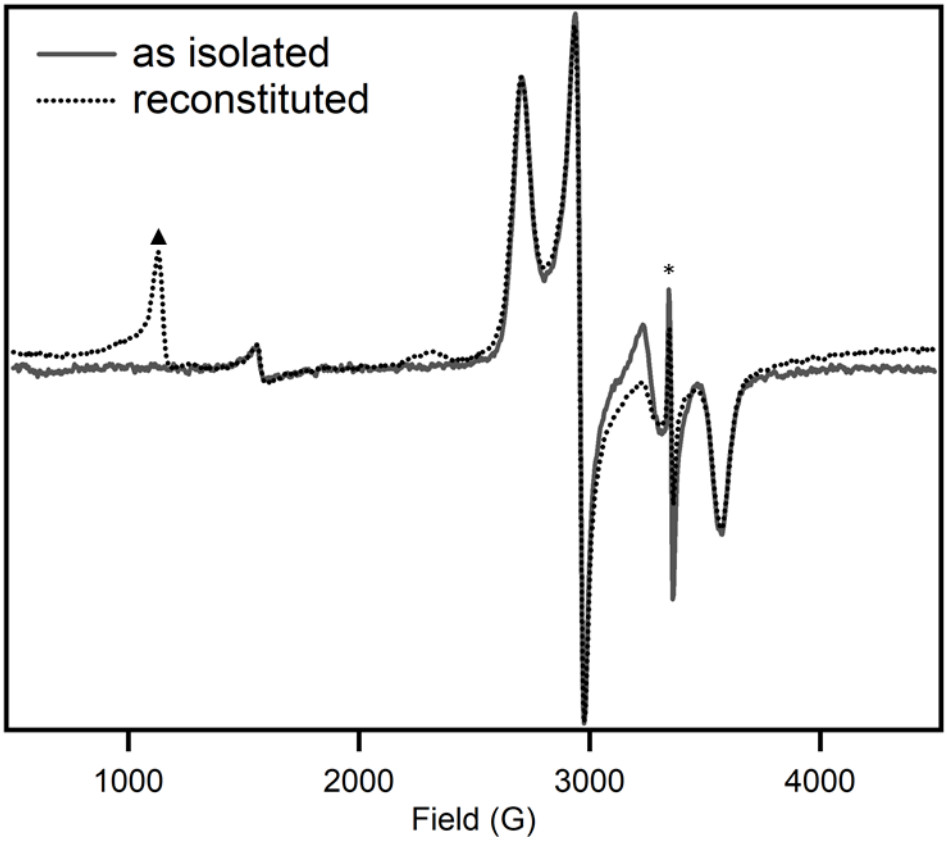


Figure 2.S3. X-band EPR spectra of WT *Px*RcoM-1 as isolated after purification and reconstituted with hemin chloride. The asterisk denotes the EPR cavity signal, while the triangle denotes the axial, high-spin Fe(III) signal indicative of a small amount of adventitiously-bound heme.

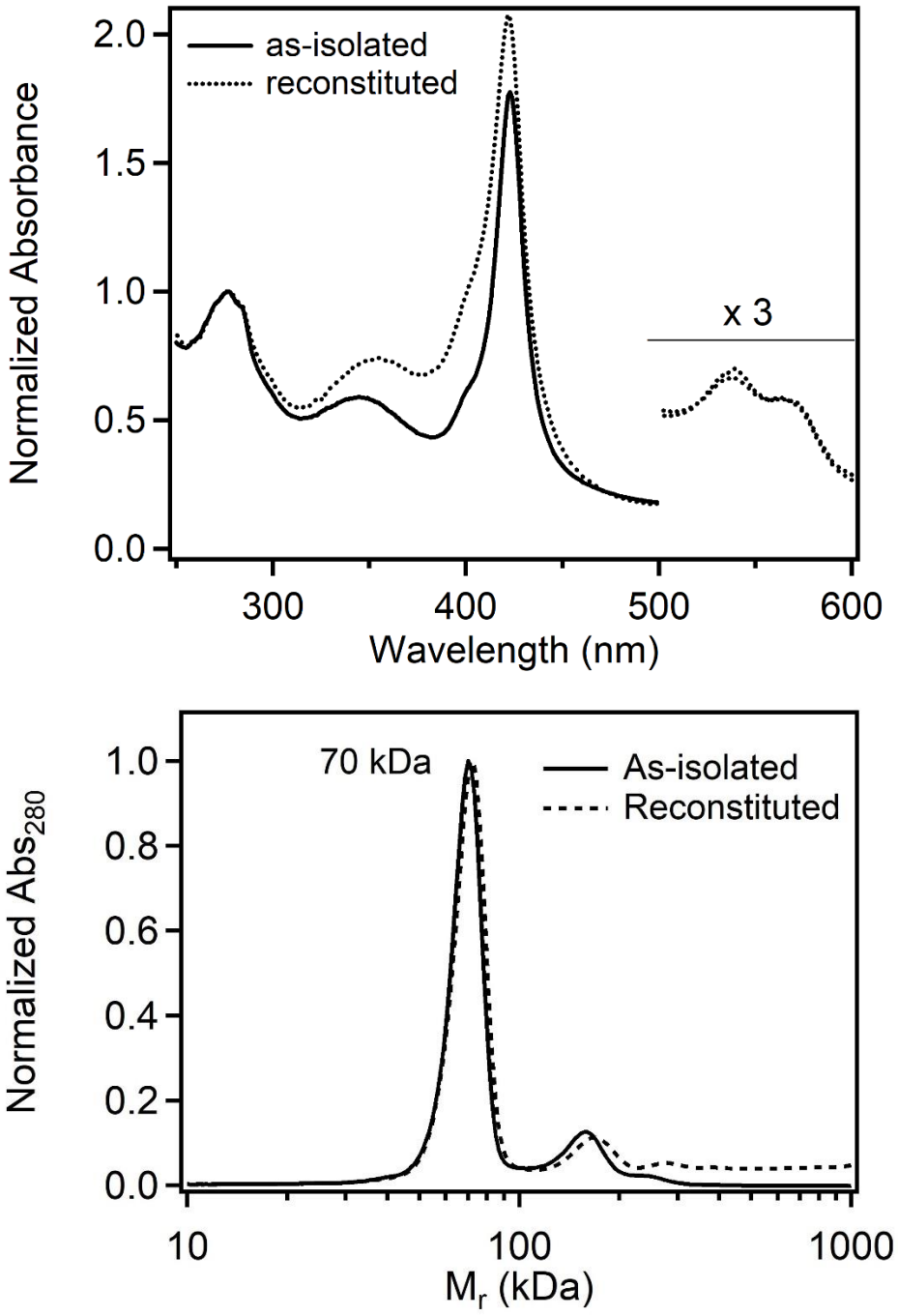


Figure 2.S4. A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography profiles (lower panel) for as-isolated and reconstituted full-length WT *Px*RcoM-1. Electronic absorption spectra are normalized to unity at 280 nm, and the most intense peaks in the SEC traces (monitored at 280 nm) are normalized to unity.

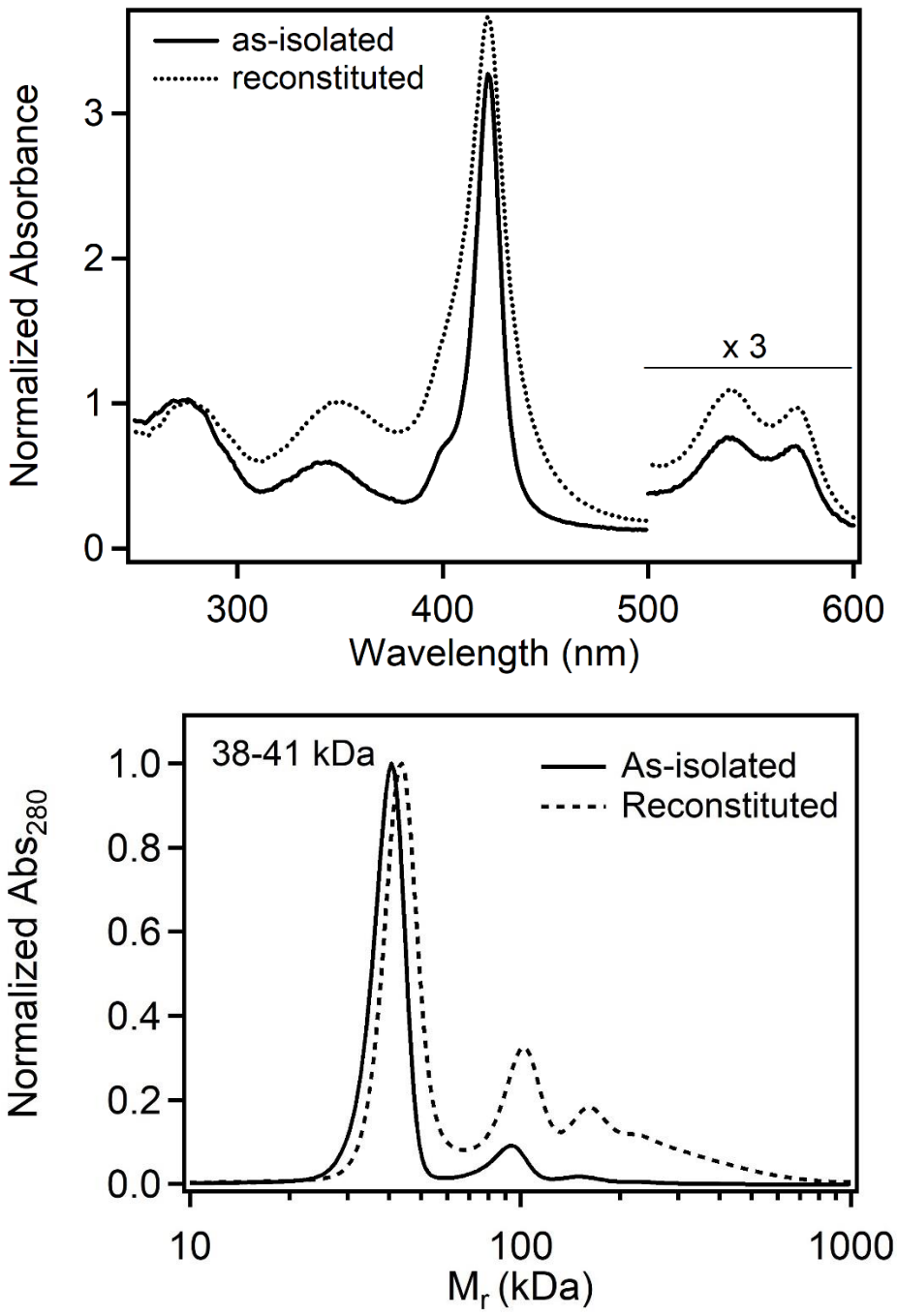


Figure 2.S5. A comparison of electronic absorption spectra (upper panel) and size exclusion chromatography profiles (lower panel) for the as-isolated and reconstituted *Px*RcoM-1 heme binding domain truncate. Electronic absorption spectra are normalized to unity at 280 nm, and SEC traces (monitored at 280 nm) are normalized to unity for the most intense peak.

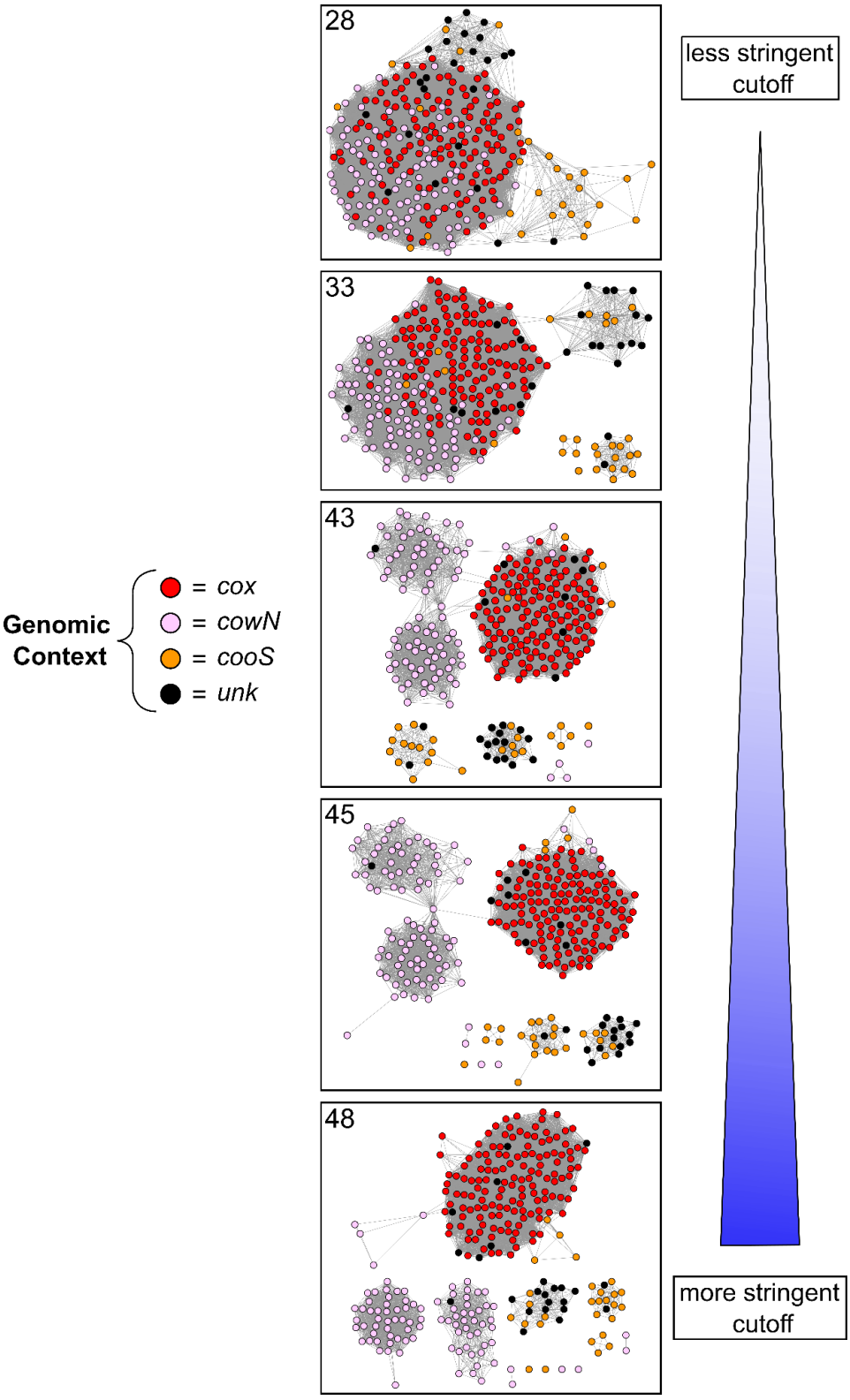


Figure 2.S6. RcoM sequence similarity network shown as a function of alignment score cutoff. Network images are labeled with alignment score cutoff in upper left corner of each image. Nodes are colored according to their genomic context classification.

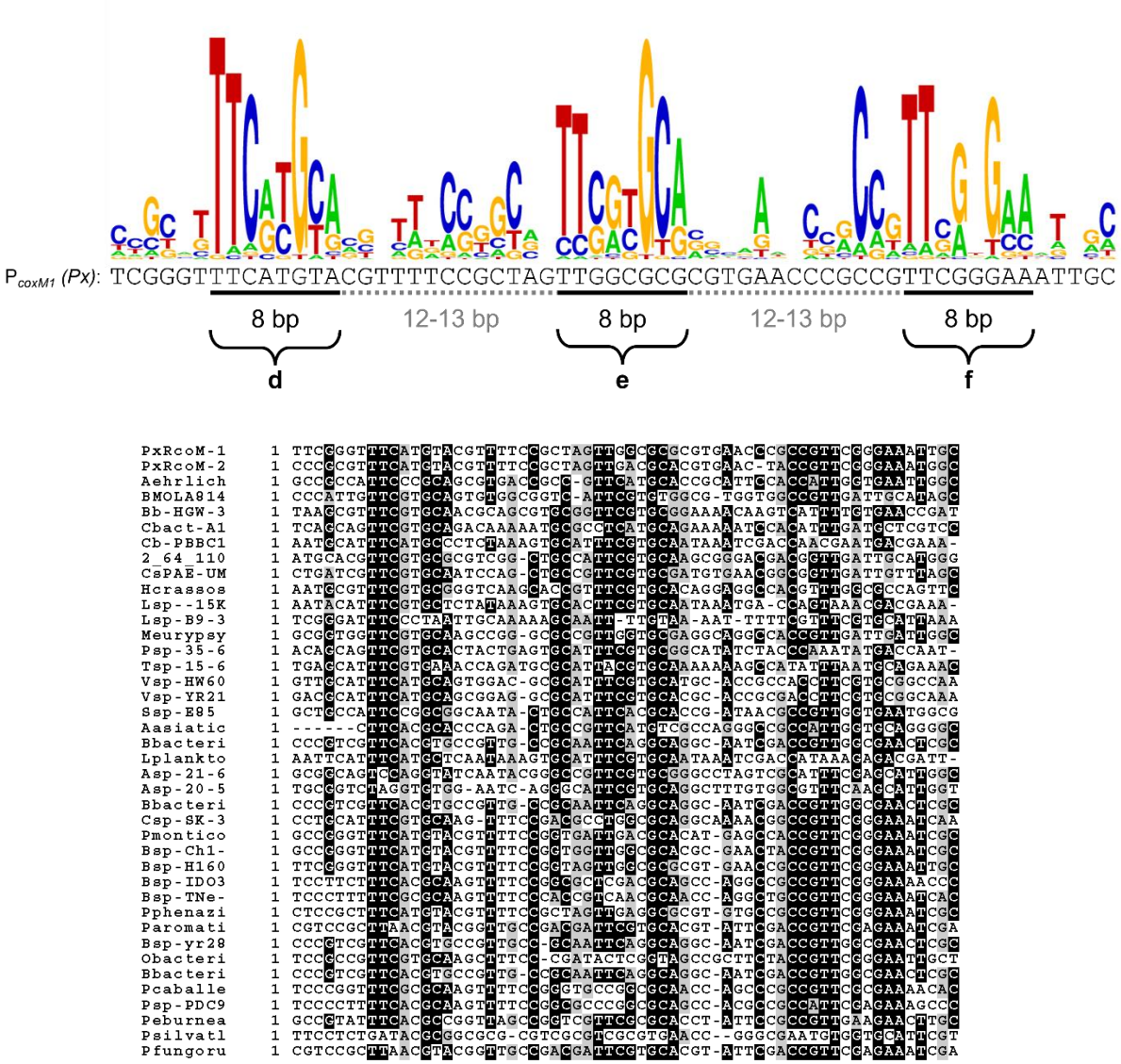


Figure 2.S7. Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cox*-associated RcoM proteins. Black boxes denote positions where sequence is identical to the consensus.

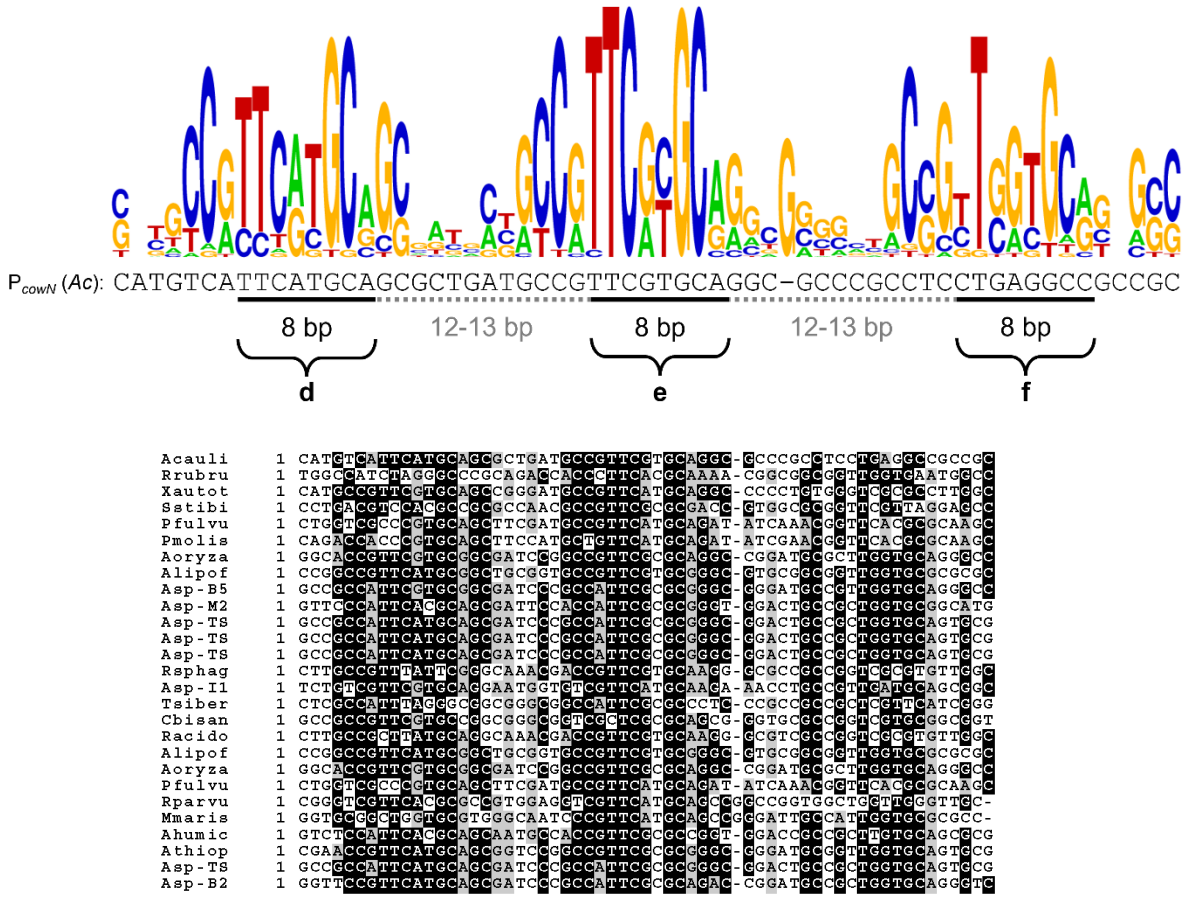
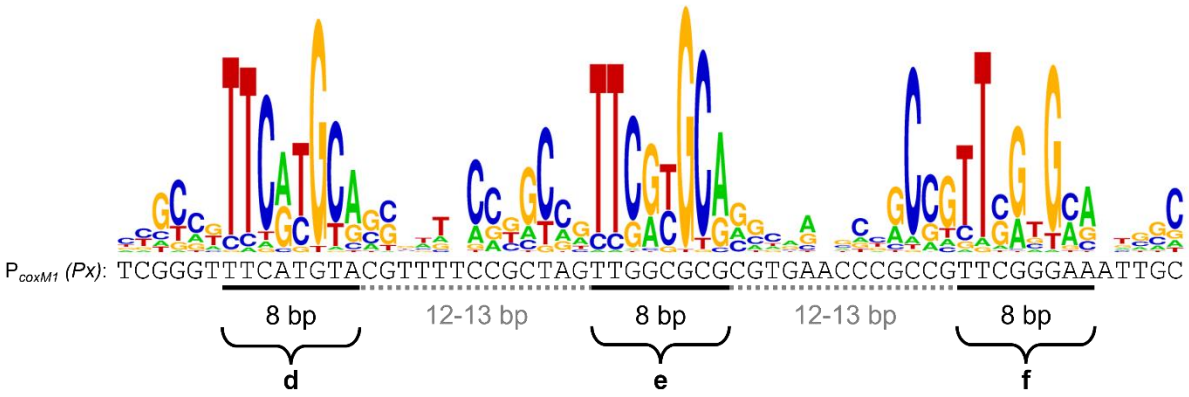


Figure 2.S8. Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cowN*-associated RcoM proteins. Black boxes denote positions where sequence is identical to the consensus.



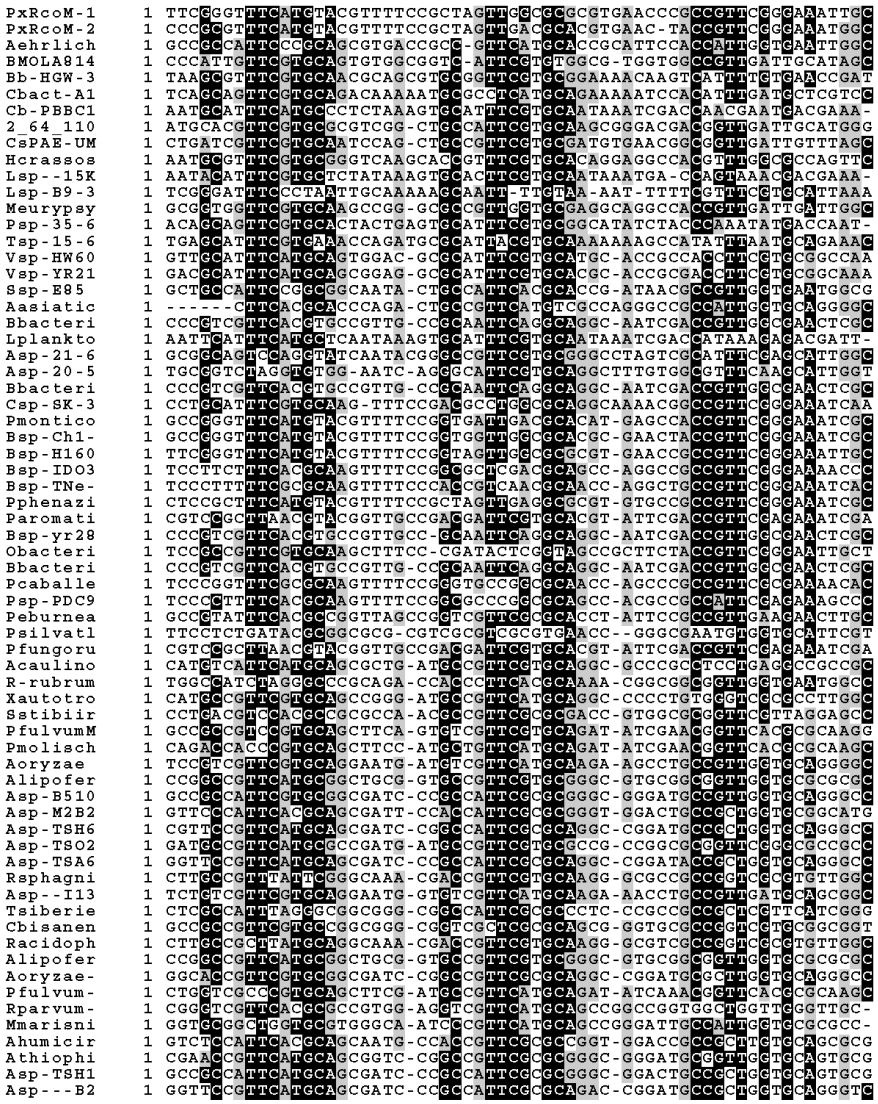


Figure 2.S9. Logo plot (top) and sequence alignment (bottom) for putative promoter regions of *cox*- and *cowN*-associated RcoM proteins. Black boxes denote positions where sequence is identical to the consensus.

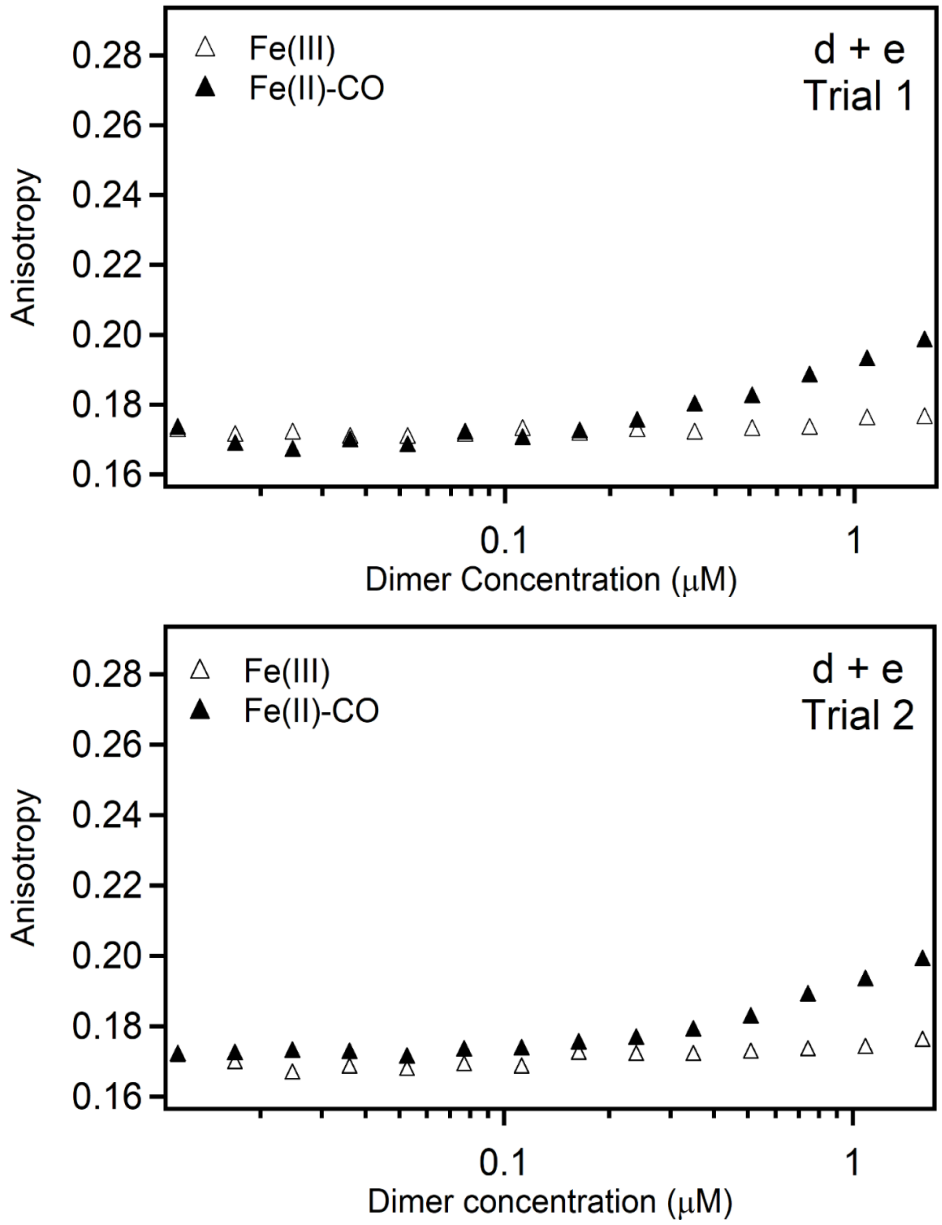


Figure 2.S10. Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the "d+e" Texas Red labeled oligonucleotide. The data are presented for two replicate experiments, with each data point representing an average of three technical repeat measurements of anisotropy. Open and closed triangles represent the Fe(III) and Fe(II)-CO bound state of RcoM, respectively.

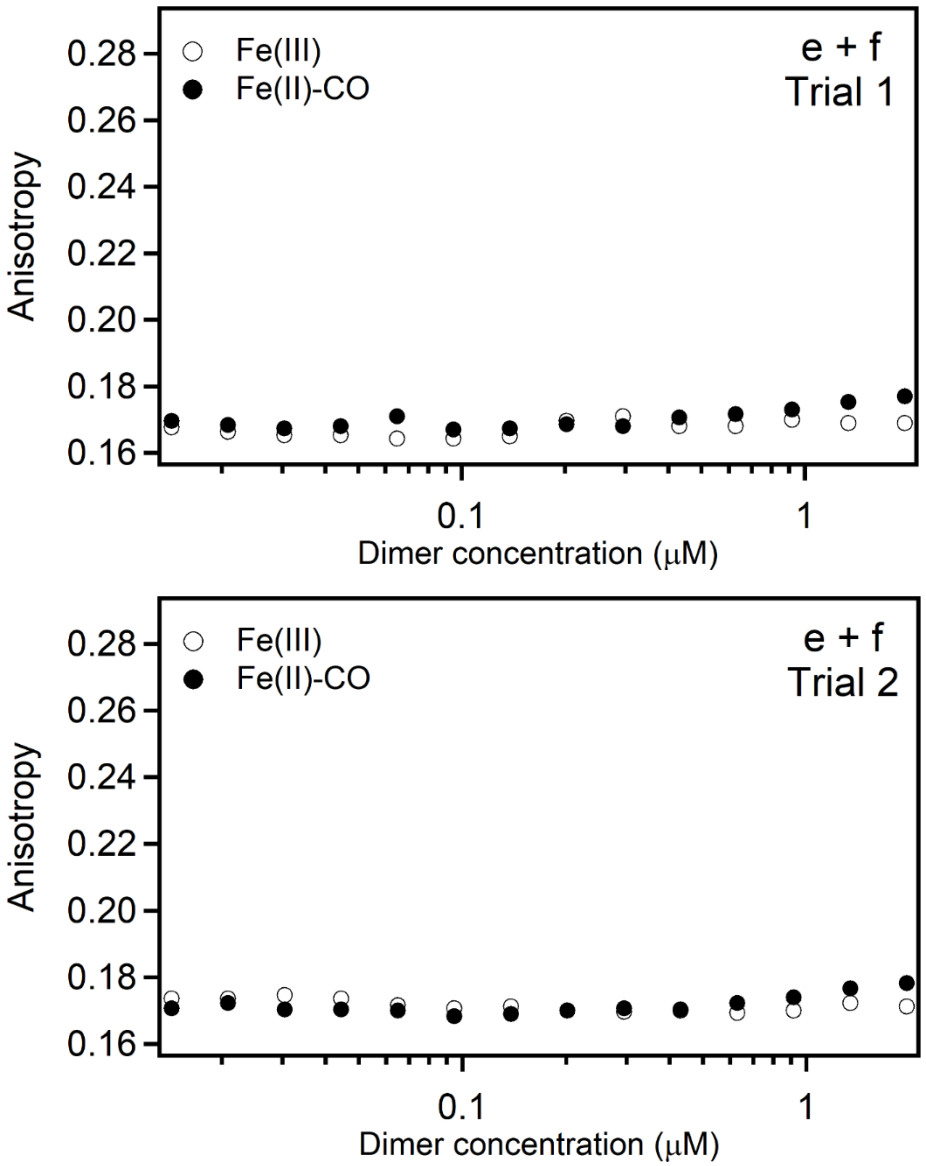


Figure 2.S11. Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the "e+f" Texas Red labeled oligonucleotide. The data are presented for two replicate experiments, with each data point representing an average of three technical repeat measurements of anisotropy. Open and closed circles represent the Fe(III) and Fe(II)-CO bound state of RcoM, respectively.

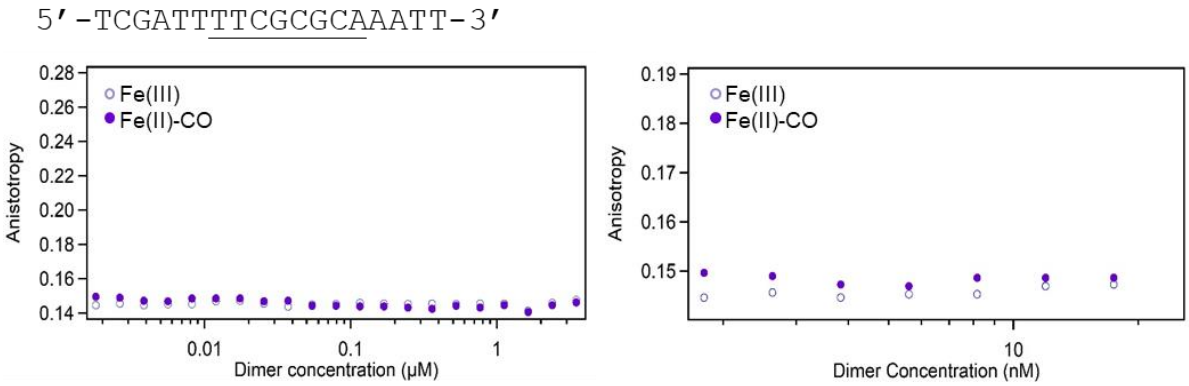


Figure 2.S12. Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the "a" site Texas Red-labeled dsDNA oligonucleotide. Experimental conditions for this *in vitro* experiment are 6.4 nM "a" site 18-mer in 50 mM Tris-HCl pH 7.6, 100 mM NaCl (No glycerol, no competitor DNA). No significant increase in anisotropy is observed upon reduction and CO-binding, therefore WT *Px*RcoM-1 does not interact significantly with a single promoter site. The 5' → 3' sense strand is displayed, and the complementary strand is omitted for clarity.

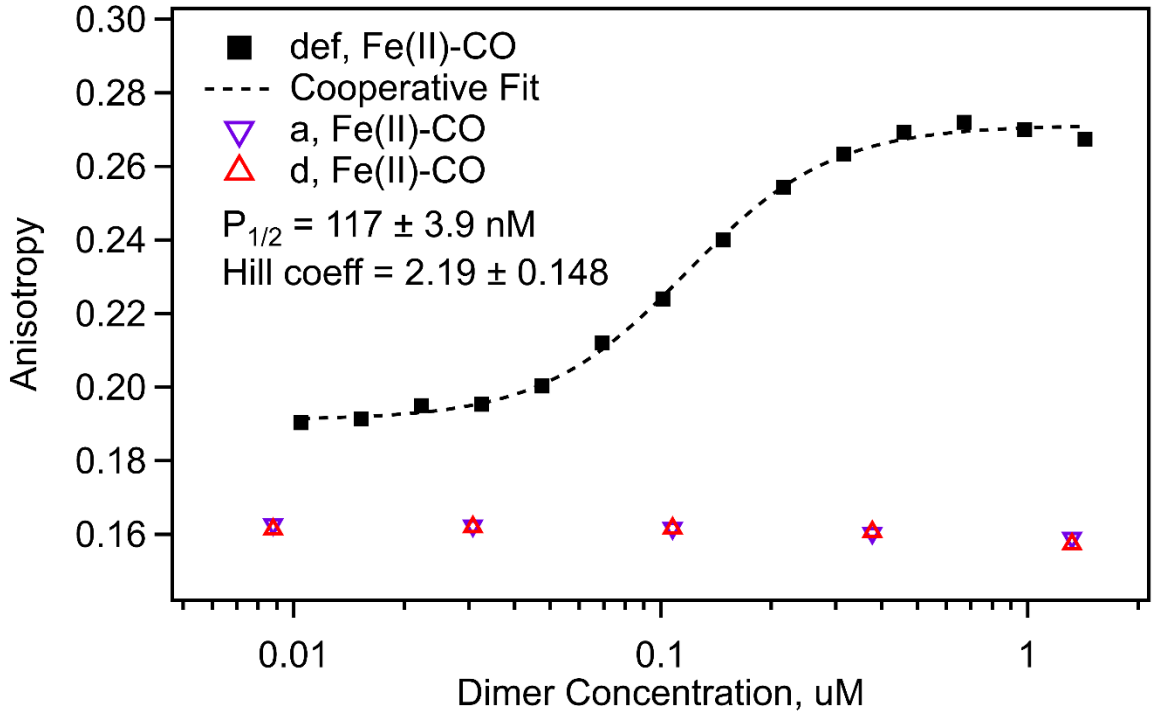


Figure 2.S13. Fluorescence anisotropy data for active, Fe(II)-CO *Px*RcoM-2 DNA binding to Texas Red-labeled dsDNA oligonucleotides bearing *coxM*$_1$ promoter elements. Experiments were carried out in the presence of 40 mM Tris·HCl pH 8.0, 20 mM KCl, 5% (v/v) glycerol, 5.3 mM TCEP. As with *Px*RcoM-1 there is no significant increase in anisotropy when PxRcoM-2 is incubated with single repeat promoters elements "a" and "d"; however a distinct, protein-dependent increase in anisotropy is observed in the presence of the triplet repeat "d + e + f" upstream of *coxM*$_1$. The dashed line represents the best fit to the Hill equation ($\chi^2 = 3.8 \times 10^{-4}$) for *Px*RcoM-2 binding to "d + e + f", as determined by nonlinear least squares analysis.

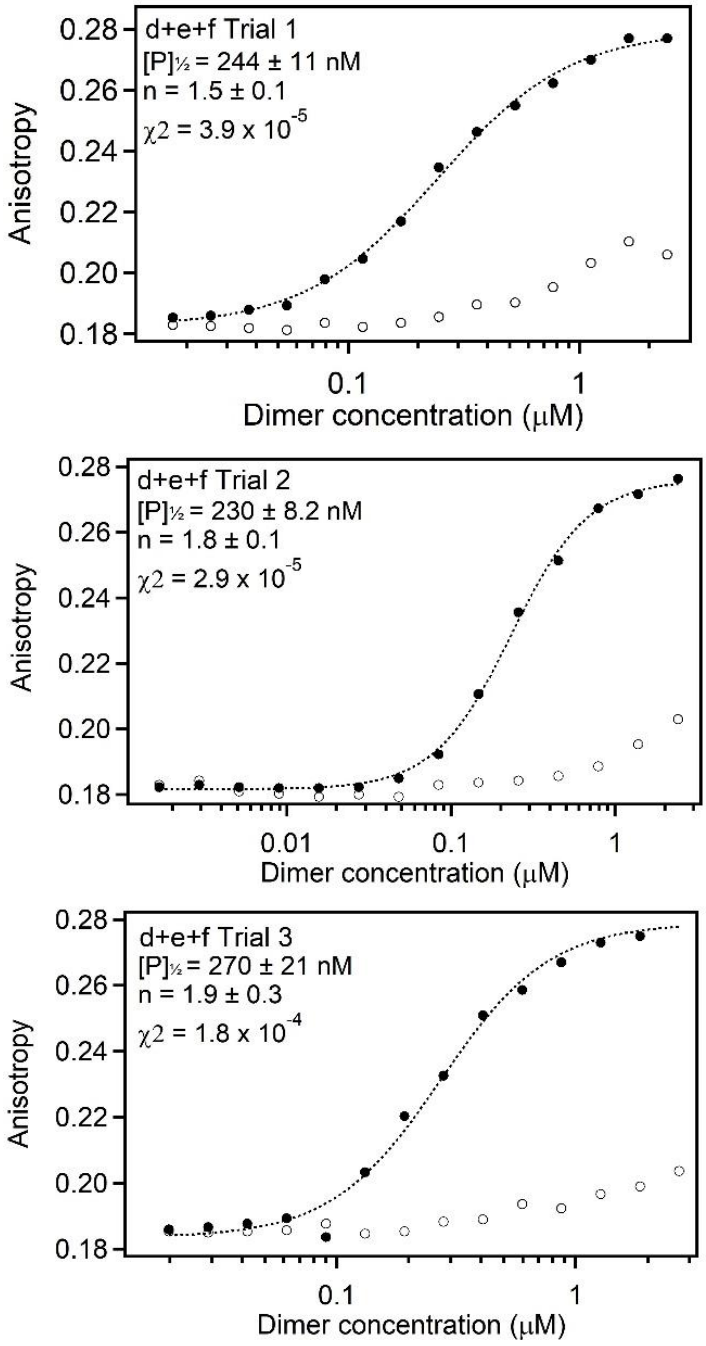


Figure 2.S14. Complete fluorescence anisotropy data for WT *Px*RcoM-1 DNA binding to the "d + e + f" Texas Red labeled oligonucleotide. The data are presented for three replicate experiments, with independent fits for each data set. Each data point represents an average of three technical repeat measurements of anisotropy. Open and closed circles represent the Fe(III) and Fe(II)-CO bound state of RcoM, respectively. The dashed line represents the best fit of the closed circle data set to the Hill equation, as determined by nonlinear least squares analysis. The $[P]_{1/2}$ and n values are displayed with standard errors.

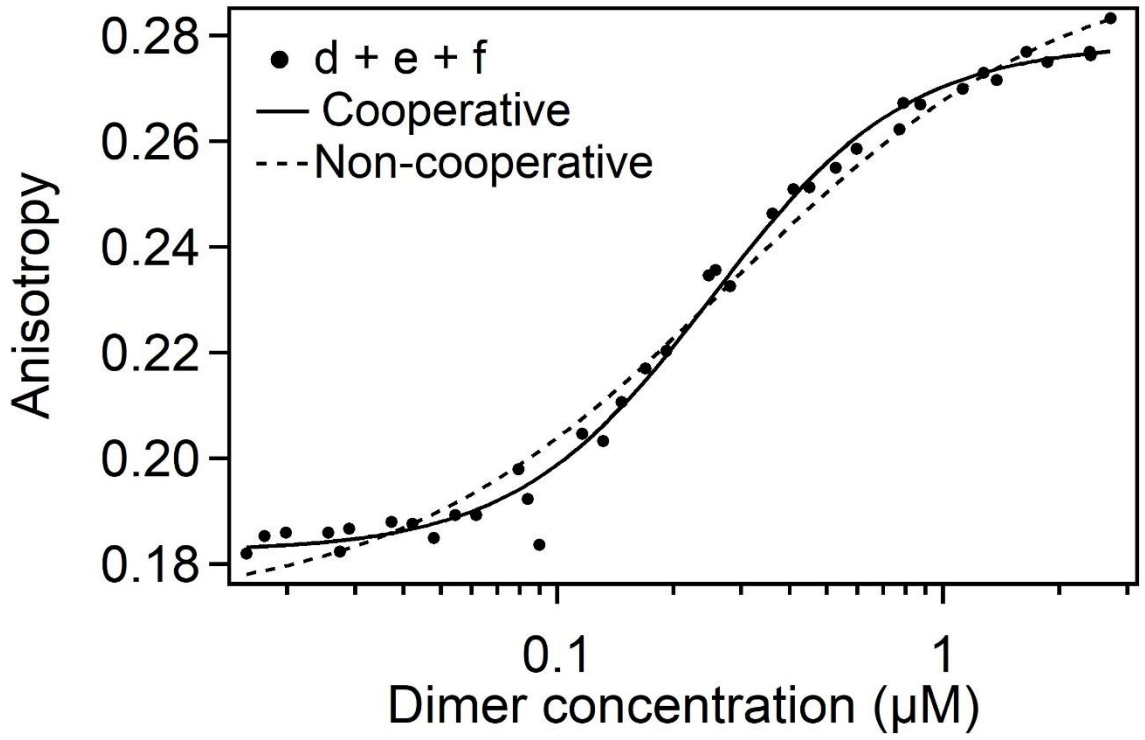


Figure 2.S15. Fluorescence anisotropy data of WT *Px*RcoM-1 DNA binding to the "d + e + f" Texas Red labeled oligonucleotide fit using a simple bimolecular interaction model ($\chi^2 = 1.0 \times 10^{-3}$, dashed line) and cooperative binding model ($\chi^2 = 3.8 \times 10^{-4}$, solid line). Fits determined by a nonlinear least squares analysis.

2.6 References

(1) King, G. M.; Weber, C. F. Distribution, Diversity and Ecology of Aerobic CO-Oxidizing Bacteria. *Nat. Rev. Microbiol.* **2007**, *5* (2), 107–118, DOI: 10.1038/nrmicro1595

(2) Oelgeschlager, E.; Rother, M. Carbon Monoxide-Dependent Energy Metabolism in Anaerobic Bacteria and Archaea. *Arch. Microbiol.* **2008**, *190* (3), 257–269, DOI: 10.1007/s00203-008-0382-6

(3) Roberts, G. P.; Youn, H.; Kerby, R. L. CO-Sensing Mechanisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68* (3), 453—473, DOI: 10.1128/mmbr.68.3.453-473.2004

(4) Shimizu, T.; Huang, D.; Yan, F.; Stranava, M.; Bartosova, M.; Fojtíková, V.; Martínková, M. Gaseous O_2, NO, and CO in Signal Transduction: Structure and Function Relationships of Heme-Based Gas Sensors and Heme-Redox Sensors. *Chem. Rev.* **2015**, *115* (13), 6491–6533, DOI: 10.1021/acs.chemrev.5b00018

(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* (8), 2157–2163, DOI: 10.1128/jb.177.8.2157-2163.1995

(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* (9), 3336–3343, DOI: 10.1128/JB.00033-08

(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* (10), 876–880, DOI: 10.1038/82820

(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* (Pt 3), 282–287, DOI: 10.1107/s0907444906051638

(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* (3), 752–756, DOI: 10.1006/bbrc.1996.1727

(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* (9), 2669–2678, DOI: 10.1021/bi982658j

(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* (2), 388–396, DOI:10.1021/bi991378g

(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* (6), 1515–1524, DOI: 10.1128/jb.178.6.1515-1524.1996

(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* (7), 2980–2987, DOI: 10.1016/S0021-9258(18)61456-5

(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* (21), 6200–6208, DOI: 10.1128/jb.178.21.6200-6208.1996

(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* (34), 9016–9028, DOI:10.1021/bi800486x

(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* (7), 1071–1082, DOI: 10.1007/s00775-012-0920-1

(17) Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met104is the CO-Replaceable Ligand at Fe(II) Heme in the CO-Sensing Transcription Factor *Bx*RcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21* (4), 559–569, DOI: 10.1007/s00775-016-1368-5

(18) 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* (21), 5803–5816, DOI: 10.1128/JB.01024-12

(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* (41), 10686–10694, DOI:10.1021/acs.jpcb.6b08160

(20) Salman, M.; Villamil Franco, C.; Ramodiharilafy, R.; Liebl, U.; Vos, M. H. Interaction of the Full-Length Heme-Based CO Sensor Protein RcoM-2 with Ligands. *Biochemistry* **2019**, *58* (39), 4028-4034, DOI: 10.1021/acs.biochem.9b00623

(21) 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* (1), 115–124, DOI: 10.1016/S0378-1119(99)00245-0

(22) 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* (9–10), 865–876, DOI: 10.1515/bc.2000.108

(23) 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 CO_2. *Gene* **2003**, *322*, 67–75, DOI: 10.1016/j.gene.2003.08.027

(24) Kerby, R. L.; Roberts, G. P. Sustaining N_2-Dependent Growth in the Presence of CO. *J. Bacteriol.* **2011**, *193* (3), 774–777, DOI: 10.1128/jb.00794-10

(25) Hoffmann, M.-C.; Pfänder, 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. **2014**, *96* (1), 3496-3502, DOI: 10.1128/JB.01754-14

(26) Henry, J. T.; Crosson, S. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu. Rev. Microbiol.* **2011**, *65* (1), 261–286, DOI: 10.1146/annurev-micro-121809-151631

(27) 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.* **1998**, *95* (26), 15177–15182, DOI:

10.1073/pnas.95.26.15177

(28) 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* (19), 20186–20193, DOI: 10.1074/jbc.M314199200

(29) Hao, B.; Isaza, C.; Arndt, J.; Soltis, M.; Chan, M. K. Structure-Based Mechanism of O_2 Sensing and Ligand Discrimination by the FixL Heme Domain of *Bradyrhizobium japonicum*. *Biochemistry* **2002**, *41* (43), 12952–12958, DOI: 10.1021/bi020144l

(30) Park, H.; 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* (10), 2738–2746, DOI: 10.1021/bi035980p

(31) Nikolskaya, A. N.; Galperin, M. Y. A Novel Type of Conserved DNA-Binding Domain in the Transcriptional Regulators of the AlgR/AgrA/LytR Family. **2002**, *30* (11), 2453–2459, DOI: 10.1093/nar/30.11.2453

(32) 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* (9), 2993–3002, DOI: 10.1128/jb.01156-08

(33) 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* (7), 3064–3072, DOI: 10.1128/iai.00123-10

(34) 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* (21), 6020–6031, DOI: 10.1128/jb.05436-11

(35) 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* (5), 727–735, DOI: 10.1016/j.str.2008.02.011

(36) 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* (4), 817–827, DOI: 10.1046/j.1365-2958.1999.01528.x

(37) 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* (1), 57–66, DOI: 10.1128/jb.182.1.57-66.2000

(38) 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* (17), 4696–4703, DOI: 10.1128/jb.182.17.4696-4703.2000

(39) 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* (15), 4472–4483, DOI: 10.1128/jb.178.15.4472-4483.1996

(40) Risøen, 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* (1),

198–206, DOI: 10.1007/s004380000397

(41) Knutsen, E.; Ween, O.; Håvarstein, L. S. Two Separate Quorum-Sensing Systems Upregulate Transcription of the Same ABC Transporter in *Streptococcus pneumoniae*. *J. Bacteriol.* **2004**, *186* (10), 3078–3085, DOI: 10.1128/jb.186.10.3078-3085.2004

(42) 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* (3), 1278–1283, DOI: 10.1128/jb.171.3.1278-1283.1989

(43) 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* (11), 6576–6580, DOI: 10.1128/jb.172.11.6576-6580.1990

(44) Mohr, C. D.; Hibler, N. S.; Deretic, V. AlgR, a Response Regulator Controlling Mucoidy in *Pseudomonas aeruginosa*, Binds to the FUS Sites of the a*lgD* Promoter Located Unusually Far Upstream from the mRNA Start Site. *J. Bacteriol.* **1991**, *173* (16), 5136–5143, DOI: 10.1128/jb.173.16.5136-5143.1991

(45) Mohr, C. D.; Leveau, J. H. J.; 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* (20), 6624–6633, DOI: 10.1128/jb.174.20.6624-6633.1992

(46) Straume, D.; Johansen, R. F.; Bjøras, 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*. https://doi.org/10.1186/1471-2091-10-17.

(47) Stewart, V.; Lu, Y.; Darwin, A. J. Periplasmic Nitrate Reductase (NapABC Enzyme) Supports Anaerobic Respiration by *Escherichia coli* K-12. *J. Bacteriol.* **2002**, *184* (5), 1314–1323, DOI: 10.1128/JB.184.5.1314-1323.2002

(48) Dykxhoorn, D. M.; St. Pierre, R.; Linn, T. A Set of Compatible t*ac* Promoter Expression Vectors. *Gene* **1996**, *177* (1–2), 133–136, DOI: 10.1016/0378-1119(96)00289-2

(49) Berry, E. A.; Trumpower, B. L. Simultaneous Determination of Hemes *a*, *b*, and *c* from Pyridine Hemochrome Spectra. *Anal. Biochem.* **1987**, *161* (1), 1–15, DOI: 10.1016/0003-2697(87)90643-9

(50) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571–607, DOI: 10.1385/1-59259-890-0:571

(51) Ernst, O.; Zor, T. Linearization of the Bradford Protein Assay. **2010**, No. 38, e1918, DOI: 10.3791/1918

(52) 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 Proteomics* **2015**, *1854* (8), 1019–1037; DOI: 10.1016/j.bbapap.2015.04.015

(53) 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* (11), 2498–2504, DOI: 10.1101/gr.1239303

(54) Gerlt, J. A. Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-

Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **2017**, *56* (33), 4293–4308, DOI: 10.1021/acs.biochem.7b00614

(55) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169–4182, DOI: 10.1021/acs.biochem.9b00735

(56) Notredame, C.; Higgins, D. G.; Heringa, J. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* **2000**, *302* (1), 205–217, DOI: 10.1006/jmbi.2000.4042

(57) Erb, I.; González-Vallinas, J. R.; Bussotti, G.; Blanco, E.; Eyras, E.; Notredame, C. Use of ChIP-Seq Data for the Design of a Multiple Promoter-Alignment Method. *Nucleic Acids Res.* **2012**, *40* (7), e52–e52, DOI: 10.1093/nar/gkr1292

(58) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190, DOI: 10.1101/gr.849004

(59) Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17* (10), 1282–1294, DOI: 10.1016/j.str.2009.08.011

(60) Ayers, R. A.; Moffat, K. Changes in Quaternary Structure in the Signaling Mechanisms of PAS Domains. *Biochemistry* **2008**, *47* (46), 12078–12086, DOI: 10.1021/bi801254c

(61) 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 Studies. *J. Mol. Biol.* **2000**, *301* (2), 415–431, DOI: 10.1006/jmbi.2000.3954

(62) 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*, 2889–2901, DOI: 10.1021/acs.biochem.8b00128

(63) 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* (17), 5134–5144, DOI: 10.1128/JB.183.17.5134-5144.2001

(64) Galperin, M. Y. Telling Bacteria: Do Not LytTR. *Structure* **2008**, *16* (5), 657–659, DOI: 10.1016/j.str.2008.04.003

(65) Galperin, M. Y. Structural Classification of Bacterial Response Regulators: Diversity of Output Domains and Domain Combinations. *J. Bacteriol.* **2006**, *188* (12), 4169–4182, DOI: 10.1128/JB.01887-05

(66) Boudes, M.; Sanchez, D.; Graille, M.; Van Tilbeurgh, H.; Durand, D.; Quevillon-Cheruel, S. Structural Insights into the Dimerization of the Response Regulator ComE from *Streptococcus pneumoniae*. *Nucleic Acids Res.* **2014**, *42* (8), 5302–5313, DOI: 10.1093/nar/gku110

(67) Ortiz-Guerrero, J. M.; Polanco, M. C.; Murillo, F. J.; Padmanabhan, S.; Elías-Arnanz, M. Light-Dependent Gene Regulation by a Coenzyme B_{12}-Based Photoreceptor. *Proc. Natl. Acad. Sci.* **2011**, *108* (18), 7565–7570, DOI: 10.1073/pnas.1018972108

(68) Jost, M.; Fernández-Zapata, J.; Polanco, M. C.; Ortiz-Guerrero, J. M.; Chen, P. Y. T.; Kang, G.; Padmanabhan, S.; Elías-Arnanz, M.; Drennan, C. L. Structural Basis for Gene Regulation by a B_{12}-Dependent Photoreceptor. *Nature* **2015**, *526* (7574), 536–541, DOI: 10.1038/nature14950

Chapter 3

Insight into the protein:DNA complex of heme-containing transcription factor, *Px*RcoM-1

A version of this chapter will be submitted as:

Roberts, M. G.; Sholes, A. E.; Riney, M. B.; Berndsen, C. E.; Burstyn J. N. Insight into the protein:DNA complex of heme-containing transcription factor, *Px*RcoM-1.

MGR generated and purified RcoM and variants to prepare SAXS samples and performed glutaraldehyde crosslinking experiments, in the lab of JNB. AES and MBR, under the direction of CEB, analyzed the SAXS data. MGR wrote the majority of the manuscript with input from CEB and JNB.

3.1 Introduction

Carbon monoxide (CO) is a critical source of energy and carbon to a diverse array of microbes. To regulate the production of the cellular machinery needed to metabolize CO, bacteria use one of two biochemically characterized nonhomologous transcription factors, CO Oxidation Activator (CooA) or Regulator of CO Metabolism (RcoM).[1] RcoM has been identified in the context of genes controlling three biological processes: aerobic CO oxidation (*cox* genes), anaerobic CO oxidation (*coo* genes), and the protection of nitrogenase from CO inhibition (*cowN* gene).[2,3] RcoM-1 was originally identified upstream of the *cox* operon in *Paraburkholderia xenovorans*, and it is one of two homologs in the *P. xenovorans* genome, thought to arise from a gene duplication event.[4] The two proteins (88% sequence similarity) have been studied interchangeably; no structural or biochemical differences have been observed between the two proteins. RcoM is the only single-component fusion of a Per-Arnt-Sim (PAS)-type heme binding sensory domain and a LytTR-type DNA binding domain, with the two domains connected by a 21 residue, unstructured linking region. Given this architecture, RcoM is predicted to display a unique regulatory mechanism.

Over 200,000 proteins are annotated to contain a PAS fold (Interpro IPR013767). This versatile sensory domain is known to respond to diverse chemical signals as a part of single-component and two-component systems. Some PAS domains use a cofactor, including metal ions, heme, and flavin, to regulate the activity of diverse effector domains, including DNA binding domains.[5] Despite relatively low sequence similarity (below 20% on average) and in contrast with their functional diversity, PAS domains are remarkably structurally similar.[6] PAS-heme proteins are primarily dimeric though a very plastic dimerization interface, formed by a patch of hydrophobic residues on the outer face of the β-sheets. A number of quaternary structure orientations are formed by PAS domains, even multiple orientations of the same oligomeric state

in the same solution.[7,8] Signal binding-dependent change in oligomeric state is a necessary component of the function of many PAS proteins. [7,9–13] RcoM was recently shown be primarily dimeric in solution through its PAS domains, though it has not been established if the oligomeric state changes in response to CO binding.[3]

RcoM undergoes a ligand-switching mechanism at heme to sense CO (Figure 3.1A). In the Fe(III) state, heme is bound by Cys94 and His74.[2,4,14,15] The reduction of Fe(III) heme causes the charged cysteine thiolate ligand to be replaced by neutral Met104, then methionine is replaced by the incoming CO.[2,16] CO binding activates RcoM to bind its promoter DNA through allosteric coupling to the LytTR domain.[3,17] Unlike archetypical CO-sensor CooA, RcoM displays an extremely high affinity for CO in the presence of oxygen, consistent with its primary biological role as an aerobic CO sensor. In addition to the observation made in several studies that heterologous protein expression yields CO-bound RcoM, spectroscopic characterization of CO binding and dissociation kinetics found a nanomolar range CO affinity (K_d = 4 nM) for RcoM-2.[17,18] While the details of CO binding to heme have been well-studied, little is known the mechanism by which the signal of CO binding is propagated across the protein to the LytTR domains to modulate DNA binding. While the specific residues involved in forming an allosteric network between heme and the LytTR domains will likely remain unknown until high-resolution structures can be achieved, there is still much that can be learned about large scale CO-dependent conformation changes that contribute to RcoM allostery.

The LytTR family of proteins are a class of response regulators that primarily regulate bacterial virulence, and RcoM is unique among this group as, to our knowledge, the only biochemically characterized single-component transcription factor.[19] A crystal structure of one LytTR domain of truncated AgrA bound to a 15 bp piece of DNA revealed 10 β-strands, organized

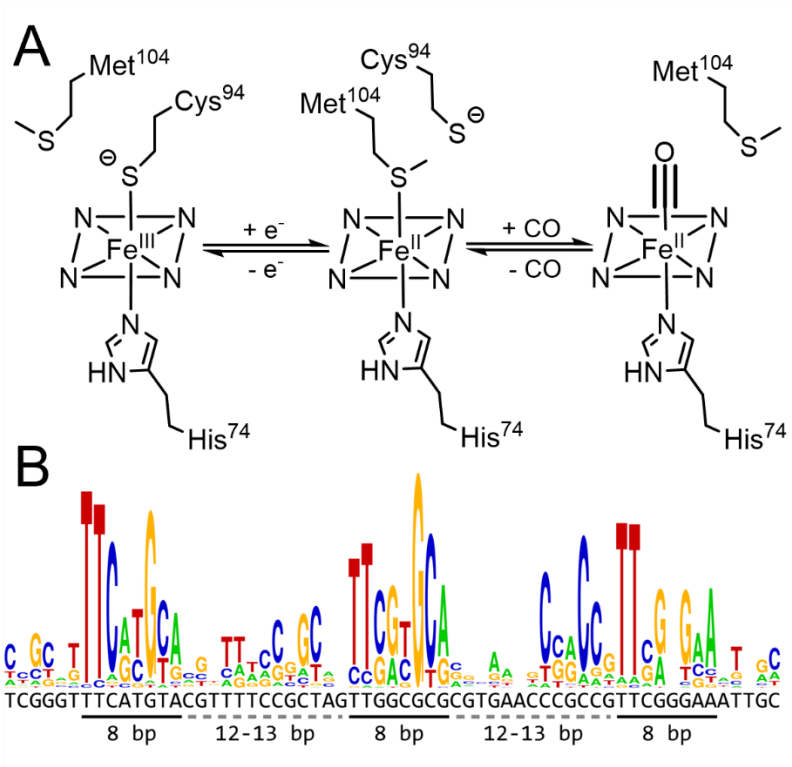


Figure 3.1. A. *Px*RcoM-1 heme coordination states showing the Fe(III), Fe(II), and active Fe(II)-CO state that is able to interact with the promoter region. B. Logo plot derived from the alignment of 40 *coxM* promoters, shows conservation of the triplet direct repeat motif.[3] Below the logo plot is the native sequence from the *Paraburkholderia xenovorans* genome that is recognized by RcoM-1.

into three antiparallel β-sheets, with a small two-turn α-helix.[20] LytTR domains exhibit a unique DNA binding motif, where the loops between β-strands form specific interactions with two major grooves of the DNA with only two amino acids forming base-specific DNA contacts.[20,21] LytTR proteins tend to bind a pair of 8-11 bp imperfect direct repeat sequences, with 12-13 bp spacing regions, upstream of the -35 region of the relevant operon.[19,22–27] Dimerization is required for recognition of the doublet promoters, though some LytTR proteins bind DNA in a stepwise manner as monomers and others form dimers upon activation prior to DNA binding.[21,28] The crystal structure of a constitutively active variant of full-length ComE shows that the active state of the protein is an asymmetric dimer. ComE dimerizes through its REC receiver domains that exhibit rotational symmetry, with the conformation of a long flexible linking region between the REC domains and the LytTR domains allowing a tandem arrangement of the DNA binding domains consistent with the two direct repeats in its promoter sequence.[21]

The RcoM consensus promoter sequence was identified by bioinformatic analysis and *in vitro* fluorescence anisotropy-based DNA binding assays as a set of three 8 bp imperfect direct repeats of sequence "TTnnnGCA", separated by 13 bp, which RcoM binds with a moderate apparent affinity, $[P]_{1/2}$ = 250 nM and a Hill coefficient n = 1.7 (Figure 3.1B).[3] This study was able to eliminate another triplet direct repeat site, found just downstream of *rcoM1* and about 100 bp upstream of the *coxM* promoter, as necessary for *coxM* transcription.[29] The role of this second set of triplet direct repeat sequences is unknown. The existence of a triplet direct repeat promoter was dissimilar to the promoters of other LytTR proteins, that typically bind sets of two imperfect direct repeats with a dimeric activated state, so this set up significant questions for how RcoM functions differently than other characterized LytTR proteins. The *coxM* promoter organization implies that two asymmetric RcoM dimers are involved in the interaction, with a parallel organization of PAS

domains and tandem LytTR domains similarly to that observed in ComE. We hypothesize that CO binding at heme induces rearrangement of the PAS domain that allows a dimer of dimers to bind the triplet direct repeat promoter.

In light of the absent structural data for RcoM and open questions about the RcoM activation mechanism, we sought to better understand the CO-dependent structural changes that lead to interaction with DNA. In this paper we will present a structural model of the RcoM dimer by combining protein structure prediction and small-angle X-ray scattering (SAXS). Using size-exclusion chromatography (SEC) and fluorescence anisotropy-based DNA binding assays, we assess DNA binding properties of the RcoM:*pcoxM* system to produce a higher-affinity, more stable complex. Then through chemical crosslinking and SEC-SAXS, we are able to achieve a low-resolution structural model of the RcoM:DNA interaction. Solution-based structures are an essential first step in understanding CO-dependent DNA binding of RcoM. Together, these data provide additional support for the unique DNA binding mode of RcoM: two dimers bound to a triplet direct repeat DNA motif, which we believe to be highly dynamic.

3.2 Materials and Methods

Materials

All chemicals used in buffer and media preparation (99% purity or greater), sodium dithionite (85% purity, Fluka, stored under Ar(g) at -20°C), 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. Cell stocks of *Escherichia coli* containing wild type (WT) *Px*RcoM was originally provided by Kerby (University of Wisconsin−Madison, Department of Bacteriology).[4] Texas-Red-X labeled oligonucleotides for fluorescence anisotropy DNA-binding assays were synthesized by Integrated

DNA Technologies (IDT). Un-labeled complementary strands underwent a standard desalting purification, whereas PAGE purification was performed on labeled strands.

Protein expression, purification of RcoM-1 and RcoM-1 variant, H74Y

*Px*RcoM-1 and H74Y variant was expressed and purified by methods previously described.[3,4,15,16,29] *E. coli* BL21(DE3) containing the pEXT20 expression vector for *Px*RcoM-1 was grown in Lennox-LB medium. Dense overnight culture was diluted 1:250 into Terrific Broth supplemented with ferric citrate (20 mg/mL) and grown to an OD_{550} of 0.8-1 at 37° and 220 rpm. Isopropyl β- D-thiogalactopyranoside (1 mM, IPTG) was added to induce expression with δ-aminolevulinic acid (50 μM, δ-ALA) to improve heme loading for 19− 20 h at 30 °C with shaking. Cells were harvested by centrifugation, resuspended in binding buffer [50 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 500 mM NaCl, 10 mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), pH 7.4], and lysed by sonication. A HisTrap HP column (Cytiva, 5 mL column volume) was pre-equilibrated on an ÄKTA Prime FPLC with binding buffer. The cell supernatant was applied to the column and washed with binding buffer until the absorbance at 280 nm reached baseline. Additional washing of the bound *Px*RcoM-1 was achieved by adding four column volumes of 15:85, 25:75, and 35:65 (% v/v) mixtures of elution buffer (50 mM MOPS, 150 mM NaCl, 220 mM imidazole, 1mM TCEP, pH 7.4) and binding buffer to the column. Elution of bound *Px*RcoM-1 was achieved by washing with four column volumes of the elution buffer. The red-colored fractions were pooled and incubated in 45% (v/v) saturated ammonium sulfate on ice for 15 min. Aggregated protein was harvested by centrifugation, and the resulting protein pellet was resuspended in storage buffer (50 mM MOPS, 150 mM NaCl, 1 mM TCEP, pH 7.4) and flash frozen with liquid nitrogen before being stored at -80°C. Total protein concentration was assessed with Pierce 660 nm assay (Pierce- Thermo Fisher) using bovine serum

albumin as a calibration standard. Protein purity was greater than 90%, as assessed by SDS-PAGE. Heme-loading was determined by pyridine hemochromagen assay.[30]

Protein Structure Prediction

The dimeric protein structure was predicted by ColabFold with the expressed *Px*RcoM-1 (RCOM1_PARXL) sequence using a selected portion of the BFD database and no homologous structures.[31–33]. No prior information was provided for building the monomer or dimer structure. The predicted structures were used without minimization and only altered based on SAXS data as described below.

Reoxidation, Reduction and CO-binding of RcoM

This method is used to prepare protein for fluorescence anisotropy DNA binding assays and SAXS experiments.

Fe(III) *Px*RcoM-1 was prepared by incubation with 3mM potassium ferricyanide ($K_3[Fe(CN)_6]$) on ice for 30 min before desalting by passing through a Sephadex G-25 gravity column pre-equilibrated with appropriate buffer (buffer contents listed in figure legends). The desalted protein was concentrated using Amicon spin concentrators (Millipore, 30 kDa MWCO, 500 μL).

Fe(II)-CO *Px*RcoM-1 was prepared by adding a small amount of solid sodium dithionite (DTH) to protein, allowing to reduce for 2 minutes, and passing through a gravity desalting column pre-equilibrated with CO-sparged buffer (buffer contents dependent on experiment, described in figure legends). The desalted protein was concentrated using Amicon spin concentrators.

Fluorescence Anisotropy DNA Binding Assay

For each fluorescence polarization assay, Fe(III) *Px*RcoM-1 was mixed with a double-stranded DNA (dsDNA) oligonucleotide containing a 5′ Texas Red-X (TRX) fluorescent label and a putative *Px*RcoM-1 binding site (Table 3.1). Final protein concentrations varied from 0.001 to 3 μM, and assay conditions in each 500 μL sample were as follows: 40 mM Tris·HCl pH 8.0, 20 mM KCl, 5% (v/v) glycerol, 5.0 mM DTT, 48 μg sheared salmon sperm DNA, and 3.2 nM TRX-labeled dsDNA. Each Fe(III) *Px*RcoM sample was loaded into a septum-sealed glass culture tube (6 × 50 mm dimensions, VWR). Anisotropy values were recorded using an ISS PC1 fluorimeter. The headspace of each tube was purged with Ar(g) for 10 min, and protein samples were reduced by anaerobic addition of 10 μL of 100 mM sodium dithionite, prepared in 40 mM Tris·HCl pH 8.0, 20 mM KCl, 5% (v/v) glycerol. Fe(II)−CO adducts were generated by introducing 100 μL CO(g) to the headspace of each tube using a gas-tight syringe and gently vortexing the sample. Anisotropy values were recorded immediately after the introduction of CO and after a 40 min incubation period at room temperature to ensure complete conversion to the Fe(II)−CO species and equilibration of DNA binding. Each data point represents an average of three technical repeat measurements of anisotropy.

The Hill equation was used to fit a cooperative-binding model to fluorescence anisotropy data taken of the Fe(II)−CO species after the 40 min incubation period.

$$A = A_f + \frac{(A_b - A_f)}{1 + \left(\frac{[P]_{1/2}}{[P]}\right)^n}$$

Where A is the observed fluorescence anisotropy, A_f is the baseline fluorescence anisotropy in the absence of binding, A_b is the maximum fluorescence anisotropy (i.e., when all protein is bound to DNA), $[P]_{1/2}$ is the protein concentration at half- saturation, $[P]$ is the total protein concentration, and n is the Hill coefficient.

Table 3.1. Double stranded oligonucleotide sequences with 8 bp binding sites underlined and changes from the native sequence bolded.

Name	Length	Sequence[a,b]
"def"	62	CGGGTTTCATGTACGTTTTCCGCTAGTTGGCGCGCGTGAACCCGCCGTTCGGGAAATTGCGT
"de+f"	62	CGGGTTTCATGTACGTTTTCC**TAC**AGTT**CA**CG**CA**CGTGAACCCGCCGTTCGGGAAATTGCGT
"de-f"	62	CGGGTTTCATGTACGTTTTCCGCTAG**GC**GGC**TT**GCGTGAACCCGCCGTTCGGGAAATTGCGT
"def-"	62	CGGGTTTCATGTACGTTTTCCGCTAGTTGGCGCGCGTGAACCCGCCG**GA**CGG**TT**AATTGCGT

a. Only 5' → 3' sense strand displayed for clarity.

b. Oligonucleotides used in fluorescence-based DNA binding assays were labeled with a 5'-Texas-Red-X fluorophore on the sense strand.

Chemical crosslinking with Glutaraldehyde

*Px*RcoM-1 (60 μM), dsDNA (30 μM ,"def" or "de+f", Table 1), and DTH (1.7 mM) were incubated at room temperature for 15 min in CO-sparged DNA binding buffer (40 mM MOPS, 75 mM KCl, 5 mM DTT pH 7.4) to form the RcoM:DNA complex. To initiate crosslinking, 0.1% glutaraldehyde was added at 4°C. Crosslinking was stopped after 10 or 20 minutes by diluting 50x with SEC buffer (50 mM MOPS, 150 mM NaCl, 5 mM TCEP pH 7.4). The crosslinked protein solution was passed through a gravity desalting column equilibrated with SEC buffer and concentrated with Amicon spin concentrators.

Small Angle X-ray Scattering: Data Collection and Analysis

Size-Exclusion Chromatography coupled Small-Angle X-ray Scattering (SEC-SAXS)

Fe(III) and Fe(II)-CO RcoM (10 mg/mL, 50 mM MOPS, 150 mM NaCl, 10 mM TCEP pH 7.4) were shipped overnight at 4°C to SIBLYS Beamline 12.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA.[34–36] Samples were separated on either a Shodex KW-802.5 or KW-803 column at a flow rate of 0.5 ml/min at 10 °C, and eluate was measured in-line with UV/vis absorbance at 280nm, Multi-Angle X-ray Scattering (MALS), and SAXS. The incident light wavelength was 1.03 Å at a sample to detector distance of 1.5 m. This setup results in scattering vectors, q, ranging from 0.013 $Å^{-1}$ to 0.5 $Å^{-1}$, where the scattering vector is defined as $q = \frac{4\pi sin\theta}{\lambda}$ and 2θ is the measured scattering angle. The data collection statistics are available in Table 3.2.

SAXS Data Analysis of RcoM Dimer

Radially averaged SAXS data files were processed and analyzed in Scatter IV and RAW to determine the dimensions of the protein, molecular weight, and pair-distance distribution function

(PDDF) .[37] The resulting data statistics are available in Table 3.2. The electron density for the dimer was calculated using DENSS within the RAW software.[38] The default settings were used with a two-fold symmetry axis through the second longest dimension.

We used ColabFold and the expressed sequence of the RcoM protein to model the atomic coordinates of the RcoM dimer.[39] During modeling, we only used sequences that covered >75% of the input sequence. The top scoring model was then fitted to the electron density from DENSS within RAW and the structure refined based on this alignment in YASARA structure. During refinement the LytTR domains were rotated 180° to put the likely DNA binding residues away from the heme-binding domain followed by a 30° rotation of the entire domain toward the heme-binding domain. The model and electron density were then realigned.

SAXS Data Analysis of RcoM:DNA Complex

For the crosslinked mixture of RcoM with DNA, radially averaged SAXS data files were processed and analyzed in Scatter IV and RAW. Peaks containing potential RcoM:DNA complexes were identified by comparing plots of integrated intensity vs. elution volume to similar plots for RcoM and DNA alone. Similar frames from the peaks were averaged to generate the data to determine the dimensions of the protein, molecular weight, and PDDF.

To produce models for potential RcoM:DNA complexes, we used HADDOCK, which can generate several possible complexes from 3-D models and suggest interacting residues.[40,41] We used the DNA sequence to structure server to model the 62 bp dsDNA.[42] For modeling the complex, we provided HADDOCK with the refined RcoM dimer model and a model of free dsDNA. Interacting residues in RcoM were arginine 208, 216, and 219 and passive residues were allowed using the default HADDOCK settings. Complexes were modeled in 1:1, 2:1, or 3:1

RcoM:DNA. The 1:1 and 2:1 complexes were modeled with RcoM occupying all of the possible combinations interactions with the d, e, or f sites. Representative models from each of the HADDOCK results were randomly selected for fitting to the SAXS data in FoXS.[43] This software predicts the SAXS data for 3-D models and can fit these predicted data to experimental data. All the representative HADDOCK models were combined into a single compressed file along with models of RcoM and DNA alone. This file was provided to the FoXS server for fitting to the SAXS data of each of the RcoM:DNA data sets.

3.3 Results

SEC-SAXS refines the predicted dimeric structure of PxRcoM-1

A reliable structural model of the RcoM dimer was needed in order to describe the molecular interactions between RcoM and DNA. Given the lack of a high-resolution crystal structure and no full-length structural homologs consisting of a PAS-type heme-binding domain and a LytTR-type DNA binding domain, ColabFold was used to predict the structure of dimeric *Px*RcoM-1 (Figure 3.2A). The predicted local distance difference test (pLDDT) score of 67.8 and the predicted template modeling (pTM) score of 0.59 indicates reasonable confidence in the accuracy of the model (Figure 3.S1). The predicted structure displays each monomeric unit interacting through parallel PAS domains, via the N-terminal helices. Heme-binding residues His74 and Met104 are positioned within a pocket of the PAS domains which could accommodate a heme group, with the third heme ligand, Cys94, on a nearby loop. The PAS domain connects to the LytTR domain through a 21 residue unstructured linking region, potentially allowing for mobile LytTR domains while the PAS domains are dimerized. Lastly, the DNA binding domains are modeled with high structural similarity to the only two structures of LytTR domains, AgrA and ComE. This model of RcoM positions the DNA binding domains with the likely DNA-contacting residues (His218 and

Arg219) on the outside face of the model, at an angle that is unlikely to allow for interaction with DNA. The presence of the unstructured linkers and the potentially non-functional orientation of the DNA binding domains meant that this model needed further validation before we could use it to describe DNA binding by RcoM.

Size-exclusion chromatography demonstrates that RcoM remains primarily dimeric across all allosteric states. *Px*RcoM-1 has a monomeric sequence molecular weight of 30.8 kDa, including the 6-His tag and one heme *b* molecule. The UV (A280nm) and MALS detector at the SIBYLS beamline allows for collecting SEC data on the same samples, Fe(III) and Fe(II)-CO RcoM at low and high concentrations, submitted for SAXS structural analysis. The apparent molecular weight of the major species in solution, calculated from in-line MALS, is 68.3 and 67.5 kDa for the Fe(III) and Fe(II)-CO species respectively (Figure 3.3A). All samples show that the primary species in solution is RcoM dimer, and there is a smaller proportion of a larger oligomeric state present. Fe(III) and Fe(II)-CO RcoM, in both the lower and higher concentration samples, both contain the same proportion of dimer and larger oligomer. These data suggest that RcoM does not employ an effector-dependent change in oligomeric state as the allosteric mechanism.

Size-exclusion chromatography coupled small-angle X-ray scattering data allow us to refine the ColabFold model of the RcoM dimer. SAXS data from frames corresponding to dimer were used for proceeding with low-resolution solution structural characterization (Figure 3.3B, C). SAXS data from the most prominent peak for Fe(III) RcoM (5 mg/mL) has an Rg value of 29.3 ± 0.8 Å and a molecular mass of 59.5 [94% C.I. 55 to 60.2] kDa using the Bayes method, indicating that the dimer is the primary species in solution.[44] Analysis of Fe(II)-CO RcoM (5 mg/mL) showed a similar distribution of species and the Rg of the most prominent peak was 30.2 ± 0.6 Å, suggesting that the protein is in a similar conformation. Kratky plots of the dimeric species show

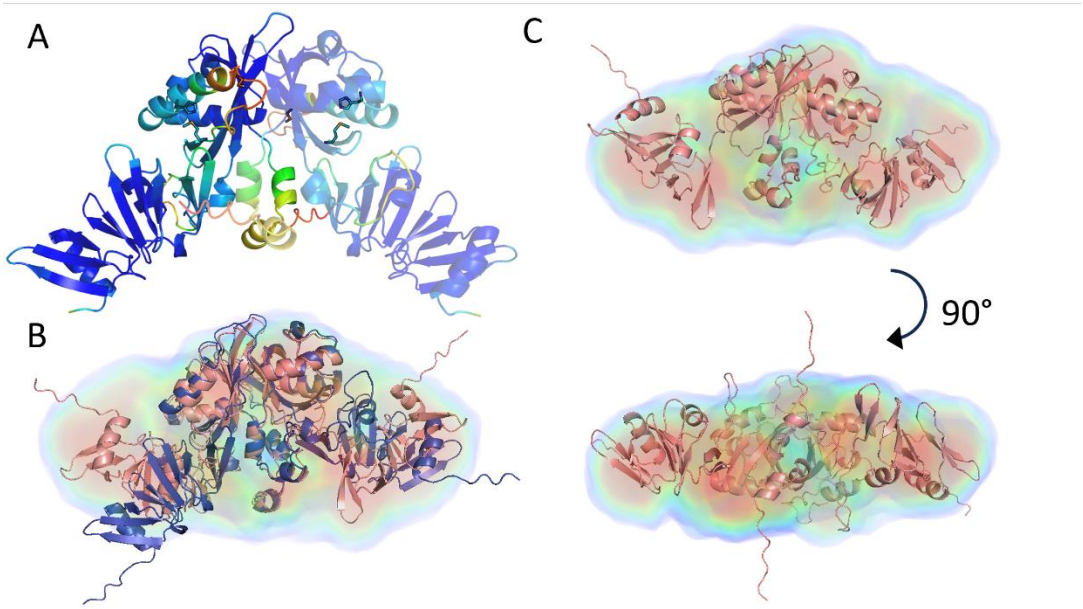


Figure 3.2. A. ColabFold model of *Px*RcoM-1 dimer, colored with pLDDT confidence values. Residues predicted with high confidence (pLDDT > 90) are dark blue, confidence (90 > pLDDT > 70) are light blue, low confidence (70 > pLDDT > 50) are yellow, and very low confidence (pLDDT < 50) are orange. Heme ligands, His74, Cys94, and Met104 are shown as sticks within the PAS domain. One monomer is transparent. B. Overlayed ColabFold model (blue) and SAXS refined model (pink), displaying how LytTR domains were rotated to better fill the experimental SAXS electron density. C. SAXS refined model of RcoM dimer (pink) overlayed with experimental SAXS electron density, shown from two angles.

that the Fe(III) and Fe(II)-CO samples were folded, globular proteins with similar shapes (Figure 3.3D). These initial data were sufficient for basic structural descriptions of RcoM, but insufficient for refining the structural model of RcoM. Therefore, we collected further SAXS data on Fe(II)-CO RcoM at a higher concentration. This new data set had a similar Rg value (30.1 ± 0.1 Å) and molecular weight (62.4 kDa [90.4% C.I. 61.6 to 67.9]) as the lower concentration data sets.

Using both the ColabFold model and the SAXS data allowed for the development of a refined model of the RcoM dimer. We used DENSS to calculate the electron density from the Fe(II)-CO RcoM (10 mg/mL) data. The calculated density had a real space correlation of 0.77 ± 0.02 and a Fourier shell correlation resolution of 38.5 ± 4.2 Å, both of which were reasonable values for SAXS. The ColabFold model was then fitted to the electron density; however, the LytTR domains fell outside the density. We then adjusted the position of the LytTR domains by moving each domain individually to fit the density (Figure 3.2B). Then we had to rotate the domains to keep the putative DNA binding residues accessible given that we had no evidence to indicate a change in structure between the active, CO-bound and inactive forms. The final refined model of RcoM is the first experimentally supported structure for this class of protein and can be used to provide insight into the RcoM DNA binding mechanism (Figure 3.2C).

Enhancing RcoM DNA binding

With a model of RcoM dimer available, we next turned to identify conditions to increase the likelihood that a RcoM:DNA species could be observed. Biochemical techniques were used to explore the formation of the RcoM:DNA complex with the goal of maximizing and stabilizing the formation of a complex in solution for structural characterization. *In vitro* fluorescence anisotropy assays had showed a relatively low apparent affinity of active RcoM for its native promoter sequence, challenging the study of the complex in solution. We aimed to identify

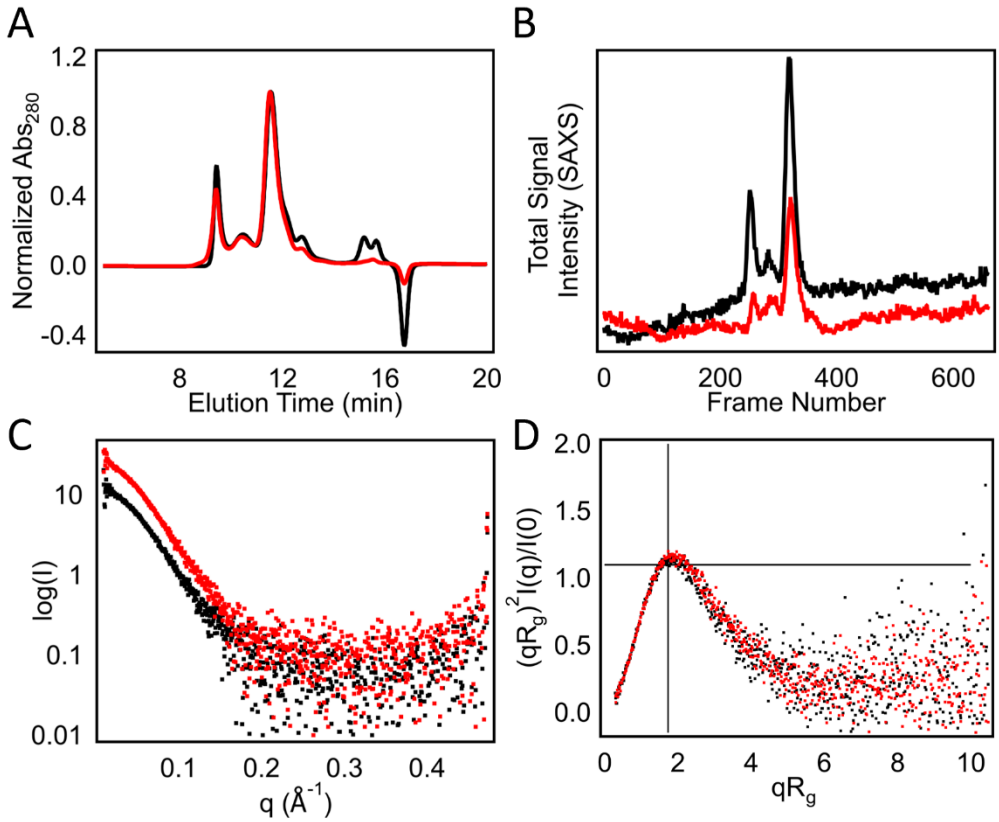


Figure 3.3. SEC-SAXS data for 5 mg/mL Fe(III) (black) and Fe(II)-CO (red) *Px*RcoM-1 in 50 mM MOPS, 150 mM NaCl, 10 mM TCEP pH 7.4. A. SEC chromatograms of Fe(III) and Fe(II)-CO *Px*RcoM-1, shown as normalized absorbance at 280 nm and detected by total SAXS signal intensity. B. SAXS data frames corresponding to RcoM dimer were used for proceeding SAXS analysis. C. Log of intensity versus momentum transfer for Fe(III) and Fe(II)-CO *Px*RcoM-1 dimer. C. Kratky plot of Fe(III) and Fe(II)-CO *Px*RcoM-1 dimer.

strategies to minimize protein preparation, increase DNA affinity, and prevent protein aggregation. We first characterized RcoM variant H74Y. Despite containing a perturbed heme environment that cannot bind heme, H74Y RcoM is constitutively active to bind DNA.[4] The heme-free variant was evaluated for its activity in the presence of the native promoter with fluorescence anisotropy-based DNA binding assays; H74Y RcoM displays apparent affinity $[P]_{1/2} = 118 \pm 7.58$ and $n = 2.2 \pm 0.2$ (Figure 3.S2). Size exclusion chromatography traces of H74Y RcoM shows that the protein is dimeric, and mixtures of the H74Y RcoM with promoter DNA reveal differences in peak width and peak tailing (Figure 3.S2). These results reveal that there is not a clean transition from dimeric RcoM to RcoM:DNA complex, though a change in protein shape and mobility can be seen in the samples with DNA. Studying H74Y RcoM does not require evaluations of heme incorporation or labor-intensive sample preparation to ensure consistent heme-functional state, making this variant an interesting candidate for structural investigation.

We further found that the affinity of the native promoter could be increased by altering the middle repeat sequence to look more like consensus, resulting in the ability to produce a more stable DNA complex in solution. Kerby et al, 2012 report a 62-bp sequence with roughly twice the affinity as the native promoter sequence.[29] This "improved" sequence alters the middle binding site sequence to more closely resemble the 8-bp consensus sequence (Table 3.1, Figure 3.1).[3] By *in vitro* fluorescence anisotropy assays, we find the "improved" "de+f" oligonucleotide binds Fe(II)-CO RcoM with $[P]_{1/2} = 92 \pm 5$ nM with a $n = 1.7 \pm 0.1$, compared to the native sequence "def" oligonucleotide of $[P]_{1/2} = 250 \pm 10$ nM with a $n = 1.7 \pm 0.1$, indicating more than twice the apparent binding affinity as the native sequence (Figure 3.4A).[3]

Altered DNA sequences that sequentially drop out binding sites while maintaining a consistent length display significantly reduced interaction with active RcoM, showing that all three

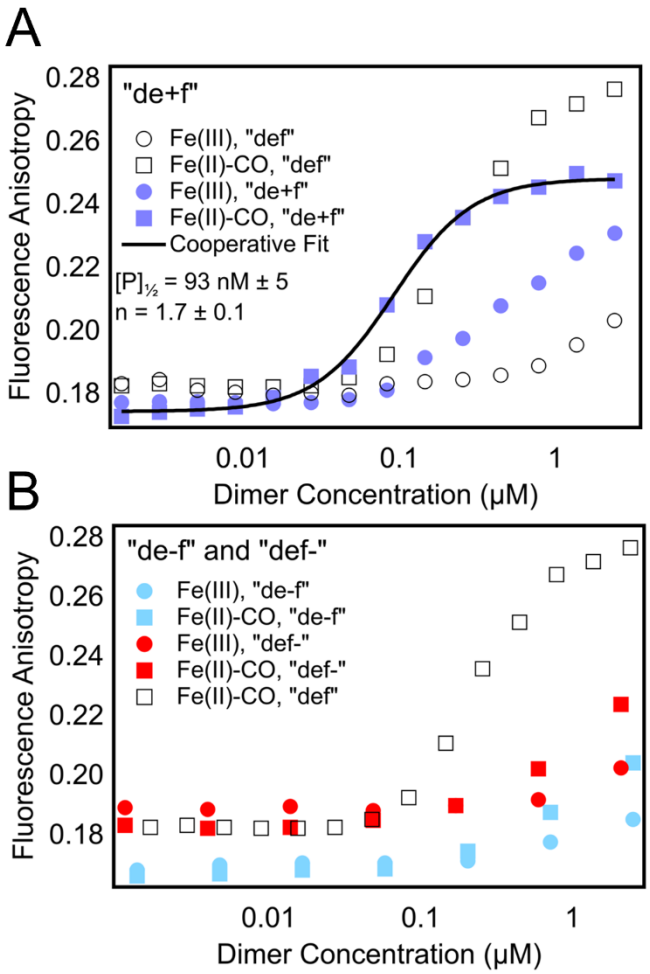


Figure 3.4. A. Fluorescence polarization data for active (squares) and inactive (circles) RcoM with the higher-affinity "de+f" dsDNA oligonucleotide (blue), a 62 bp piece of dsDNA with the middle triplet repeat site altered to more closely resemble the consensus binding site, and the native 62 bp triplet repeat sequence "def" dsDNA (open shapes). The solid line shows the best fit of the "de+f" data to the Hill Equation (χ^2= 4.70699 x 10^{-5}). B. Fluorescence polarization data for active (squares) and inactive (circles). RcoM sequentially dropping out binding sites from the promoter sequence. "de-f" oligonucleotide data with the middle site dropped out is shown in light blue, and the "def-" oligonucleotide with the last site dropped out is shown in red, with the native sequence shown as an open square to facilitate comparison.

binding sites must be included to observe binding between RcoM and the triplet direct repeat promoter sequence. Binding sites were minimally altered by switching critical base pairs of the consensus sequence, the first two TT base pairs, and the middle two base pairs, GC/A (Figure 3.1). These changes produced two new 62 bp oligonucleotides, dropping out the middle site, called "de-f", and dropping out the last site, called "def-" (Table 3.1). Dropping out either site eliminates the strong, specific interaction seen between Fe(II)-CO RcoM and the native promoter sequence (Figure 3.4B), confirming that all three sites must be involved in forming the RcoM:DNA complex.

Screening DNA binding conditions with high-throughput SAXS (HT-SAXS) resulted in identifying conditions that allowed for the formation of RcoM:DNA complex. The HT-SAXS screened a wide range of concentration ratios of WT and H74Y RcoM, native "def" and higher affinity "de+f" DNA sequences, in a variety of buffer conditions to identify those which could produce stabilized the complex. We did not observe any conditions which resulted in complete or near complete complex formation for structural modeling. However, we found that a 2:1 ratio of RcoM:DNA resulted in data which we posited was a strong indicator for complex formation for WT RcoM with both the native "def" and high affinity "de+f" oligonucleotides (Figure 3.S3). We did not observe significant enhancement of complex formation by the H74Y variant. The heterogeneity of samples prohibited successful modeling of electron density or the molecular envelope from these data, and we were unable to use deconvolution methods to separate species.

Using the conditions identified by HT-SAXS to provide the best complex formation, glutaraldehyde crosslinking of WT RcoM with DNA generated a stabilized complex species. A solution containing 0.1% glutaraldehyde was identified to provide the best conditions for chemical crosslinking of the RcoM:DNA complex. We planned to use SEC-SAXS to separate the complex

from other species, so the final crosslinking conditions were chosen to minimize over-crosslinking into a non-biologically relevant aggregate. Ultimately, the RcoM:DNA complex samples submitted for analysis with SEC-SAXS were crosslinked with either the “def” or “de+f” DNA sequence in a 0.1% glutaraldehyde solution for 10 or 20 minutes, resulting in the formation of a new species roughly 100 kDa, visible in SDS-PAGE (Figure 3.S3)

A trimeric RcoM:DNA complex

Chemical crosslinking of RcoM with “def” and “de+f” DNA allowed us to collect SEC-SAXS data on the RcoM:DNA complex. The SEC-SAXS data of crosslinked RcoM:DNA samples show peaks corresponding to free dimeric RcoM and several RcoM or RcoM:DNA complexes, including a large aggregate peak (Figure 3.5A). We then focused on a peak, which was not observed in the RcoM or DNA alone data collection. This peak appeared across all four samples with a Guinier Rg value between 42 and 43.8 Å and had a distinct Kratky plot shape from RcoM alone showing two peaks instead of one, which can be indicative of multi-subunit complexes (Figure 3.5B). Molecular weight estimates ranged from 109 to 147 kDa, corresponding to a 1:1 to 2:1 RcoM:DNA complex, which would have an expected masses 97 and 157 kDa, respectively. However, we only cautiously use these mass values as the Bayes method has not been optimized for protein-DNA complexes. For all four of the datasets, we observed similar shapes in Kratky plots and log(I) vs. q plots, with the native “def” DNA and higher-affinity sequence “de+f” clustering separately, suggesting to us a similar structure.

Fitting experimental SAXS data with the refined RcoM dimer model provides further evidence that two RcoM dimers form the complex with DNA. We used the HADDOCK webserver

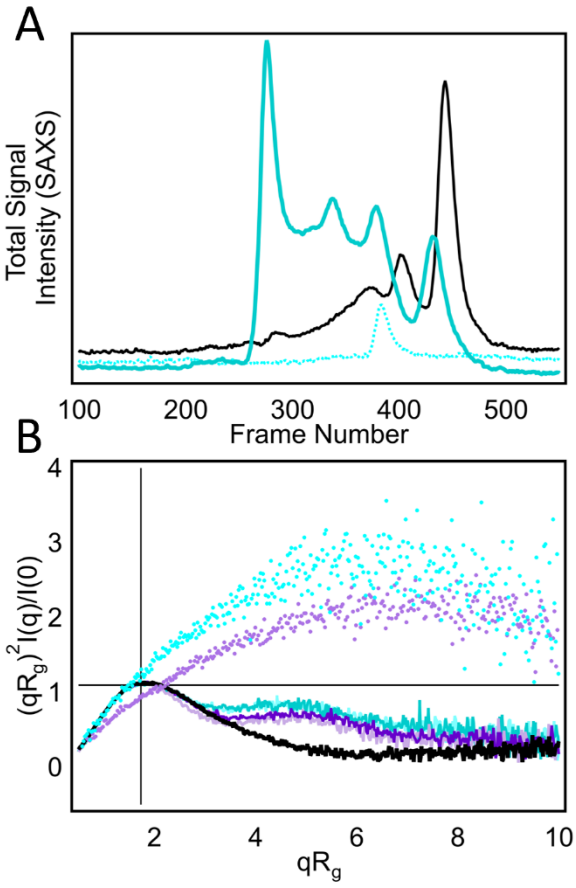


Figure 3.5. SEC-SAXS data for crosslinked 60 μM Fe(II)-CO RcoM and and 30 μM DNA, either the native sequence "def"(cyan) or enhanced affinity sequence "de+f" (purple). A. Representative SEC-SAXS trace of crosslinked Fe(II)-CO protein with "def" DNA (cyan solid line), "def" ds oligonucleotides (cyan dotted line) alone, and Fe(II)-CO alone (black line) . B. SAXS data frames corresponding to newly formed crosslinked species were used for further analysis. Kratky plot of Fe(II)-CO RcoM (black), 25 μM "def" DNA alone (cyan dots), 25 μM "de+f" DNA alone (purple dots), RcoM + "def" ds oligonucleotide crosslinked for 10 or 20 min (light and dark cyan lines, respectively) , RcoM + "de+f" ds oligonucleotide crosslinked for 10 or 20 min (light or dark purple lines, respectively).

to predict the conformations of RcoM and DNA in several ratios based on known predictions of how LytTR domains interact with DNA. We also modeled configurations where a single RcoM dimer interacted with multiple sites on the DNA and configurations where multiple RcoM dimers interacted with multiple sites to encompass all the plausible complexes. We fitted representative samples from each ratio and configuration to all four of the datasets using FoXS.[43,45] Pooling of the data from all four fittings showed that 2:1 RcoM:DNA complexes as a group fitted better to the datasets (Figure 3.6A) . When we looked into the best 2:1 RcoM:DNA models from each fitting, two models from the same run appeared as the best fitting structures (Figure 3.6B). The X^2 values could be modestly improved by fitting an ensemble of a 2:1 RcoM:DNA complex and a RcoM tetramer (dimer of dimers, which would have similar mass) in Multi-FoXS. In further analysis of the complex data, we observed that both the Rg and the Dmax for the complex were smaller than that for the free DNA. The basis for this apparent reduction in dimensions is not clear but may suggest a bending of the DNA when bound to RcoM. Overall, these data suggest RcoM forms a 2:1 complex with two RcoM dimers bound to DNA.

The best HADDOCK model of the RcoM:DNA complex leaves questions remaining about the orientations of RcoM dimer on DNA and necessary quaternary structural interactions. The HADDOCK derived model of the RcoM:DNA complex that best fits the SAXS data of crosslinked RcoM and “def” DNA shows two dimers interacting with the DNA binding sites of the native promoter (Figure 3.6B). The LytTR domains of one dimer are shown interacting with the first two DNA binding sites, and the second dimer is shown with one LytTR domain close to the middle DNA binding site and the second LytTR domain unbound to DNA. Interestingly, the two dimers are not interacting with each other or the third DNA binding site. The fit to experimental data was

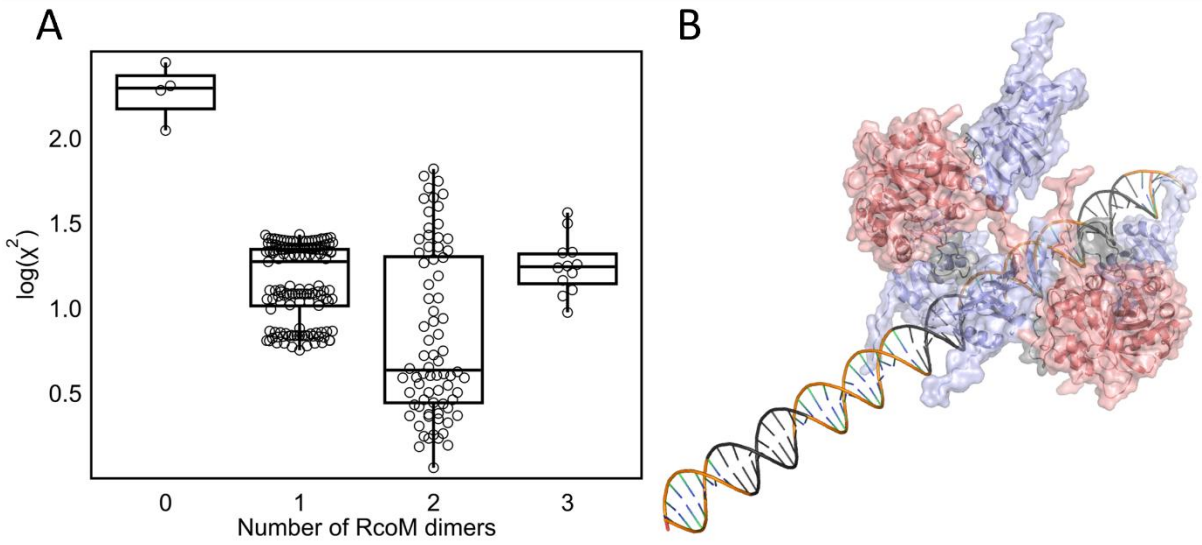


Figure 3.6. A. Box plot representing HADDOCK generated models of the RcoM:DNA complex. Plot shows the fit to the experimental SAXA data as a function of the number of RcoM dimers in that complex. B. The best HADDOCK generated model of the RcoM:DNA complex involves two RcoM dimers interacting with a linear piece of DNA, where the 8 bp binding sites are colored black. RcoM dimers are colored with PAS domains in pink, unstructured linking region in black, and LytTR domains in blue. The cartoon is shown overlayed with the protein surface. Table 2. Characteristic Parameters of SAXS experimental curves and theoretical models. Dashes are used for values that do not apply to that sample.

not improved by inducing a bend in the DNA, as is consistent with other structurally characterized LytTR proteins.[20,46]

3.4 Discussion

In this work, we build the first experimentally-supported models of RcoM dimer and RcoM:DNA complex, while providing insight into the functional properties of RcoM that allow for interaction with its promoter. We produced a model of the RcoM dimer refined with experimental SEC-SAXS data. This model, in agreement with previous analysis, shows RcoM PAS domains forming a parallel dimer through the N-terminal helices, and the LytTR domains are noninteracting. Creation of a low-resolution structural model of the RcoM:DNA complex was only achievable by crosslinking with glutaraldehyde prior to SEC-SAXS. We identified a likely structural configuration of the RcoM:DNA complex based on fitting SAXS data to a library of possible structures, indicating that the complex is best modeled by two RcoM dimers. These findings provide new information on the structure of this class of protein and a starting point to describe the binding mechanism of RcoM to its promoter.

The ColabFold structural model of the RcoM dimer could be refined to produce the first experimentally supported structure of the RcoM dimer. In fitting the ColabFold model to the SAXS electron density, we rotated the LytTR domains 180° to better fill the experimental electron density. The rotation places the likely DNA recognizing residues on an outer face where they could interact with DNA.[4] Viewing the electrostatic data of the final refined model shows that the outer face of each LytTR domain is a positively charged patch, making good DNA targets (Fig. 8). The unstructured 21 residue linking region allows for significant flexibility in the positioning of the LytTR domains relative to the PAS domains. Our predictions about requiring an asymmetric dimer in order to bind direct repeats are not yet satisfied by this model, but again we note that the long

Table 3.2. Characteristic Parameters of SAXS experimental curves and theoretical models. Dashes are used for values that do not apply to that sample.

	Fe(III) RcoM	**Fe(II)-CO RcoM**	**"def" dsDNA**	**"de+f" dsDNA**	**XL RcoM:"def"**	**XL RcoM:"de+f"**
Wavelength (Å)	1.127	1.127	1.127	1.127	1.127	1.127
q range ($Å^{-1}$)	0.0099 - 0.4734	0.0099 - 0.4734	0.0114 - 0.4729	0.0114 - 0.4729	0.0114 - 0.4729	0.0114 - 0.4729
Concentration (mg/mL)	5	5	1	1	9 mg/mL protein + 3 mg/mL DNA	9 mg/mL protein + 3 mg/mL DNA
Temperature (K)	283	283	283	283	283	283
I(0) (A.U., P(r))	22.89±0.3	37.45±0.1	21.44±2.12	28.06±1.4	60.73±0.5	61.31 ± 0.4
R_g (Å, P(r))	29.99±0.3	30.18±0.1	65.9±5.9	67.6±2.3	42.9±0.3	43.87 ± 0.3
I(0) (A.U., Guinier)	22.69±0.4	37.32±0.3	18.01±1.8	25.96±2.0	60.15 ± 0.72	61.32 ± 0.37
R_g (Å, Guinier)	29.34±0.6	29.65±0.3	50.1±5.6	57.81±3.5	42.04 ± 0.48	43.54 ± 0.19
D_{max} (Å)	101	98	236	233	144	148
Porod volume ($Å^3$, V_p)	$80.9*10^3$	$85.5*10^3$	$8.29x10^4$	$1.66x10^5$	$1.51x10^5$	$1.99x10^5$
M_w (kDa, Bayes)	59.5	62.4	47.4 (46.1 to 53.8)	85.7 (84.3 to 95.8)	109.1 (99.2 to 111.2)	130.9 (127.5 to 151.4)
M_r (kDa, Porod)	67.1	70.9	68.8	137.4	125.2	165.4
Sequence M_w (Da)	61793 (dimer)	61793 (dimer)	37816	37816	161401 (2 dimers, 1 dsDNA)	161401 (2 dimers, 1 dsDNA)
SASBDB ID						
Simple Scatter dataset code	-	-	XSXID7SG	XSHPCWEC	XSCZDQFI	XSKHCUYA
X^2 FOXS fit to model	-	-	-	-	-	-
DENSS Real Space Correlation	-	-	-	-	0.6928	0.7098
DENSS Reconstruction resolution	-	-	-	-	50.12 ± 7.20	54.16 ± 9.54
DENSS χ^2	-	-	-	-	0.00858	0.01316
DENSS R_g	-	-	-	-	42.12	43.25
Primary data reduction	RAW	RAW	RAW	RAW	RAW	RAW
Data processing	RAW	RAW	RAW	RAW	RAW	RAW
Rigid body modeling	-	-	-	-	DENSS	DENSS
Computation of model intensities	-	-	-	-	-	-

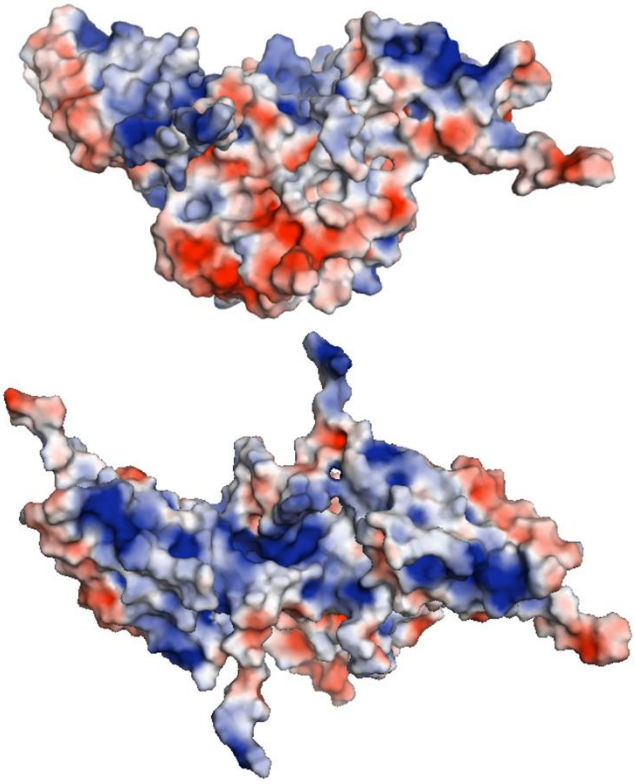


Figure 3.8. APBS Electrostatics mapped onto the SAXS-refined structural model of the RcoM dimer. Blue represents positive charges, red represents negative charges.

flexible linking region would impart the required flexibility to situate the DNA binding domains as required to bind the promoter.

The RcoM regulatory mechanism does not involve a change in oligomeric state, as is common for other PAS-type proteins. Size exclusion chromatographic data suggest that the protein remains primarily dimeric in both the Fe(III) and Fe(II)-CO states (Figure 3.3A). Fe(III) and Fe(II)-CO RcoM both display comparable molecular weights of the dimeric major species, and CO-binding does not favor formation of the larger oligomer. The triplet direct repeat promoter and the implication that two RcoM dimers are involved in the DNA interaction had led us to hypothesize that CO binding leads to increased formation of tetramer, enabling four LytTR domains to saturate the promoter. Ruling out an effector-dependent change in oligomeric state as the allosteric mechanism, RcoM dimer must undergo changes in conformation and/or protein dynamics that allow each dimer to bind the promoter separately. Large scale changes in conformation between Fe(III) an Fe(II)-CO dimer would also be observable by SAXS, and we do not observe significant differences between the two species.

Fitting SEC-SAXS data of crosslinked RcoM and DNA to our refined model of dimeric RcoM leads us to conclude that two RcoM dimers are involved in formation of the complexed species. FoXS, a web server for computing SAXS profiles based on a molecular structure and demonstrated to be useful in the structural characterization of flexible proteins and the assembly of multi-protein complexes, indicated that two RcoM dimers and one piece of DNA displayed the best fit to the experimental data compared to other stoichiometries tested (Figure 3.7A). The best FoXS model shows one dimer interacting with two adjacent DNA binding sites and a second dimer interacting with the middle site with its other LytTR domain left non-interacting (Figure 3.7B). The third DNA binding site is, puzzlingly, left unbound to protein. It is easy to speculate that the

free LytTR domain could bind to the third site, possibly in concert with a bending of the DNA. The pDDF fitting of the SAXS data does indicate that the Dmax for the complex is smaller than the free DNA, perhaps indicative of changes to DNA conformation upon binding to RcoM, however we cannot support this scenario further with the present data. In the case of LytTR protein, ComE, a SAXS model was developed of the ComE:ComD:promoter system by imposing an 85° bend in the DNA.[46] We attempted to replicate this technique, imposing an 85° bend on the DNA used to model the complex in HADDOCK, but this technique does not improve the fit to experimental data (data not shown).

Despite the SAXS model suggesting that the third DNA binding site is not involved in the interaction with RcoM, biochemical characterization of DNA binding supports that all three sites of the native promoter form protein interactions. The RcoM triplet repeat promoter is unique among LytTR proteins and especially surprising for a dimeric protein. We created altered DNA sequences to verify that all three sites were critical in the RcoM:DNA interaction by sequentially dropping out the middle and flanking sites. Eliminating either the middle (e) or flanking (f) site resulted in a significant reduction of CO-dependent RcoM binding to DNA (Figure 3.4B). Thus, a SAXS model where each DNA binding site is not bound to protein is inconsistent with our biochemical characterization. Unfortunately, our ability to analyze SAXS data on protein:DNA complexes are currently limited by the ability of data analysis programs to handle flexible complexes and DNA. While the lifetime of the complex *in vivo* is not clear, our observations might suggest a rapid on/off situation or they may be consistent with needing other cellular components to stabilize the interaction. Our model is the first structural description of the RcoM:DNA complex and presents new questions for how the system may be regulated and stabilized further during the initiation of transcription.

3.5 Supplementary Information

High Throughput Small-Angle X-ray Scattering (HT-SAXS)

High-Throughput Small Angle X-ray Scattering (HT-SAXS) data were collected at the SIBLYS Beamline 12.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA.[35,36,47] Prior to data collection, the plate was spun at 3700 rev min^{-1} for 10 min. Samples were held at 10 ℃ during collection. The exposure was 10 s, with frames collected every 0.3 s for a total of 33 frames per sample. The incident light wavelength was 1.03 Å at a sample to detector distance of 1.5 m. This setup results in scattering vectors, q, ranging from 0.013 $Å^{-1}$ to 0.5 $Å^{-1}$, where the scattering vector is defined as $q = \frac{4\pi sin\theta}{\lambda}$ and 2θ is the measured scattering angle. Matching buffer exposures were collected before and after samples to ensure there was no difference in the scattering owing to contamination of the sample cell.

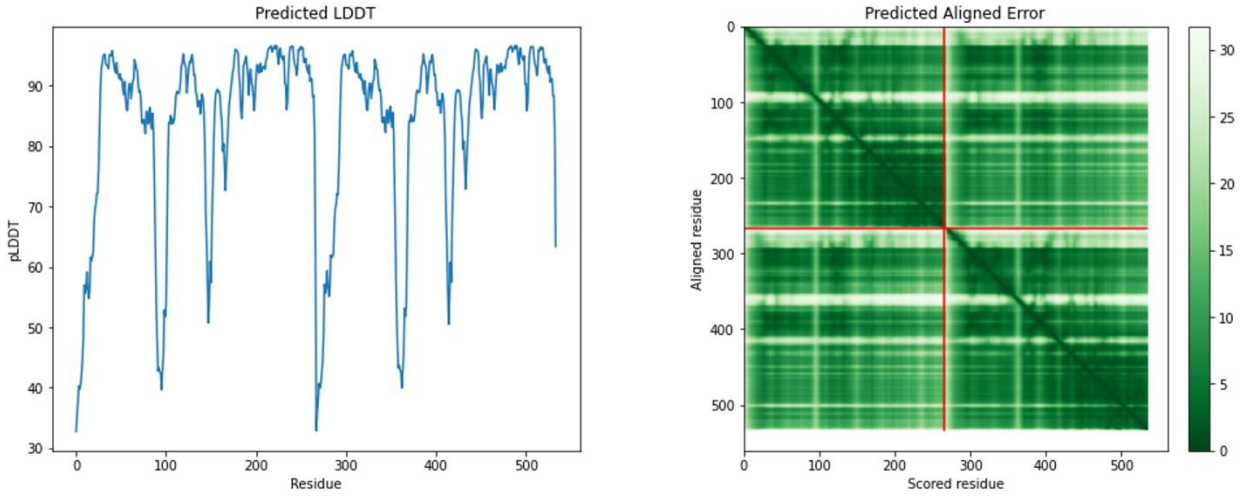


Figure 3.S1. Left: Plot with predicted local distance difference test (pLDDT) values by residue in *Px*RcoM-1 sequence. Right: Predicted Alignment Error (PAE) plot, which displays the absolute error (Å) of the relative position between residues, with dark green indicating low error and white showing increasing error.

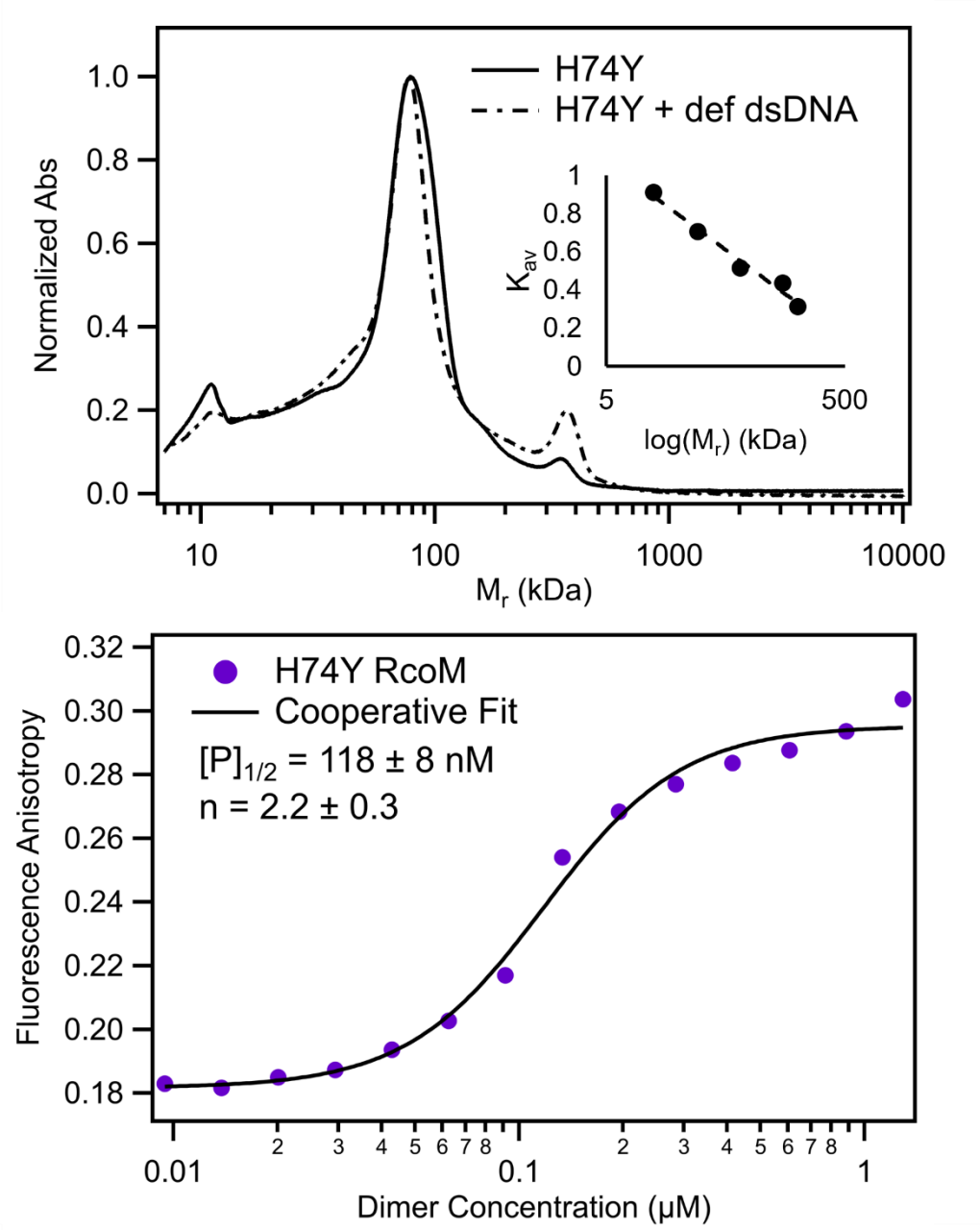


Figure 3.S2. Top: Size-exclusion chromatograms of 50uM H74Y and 50uM H74Y plus equimolar "def" ds oligonucleotide. Inset is the standard curve plot of K_{av} versus log(M_r) made from molecular weight standards used to convert elution time to molecular weight (M_r). Standards data were fit to a logarithmic function: y = -0.204Ln(x) + 1.4136, R^2 = 0.9794. This calibration curve was used to convert elution time to M_r for the entire H74Y dataset. Bottom: Fluorescence polarization data for H74Y RcoM with native "def" dsDNA. The solid line shows the best fit to the Hill Equation (χ^2= 2.46809 x 10^{-4}).

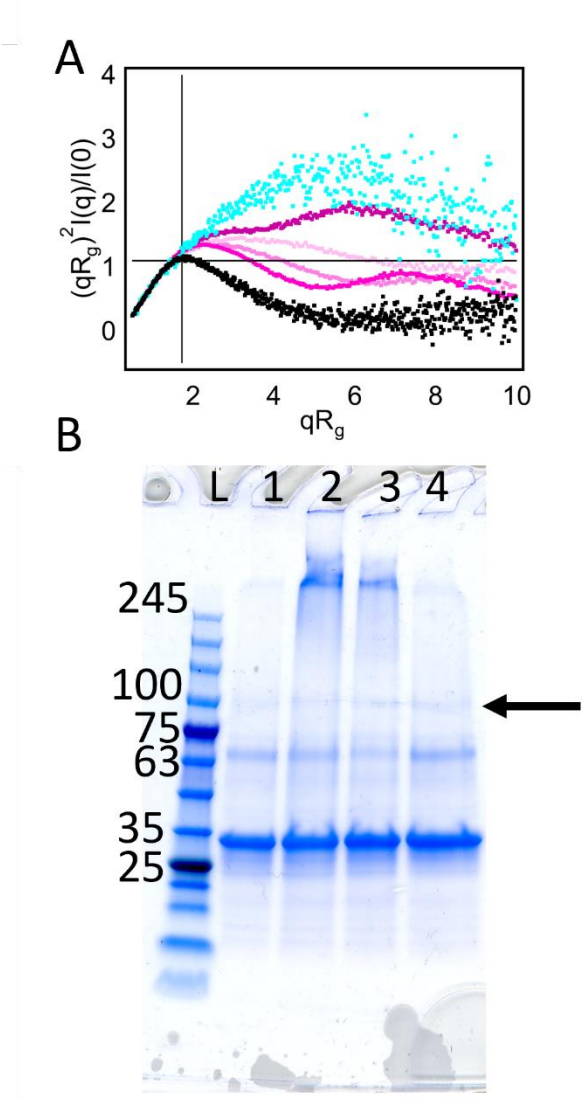


Figure 3.S3. A. HT-SAXS data demonstrating the formation of RcoM:DNA complex at increasing RcoM:DNA ratio. Kratky plot of *Px*RcoM-1 dimer from SEC-SAXS (black), 25 uM "def" native oligonucleotide (blue), and mixtures of 96 uM Fe(II)-CO RcoM with 12.5 uM DNA (lightest pink), 25 uM, 50 uM, and 100 uM (darkest pink). B. SDS-PAGE of crosslinked RcoM:DNA with 0.1% glutaraldehyde with the native "def" ds oligonucleotide for 10 and 20 min (Lanes 1 and 2), and with the enhanced affinity "de+f" ds oligonucleotide for 10 and 20 min (Lanes 3 and 4), against GoldBio BlueStain2 protein ladder (Lane L). An arrow points to a new species generated slightly above the 100 kDa MW marker.

3.6 References

(1) Dent, M. R.; Weaver, B. R.; Roberts, M. G.; Burstyn, J. N. Carbon Monoxide-Sensing Transcription Factors: Regulators of Microbial Carbon Monoxide Oxidation Pathway Gene Expression. *J. Bacteriol.* **2023**, *205* (5), e0033222. https://doi.org/10.1128/JB.00332-22/SUPPL_FILE/JB.00332-22-S0001.PDF.

(2) 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* (21), 5803–5816. https://doi.org/10.1128/jb.01024-12.

(3) Dent, M. R.; Roberts, M. G.; Bowman, H. E.; Weaver, B. R.; McCaslin, D. R.; Burstyn, J. N. Quaternary Structure and Deoxyribonucleic Acid-Binding Properties of the Heme-Dependent, CO-Sensing Transcriptional Regulator PxRcoM. *Biochemistry* **2022**, *61* (8), 678–688. https://doi.org/10.1021/ACS.BIOCHEM.2C00086/ASSET/IMAGES/LARGE/BI2C00086_0009.JPEG.

(4) Kerby, R. L.; Youn, H.; Roberts, G. P. RcoM: A New Single-Component Transcriptional Regulator of CO Metabolism in Bacteria. *J. Bacteriol.* **2008**, *190* (9), 3336–3343. https://doi.org/10.1128/JB.00033-08.

(5) Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17*, 1282–1294. https://doi.org/10.1016/j.str.2009.08.011.

(6) Henry, J. T.; Crosson, S. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu. Rev. Microbiol.* **2011**, *65* (1), 261–286. https://doi.org/10.1146/annurev-micro-121809-151631.

(7) Ayers, R. A.; Moffat, K. Changes in Quaternary Structure in the Signaling Mechanisms of PAS Sensor Domains. *Biochemistry* **2009**, *47*, 12078–12086. https://doi.org/10.1016/j.bpj.2008.12.2692.

(8) Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **2009**, *17* (10), 1282–1294. https://doi.org/https://doi.org/10.1016/j.str.2009.08.011.

(9) 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 Studies. *J. Mol. Biol.* **2000**, *301* (2), 415–431. https://doi.org/10.1006/jmbi.2000.3954.

(10) Möglich, A.; Moffat, K. Structural Basis for Light-Dependent Signaling in the Dimeric LOV Domain of the Photosensor YtvA. *J. Mol. Biol.* **2007**, *373* (1), 112–126. https://doi.org/10.1016/J.JMB.2007.07.039.

(11) Key, J.; Hefti, M.; Purcell, E. B.; Moffat, K. Structure of the Redox Sensor Domain of Azotobacter Vinelandii NifL at Atomic Resolution: Signaling, Dimerization, and Mechanism. *Biochemistry* **2007**, *46* (12), 3614–3623. https://doi.org/10.1021/BI0620407/ASSET/IMAGES/MEDIUM/BI0620407N00001.GIF.

(12) Smith, A. T.; Majtan, T.; Freeman, K. M.; Su, Y.; Kraus, J. P.; Burstyn, J. N. Cobalt Cystathionine β-Synthase: A Cobalt-Substituted Heme Protein with a Unique Thiolate Ligation Motif. *Inorg. Chem.* **2011**, *50* (10), 4417–4427. https://doi.org/10.1021/ic102586b.

(13) Yang, Y.; Inouye, M. Intermolecular Complementation between Two Defective Mutant Signal- Transducing Receptors of Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (24), 11057–11061. https://doi.org/10.1073/PNAS.88.24.11057.

(14) 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* (34), 9016–9028. https://doi.org/10.1021/bi800486x.

(15) 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* (7), 1071–1082. https://doi.org/10.1007/s00775-012-0920-1.

(16) Bowman, H. E.; Dent, M. R.; Burstyn, J. N. Met104is the CO-Replaceable Ligand at Fe(II) Heme in the CO-Sensing Transcription Factor BxRcoM-1. *J. Biol. Inorg. Chem.* **2016**, *21* (4), 559–569. https://doi.org/10.1007/s00775-016-1368-5.

(17) Salman, M.; Villamil Franco, C.; Ramodiharilafy, R.; Liebl, U.; Vos, M. H. Interaction of the Full-Length Heme-Based CO Sensor Protein RcoM-2 with Ligands. *Biochemistry* **2019**, *58* (39), 4028–4034. https://doi.org/10.1021/acs.biochem.9b00623.

(18) 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* (41), 10686–10694. https://doi.org/10.1021/acs.jpcb.6b08160.

(19) Nikolskaya, A. N.; Galperin, M. Y. A Novel Type of Conserved DNA-Binding Domain in the Transcriptional Regulators of the AlgR/AgrA/LytR Family. **2002**, *30* (11), 2453–2459.

https://doi.org/10.1093/nar/30.11.2453.

(20) 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* (5), 727–735. https://doi.org/https://doi.org/10.1016/j.str.2008.02.011.

(21) Boudes, M.; Sanchez, D.; Graille, M.; Van Tilbeurgh, H.; Durand, D.; Quevillon-Cheruel, S. Structural Insights into the Dimerization of the Response Regulator ComE from Streptococcus Pneumoniae. *Nucleic Acids Res.* **2014**, *42* (8), 5302–5313. https://doi.org/10.1093/nar/gku110.

(22) 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* (21), 6020–6031. https://doi.org/10.1128/jb.05436-11.

(23) 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* (4), 817–827. https://doi.org/10.1046/j.1365-2958.1999.01528.x.

(24) 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* (1), 57–66. https://doi.org/10.1128/jb.182.1.57-66.2000.

(25) 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* (17),

4696–4703. https://doi.org/10.1128/jb.182.17.4696-4703.2000.

(26) 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* (15), 4472–4483. https://doi.org/10.1128/jb.178.15.4472-4483.1996.

(27) 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* (1), 198–206. https://doi.org/10.1007/s004380000397.

(28) Behr, S.; Heermann, R.; Jung, K. Insights into the DNA-Binding Mechanism of a LytTR-Type Transcription Regulator. *Biosci. Rep.* **2016**, *36*, e00326. https://doi.org/10.1042/BSR20160069.

(29) 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* (21), 5803–5816. https://doi.org/10.1128/JB.01024-12.

(30) Berry, E. A.; Trumpower, B. L. Simultaneous Determination of Hemes a, b, and c from Pyridine Hemochrome Spectra. *Anal. Biochem.* **1987**, *161*, 1–15.

(31) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction

with AlphaFold. *Nat. 2021 5967873* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(32) Steinegger, M.; Söding, J. Clustering Huge Protein Sequence Sets in Linear Time. *Nat. Commun.* **2018**, *9*, 2542. https://doi.org/10.1038/s41467-018-04964-5.

(33) Steinegger, M.; Mirdita, M.; Söding, J. Protein-Level Assembly Increases Protein Sequence Recovery from Metagenomic Samples Manyfold. *Nat. Methods* **2019**, *16*, 603–606. https://doi.org/10.1038/s41592-019-0437-4.

(34) Rosenberg, D. J.; Hura, G. L.; Hammel, M. Size Exclusion Chromatography Coupled Small Angle X-Ray Scattering with Tandem Multiangle Light Scattering at the SIBYLS Beamline. *Methods Enzymol.* **2022**, *677*, 191–219. https://doi.org/10.1016/BS.MIE.2022.08.031.

(35) Classen, S.; Hura, G. L.; Holton, J. M.; Rambo, R. P.; Rodic, I.; McGuire, P. J.; Dyer, K.; Hammel, M.; Meigs, G.; Frankel, K. A.; Tainer, J. A. Implementation and Performance of SIBYLS: A Dual Endstation Small-Angle X-Ray Scattering and Macromolecular Crystallography Beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **2013**, *46* (1), 1–13. https://doi.org/10.1107/S0021889812048698/HE5572SUP1.TXT.

(36) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q. Rev. Biophys.* **2007**, *40* (3), 191–285. https://doi.org/10.1017/S0033583507004635.

(37) Hopkins, J. B.; Gillilan, R. E.; Skou, S. BioXTAS RAW: Improvements to a Free Open-Source Program for Small-Angle X-Ray Scattering Data Reduction and Analysis. *J. Appl.*

Crystallogr. **2017**, *50* (5), 1545–1553. https://doi.org/10.1107/S1600576717011438.

(38) Grant, T. D. Ab Initio Electron Density Determination Directly from Solution scattering Data. *Nat. Methods* **2018**, *15* (3), 191–195. https://doi.org/10.1038/nMeth.4581.

(39) Mirdita, M.; Ovchinnikov, S.; Steinegger, M.; Schütze, K.; Moriwaki, Y.; Heo, L. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

(40) Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. M. J. J. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8* (July), 1–7. https://doi.org/10.3389/fmolb.2021.729513.

(41) 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.; Sternberg, M. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2015**, *428*, 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

(42) Arnott, S.; Campbell-Smith, P. J.; Chandrasekaran, R. No Title. In *Handbook of Biochemistry and Molecular Biology, 3rd ed. Nucleic Acids--Volume II*; Fasman, G. P., Ed.; CRC Press: Cleveland, 1976; pp 411–422.

(43) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. FoXS, FoXSDock and MultiFoXS: Single-State and Multi-State Structural Modeling of Proteins and Their Complexes Based on SAXS Profiles. *Nucleic Acids Res.* **2016**, *44*. https://doi.org/10.1093/nar/gkw389.

(44) Hajizadeh, N. R.; Franke, D.; Jeffries, C. M.; Svergun, D. I. Consensus Bayesian Assessment of Protein Molecular Mass from Solution X-Ray Scattering Data OPEN. *Nat. Sci. Reports* **2018**, *8*, 7204. https://doi.org/10.1038/s41598-018-25355-2.

(45) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105*, 962–974. https://doi.org/10.1016/j.bpj.2013.07.020.

(46) Sanchez, D.; Boudes, M.; Van Tilbeurgh, H.; Durand, D.; Quevillon-Cheruel, S. Modeling the ComD/ComE/Comcde Interaction Network Using Small Angle X-Ray Scattering. *FEBS J.* **2015**, *282* (8), 1538–1553. https://doi.org/10.1111/FEBS.13240.

Chapter 4

Probing conformational dynamics of DNA binding by CO-sensing transcription factor, CooA, with site-directed spin label electron paramagnetic resonance spectroscopy

A version of this chapter will be published as:

Roberts, M. G.; Dent, M. R.; Burstyn, J. N. Probing conformational dynamics of DNA binding by CO-sensing transcription factor, CooA, with site-directed spin label electron paramagnetic resonance spectroscopy.

MGR and MRD carried out experiments and analyzed the data. JNB advised project progress. MGR wrote the original manuscript with input from MRD and JNB.

4.1 Introduction

Allostery is a widely used mechanism for the regulation of protein function in response to environmental stimuli, where binding of an effector molecule changes the functional activity (e.g. changing binding affinity or catalytic efficiency) of the protein at a distant site.[1,2] As such, an active research area for allosterically regulated proteins involves understanding how effector-binding at a distant site leads to functional changes at the active site; historically these changes have been viewed as largely structural in origin.[3,4] A modern view of allosteric signal propagation involves a combination of structural rearrangement as well as protein sidechain and backbone dynamics, though difficulty in studying changes in dynamics meant this contribution was often overlooked until experimental techniques allowed for monitoring protein dynamic motions.[5–9] It was long speculated that protein dynamics could control allostery in the absence of structural changes, and recently this type of regulation has been identified and extensively described in CRP (cAMP receptor protein, also known as CAP, Catabolite Activator Protein).[10–14] However, the existence of allosterically induced changed in protein dynamics in other members of the CRP/FNR structural superfamily, including CooA (CO Oxidation Activator) remain unexplored.

In WT CRP, a structural mechanism exists that explains DNA binding after the effector cAMP binds, but protein dynamics play a key role in the negative cooperativity of cAMP binding. Crystal structures of CRP demonstrate that two molecules of cAMP binding induce a repositioning of the DNA binding domains that allow for interaction with DNA.[15–17] Modulation of protein dynamics is observed in the sequential, negatively cooperative binding of two cAMP molecules, with binding of the first cAMP enhancing slow timescale (μs-ms) protein motion and the binding of the second molecule completely quenching fast (ps-ns) motions.[18] Studies of CRP variants that constitutively bind DNA in the absence of cAMP or the structural changes elicited by effector

binding, have established that changes in fast (ps-ns) dynamics alone can mediate allosteric activity.[12,13] These experiments establish that changes in fast internal motions result in a large difference in conformational entropy that can describe allosteric regulation without enthalpic contributions from structural rearrangement.[19–22] Studies that include slow timescale protein motions, indicative of exchange between conformation states, indicated that allostery can be modulated by a redistribution of conformational substates, without necessarily altering the mean structure, underlining the role of μs-ms protein motions in allosteric mechanisms.[14]

For CooA, a carbon monoxide (CO)-sensing transcription factor in the CRP/FNR superfamily, a structural approach to allostery has established that CO binding at heme leads to conformational rearrangement that allows for interaction with DNA (Figure 4.1). Similarly to CRP, CooA is a homodimer in solution consisting of an effector binding domain, that bind heme and CO, and a helix-turn-helix type DNA binding domain.[23] CooA heme is a dual redox and CO sensor.[24] Reduction of the Fe(III) heme leads to His^{77} replacing the Cys^{75} thiolate ligand, then the incoming CO ligand replaces Pro^{2}.[25–30] The transition to Fe(II)-CO heme corresponds with a large change in the hinge region, flipping open each DNA binding domain exposing the DNA-binding F-helices, like a "swinging hinge".[31,32] High-resolution structures of *Rhodospirilum rubrum* (*Rr*) Fe(II) CooA and a constitutively active variant of *Carboxydothermus hydrogenoformans* (*Ch*) Fe(II)-CO CooA led to an understanding of the reversible transitions between discrete conformations that allow CooA to bind its promoter, pCooF.[23,31] However, asymmetry in both crystal structures and the lack of a structures in the Fe(III) "locked-off" state or DNA bound state limits understanding of the entire mechanism from crystal structures alone. Solution small-angle X-ray data comparing the Fe(II) and Fe(II)-CO states allowed for development of the structural model of CooA activation that suggests the swinging open open of the DNA binding domains

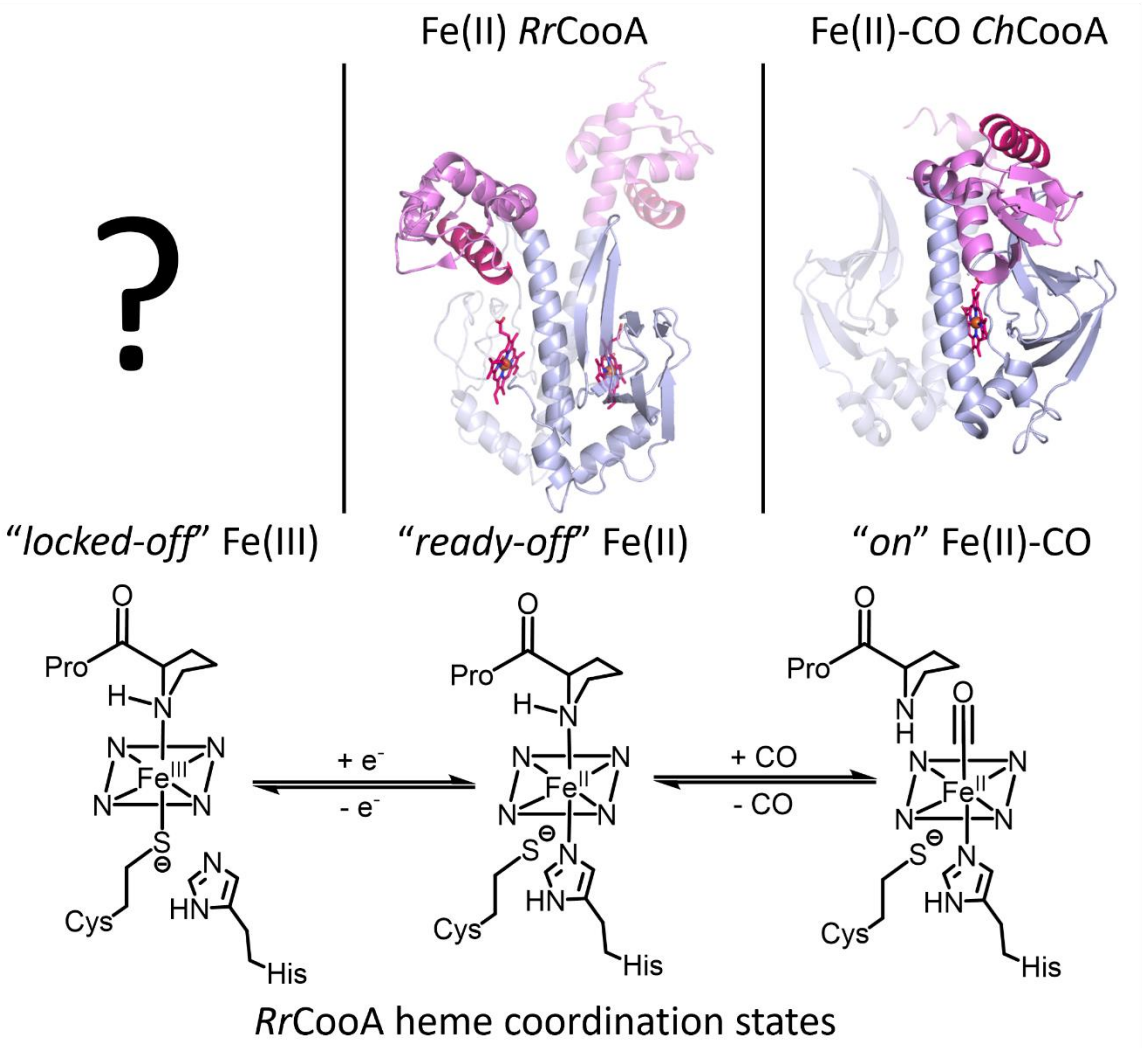


Figure 4.1. Three functional states of CooA, showing the redox-mediated ligand switching mechanism of heme and the corresponding structural data. For each crystal structure, the heme binding domain is shown in light blue with the heme shown as magenta sticks, the DNA binding domain is shown in purple with the recognition helix highlighted in magenta. One monomer is shown as transparent. Left: No structure of Fe(III) CooA exists to date. Middle: Fe(II) *Rr*CooA (PDB 1FT9), Right: Fe(II)-CO *Ch*CooA N127L/S128L (PBD 2HKX).

occurs via changes in the hinge region.[33]

The importance of fast and slow timescale protein motions to the activation mechanism of CRP, archetypical protein of the CRP/FNR superfamily, gives reason to believe that dynamics could play an important role in CooA allosteric regulation. Previously, Hines et al. used site-directed spin label electron paramagnetic resonance spectroscopy (SDSL-EPR) to demonstrate that CooA exhibits conformational heterogeneity in the effector-free Fe(III) "locked off" state.[34] SDSL-EPR is a method used to study local site dynamics of conformational changes of biological macromolecules.[35–37] Nitroxide radical labels are attached to the protein at specific introduced cysteine residues and continuous wave EPR spectra are collected (Figure 4.2). An EPR spectrum is highly sensitive to nitroxide mobility; reduced mobility of the label results in spectrum broadening. The rotational motion of the label itself, motion of the protein backbone, and tumbling motion of the entire protein in solution all contribute to the spectrum, and careful choice of experimental conditions can distinguish between types of motions.[37] The addition of 25% w/w Ficoll-70 eliminates the contribution of protein rotary diffusion from the spectrum, and the contribution of label rotational motion and protein backbone motions can be distinguished with osmolyte perturbation studies using sucrose.[38]

This work aims to build on our earlier experiments with Fe(III) CooA, showing that the populations of more and less mobile conformations change throughout the allosteric process, broadening our understanding of CooA regulation to include protein dynamics. A series of CooA variants reintroduce cysteines into key regions of the protein, the heme binding domain, the 4/5-loop, the hinge region, and the DNA binding domain adjacent to the F-helix, to allow for site-selective probing of dynamics in each region. These variants were evaluated for their ability to

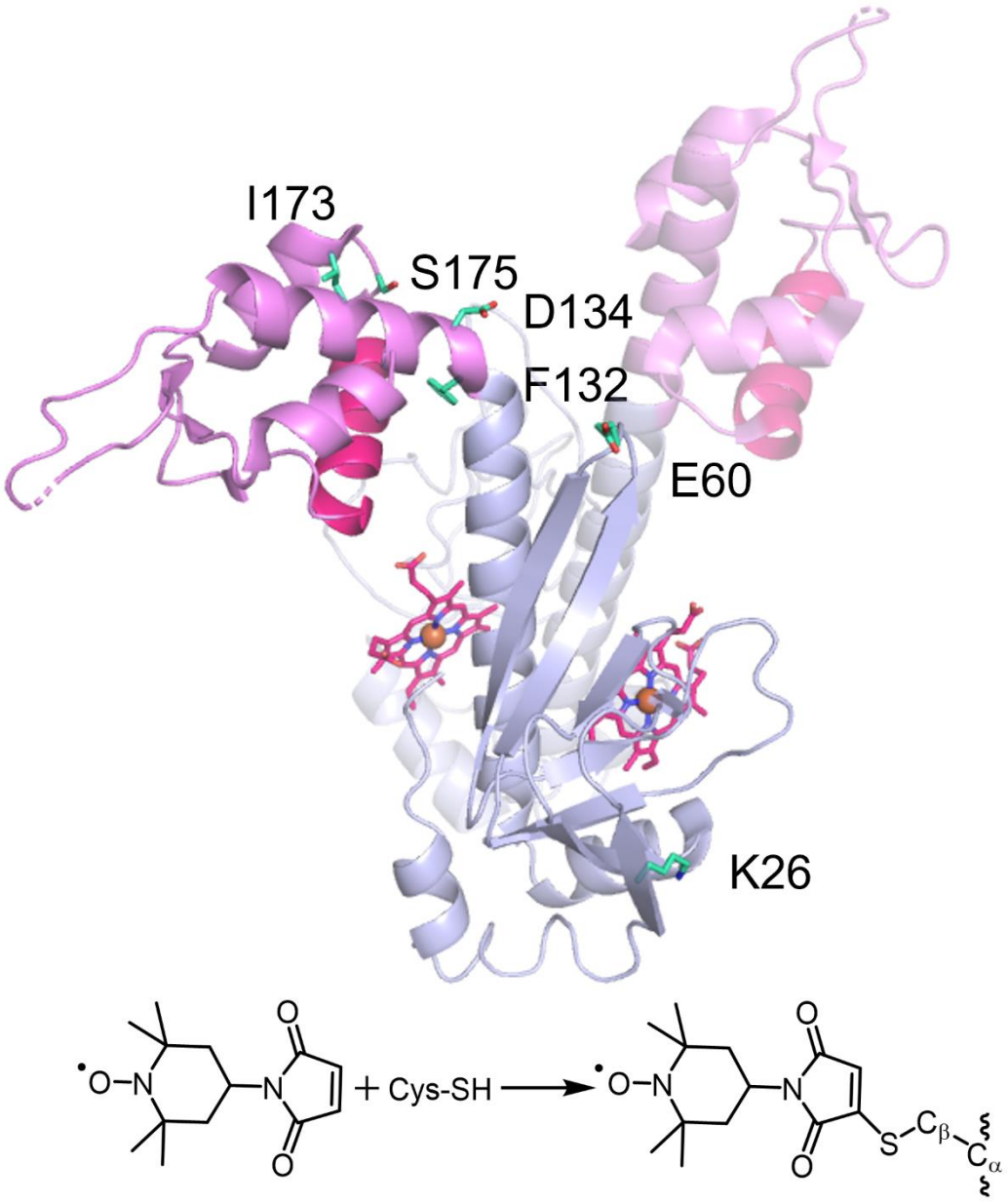


Figure 4.2. TOP: Crystal structure of Fe(II) *Rr*CooA (PDB 1FT9) with the location of six cysteine substitution sites highlighted, K26, E60, F132, D134, I173, and S175. Protein is shown with the heme binding domain in light blue, heme in magenta, the DNA binding domain in purple, the recognition helix in magenta, and the residues substituted to facilitate labeling in cyan. One monomer is shown as transparent. BOTTOM: The reaction of MAL-6 with introduced cysteine to form site-specific redox-stable linkages with protein.

retain DNA binding ability before and after labeling with MAL-6 to ensure variants could accurately represent the behavior of that region when bound to DNA. Then room temperature CW (continuous wave) EPR experiments were performed that demonstrate global changes in protein dynamics upon effector binding. We observe that once DNA is bound, complementary shifts in the hinge region and DNA binding domain suggest that the hinge becomes more flexible upon DNA binding to compensate for restricted motion in the DNA binding domain. This study provides insight into the behavior of CooA in solution before and after CO-binding and upon interaction with DNA, with detail of the slow timescale protein motions unavailable in crystal structures.

4.2 Materials and Methods

Materials

MAL-6 (N-(2,-2,6,6-tetramethylpiperidin-4-yl-1-oxyl)maleimide) was purchased from Sigma Aldrich and used without further purification. CO(g) (≥99.0% purity) and Ar(g) (100% purity) were obtained from Airgas and used as received. All other chemicals were obtained from Sigma Aldrich or Fisher Scientific.

Vector Construction

K26C, E60C, F132C, D134C, I173C, and S175C ΔCys_4 CooA) in *R. rubrum cooA* were created in a pEXT-20-based vector containing the gene for ΔCys_4CooA (C35A C80A C105Y C123I) as described in previous work.[34]

Protein Expression and Purification

Expression of variant CooA proteins followed the protocol previously reported.[34] All cells were grown in Terrific Broth media supplemented with 20 mg/L ferric citrate. A dense 2 mL culture was added to 500 mL media and incubated at 37°C. At OD_{550} = 0.8-1.0, expression was induced

with 1 mM IPTG, followed by growth for 20 h at 30°C. Following expression, cells were harvested by centrifugation, incubated with 2 mg/mL lysosyme in "Binding Buffer" (50 mM MOPS (3-morpholinopro- pane-1-sulfonic acid), 10 mM imidazole, 500 mM KCl pH 7.4), and lysed by sonication. The lysate was cleared by centrifugation and added to a Ni-NTA column (GE Life Sciences, 5 mL column volume) pre-equilibrated with binding buffer. The column was washed with binding buffer until A_{280} reached, followed by additional washing with three column volume rinses with 15:85, 25:75, and 35:65 mixtures of elution buffer (50 mM MOPS, 220 mM imidazole, 500 mM NaCl pH 7.4) and binding buffer. Bound CooA was eluted with four column volumes of elution buffer. Eluted CooA was incubated with 45% saturated ammonium sulfate on ice for 30 min. Precipitated CooA was redissolved in 100 mM MOPS, 500 mM NaCl pH 7.4 and flash frozen on LN_2 prior to storage. Total protein concentration was assessed via Pierce660 assay (Pierce-Thermo Fisher) using bovine serum albumin as a calibration standard, and heme concentration was measured with pyridine hemochromogen assay.[39]

Fluorescence Anisotropy

For assessment of *in vitro* DNA binding of MAL-6 labeled CooA variants, 5 nmol protein was incubated with 10-fold molar excess MAL-6 at 4°C overnight. The next day, protein was concentrated and underwent three rounds of spin concentration and dilution, to a final concentration of 1000 nM dimer.

In vitro DNA binding affinity of CooA variants was measured as previous described.[34] Binding data was fit when appropriate to the Hill Equation, holding the Hill coefficient at 1 for a simple noncooperative interaction, to calculate an apparent affinity, $[P]_{1/2}$.

Single-point assays were employed when assessing protein activity. The fluorescence anisotropy of Fe(III) 1000 nM dimer CooA, either unlabeled or MAL-6 labeled, with 6.4 μM pCooF was measured. Fe(II)-CO protein was achieved by the anaerobic addition of 10 μL of a 100 mM sodium dithionite solution and a 100 μL bolus of CO(g). The samples were gently vortexed and allowed to sit at room temperature for 30 min to ensure complete conversion to the Fe(II)-CO state and equilibration of the DNA bound form, before measuring the fluorescence anisotropy again. Each point represents an average of three technical repeats. Then a $\Delta_{anisotropy}$ was calculated by subtracting the initial anisotropy in the Fe(III) state from the final anisotropy in the Fe(II)-CO state.

Site-Directed Spin Labeling of CooA Variants

K26C, E60C, F132C D134C, I173C and S175C Δ_4CysCooA proteins were labeled by incubating Fe(III) protein with 10-fold molar excess of MAL-6 label. MAL-6 (10 mg/mL in acetonitrile) was added to 50-75 nmol CooA in 2000 μL labeling buffer (100 mM MOPS, 500 mM NaCl, pH 7.4) and allowed to react at 4°C overnight. Labeled samples were dialyzed against labeling buffer in three rounds over 24 hours to remove excess free label. The resulting labeled CooA proteins were buffer exchanged into Ar(g)-sparged buffer (final buffer contents in figure legends) using a PD-10 desalting column, then concentrated using a 30,000 MWCO spin concentrator (Amicon Ultra).

For Fe(III) samples, 100-200 μM protein was loaded into capillary tubes (Drummond Scientific Company, 50 μL) for EPR analysis.

For Fe(II)-CO and Fe(II)-CO+DNA samples, 50 uL 100-200 uM protein (mixed with equimolar dsDNA containing pCooF, as appropriate) was transferred to a PCR tube and the headspace was purged with Ar(g). Reduction and CO binding as achieved by purging headspace with CO(g)

adding three-times molar concentration DTH solution and gently mixing. Prior to loading into capillary tubes, the samples were uncapped, gently blown into, and allowed to sit for 30 s to reoxidize the MAL-6 labels.

EPR data were acquired on a Bruker Elexsys 500 system at 10.0 mW power, 200 G scan width, 2 G modulation amplitude, and 100 kHz modulation frequency. The number of scans ranged from 10-200 based on protein concentration.

4.3 Results

Five cysteine substitution variants on a Δ_4CysCooA (C35A, C80A, C105Y, C123I) background were previously created: K26C, E60C, F132C, D134C, and S175C Δ_4CysCooA. These variants sequentially add Cys back in key locations for site-specific MAL-6 labeling—the DNA binding domain (S175C), the hinge region (F132C and D134C), the 4/5 loop (E60C), and in the heme-binding domain (K26C) (Figure 4.2). From here, these variants will be referred to as K26C, E60C, F132C, D134C, and S175C, though all variants also bear the additional cysteine deletions of the Δ_4CysCooA background.

DNA Binding of MAL-6 Labeled Δ_4CysCooA Variants

Prior to observing protein dynamics of CooA variants, all variants were evaluated for their ability to bind DNA after undergoing labeling with MAL-6. All cysteine variants, with the exception of F132C, were previously shown to exhibit similar apparent affinities for DNA as WT CooA, but it was still unknown if appending a label to each location would affect DNA binding.[34] Fluorescence anisotropy-based DNA binding assays monitor the increase in anisotropy of Texas Red-X labeled double stranded oligonucleotide containing pCooF, as a function of protein concentration. As protein concentration increases, more protein binds pCooF and anisotropy increases, allowing for

binding parameters to be calculated from a binding curve. Single-point assays monitor the change in anisotropy upon activation of the protein at a single concentration at the top of the binding curve, allowing for a rapid measure of protein activity. Single-point assays were used to determine if labeling CooA variants with MAL-6 prevents interaction with pCooF (Table 4.1). K26C, E60C, and D134C were all shown to retain DNA binding affinity after labeling with MAL-6. However, S175C does not exhibit a CO-dependent change in anisotropy after MAL-6 labeling. The location of S175C, directly adjacent to the F-helix that recognizes pCooF, perhaps means that the bulky label sterically limits the specific interactions between the CooA F-helix and DNA.

A new cysteine variant was produced to label the DNA binding domain, moving the labeling site away from the F-helix, resulting in a variant that allows for labeling of the DNA binding domain and retains its activity. I173C was created with the goal of minimally moving the labeling site from S175 but allowing enough distance from the F-helix to make room for the bulky label. UV/visible absorbance spectroscopy shows that I173C retains the same ability to undergo its native ligand switching mechanism as WT CooA (Figure 4.3A). The band shapes and locations of Fe(III), Fe(II), and Fe(II)-CO I173C are virtually indistinguishable from WT, indicating that like the other cysteine substitution variants, the ligand switching and CO-binding functionalities are unaffected by the mutations. Then fluorescence anisotropy DNA binding assays were used to calculate binding parameters for I173C before and after MAL-6 labeling (Figure 4.3B). Apparent affinity, $[P]_{1/2}$ = 36.3 ± 11.1 nM for the unlabeled protein is in the same range as the other variants.[34] The MAL-6 labeled protein displays a reduced affinity ($[P]_{1/2}$ = 140.6 ± 24.5 nM) compared to unlabeled protein but will still bind DNA at the concentrations used for EPR experiments. A new Cys substitution variant, I173C, allows for MAL-6 labeling of the DNA binding domain and monitoring of the dynamics of this region after DNA binding with EPR spectroscopy.

Table 4.1. *In vitro* fluorescence anisotropy single-point DNA binding assays of 1000 nM Δ_4CysCooA variants, both before and after overnight labeling with MAL-6, $\Delta_{anisotropy}$ is the difference in the fluorescence anisotropy of the inactive Fe(III) state and after conversion to the active Fe(II)-CO state. The "DNA Binding" column indicates if the variant can bind DNA in the conditions required for EPR experiments.

Variant	$\Delta_{anisotropy}$	$\Delta_{anisotropy}$, after MAL-6 labeling	DNA Binding
K26C	0.046	0.045	+
E60C	0.041	0.041	+
F132C	-0.005	–	-
D134C	0.034	0.032	+
S175C	0.029	0.005	-
I173C	0.047	0.045	+

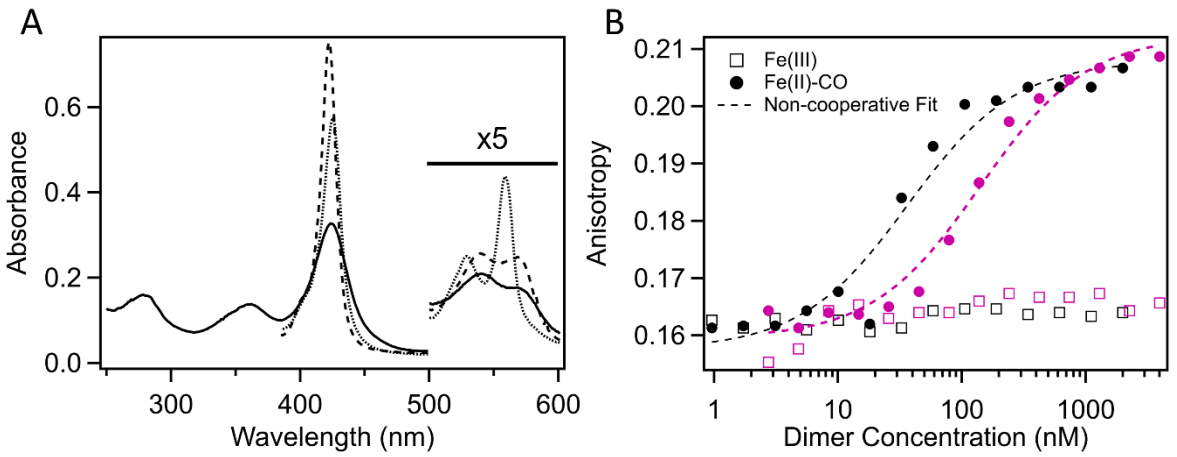


Figure 4.3. A. Electronic absorption spectra of I173C Δ_4CysCooA in the Fe(III) (solid line), Fe(II) (dotted line), and Fe(II)-CO (dashed line) heme coordination states. B. Fluorescence polarization data for Fe(III) (squares) and Fe(II)-CO (circles) I173C Δ_4CysCooA, with (magenta) and without (black) MAL-6 labeling, with ds pCooF. The dashed lines shows the best fit of active CooA to a non-cooperative binding model (Unlabeled: $[P]_{1/2}$ = 36.3 ± 11.1 nM, n = 1, χ^2= 2.29202 x 10^{-4}; Labeled: $[P]_{1/2}$ = 140.6 ± 24.5 nM, n = 1, χ^2= 7.98032 x 10^{-5}).

Comparison of EPR Spectra of MAL-6 Labeled Fe(III), Fe(II)-CO, and Fe(II)-CO + pCooF Δ4CysCooA Variants

EPR spectra of MAL-6 labeled CooA variants show a multicomponent signal that changes between the Fe(III) and Fe(II)-CO states. Previous work demonstrated that Fe(III) CooA variants exhibit a multicomponent EPR signal that consists of a narrower signal corresponding to a more mobile component, and a wider signal corresponding to a less mobile state.[34] EPR spectra of Fe(II)-CO CooA variants show that reduction and CO binding redistributes the populations of more and less mobile states in almost all variants, particularly for E60C, D134C, and I173C (Figure 4.4). These three variants, corresponding to label locations in the 4/5-loop, hinge region, and adjacent to the DNA binding F-helix, show significant changes in dynamics in response to effector binding. For E60C and D134C, the EPR spectra demonstrate a reduction in the more mobile conformation, while the I173C spectra show a shift to the more mobile state. In comparing multicomponent EPR spectra of Fe(III) "locked off" and Fe(II)-CO "on" CooA variants, effector binding causes a redistribution of more and less mobile conformations in the regions that are known to undergo CO-dependent structural changes, the DNA binding domain, hinge region, and 4/5-loop.

Key regions of CooA, the 4/5-loop, hinge region, and adjacent to the DNA-binding F-helix, undergo changes in conformational dynamics in response to DNA binding. Variants that were identified to bind DNA when labeled and demonstrated significant changes between the Fe(III) and Fe(II)-CO EPR spectra were chosen for analysis in the DNA bound state—E60C, D134C, and I173C. Buffer conditions were chosen for this set of samples based on the known requirements for CooA DNA binding, including the addition of divalent cation, Ca^{2+}, minimal buffer and salt, and increased pH. All CooA variants display changes in conformational heterogeneity between the Fe(III), Fe(II)-CO, and Fe(II)-CO bound to DNA samples in DNA binding conditions (Figure 4.5).

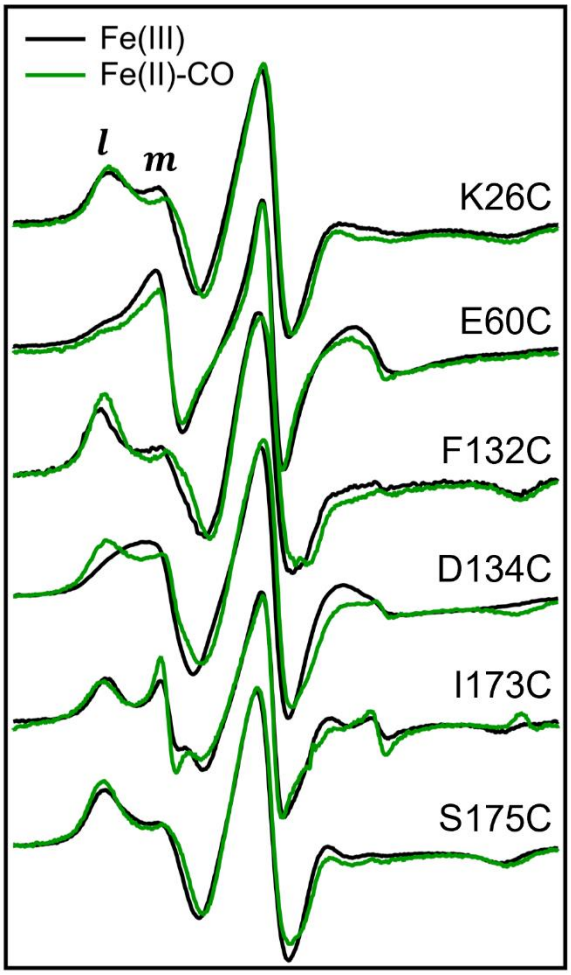


Figure 4.4. EPR spectra of MAL-6 labeled Fe(III) (black) and Fe(II)-CO (green) Δ_4CysCooA variants in 100 mM MOPS, 500 mM NaCl pH 7.4, 25% w/w Ficoll-70. The spectra are normalized to the central amplitude. More (*m*) and less (*l*) mobile components are labeled.

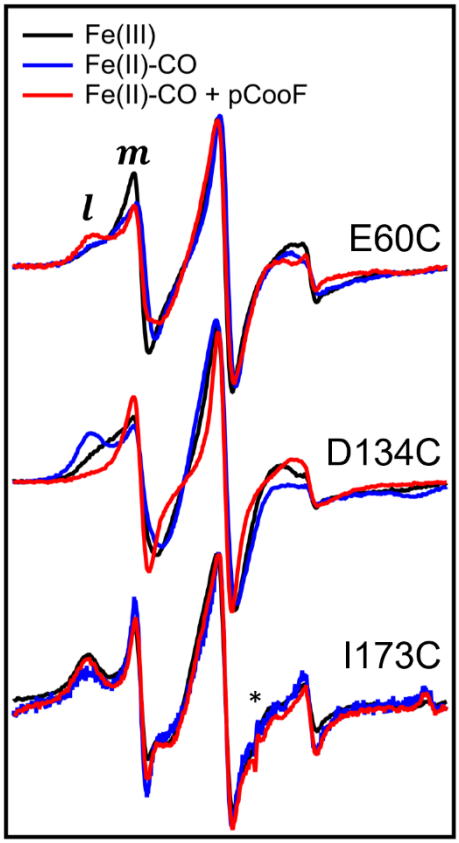


Figure 4.5. EPR spectra of MAL-6 labeled Fe(III) (black), Fe(II)-CO (blue), Fe(II)-CO + equimolar pCooF (red) Δ_4CysCooA variants in 40 mM MOPS, 50 mM NaCl, 6.4 mM $CaCl_2$, pH 8.0, and 25% w/w Ficoll-70. The spectra are normalized to the central amplitude. More (*m*) and less (*l*) mobile components are labeled. An asterisk marks a small cavity signal that appeared in the course of our data collection.

E60C and D134C show a shift to the less mobile state upon reduction and CO-binding, while I173C does the opposite, shifting to the more mobile state in response to effector binding. Upon the addition of equimolar pCooF to Fe(II)-CO "on" CooA, D134C displays a complete shift to the more mobile state, with the broader EPR signal disappearing. Conversely, addition of pCooF to Fe(II)-CO I173C reduces the population of the more mobile state. These results show that local protein dynamics, in locations far from the heme CO-binding site, change as a function of protein activity state in CooA.

4.4 Discussion

EPR spectra of MAL-6 labeled CooA variants in the Fe(III) "locked off", Fe(II)-CO "on", and Fe(II)-CO bound to DNA states show that protein dynamics change in key functional regions throughout the allosteric process. Previous work demonstrated that there is global conformational in the Fe(III) "locked off" state of CooA, but did not demonstrate how these protein dynamics may change in response to effector or DNA binding. This study examines the multicomponent EPR spectra of CooA variants labeled with redox-stable MAL-6 in key functional regions: the heme binding domain (K26C), 4/5-loop (E60C), hinge region (F132C and D134C), and DNA binding domain adjacent to the F-helix (I173C and S175C). We observe that the distribution of population of more and less mobile states shifts in response to changing functional state in all variants. Most notably, comparison of EPR spectra of D134C and I173C, representing the hinge region and DNA binding domain, allows for the development of a proposal for how protein dynamics contribute to CooA DNA binding. The data indicate that while μs-ms protein motions decrease in the DNA binding domain upon binding the cognate DNA sequence, there is a quenching of the less mobile conformation in the hinge region, shifting entirely to the more mobile conformation upon DNA binding.

We observe a global change in conformational dynamics in response to effector binding in CooA. Comparison of multicomponent EPR spectra of CooA variants in the Fe(III) "locked-off" state versus the Fe(II)-CO "on" state shows changes in protein dynamics across the protein, far from the effector binding site (Fig. 3). The variants that display the largest changes are E60C (representing the 4/5-loop), D134C (representing the hinge region), and I73C (representing the DNA binding domain adjacent to the hinge region). E60C, displays a reduction in the more mobile component of the EPR signal in response to effector binding. The 4/5-loop makes contacts with the DNA binding domain in the Fe(II)-CO form, and it has been suggested due to high sequence conservation among CooA homologs and its role in CRP as the contact with the sigma subunit of RNA polymerase, that this region is also involved in propagation of effector binding signal. D134C and I173C, found in the hinge region and the DNA binding domain, show CO-dependent changes in dynamics consistent with their role in the "swinging hinge" structural model of CooA allosteric activation.

To observe the impact of DNA binding on protein dynamics, we chose to compare EPR spectra of relevant regions to the DNA interaction, the 4/5-loop, the hinge region, and the DNA binding domain adjacent to the F-helix. The Fe(III) "locked-off", Fe(II)-CO "on", and Fe(II)-CO bound to pCooF all displayed distinct multicomponent EPR spectra in each region. E60C shows a reduction in the more mobile state in both the Fe(II)-CO that does not significantly change when bound to DNA, which is unsurprising for this location that is suggested to change structurally in response to effector binding, but the main role of the 4/5-loop in DNA binding is to contact RNA polymerase.[40] D134C also exhibits a CO-dependent reduction in the more mobile component, but upon DNA binding, the signal shifts entirely back to the more mobile component. In contrast to these two locations, the label in the DNA binding domain exhibits opposite changes in response

to CO and DNA, where effector binding leads to an increase in the more mobile state that is reduced after binding to DNA. This paired change in protein motions of the hinge region and DNA binding domain lead us to propose a role of protein dynamics in the CooA allosteric mechanism similar to that of CRP, where increased motion of the hinge region is paired with restricted motion of the DNA binding domain once bound to the promoter.

Our proposal broadens the understanding of the CooA allosteric mechanism, adding changes in conformation dynamics to the structural mechanism. In comparing the mechanism of CooA to that of other archetypical proteins of the same structural superfamily, CPR and FNR, CooA seems to more closely mirror the mechanism of CRP, where slow protein motions contribute to allostery. The regulatory mechanism of FNR is O_2- dependent dimerization. When O_2 is not present, the protein is a dimer and specifically binds to DNA. In the presence of O_2, degradation of [4Fe-4S] clusters lead to FNR monomerization and DNA disassociation.[41,42] CooA is not regulated by an effector-dependent change in oligomeric state; instead it appears to rely on a combination of effector-induced structural rearrangement of its hinge region and complimentary changes in flexibility of the DNA binding domain and hinge region for CO-dependent DNA binding, more similar to what is observed in CRP. Extensive study of the enthalpy-entropy compensation involved in DNA binding of CRP has found that both fast and slow timescale dynamics contribute to allostery.[14,43] While fast timescale motions result in changes to conformational entropy, slow protein motions are related to the population of conformational states; both can affect changes in protein function without changes in the ground-state structure. Further analysis with NMR-based techniques and thermodynamic analysis with calorimetric methods is needed to detect the fast internal motions that are related to conformational entropy and

directly measure the enthalpic and entropic components of the free energy, to be able to conclude the relative contributions to DNA binding of dynamics and structure change.

4.5 References

(1) Liu, J.; Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLOS Comput. Biol.* **2016**, *12* (6), e1004966. https://doi.org/10.1371/JOURNAL.PCBI.1004966.

(2) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585–609. https://doi.org/10.1146/annurev-biophys-050511-102319.

(3) Changeux, J.-P.; Edelstein, S. J. Allosteric Mechanisms of Signal Transduction. *Science (80-.).* **2005**, *308*, 1424–1428.

(4) Ribeiro, A. A. S. T.; Ortiz, V. A Chemical Perspective on Allostery. *Chem. Rev.* **2016**, *116* (11), 6488–6502. https://doi.org/10.1021/acs.chemrev.5b00543.

(5) Wand, A. J. Dynamic Activation of Protein Function: A View Emerging from NMR Spectroscopy. *Nat. Struct. Biol. 2001 811* **2001**, *8* (11), 926–931. https://doi.org/10.1038/nsb1101-926.

(6) Hawkins, R. J.; McLeish, T. C. B. Coarse-Grained Model of Entropic Allostery. *Phys. Rev. Lett.* **2004**, *93* (9), 7–10. https://doi.org/10.1103/PhysRevLett.93.098104.

(7) Kalodimos, C. G. Protein Function and Allostery: A Dynamic Relationship. *Ann. N. Y. Acad. Sci.* **2012**, *1260* (1), 81–86. https://doi.org/10.1111/J.1749-6632.2011.06319.X.

(8) Mittermaier, A.; Kay, L. E. New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science (80-.).* **2006**, *312*, 224. https://doi.org/10.1126/science.1124618.

(9) Hubbell, W. L.; Cafiso, D. S.; Altenbach, C. Identifying Conformational Changes with Site-Directed Spin Labeling. *Nat. Struct. Biol.* **2000**, *7* (9), 735–739.

(10) Cooper, A.; Dryden, D. T. F. Allostery without Conformational Change - A Plausible Model. *Eur. Biophys. J.* **1984**, *11* (2), 103–109. https://doi.org/10.1007/BF00276625.

(11) Tsai, C.-J.; Del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *J. Mol. Biol.* **2008**, *378*, 1–11. https://doi.org/10.1016/j.jmb.2008.02.034.

(12) Tzeng, S.-R.; Kalodimos, C. G. Dynamic Activation of an Allosteric Regulatory Protein. *Nature* **2009**, *462*, 368–374. https://doi.org/10.1038/nature08560.

(13) Tzeng, S.-R.; Kalodimos, C. G. Protein Activity Regulation by Conformational Entropy. *Nature* **2012**, *488*, 236–240. https://doi.org/10.1038/nature11271.

(14) Tzeng, S.-R.; Kalodimos, C. G. The Role of Slow and Fast Protein Motions in Allosteric Interactions. *Biophys. Rev.* **2015**, *7* (2), 251–255. https://doi.org/10.1007/s12551-015-0172-8.

(15) 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.* **2009**, *106* (17), 6927–6932. https://doi.org/10.1073/pnas.0900595106.

(16) Passner, J. M.; Schultz, S. C.; Steitz, T. A. Modeling the CAMP-Induced Allosteric Transition Using the Crystal Structure of CAP-CAMP at 2.1Åresolution. *J. Mol. Biol.*

2000, *304* (5), 847–859. https://doi.org/http://dx.doi.org/10.1006/jmbi.2000.4231.

(17) Harman, J. G. Allosteric Regulation of the CAMP Receptor Protein. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **2001**, *1547* (1), 1–17. https://doi.org/10.1016/S0167-4838(01)00187-X.

(18) Popovych, Nataliya; Sun, Shangjin; Ebright, Richard; Kalldimos, C. Dynamically Driven Protein Allostery. *Nat. Struct. Mol. Biol.* **2006**, *13* (9), 831–838. https://doi.org/10.1038/nsmb1132.Dynamically.

(19) Forman-Kay, J. D. The "dynamics" in the Thermodynamics of Binding. *Nat. Struct. Biol.* **1999**, *6* (12), 1086–1087. https://doi.org/10.1038/70008.

(20) Jarymowycz, V. A.; Stone, M. J. Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev.* **2006**, *106* (5), 1624–1671. https://doi.org/10.1021/cr040421p.

(21) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448* (7151), 325–329. https://doi.org/10.1038/nature05959.

(22) Tzeng, S.-R.; Kalodimos, C. G. Allosteric Inhibition through Suppression of Transient Conformational States. *Nat. Chem. Biol.* **2013**, *9* (7), 462–465. https://doi.org/10.1038/nchembio.1250 http://www.nature.com/nchembio/journal/v9/n7/abs/nchembio.1250.html#supplementary-information.

(23) 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* (10), 876–880. https://doi.org/10.1038/82820.

(24) 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* (40), 25757–25764.

(25) 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* (9), 2669–2678. https://doi.org/10.1021/bi982658j.

(26) Yamamoto, K.; Ishikawa, H.; Takahashi, S.; Ishimori, K.; Morishima, I.; Nakajima, H.; Aono, S. Binding of CO at the Pro2 Side Is Crucial for the Activation of CO-Sensing Transcriptional Activator CooA. 1H NMR Spectroscopic Studies. *J. Biol. Chem.* **2001**, *276* (15), 11473–11476. https://doi.org/10.1074/jbc.C100047200.

(27) 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* (41), 37895–37899. https://doi.org/10.1074/jbc.M105429200.

(28) 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* (37), 33616–33623. https://doi.org/10.1074/jbc.M203684200.

(29) 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* (2), 345–350. https://doi.org/https://doi.org/10.1016/j.bbrc.2006.06.200.

(30) 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* (44), 14149–14160. https://doi.org/10.1021/bi0487948.

(31) 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* (3), 282–287. https://doi.org/10.1107/S0907444906051638.

(32) 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* (4), 809–823.

(33) 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* (3), 651–668. https://doi.org/http://dx.doi.org/10.1016/j.jmb.2004.06.040.

(34) 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. *Protein Sci.*

2018, *27* (9), 1670–1679. https://doi.org/10.1002/PRO.3449.

(35) Hubbell, W. L.; Mchaourab, H. S.; Altenbach, C.; Lietzow, M. A. Watching Proteins Move Using Site-Directed Spin Labeling. *Structure* **1996**, *4* (7), 779–783. https://doi.org/10.1016/S0969-2126(96)00085-8.

(36) López, C. J.; Oga, S.; Hubbell, W. L. Mapping Molecular Flexibility of Proteins with Site-Directed Spin Labeling: A Case Study of Myoglobin. *Biochemistry* **2012**, *51* (33), 6568–6583. https://doi.org/10.1021/bi3005686.

(37) Klug, C. S.; Feix, J. B. Methods and Applications of Site-Directed Spin Labeling EPR Spectroscopy. In *Biophysical Tools for Biologists, Volume One: In Vitro Techniques*; 2008; pp 617–658. https://doi.org/10.1016/S0091-679X(07)84020-9.

(38) Mchaourab, H. S.; Berengian, A. R.; Koteiche, H. A. Site-Directed Spin-Labeling Study of the Structure and Subunit Interactions along a Conserved Sequence in the α-Crystallin Domain of Heat- Shock Protein 27. Evidence of a Conserved Subunit Interface. *Biochemistry* **1997**, *36* (48), 14627–14634. https://doi.org/10.1021/bi971700s.

(39) Berry, E. A.; Trumpower, B. L. Simultaneous Determination of Hemes a, b, and c from Pyridine Hemochrome Spectra. *Anal. Biochem.* **1987**, *161*, 1–15.

(40) 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* (43), 39968–39973. https://doi.org/10.1074/jbc.M105758200.

(41) Kiley, P. J.; Beinert, H. Oxygen Sensing by the Global Regulator, FNR: The Role of the

Iron-Sulfur Cluster. *FEMS Microbiol. Rev.* **1998**, *22* (5), 341–352. https://doi.org/10.1111/j.1574-6976.1998.tb00375.x.

(42) Volbeda, A.; Darnault, C.; Renoux, O.; Nicolet, Y.; Fontecilla-Camps, J. C. The Crystal Structure of the Global Anaerobic Transcriptional Regulator FNR Explains Its Extremely Fine-Tuned Monomer-Dimer Equilibrium. *Sci. Adv.* **2015**, *1* (11), e1501086. https://doi.org/10.1126/sciadv.1501086.

(43) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Sructural and Energenic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585-609. https://doi.org/10.1146/annurev-biophys-050511-102319

Chapter 5

Opportunities for Implementing Natural Climate Solutions in Wisconsin Agriculture

A version of this chapter was written as a requirement for the WISCIENCE Public Service Fellowship:

MGR conducted the literature review and wrote the manuscript, with input from a community partner mentor.

5.1 Introduction

Wisconsin has long been at the epicenter of the modern movement for environmental conservation, though the State's ability to address current and future environmental crisis are politically uncertain. Beginning in the 1960s, Wisconsin Governor of Gaylor Nelson pioneered legislation to protect the state's natural features and ensure all residents could enjoy nature. During his time in the U.S. Senate, Nelson's efforts to amend the Constitution to ensure that "Every person has the inalienable right to a decent environment" led to the first Earth Day in 1970.[1] Nelson led during a period of major progress now called the "Environmental Decade" that included legislative milestones such as the creation of the Environmental Protection Agency and passage of the Clean Air Act, Clean Water Act, and Endangered Species Act. Today, the current governor is working to address another challenging problem, combatting the effects of climate change. In 2019, Gov. Tony Evers and Lt. Gov. Mandela Barnes signed Executive Order 52, establishing The Governors Task Force on Climate Change.[2] This group was tasked with identifying current and future impacts of climate change on Wisconsin communities, as well as working to mitigate the worst effects of a changing climate on critical Wisconsin industries. As reports are released from the taskforce, implementation of their goals must largely come from the Legislature, which has not recently viewed climate change as a legislative priority.

Data detailing the effects of the changing climate in Wisconsin has been collected by the Wisconsin Initiative on Climate Change Impacts (WICCI) since 2011. WICCI is a statewide initiative of scientists, experts, and practitioners that was directed to collect updated data on the extent and impact of climate change in Wisconsin by Executive Order 52. Their reports from 2020 and 2021 present data on how climate change is currently impacting state communities and critical industries, such as agriculture, forestry, and fishing.[3,4] More extreme weather, including

warmer winters and wetter springs, are especially problematic for the agricultural industry that relies on reliable weather patterns. Waterlogged soils and extreme heat have impacted spring planting, fall harvesting, as well as milk production and water usage by the dairy industry. Water and soil quality have also been impacted by the nutrient runoff and soil erosion caused by increased precipitation. Farmers are already seeing changing weather patterns that impact their planting, growing season, and harvesting and the wellbeing of livestock. These effects are expected to worsen over time, leading to a sense of urgency to minimize the worst impacts on Wisconsin's critical agricultural industry through sustainable practices and new technology.[5]

To limit global warming by less than 2°C, the goal laid out by the Paris Climate Accord to avoid the worst impacts of global warming, the world needs to make drastic changes to policy surrounding energy use and greenhouse gas (GHG) emissions. Wisconsin, as a part of this agreement, set several goals to meet this target as a part of the Wisconsin Clean Energy Plan, released in 2022.[6] The main goal is for all electricity consumed within the state to be 100 percent carbon-free by 2050. Policies that target reduction of carbon emissions are most often the paradigm change in global energy practices: developing carbon-free energy sources to reduce and eventually eliminate the burning of fossil fuels, increasing reliance on electric cars, and improving battery storage technology. However the reduction of emissions alone will not be enough to meet global goals as presented by the Paris Climate Accord and prevent climate catastrophe. In addition to reduction of GHG emissions, the removal and storage of CO_2 from the atmosphere is necessary.[7,8] Carbon capture is an active area of scientific research, with expensive technology to sequester carbon currently being investigated, but it is not yet feasible to implement on a large scale.

Natural climate solutions (NCS) provide an immediate and cost-effective source of carbon sequestration with potential to significantly contribute to meeting global climate targets. NCS encompass land management, restoration, and protection practices to mitigate climate change effects by sequestering carbon in ecosystems and reducing GHG emissions. The earth's soil can store a tremendous amount of carbon (3.2 times the carbon in the atmosphere[9]), as can living plant material. Specific ecosystems, forests, wetlands, grasslands, and agricultural areas, are especially capable of holding onto carbon that would otherwise enter our atmosphere. It is estimated that implementation of cost-effective natural climate solutions could contribute one third of the global mitigation (carbon sequestration + avoided GHG emissions) needed by 2030 to keep warming under 2°C.[10] In the United States, implementation of all available NCS has been estimated to generate mitigation equivalent to 21% of our net annual emissions.[11]

Within NCS, there are opportunities to target different biomes based on local climate and land use. Agricultural practices that increase carbon storage provide an important opportunity for impact in Wisconsin, where 30% of the state's land is used for agriculture. Agricultural soils generally have less carbon compared to the native ecosystems they were derived from due to reduced carbon inputs, increased soil disturbance and erosion.[12] NCS practices can serve to increase carbon in depleted agricultural soils by increasing carbon inputs and reducing GHG emissions. Additionally, the implementation of NCS on farmed land has numerous co-benefits, including improving the overall health and productivity of soil and improving local water quality through reduced erosion and fertilizer runoff. Policies that mitigate climate change through improved farmland management also create opportunities to positively involve rural landowners in climate change policy. Involving farmers by tapping their lifetime of experiential knowledge of their land and financially incentivizing their participation in climate change mitigation, while

providing them guidance on protecting their land's productivity and increasing resilience against climate change, provides a positive opportunity to engage the agricultural sector in meeting climate goals.

Global Carbon and Nitrogen Cycling Underlies Carbon Storage

Key to NCS in the agricultural sector is the intertwined cycling of carbon (C) and nitrogen (N) in the soil, since both contribute to GHG emissions (Figure 5.1). Plant leaf and root matter is deposited into the soil, where is broken down by soil organisms into organic matter that contains carbon and nitrogen, called soil organic material. The amount of soil C in managed soils is impacted by the rate of C input from plant biomass as well as the rate of C loss from decomposition by soil microbes.

Despite the abundance of N in our atmosphere in the form of N_2 gas, N is often considered to be a limiting resource in agriculture because the N_2 form cannot be used by plants and animals. The process of converting N_2 to a version that can be used by plants and animals is called nitrogen fixation. Only a select group of organisms are able to perform this reaction, including soil-dwelling bacteria that associate with the roots of certain types of plants (legumes). These bacteria provide N to the soil in a form usable to the plants and animals. Other soil bacteria can also carry out other important steps in the nitrogen cycle including the production of nitrate (NO_3^-), a form of nitrogen taken up by plants, and eventual conversion back to N_2, returning nitrogen to the atmosphere. An important side product of this cycling is nitrous oxide (N_2O), a greenhouse gas 298 times as potent as CO_2.[13] N_2O is produced more in water-logged soil, but production is also influenced by soil oxygen content, pH, and redox potential. The implications of the inherent interdependencies of soil C and N means that efforts to increase SOC also affect soil N and impact N_2O emissions.

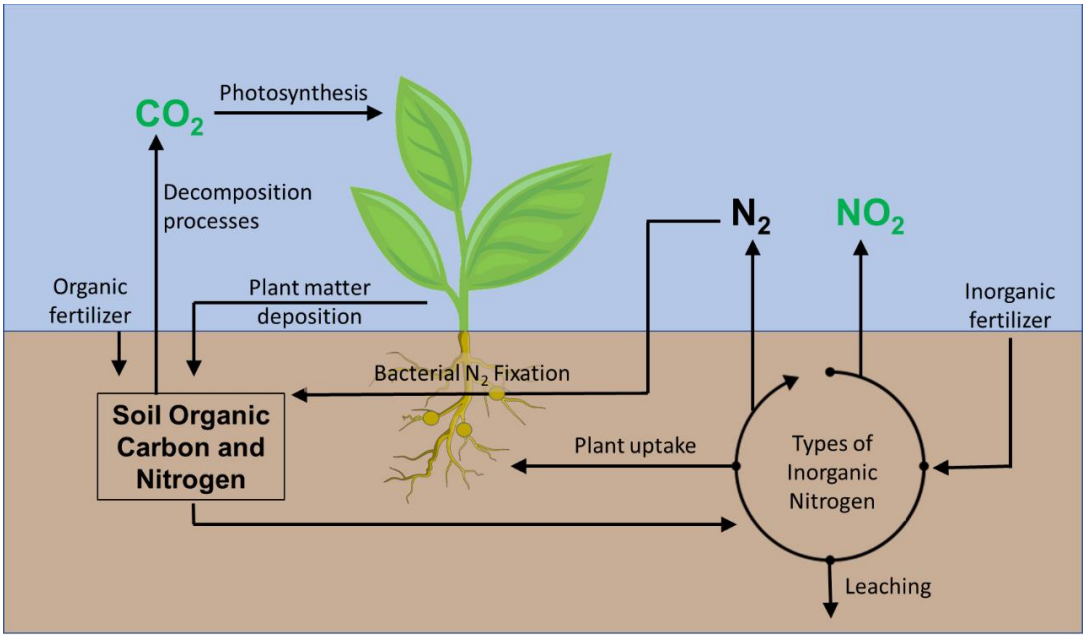


Figure 5.1. A schematic of simplified terrestrial C and N cycles and the interactions between cycles.[13] Each process listed is impacted by soil properties such as water content, redox potential, temperature, clay content, soil type, and pH. Greenhouse gases (CO_2 and N_2O) are shown in green, but other emitted gases and GHGs not mentioned in this report, including methane (CH_4), NOx, and NH_3, are not shown.

Excess nitrogen in the soil from overuse of nitrogen fertilizer, nitrogen fixation by legume crops, manure addition, or plant residues can be "lost" through several pathways, including leaching. N leaching occurs when water flows through the soil, causing the excess nitrate to enter groundwater aquifers and waterways. This process results in decreased nitrogen availability to plants and increased "indirect" N_2O emissions, as leached NO_3^- is converted into N_2O in local waterways.[13] Therefore, efforts to mind nitrogen input into the soil and prevent over-application of nitrogen fertilizers can prevent the emission of GHG, NO_2, into the atmosphere.

5.2 Review of the Scientific Literature on Natural Climate Solutions in Agriculture

Natural Climate Solutions Practices

Discussion of well-researched NCS is shown in Table 5.1. Increasing soil C can happen through reducing soil disturbance, increasing C inputs into the soil, improving soil microbe biodiversity, and maintaining plant cover on soil year-round. These practices also have numerous co-benefits that improve soil health. Each practice will be defined and discussed individually, paying particular attention to scientific debate and broad utility of each technique.

AGROFORESTRY

The term agroforestry is used to encompass practices that incorporate trees into agricultural land. Wind breaking describes surrounding fields with a strip of trees, 1-4 trees thick. Similarly, alley cropping describes planting crops between rows of trees. Both of these practices protect crops from wind (reducing soil erosion) and absorb excess N, while the trees themselves also act as a large C sink.[13] However, crop yield is impacted due to shade and reducing field area, and some farming equipment can be difficult to navigate between trees. The significant reduction to

Table 5.1. Six on-field agricultural NCS are presented. The solutions are listed alphabetically, with co-benefits and a short summary of the strength of scientific support for GHG emission reductions of the practice included.

Practice	Effect on GHG Balance and Co-benefits	State of Science
Agroforestry	• Carbon sequestration—increased C storage in plant biomass • Reduce N_2O emissions • Protect crops from wind	Strong evidence of net GHG benefit
Conversion to Perennial Fields	• Carbon sequestration—increased C input and decreased C loss • Reduce erosion • Reduce fertilizer runoff and N emissions	Strong evidence of net GHG benefit
Cover crops	• Carbon sequestration—increased C inputs • Reduce erosion • Reduce nitrogen leaching by reducing fertilizer need • Increase biodiversity • Improve water retention	Strong evidence of GHG benefit
Improved fertilizer management	• Reduce N_2O emissions and fertilizer runoff • Reduce GHG emissions from fossil fuels used in fertilizer production	Strong evidence of GHG benefit
Regenerative Grazing Strategies	• Carbon sequestration—increased C inputs • Reduce erosion • Increase N retention	Mixed evidence of GHG benefit
Tillage/No-till	• Modest carbon sequestration likely offset by increased N_2O emissions • Reduced C losses • Improve water content • Possibly less labor intensive	Mixed evidence of GHG benefit

farmers' yields makes agroforestry less appealing to implement widely, though there would be great potential to sequester carbon.[11] Silvopasture is the combination of forest and grassland for livestock grazing. Planting trees in pastureland can increase soil health and reduce erosion, and the trees are able to shade the livestock. However, there are concerns about implementing silvopasture as it is difficult to establish with significant time required for new trees to reach maturity before livestock can be introduced, and there is increased risk to the animals from falling branches and lightning attraction.

PERENNIAL FIELDS

Annual crops are sown and harvested within one year often using conventional tillage techniques that enhance soil erosion. Perennial plants are left to grow for years, limiting the need for tillage and allowing for well-established root systems which limit erosion and promote soil organic carbon storage. Conversion of less productive cropland to fields of perennial grasses that would be left unharvested are well-supported to sequester large amounts of carbon deeper into the soil through their deeper, more mature root systems, and are able to better store carbon due to the lack of soil disturbance by plowing, replanting, etc.[12]

COVER CROPS

Cover crops are planted after the main harvest of cash crops with the goal of covering the bare soil, preventing erosion, increasing water retention, improving biodiversity, and returning important nutrients to the soil after being depleted by cash crops. The practice of planting cover crops increases SOC by increasing C input that returns to the soil. There is debate among researchers as to the best cover crops to be used for SOC sequestration, especially non-legume

vs. legume plants and the extent to which increased N_2O emissions could offset any gains in sequestered C. Meta-analyses have found that cover crops increase SOC independent of tillage method used, climate zone, or plant type, though it does depend on the duration and frequency of inclusion in a crop rotation.[13,14] The effect of planting cover crops on N_2O emissions is highly variable between studies based on the use of extra N fertilizer, whether the cover crops are left to decay or if they are plowed into the soil, and type of cover crop used. A recent meta-analysis found that cover crops on average resulted in a small increase in "direct" N_2O emissions, from soil to the atmosphere, that was not large enough to offset gains in SOC.[13,15] However, an important indirect effect of cover crop planting is the reduction of nitrate leaching, which is beneficial for local water quality and decreases "indirect" N_20 emissions in waterways from the leaked NO_3^-.[15,16] Overall, effects on N_2O emissions due to planting cover crops may be highly site specific, but likely would not be great enough to offset the positive benefits to SOC.[13,16]

IMPROVED FERTILIZER MANAGEMENT

Improved fertilizer management is another opportunity to reduce GHG emissions and provide positive co-benefits. More efficient use of nitrogen fertilizers would result in less N_2O emissions on fields as well as reduce the fossil fuel emissions associated with the production of N fertilizers via the Haber-Bosch Process (globally, 52 Tg CO_2e $year^{-1}$).[11] Precision agriculture, which involves applying only the precise amount of fertilizer required by each part of a field with thoughtful application based on crop demand and soil content, would allow for less fertilizer to be used while maintaining crop yields. Additional practices to use fertilizer more efficiently include minimizing water movement across fields after fertilizer application, using slow-release fertilizers and additives to prevent mineralization, and using cover crops and deep-rooted crops

to scavenge residual nitrogen. There is also an important water quality co-benefit of improved fertilizer management because nitrates are Wisconsin's most widespread groundwater contaminant.[17]

REGENERTIVE GRAZING STRATEGIES

The term "regenerative grazing strategies" is used here to describe practices that consider how to maximize sustainability in livestock grazing, including adaptive multi-paddock grazing. Sustainable grazing strategies have been researched for their potential to sequester carbon due to the ability of livestock to redistribute carbon throughout a pasture, in the form of manure, increasing C inputs and stored C in soil. Grazing strategies to prevent overgrazing ensure that pasture plants have continuously high biomass. Adaptive multi-paddock (AMP) grazing is a type of rotational grazing where livestock rotates through multiple fenced paddocks for short periods of time while monitoring plant consumption.[18] This practice ensures that there is time for plant growth to recover between grazings, reduces fertilizer consumption, increases biodiversity, decreases erosion, and is better for livestock health. Rotational grazing practices like AMP increase soil organic carbon, potentially by increasing amount and quality of carbon inputs and retaining more N, which allows soil to sequester more carbon than conventionally grazed soils.[18]

NO-TILL

No or low-tillage agriculture has been a well-researched NCS since the 1980s. As well as leading to a modest increase in SOC in the top 10 cm of soil, it has been shown to have other positive benefits including reducing erosion and increasing water content while being less labor intensive. However, more recent research raises questions about whether low/no-till practices actually

increase the total amount of stored soil carbon, or if it only redistributes it to the top 10 cm of soil.[19] Also, low/no-till practices create conditions that increase N_2O emissions. While the increase in N_2O emissions is relatively small, it is enough to approximately offset any increased carbon accumulated by low/no-till practices, and therefore low/no-till agriculture does not appear to lead to a net reduction GHG emissions in most cases.[13]

DEVELOPING TECHNOLOGIES

Several frontier technologies may be useful in the future but require additional research to better understand their quantifiable effects on soil C storage and to make widespread adoption economically viable.[12]

1) Biochar- A C-rich solid produced from the pyrolysis of plant material. Adding biochar to soils can improve soil C content because it is largely resistant to decomposition and influences crop productivity. However, it is still unknown how biochar addition to soils impacts the native microbial processes that control soil C and N cycling, and thus it is unknown how biochar effects decomposition of SOC and N_2O emissions, though there is an emerging consensus that biochar can reduce N_2O. A full analysis of the biochar lifecycle, involving the production and transport of biochar involves many sources of GHGs, so the location and system design would need to be better understood prior to large-scale deployment.
2) Conversion to perennial grain crops— Kernza (trademarked) and alfalfa, are two alternatives to annuals that could be reasonably planted in Wisconsin, as these are both capable of surviving winters and are widely researched.[20,21] Both alfalfa and Kernza have deep root systems and reduce the need to plow soil yearly, leading to a reduction of

erosion and N leaching. Both plants require less fertilizer, herbicide, and pesticide application. However, there are few long-term studies to quantify the potential for C sequestration in fields that convert to perennial wheat crops. Reduced yields and economic viability are the primary barriers to implementation of perennial crops versus the high yield, traditional annual crops.[22]

3) Research into breeding annual crops with deeper root systems to sop up excess N and deposit C deeper into soil is currently being pursued.

5.3 Limitations and Continued Debate

Limitations to NCS Practices

Although the practices discussed above are well-supported by science in a wide variety of conditions, the quantitative potential for C sequestration and minimizing N_2O emissions depends heavily on individual field conditions. Soil temperature, moisture, drainage, O_2 content, pH, texture and clay content impact decomposition rates in the soil as well as the ability to "stabilize" SOC from decomposition, each impact the amount of C sequestered in soil. Local climate, cropping regimes, and management practices all influence SOC increases as well. Thus, any implementing policy to increase adoption of land management practices that increase soil C– particularly any that would pay on the basis of carbon stored— there would need to be a robust system for measuring, reporting, and verifying increases in soil C to ensure that policy goals are being met.

There is a ceiling as to how much C can be sequestered in soils. Gains in soil C will attenuate over time as the soil approaches that new equilibrium value. Meaning that in the beginning of implementation, there will be larger gains and become increasingly small over time.[23] Most managed soils are well below their saturation level, but it does mean that soils with

higher soil organic matter may not be able to sequester much additional C. Many studies of carbon storage capacity of various practices only evaluate carbon storage over the first few years, when storage increases are the greatest. Thus, extrapolations from such studies will likely overestimate storage potential over the course of decades. Notably, some NCS, such as improved fertilizer management, that are based on avoided GHG emissions produce benefits that do not saturate over time. Soil C storage is governed by biotic processes, and thus soil C storage is potentially reversible. Practices to increase soil C must be continued long-term to avoid losing the stored C.[23]

The ratio of C to N is narrowly maintained in soils to range from 8-20, with many agricultural soils being 10-12.[12] To maintain this balance of C:N while increasing C, N would also have to be significantly increased. This underlines the importance of incorporating legumes in many of the practices discussed above, as using legumes to increase soil N would avoid the use of many more tons of N fertilizers. Improved management of soil N with the goal of maintaining crop yields, increasing soil organic N, while preventing N_2O emissions from excess N loss, are an important component of reducing GHG emissions from soil.

Areas of Uncertainty in the Scientific Literature

The overall impact of no-till practices over time is debated. Some studies show that no-till land undergoing an occasional very deep till, where the C rich soil in the top few cm can be redistributed to 60-80cm and pulling the less C rich soil to the top can allow for better storage of the C rich soil. The depth helps to reduce the CO_2 lost to decomposition, while allowing the newly exposed soil to be subject to traditional practices to increase soil C- cover crops, no-till, etc.[12]

The best practices for grazing, well-managed continuous grazing vs AMP/rotational grazing depend heavily on context. Debate as to the increases in SOC by AMP grazing and to the body of science supporting rotational grazing.[12] It is difficult to produce a replicable experiment studying these practices due to varying weather and environmental conditions that effect grazing lands. More research needed to decide optimal conditions and practices for increasing SOC and minimizing other GHG emissions.

The productivity of agricultural land is the top priority to farmers in implementing new practices. NCS practices seek to improve the overall health of soil and the ability of the land to withstand the negative impacts of climate change, however some of the practices discussed above would remove part of the field from being planted with cash crops or reduce fertilizer application that may reduce yield. If productivity losses require agricultural expansion elsewhere, the climate gains are likely eliminated since conversion from natural habitat to cultivated land is associated with substantial carbon losses.

5.4 Policy Pathways

The Governors Task Force on Climate Change proposed wide policy recommendations based on the results of the WICCI studies, with particular focus on the needs of agriculture. They recommend four main policy pathways: 1) Support farmer-led watershed groups that promote conservation practices through grants, 2) Pay farmers to increase soil carbon storage in agricultural and working lands, 3) Avoid conversion of natural working lands by protecting productive agricultural, forested, and grassland areas from development, and 4) Make managed grazing livestock production systems an agricultural priority to reduce emission of GHG. These priorities fall well within the scientific consensus to build healthy soils and protect productive agricultural land from the impacts of climate change while creating a natural carbon storage sink.

These policy pathways largely require cooperation from executive agencies such as the Department of Agriculture, Trade, and Consumer Protection (DATCP), community partner organizations and farmers, and legislative support to make significant progress. Decisions about which evidence-based sustainable practices to pay farmers for, how to navigate monitoring soil C storage, and measurement of the reduced GHG emissions will require partnerships with Wisconsin scientists and farmers to implement specific, measurable policy goals.

5.5 Summary and Conclusions

Soil C sequestration on agricultural lands are attractive practices because they do not require converting land back to forest to have a significant climate impact, both in protecting key a key Wisconsin industry from the worst effects of climate change and in storing carbon. There is strong scientific support for at least some agricultural management practices to sequester large amounts of soil C in the next several decades. Importantly, the management practices described here have indisputable important co-benefits of improving soil health and local water quality. Thus, investments in such policies are virtually guaranteed to provide some benefit, providing for a “no-regret” opportunity in climate change mitigation even if the carbon storage capacity is not as great as expected.

5.6 References

(1) Library of Congress, "Senator Gaylord Nelson", Congress.gov.

(2) State of Wisconsin Executive Order No. *52*; **2019**.

(3) Wisconsin Initiative on Climate Change Impacts. Report to the Governor's Task Force on Climate Change: Strategies to Improve Wisconsin's Climate Resilience and Readiness. **2020**, 4–10.

(4) Wisconsin Initiative on Climate Change Impacts. *Wisconsin's Changing Climate: Impacts and Solutions for a Warmer Climate*; Madison, WI, **2021**.

(5) Motha, R. P. Chapter 30: The Impact of Extreme Weather Events on Agriculture in the United States, **2011**. Publications from USDA-ARS / UNL Faculty. 1311. https://digitalcommons.unl.edu/usdaarsfacpub/1311

(6) Wisconsin Office of Sustainability and Clean Energy, Great Plains Institute, and Slipstream, Inc. Clean Energy Plan. **2022**, No.1.

(7) IPCC, 2021: Summary for Policymakers. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.

(8) Post, W. M.; Kwon, K. C. Soil Carbon Sequestration and Land-Use Change: Processes and Potential. *Glob. Chang. Biol.* **2000**, *6* (3), 317–327. https://doi.org/10.1046/J.1365-

2486.2000.00308.X.

(9) Zomer, R. J.; Bossio, D. A.; Sommer, R.; Verchot, L. V. Global Sequestration Potential of Increased Organic Carbon in Cropland Soils. *Sci. Reports* **2017**, *7* (1), 1–8. https://doi.org/10.1038/s41598-017-15794-8.

(10) Griscom, B. W.; Adams, J.; Ellis, P. W.; Houghton, R. A.; Lomax, G.; Miteva, D. A.; Schlesinger, W. H.; Shoch, D.; Siikamäki, J. V.; Smith, P.; Woodbury, P.; Zganjar, C.; Blackman, A.; Campari, J.; Conant, R. T.; Delgado, C.; Elias, P.; Gopalakrishna, T.; Hamsik, M. R.; Herrero, M.; Kiesecker, J.; Landis, E.; Laestadius, L.; Leavitt, S. M.; Minnemeyer, S.; Polasky, S.; Potapov, P.; Putz, F. E.; Sanderman, J.; Silvius, M.; Wollenberg, E.; Fargione, J. Natural Climate Solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (44), 11645–11650. https://doi.org/10.1073/pnas.1710465114.

(11) Fargione, J. E.; Bassett, S.; Boucher, T.; Bridgham, S. D.; Conant, R. T.; Cook-Patton, S. C.; Ellis, P. W.; Falcucci, A.; Fourqurean, J. W.; Gopalakrishna, T.; Gu, H.; Henderson, B.; Hurteau, M. D.; Kroeger, K. D.; Kroeger, T.; Lark, T. J.; Leavitt, S. M.; Lomax, G.; McDonald, R. I.; Patrick Megonigal, J.; Miteva, D. A.; Richardson, C. J.; Sanderman, J.; Shoch, D.; Spawn, S. A.; Veldman, J. W.; Williams, C. A.; Woodbury, P. B.; Zganjar, C.; Baranski, M.; Elias, P.; Houghton, R. A.; Landis, E.; McGlynn, E.; Schlesinger, W. H.; Siikamaki, J. V.; Sutton-Grier, A. E.; Griscom, B. W. Natural Climate Solutions for the United States. *Sci. Adv.* **2018**, *4* (11), 1–15. https://doi.org/10.1126/sciadv.aat1869.

(12) Paustian, K.; Larson, E.; Kent, J.; Marx, E.; Swan, A. Soil C Sequestration as a Biological Negative Emission Strategy. *Front. Clim.* **2019**, *1* (8), 1–11. https://doi.org/10.3389/fclim.2019.00008.

(13) Guenet, B.; Gabrielle, B.; Chenu, C.; Arrouays, D.; Balesdent, J.; Bernoux, M.; Bruni, E.; Caliman, J. P.; Cardinael, R.; Chen, S.; Ciais, P.; Desbois, D.; Fouche, J.; Frank, S.; Henault, C.; Lugato, E.; Naipal, V.; Nesme, T.; Obersteiner, M.; Pellerin, S.; Powlson, D. S.; Rasse, D. P.; Rees, F.; Soussana, J. F.; Su, Y.; Tian, H.; Valin, H.; Zhou, F. Can N_2O Emissions Offset the Benefits from Soil Organic Carbon Storage? *Glob. Chang. Biol.* **2021**, *27* (2), 237–256. https://doi.org/10.1111/gcb.15342.

(14) McClelland, S. C.; Paustian, K.; Schipanski, M. E. Management of Cover Crops in Temperate Climates Influences Soil Organic Carbon Stocks: A Meta-Analysis. *Ecol. Appl.* **2021**, *31* (3), 1–19. https://doi.org/10.1002/eap.2278.

(15) Abdalla, M.; Hastings, A.; Cheng, K.; Yue, Q.; Chadwick, D.; Espenberg, M.; Truu, J.; Rees, R. M.; Smith, P. A Critical Review of the Impacts of Cover Crops on Nitrogen Leaching, Net Greenhouse Gas Balance and Crop Productivity. *Glob. Chang. Biol.* **2019**, *25* (8), 2530–2543. https://doi.org/10.1111/gcb.14644.

(16) Kaye, J. P.; Quemada, M. Using Cover Crops to Mitigate and Adapt to Climate Change. A Review. *Agron. Sustain. Dev.* **2017**, *37* (1). https://doi.org/10.1007/s13593-016-0410-x.

(17) *Wisconsin Groundwater Coordinating Council Report to the Legislature-2021*; **2021**.

(18) Mosier, S.; Apfelbaum, S.; Byck, P.; Calderon, F.; Teague, R.; Thompson, R.; Cotrufo, M. F. Adaptive Multi-Paddock Grazing Enhances Soil Carbon and Nitrogen Stocks and Stabilization through Mineral Association in Southeastern U.S. Grazing Lands. *J. Environ. Manage.* **2021**, *288*, 112409. https://doi.org/10.1016/j.jenvman.2021.112409.

(19) Powlson, D. S.; Stirling, C. M.; Jat, M. L.; Gerard, B. G.; Palm, C. A.; Sanchez, P. A.; Cassman, K. G. Limited Potential of No-till Agriculture for Climate Change Mitigation.

Nat. Clim. Chang. **2014**, *4* (8), 678–683. https://doi.org/10.1038/nclimate2292.

(20) UW-Madison coalition aims to adopt first US perennial grain crop https://www.wisfarmer.com/story/news/2020/09/29/uw-madison-coalition-aims-adopt-first-us-perennial-grain-crop/3568717001/ (accessed Jan 23, 2022).

(21) Putnam, D.; Russelle, M.; Orloff, S.; Kuhn, J.; Fithugh, L.; Godfrey, L.; Kiess, A.; Long, R. *The Importance and Benefits of Alfalfa in the 21st Century*; Novanto, CA, 2001.

(22) Ryan, M. R.; Crews, T. E.; Culman, S. W.; Dehaan, L. R.; Hayes, R. C.; Jungers, J. M.; Bakker, M. G. Managing for Multifunctionality in Perennial Grain Crops. *Bioscience* **2018**, *68* (4), 294–304. https://doi.org/10.1093/biosci/biy014.

(23) Bossio, D. A.; Cook-Patton, S. C.; Ellis, P. W.; Fargione, J.; Sanderman, J.; Smith, P.; Wood, S.; Zomer, R. J.; von Unger, M.; Emmer, I. M.; Griscom, B. W. The Role of Soil Carbon in Natural Climate Solutions. *Nat. Sustain.* **2020**, *3* (5), 391–398. https://doi.org/10.1038/s41893-020-0491-z.