BIOCHEMICAL CHARACTERIZATION OF PHYLOGENETICALLY DIVERSE HOMOLOGS OF THE ANTIBIOTIC RESISTANCE PROTEIN CFR AND RELATED GENOME NEIGHBORS

A Dissertation in
Chemistry
by
James David Gumkowski

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

May 2020
The dissertation of James David Gumkowski was reviewed and approved by the following:

Amie K. Boal
Associate Professor of Chemistry
Associate Professor of Biochemistry and Molecular Biology
Dissertation Advisor
Chair of Committee

Squire J. Booker
Howard Hughes Medical Investigator
Professor of Chemistry
Professor of Biochemistry and Molecular Biology
Eberly Distinguished Chair in Science

Carsten Krebs
Professor of Chemistry
Professor of Biochemistry and Molecular Biology

Kenneth Keiler
Professor of Biochemistry and Molecular Biology
Associate Department Head for Graduate Education

Philip Bevilacqua
Department Head, Chemistry
Distinguished Professor of Chemistry
Distinguished Professor of Biochemistry and Molecular Biology
ABSTRACT

Radical S-adenosylmethionine (SAM) enzymes are one of the largest protein superfamilies identified to date, with more than 500,000 annotated members. While the functions of these enzymes vary widely, one of the most challenging chemical transformations that radical SAM enzymes are known to perform is methylation of unactivated carbon or phosphorous centers. The enzymes that catalyze these difficult reactions are known as the Class A-D radical SAM methyltransferases. The best characterized of the radical SAM methyltransferases are Class A, composed of RlmN and Cfr. RlmN is a housekeeping enzyme found in many bacteria that methylates rRNA and some tRNAs and is thought to promote translational fidelity. In contrast, Cfr is found in relatively few bacteria, largely Firmicutes, and its modification of rRNA has been shown to confer antibiotic resistance to agents which target the exit tunnel of the peptidyl transferase center. It accomplishes this by appending a methyl group to the C8 position of A2503 (Escherichia coli numbering). cfr is of particular concern presently because it has been found on transposable elements and is appearing with increasing frequency in clinical settings.

The original aim of this work was to structurally characterize a member of the Cfr family, a goal which required finding sequences which diverge from model system Staphylococcus aureus Cfr. In chapter 2, four phylogenetically diverse Cfr homologs were isolated and characterized. Though none proved amenable to structural studies, biophysical and spectroscopic analyses revealed that all four homologs contained the radical SAM [4Fe-4S] cluster and were able to perform C8 methylation of a 155mer substrate mimic with varying efficiencies. These results suggest that Cfr-like activity
associated with drug resistance can be found in a diverse set of organisms, some of which are human pathogens. I also showed that the identity of the iron-sulfur cluster reductant can have a significant impact on activity. The small molecule reductant dithionite yielded more dimethylated product compared to a flavodoxin protein-based reducing system. This work underscores the importance of studying Cfr activity with native reductants, because reaction outcome can be influenced by the identity of the reducing system and the rate at which it operates relative to RNA product release.

In this dissertation I show that a clostridial Cfr homolog is an active rRNA methylase \textit{in vitro}, a surprising observation given the lack of activity in other homologs from the clostridial clade. Additionally, the Mössbauer spectrum of \textit{Clostridioides difficile} Cfr demonstrated the presence of a mononuclear iron cofactor in the C-terminal domain. In chapter 3, I use truncation variants to verify the biophysical and functional properties of this domain. When isolated separately, the domain binds a single iron (II) ion with rubredoxin-like properties. Addition of this domain \textit{in trans} to a clostridial Cfr variant containing only the radical SAM domain did not affect activity. However, mismetallation of the domain or mutation of a predicted Cys ligand nearly eliminates methylation of a 155mer RNA substrate, suggesting an essential role for iron when the domain is present. Co-elution of the domain with RNA could also indicate a function related to RNA substrate binding. These results raise the possibility that the limited activity demonstrated by some clostridial Cfrs could be due to a lack of additional protein factors available in the native host organisms but missing when the RNA methylase is overexpressed by itself in a heterologous host.
The C-terminal domain identified on *Clostridioides difficile* Cfr is homologous to members of a family of small, uncharacterized proteins containing two pairs of CXXC motifs named Cys-rich KTR. In chapter 4, I isolate a standalone KTR from *Staphylococcus aureus* and demonstrate that it, like the C-terminal domain of the clostridial Cfr, coordinates a single iron and binds RNA. This result suggests that the members of this family could share a requirement for iron and that they may all function in RNA-binding or electron transfer. Bioinformatic analysis of the clostridial Cfrs revealed that all conserve either a Cys-rich KTR or a Lsa family ABC-F type ribosomal protection protein. This result suggests links between RNA methylases, ribosomal protection proteins, and a new family of metalloprotein, cys-rich KTRs, all of which may be necessary for efficient RNA methylation activity and drug resistance in these systems.

The work in this dissertation expands the understanding of Cfr activity from diverse phylogenetic sources, especially those from clostridial organisms. It also provides a basis for further investigation of the connection between Cfr, Lsa, and KTR function in clostridial pathogens and commensal organisms. Structure determination of a Cfr remains a top priority because it would allow for rational inhibitor design. Structural characterization of a member of the newly discovered metalloprotein family Cys-rich KTR would also validate many of the findings in this dissertation and give further insight into the genomic proximity to antibiotic resistance determinants.
TABLES OF CONTENTS

LIST OF FIGURES................................................................. viii
LIST OF TABLES......................................................................... xi

ACKNOWLEDGEMENTS.............................................................. xiii

Chapter 1: Introduction.................................................................. 1
  Project statement......................................................................... 2
  RNA modification and its importance in ribosome function........... 3
  Radical SAM enzymes............................................................... 4
  Radical SAM methylases.......................................................... 10
  Mechanistic characterization of RlmN and Cfr.............................. 15
  Project objectives....................................................................... 25
  References.................................................................................. 27

Chapter 2: Phylogenetic and biochemical analyses of four radical S-adenosylmethionine
  Cfr homologs of diverse origin................................................... 33
  Abstract..................................................................................... 34
  Introduction................................................................................. 35
  Materials and Methods............................................................... 40
  Results....................................................................................... 50
  Discussion.................................................................................. 81
  References................................................................................ 91

Chapter 3: Biochemical and spectroscopic characterization of a C-terminal rubredoxin-like domain in Clostridioides difficile Cfr................................................................. 95
  Abstract..................................................................................... 96
  Introduction............................................................................... 97
  Materials and Methods............................................................. 99
  Results...................................................................................... 108
  Discussion............................................................................... 123
  References............................................................................... 131

Chapter 4: Defining the properties and functions of Clostridial Cfr-associated proteins......................................................... 135
  Abstract................................................................................... 136
  Introduction............................................................................... 137
  Materials and Methods............................................................. 139
  Results...................................................................................... 151
  Discussion............................................................................... 177
  References............................................................................... 187

Chapter 5: Summary of Results and Future Directions.......................... 191
  Conclusions............................................................................... 192
  Future directions...................................................................... 194
  References............................................................................... 197
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure 1-1.</th>
<th>Reactivity of S-adenosylmethionine</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-2.</td>
<td>Nucleic acid modification by radical SAM enzymes</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1-3.</td>
<td>Substrate and products of class A radical SAM enzymes RlmN and Cfr</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1-4.</td>
<td>Drug classes with diminished activity in the presence of Cfr</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1-5.</td>
<td>Crystal structure of RlmN with SAM bound (3RFA)</td>
<td>17</td>
</tr>
<tr>
<td>Figure 1-6.</td>
<td>Crystal structure of RlmN cross-linked to tRNA&lt;sub&gt;Glu&lt;/sub&gt; bound (3HR7)</td>
<td>19</td>
</tr>
<tr>
<td>Figure 1-7.</td>
<td>Visualization of waters in the active site of RlmN which may be involved in proton exchange with Cys&lt;sub&gt;105&lt;/sub&gt; in Cfr</td>
<td>22</td>
</tr>
<tr>
<td>Figure 1-8.</td>
<td>Proposed mechanism of Cfr</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2-1.</td>
<td>Proposed mechanism of Cfr</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2-2.</td>
<td>Comparison of previously characterized Cfrs</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2-3.</td>
<td>Phylogenetic tree of diverse Cfr homologs and diagram of domain architecture in Cfrs selected for study</td>
<td>51</td>
</tr>
<tr>
<td>Figure 2-4.</td>
<td>UV-visible spectroscopy of Cfr homologs and Mössbauer spectroscopy of Cfrs from <em>Bacillus amyloliquefaciens</em>, <em>Enterococcus faecalis</em>, and <em>Paenibacillus lautus</em></td>
<td>55</td>
</tr>
<tr>
<td>Figure 2-5.</td>
<td>Comparison of as-isolated and reconstituted Mössbauer spectra of <em>Clostridioides difficile</em> Cfr</td>
<td>57</td>
</tr>
<tr>
<td>Figure 2-6.</td>
<td>Sequence alignment of 23S rRNA sequences from Cfr hosts showing conservation of A2503</td>
<td>58</td>
</tr>
<tr>
<td>Figure 2-7.</td>
<td>LC-MS chromatograph showing retention times of Cfr reaction products and of the tryptophan internal standard</td>
<td>59</td>
</tr>
<tr>
<td>Figure 2-8.</td>
<td>LC-MS chromatograph demonstrating Cfr homologs target only the expected A2503 site on the 155mer substrate mimic</td>
<td>60</td>
</tr>
<tr>
<td>Figure 2-9.</td>
<td>LC-MS chromatograph and methylated product quantification for Cfr homologs under limiting substrate conditions using a protein-based reducing system</td>
<td>61</td>
</tr>
<tr>
<td>Figure 2-10.</td>
<td>Kinetic analysis of methylation of a 155mer rRNA substrate by Cfr enzymes under limiting substrate conditions initiated with a protein-based reductant</td>
<td>62</td>
</tr>
</tbody>
</table>
Figure 2-11. Assessment of unproductive SAM cleavage at the burst phase under limiting substrate conditions in the presence of a protein-based reducing system or dithionite.... 64

Figure 2-12. Methylation of a 155mer rRNA by Cfr enzymes with excess substrate in reactions initiated with a protein-based reductant.......................................................... 66

Figure 2-13. Kinetic analysis of methylation of a 155mer rRNA substrate by Cfr enzymes under excess substrate conditions initiated with a protein-based reductant......................... 67

Figure 2-14. Methylation of a 155mer rRNA by Cfr enzymes under limiting substrate conditions in reactions initiated with dithionite................................................................. 70

Figure 2-15. Methylation of a 155mer rRNA by Cfr enzymes at high temperature under limiting substrate conditions in reactions initiated with dithionite........................................... 74

Figure 2-16. Methylation of a 155mer rRNA by Cfr enzymes with excess substrate in reactions initiated with dithionite.............................................................................. 76

Figure 2-17. Antibiotic resistance profiles of Cfr homologs in E. coli BL21 DE3 ΔtolC and Bacillus amyloliquefaciens Cfr in E. coli BL21 DE3 ΔtolC/ΔaraAB under optimized growth conditions tested against lincomycin................................................................. 79

Figure 2-18. Antibiotic resistance profiles of Clostridioides difficile Cfr and Paenibacillus lautus Cfr in E. coli BL21 DE3 ΔtolC under optimized growth conditions tested against linezolid.......................................................................................... 80

Figure 2-19. Comparison of Cfr activity under limiting and excess substrate conditions using a protein-based reducing system, dithionite, or dithionite at high temperature.... 83

Figure 2-20. Predicted reaction outcomes of Cfr when exposed to a small chemical reductant or a protein-based reducing system................................................................. 85

Figure 3-1. Visualization of Clostridioides difficile truncation variants.................... 99

Figure 3-2. Mössbauer spectra of Cfr enzymes and of Clostridioides difficile Cfr under varying applied magnetic fields.......................................................................................... 110

Figure 3-3. Methylation of a 155mer rRNA substrate by Clostridioides difficile truncation variants under limiting substrate and excess substrate conditions initiated by a protein-based reductant........................................................................................................... 112

Figure 3-4. Methylation of a 155mer rRNA substrate by Clostridioides difficile truncation variants under limiting substrate conditions initiated by dithionite........................................ 116

Figure 3-5. Electrochemical and UV-visible characterization of Clostridioides difficile Cfr C-terminal domain. Co-elution of Clostridioides difficile Cfr radical SAM domain and Clostridioides difficile Cfr KTR domain by size exclusion. Comparison of KTR and rubredoxin sequence elements........................................................................ 119
Figure 3-6. Co-elution of 155mer RNA and *Clostridioides difficile* Cfr C-terminal domain by size exclusion................................................................. 121

Figure 3-7. Crystal structure of the rubredoxin from *D. gigas* (2DSX).......................... 123

Figure 3-8. Proposal for electron transfer from the C-terminal KTR domain of *Clostridioides difficile* Cfr to the thyl radical speculated to remain at the conclusion of turnover................................................................................................. 126

Figure 4-1. Example of fixed concentration broth microdilution screen for antibiotic resistance profiling................................................................................................................................. 145

Figure 4-2. Example of variable percentage concentration screen for antibiotic resistance profiling........................................................................................................................................ 147

Figure 4-3. Sequence similarity network and genome neighborhood network analyses of stand-alone KTR proteins.................................................................................................................. 152

Figure 4-4. Homology model of 327 amino acid Cfr-like clostridial protein created using RlmN with SAM bound (3RFA) as a template.............................................................. 154

Figure 4-5. Metagenome abundance of KTR proteins in healthy human microbiome samples........................................................................................................................................ 156

Figure 4-6. Identification of conserved genome neighbors and demonstrated activity of clostridial Cfrs........................................................................................................................... 158

Figure 4-7. Cryo-EM structure of ABC-F type ribosomal protection protein VmlR complexed to the 70S ribosome of *B. subtilis* (6HA8)................................................................. 159

Figure 4-8. Antibiotic resistance profiles of *Bacillus amyloliquefaciens* Cfr and *Enterococcus faecalis* Cfr expressed using pET26b in *E. coli* BL21 DE3 ΔtolC...... 161

Figure 4-9. Antibiotic resistance profiles of *Bacillus amyloliquefaciens* Cfr expressed using pBAD24 in *E. coli* BL21 DE3 ΔtolC in LB medium or glycerol-supplemented EZ-rich medium.................................................................................................................. 163

Figure 4-10. Antibiotic resistance profiles of *Clostridium sporogenes* Cfr and Lsa expressed using pBAD24 in *E. coli* BL21 DE3 ΔtolC................................................................. 164

Figure 4-11. Antibiotic resistance profile of *Enterococcus faecalis* Cfr expressed using pBAD24 in *E. coli* BL21 DE3 ΔtolC........................................................................................... 165

Figure 4-12. Antibiotic resistance profiles of *Clostridioides difficile* Cfr and *Paenibacillus lautus* Cfr expressed using pBAD24 in *E. coli* BL21 DE3 ΔtolC...... 166
Figure 4-13. Antibiotic resistance profiles of *Clostridioides difficile* Cfr and *Paenibacillus lautus* Cfr expressed using pBAD24 in *E. coli* BL21 DE3 Δ*tolC* in the presence of linezolid................................................................. 167

Figure 4-14. Antibiotic resistance profiles of *Staphylococcus aureus* KTR and *Clostridium botulinum* KTR expressed using pBAD24 in *E. coli* BL21 DE3 Δ*tolC*........................................................................................................ 168

Figure 4-15. Antibiotic resistance profiles of *Staphylococcus aureus* KTR and *Bacillus amyloliquefaciens* Cfr expressed using pBAD24 in *E. coli* BL21 DE3 Δ*tolC/ΔaraAB*........................................................................................................... 170

Figure 4-16. UV-visible difference spectrum of the oxidation of *Staphylococcus aureus* KTR compared to the oxidation of the *Clostridioides difficile* Cfr KTR domain. Mössbauer spectrum of *Staphylococcus aureus* KTR................................................................. 173

Figure 4-17. Co-elution of *Staphylococcus aureus* KTR with the 155mer RNA Cfr substrate during size exclusion chromatography...................................................... 174

Figure 4-18. Mass spectrometry analysis of unknown contaminant protein present during KTR purifications........................................................................................................... 176

Figure 4-19. *Staphylococcus aureus* KTR crystal hit in Classics sparse matrix screen condition B4 (0.2 M NaCl, 0.1 M sodium acetate pH 4.6, 30 % [v/v] MPD)........... 177
**LIST OF TABLES**

Table 2-1. 155mer RNA substrate sequence, Cfr homolog protein and nucleotide sequences, and primer sequences................................................................. 87

Table 2-2. Liquid chromatography elution gradient for analysis of Cfr reaction products.................................................................................................................. 45

Table 2-3. Sequence identity for Cfr homologs tested for RNA methylation........ 54

Table 2-4. Iron incorporation for Cfr homologs analyzed by Mössbauer spectroscopy.. 56

Table 2-5. Metrics for Cfr limiting substrate assays using a protein-based reducing system.................................................................................................................. 63

Table 2-6. Metrics for Cfr excess substrate assays using a protein-based reducing system................................................................................................................. 68

Table 2-7. Metrics for Cfr limiting substrate assays using a dithionite reducing system...................................................................................................................... 72

Table 2-8. Metrics for high temperature Cfr limiting substrate assays using a dithionite reducing system................................................................................... 73

Table 2-9. Metrics for Cfr excess substrate assays using a dithionite reducing system.......................................................................................................................... 75

Table 3-1. RNA sequence and associated primers for the 155mer substrate. *Clostridioides difficile* Cfr protein and nucleotide sequences and PCR primers for truncation variants......................................................... 129

Table 3-2. Liquid chromatography elution gradient for analysis of Cfr reaction products.................................................................................................................. 106

Table 3-3. Metrics for *Clostridioides difficile* Cfr variant limiting substrate assays using a protein-based reducing system................................................................. 113

Table 3-4. Metrics for *Clostridioides difficile* Cfr variant excess substrate assays using a protein-based reducing system........................................................................ 114

Table 3-5. Metrics for *Clostridioides difficile* Cfr variant limiting substrate assays using dithionite.................................................................................................................. 117

Table 4-1. KTR and Lsa protein, nucleotide, and pET26b PCR primer sequences. Cfr, KTR, and Lsa pBAD24 PCR primer sequences............................................................. 180
Table 4-2. Comparison of KTR proteins from sequence similarity network clusters 1-3 highlighting host organism, gut microbiome habitat, and pathogenicity....................... 183

Table 4-3. Properties of purified stand-alone KTR proteins.......................................... 172
ACKNOWLEDGEMENTS

I want to thank my family for all of their support during my time at Penn State. Through the good times and the bad, you gave me both the praise and the encouragement needed to persevere. To my parents, you were the first to ask about what was going on in lab each week and take an interest in the growth of my project. Your curiosity and guidance provided some much needed outside perspective to my life. To my sister, thank you for always being there for me, though I am slightly jealous that I am the second advanced degree in the family! To my desk buddies Pedro and Hayley, thank you for keeping me sane! From lunches to afternoon coffees to weekend adventures, your friendships really meant a lot to me and gave me some stability during the tough times.

With regard to my research, I want to thank my advisor Amie first and foremost. When I began my graduate experience, I was unsure of my path forward and your assistance really helped me to hit the ground running. Without your advice, I am not sure how far this difficult project would have gotten! I would also like to thank Carsten and Squire. Both provided necessary instrumentation and technical expertise which vastly expanded the possible areas of inquiry. I would like to thank John and Ken for allowing me the use of their time and equipment for exploration of live-cell experiments. No one in the lab, myself included, had any experience in the types of assays that you performed and your assistance allowed me to expand the project in a new and interesting direction. I want to thank Matt for his help at the beginning of the project. You were the resident expert on Cfr and your assistance was essential in imparting the literature and experimental knowledge required of this project.
Lastly, I would like to thank Jake, my partner on this journey. I am glad that I went to graduate school for so many reasons but you are the one I am most happy for. You were there for the highest highs and the lowest lows and made every step along the way that much better. I love you so much and I am so excited to spend the rest of my life with you!

This material is based upon work supported by the National Institutes of Health. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Institutes of Health.
CHAPTER 1

Introduction
Project Statement

Antibiotic resistance by pathogens is a growing threat.\(^1\) My project is focused on understanding one of the mechanisms by which a pathogen can become resistant to drugs that target the bacterial ribosome. Enzymatic methylation of C8 at adenosine 2503 (A2503) in ribosomal RNA provides resistance to six different antibiotic classes.\(^2\) The radical SAM enzyme, Cfr, catalyzes this challenging C-C bond forming reaction.\(^3\) The enzyme activity was first reported in 2000, but in vitro characterization of this enzyme, beyond study of the model enzyme from *Staphylococcus aureus* (*Sa*), has largely remained unexplored.\(^4\) Recent discovery of additional Cfr homologs from multidrug-resistant pathogens\(^5\)-\(^7\) drives my interest in understanding the biophysical and structural properties of this important enzyme. Here I examine four phylogenetically diverse Cfr homologs and their associated partner proteins. I show that Cfrs phylogenetically distant from *Sa* Cfr harbor iron-sulfur clusters and methylate A2503 at the C8 position with varying efficiency. I also describe efforts to determine which factors are necessary to confer antibiotic resistance *in vivo* and the roles of neighboring genes and accessory domains in ribosome methylation activity.
RNA modification and its importance in ribosome function

Ribonucleic acid (RNA) has many important functions in biological systems. In addition to its roles as a message (mRNA) and decoder (tRNA) during transcription and translation, RNA can also act as an enzyme. This functional diversity can be tuned and enhanced in part, by modification of the ribose and nucleobase moieties of RNA. Currently, more than 170 alterations to RNA are known in nature, most of which occur in tRNA. Modification of RNA has diverse functional consequences. Pseudouridine, for example, is an isomer of uridine in which N1 and C5 in the nucleobase become switched, which changes the nucleobase-sugar linkage from an N-C bond to a C-C bond. This alteration allows for greater rotational freedom of the formerly glycosidic bond. It also allows for additional hydrogen bonding capacity in the pyrimidine ring opposite the Watson-Crick face. This property increases the stability of RNA. 5'-end capping of mRNA by 7-methylguanosine is another common modification with significant impact on function. The methylated base protects against exonuclease digestion and aids in RNA-specific processes such as translation, splicing, and polyadenylation. As such, the presence of 7-methylguanosine is closely linked to gene expression.

A major target for RNA modification in the cell is the ribosome, which coordinates translation and catalyzes peptide bond formation during protein synthesis. The bacterial ribosome is comprised of three RNA strands, the 23S, 5S, and 16S, and a number of proteins which are together grouped into two subunits, the 50S large subunit and 30S small subunit. The 50S subunit contains the peptidyl transferase center (PTC) which is the site of peptide synthesis. Ribonucleotide modifications are numerous in the fully assembled ribosome. The ribosome from model organism E. coli contains
more than 35 ubiquitous modifications, some of which were discovered quite recently.\textsuperscript{16, 17} These modifications play important roles in maintaining proper structure and function of rRNA.\textsuperscript{18}

Some of the most well-characterized modifications are those that confer antibiotic resistance in bacteria. More than 50\% of known antibiotics target some portion of the ribosome.\textsuperscript{19} Many examples of altered ribonucleotides in the ribosome that promote resistance are known. Some of the most effective RNA modifications that affect drug resistance are those found in the PTC because the binding modes of several drug classes targeting that region overlap.\textsuperscript{20} Consequently, a single modified ribonucleotide could confer resistance to multiple pharmaceutical agents. This is the case for Cfr, a member of the radical SAM superfamily, which catalyzes the C8 methylation of adenosine 2503 in 23S rRNA. This one modification confers resistance to five classes of antibiotics.

**Radical SAM enzymes**

Like Cfr, several functional classes of nucleotide-modifying enzymes are members of the radical S-adenosylmethionine (RS) superfamily. As of October 2019, this group contains >583,000 unique sequences, making it one of the largest protein superfamilies identified to date.\textsuperscript{21} Nearly all RS enzymes contain a CX\(_3\)CX\(_2\)C motif, which is used to bind a [4Fe-4S]\(^{2+}\) cluster (Figure 1-1, panel A). As opposed to many other [4Fe-4S]\(^{2+}\) cluster binding proteins, RS enzymes use protein side chains to ligate three of the four iron sites. The fourth open coordination site binds a molecule of S-adenosylmethionine (SAM) in a bidentate fashion by interacting with the amine and carboxylate of the methionine moiety.\textsuperscript{22} This binding mode places the sulfonium ion
close to the open iron site allowing for facile electron donation into the sulfur-ribose bond. Reductive cleavage of this bond yields methionine and a 5’-deoxyadenosyl radical (5’-dA•) (Figure 1-1, panel B). This radical has recently been detected by EPR in two different radical SAM enzymes, pyruvate formate-lyase activating enzyme (PFL-AE) and in tyrosine lyase (HydG), via either cryophotolysis or use of a non-native substrate.23, 24 Other studies have provided evidence for accumulation of organometallic intermediate Ω, a complex in which C5 of the ribose moiety of the 5’-dA• transiently forms an iron-carbon bond with the [4Fe-4S] cluster. This structure resembles the cobalt-carbon bond found in adenosylcobalamin-dependent enzymes.25, 26 The 5’-dA• is a potent oxidant that allows for a diverse array of reaction outcomes, nearly all of which begin with substrate hydrogen atom abstraction. This mode of S-adenosylmethionine usage contrasts with that of other SAM-dependent enzymes, such as SAM-dependent methyltransferases. The latter systems typically use SAM as a direct methyl donor through SN2 nucleophilic attack upon the pendant methyl group (Figure 1-1, panel B).27
Several different classes of radical SAM enzymes modify nucleotides with distinct mechanisms and outcomes (Figure 1-2). MiaB and YqeV are methylthiotransferases that functionalize tRNA bases at $sp^2$-hybridized C-H bonds with a $-SCH_3$ group. MiaB catalyzes the conversion of N$^6$-isopentenyladenosine ($i^6$A) to 2-methylthio-N$^6$-isopentenyladenosine (ms$^2$i$^6$A) at the A37 position of *E. coli* tRNAs.$^{29}$ YqeV also modifies A37 in *E. coli* tRNAs but does so by converting N6-threonylcarbamoyladenosine (t6A) to 2-methylthio-N6-threonylcarbamoyladenosine (ms2t6A).$^{30}$ These modifications are thought to improve codon-anticodon pairing and
reduce frame-shift errors.\textsuperscript{31-34} MoaA, QueE, TYW1, and viperin are RS enzymes involved in ring formation, contraction, or modification reactions on nucleotide or nucleobase substrates. MoaA converts guanosine-5'-triphosphate to pterin, an intermediate in the synthesis of redox cofactor molybdopterin, through a complex ring formation mechanism.\textsuperscript{35, 36} QueE catalyzes the formation of 7-carboxy-7-deazaguanine from 6-carboxy-5,6,7,8-tetrahydropterin, a precursor to the unique tRNA residue queuosine, through ring contraction and concurrent loss of ammonia.\textsuperscript{37, 38} TYW1 performs the condensation of N-methylguanosine and pyruvate to form 4-demethylwyosine.\textsuperscript{39-41} 4-demethylwyosine is an intermediate compound on a pathway leading to the wybutosine derivative of residue 37 in archael tRNA\textsuperscript{Phe}, a modification that is proposed to stabilize codon-anticodon interactions.\textsuperscript{42} Viperin converts CTP to ddhCTP, a known antiviral ribonucleotide, proposed to occur through a combination of C4' hydrogen abstraction and subsequent dehydration.\textsuperscript{43}

RS enzymes can also modify damaged bases in nucleic acids by catalyzing C-C bond cleavage reactions. Spore photoproduct lyase, for example, resolves a problematic thymine dimer, 5-thyminyl-5,6-dihydrothymine, in bacterial endospores by cleaving the C-C bond linking the two base moieties.\textsuperscript{44} Lastly, C-C bond-forming reactions can be performed on nucleotides by radical SAM enzymes. ELP3 converts UTP to mcm\textsuperscript{5}UTP in the wobble position on some tRNAs.\textsuperscript{45, 46} This modification prevents base pairing to incorrect nucleotide partners and guards against mistranslation.\textsuperscript{45, 47} The modifications to nucleotides such as those above are complex in structure, and while present in many biological systems, are not as common as simple methylation. Of the 172 known RNA modifications, 72 are methyl group additions.\textsuperscript{10} These reactions are catalyzed by a variety
of enzymes depending on the target site and chemistry required. Some modifications, such as the methylation of the N6 position of adenosine, can be accomplished by direct nucleophilic attack. Others, such as the methylation of the unactivated C8 site of the same nucleobase, require more complex methods of methyl group transfer.
Figure 1-2. Examples of nucleic acid modification by radical SAM enzymes. Four types of modifications are illustrated including thiomethylation, ring modifications, carbon-carbon bond breakage, and carbon-carbon bond formation.
Radical SAM methylases

The most common mechanism of RNA methylation is nucleophilic SN2 attack of the methyl group of SAM. The positively charged sulfur acts as an electron-withdrawing group, which renders the methyl carbon a good electrophile. This type of SAM-dependent MTase activity can be observed at nucleophilic sites of nucleic acid base and sugar moieties such as the N6 position of adenine or the 2’-OH of the ribose sugar in RNA. Methyl modification of these positions in the ribosome can confer antibiotic resistance in bacteria. ErmC methylates N6 of A2058 (E. coli numbering) of 23S rRNA. This modification causes resistance to erythromycin, macrolides, lincosamide, and streptogramin B.48 The thiostrepton resistance RNA methyltransferase (Tsr) methylates the 2’-O of A1067 (Streptomyces azureus numbering) in thiostrepton producer S. azureus. This activity confers protection against the thiostrepton natural product.56, 57 The nucleophilic mechanism of SAM-dependent MTases cannot target unactivated sites, however.

Radical SAM methylases combine the SAM-dependent activities of RS enzymes and MTases to modify additional sites by using the oxidative power of the 5’-dA radical. RS methylases are divided into four classes (A-D) based on their protein folds, reaction mechanism, and cofactor(s).58 Class A enzymes are implicated in the methylation of $sp^2$ carbon centers in rRNA and tRNA.59 They require an additional two conserved cysteine residues in addition to those present in the canonical radical SAM iron-sulfur cluster binding motif, one of which is used as a nucleophile to obtain a methyl group from SAM.60 Class B enzymes, currently representing the largest class of RS methylases, are able to methylate $sp^2$ and $sp^3$ carbon centers and phosphonate phosphorous atoms on
small molecules.\textsuperscript{61-63} These enzymes contain both the typical radical SAM domain and a unique N-terminal cobalamin-binding domain. The methyl group originates on SAM and is transferred to cobalamin by SN2 nucleophilic attack to form methylcobalamin, the source of the methyl group during catalysis.\textsuperscript{64} Class C radical SAM methylases are able to add a methyl group to $sp^2$ hybridized carbons without the presence of any additional cysteine residues beyond those in the radical SAM consensus sequence and are typically involved in the biosynthesis of complex metabolites.\textsuperscript{65-67} Class C enzymes are proposed to bind two molecules of SAM simultaneously, one to form the 5'-dA radical and the other to provide the methyl group. This binding mode is unlike class A enzymes which use a single SAM binding site to bind two equivalents sequentially during turnover.\textsuperscript{68} Class D is a relatively new class of RS methylase represented by the enzyme MJ0619,\textsuperscript{69} which is thought to use $N^5, N^{10}$-methyltetrahydrofolate as a methyl source. The enzyme performs the conversion of 6-hydroxymethyl-dihydropterin to 7-methylpterin.

Class A RS enzymes are perhaps the most well-characterized of these methylases. There are only two known members of class A, RlmN and Cfr. Each enzyme methylates a different position, C2 (RlmN) or C8 (Cfr), of adenosine 2503 (\textit{E. coli} numbering) in domain V of 23S rRNA, respectively (Figure 1-3).\textsuperscript{3,70} Cfr has also been shown to methylate C2 \textit{in vitro} but only after methylation of C8. Cfr reacts solely with rRNA, but RlmN can also modify A37 in some tRNAs.\textsuperscript{59} These transformations are chemically challenging transformations due to the high bond dissociation energy of the C2/C8-H bond ( >100 kcal/mol).\textsuperscript{71}
Figure 1-3. (A) The structure of domain V of *E. coli* 23S rRNA with secondary structures and modifications indicated. The minimal substrate required for RlmN activity (7mer) and the substrate (155mer) used for *in vitro* Cfr activity assays in this dissertation are highlighted in black outline. Figure adapted from a published *Escherichia coli* 23S rRNA secondary structure diagram.72 (B) Methylation of A2503 performed by class A RS enzymes Cfr and RlmN. Cfr can also perform the C2 methylation after methylating C8 preferentially.
RlmN is widespread amongst bacteria, with more than 5800 annotated homologs, and is thought to play a role in maintaining translational fidelity.\textsuperscript{73} but the enzyme is non-essential in \textit{E. coli}.\textsuperscript{74} Cfr homologs are fewer in number,\textsuperscript{74} but are found in environmental and clinical samples tested for antibiotic resistance, such as bacterial samples obtained from infected and drug-resistant farm animals.\textsuperscript{75} The \textit{cfr} gene was first identified in animal-borne \textit{Staphylococcus sciuri} in 2000 and was found to be a chloramphenicol/florfenicol resistance gene.\textsuperscript{4} Soon after, \textit{cfr} was shown to confer resistance not just to phenicols but also to lincosamides, oxazolidinones, pleuromutilins, streptogramin As, and some macrolides (PhLOPS\textsubscript{A}, Figure 1-4).\textsuperscript{2} These drugs have been shown to bind near the exit tunnel region of the peptidyl transferase center, where A2503 is located. In 2009, the cause of resistance, methylation of A2503 in the rRNA, was determined through primer extension stop and MALDI-TOF mass spectrometry.\textsuperscript{3} The \textit{cfr} gene product does not have a significant fitness cost and is often found on mobile genetic elements, making the study of this enzyme of increasing importance.\textsuperscript{76} While \textit{cfr} was originally thought to be an evolutionary descendent of \textit{rlmN}, a recent directed evolution study suggests that \textit{E. coli rlmN} cannot be used as a starting point to evolve \textit{cfr}-like activity.\textsuperscript{77}
An early phylogenetic analysis of Cfr homologs suggested that there are at least three significant groups of prokaryotic enzymes derived from Bacilli/Staphylococci/Enterococci, Paenibacilli, and Clostridia.\textsuperscript{78} When homologs from each clade were tested for activity by primer extension, all but the clostridial Cfr representative, \textit{Clostridium sporogenes}, were able to perform methylation of A2503 in \textit{E. coli} 23S rRNA. Since that initial study, characterization of clostridial Cfrs, in particular, has produced inconsistent results. The Cfr from \textit{Clostridium phytofermentans} demonstrated no ability to perform methylation when purified and tested with a 179mer RNA substrate nor was the protein able to methylate 23S rRNA when expressed in \textit{E. coli}.\textsuperscript{77} In contrast, a Cfr from \textit{Clostridioides difficile} str. 11140508, designated Cfr(B),
was shown to both methylate the *E. coli* A2503 site by primer extension and elevate the MICs (minimum inhibitory concentration) of several Cfr-sensitive drugs in *E. coli* AS19 and *S. aureus* RN4220.\textsuperscript{5, 79} It should be noted that this gene is much closer in sequence to the Cfrs from the bacilli clade than other representatives of the clostridial clade which may explain its activity. A Cfr from *Clostridium boltiae* str. 90B3, named Cfr(C), demonstrated the ability to elevate the MICs of several Cfr-sensitive drugs in various clostridial organisms.\textsuperscript{6} This Cfr homolog differed from the others in that it contained a C-terminal extension, though removal of the accessory domain did not appear to affect the activity. Another *Clostridioides difficile* homolog, Cfr(E), was shown to methylate the A2503 site of a synthetic 179mer *E. coli* RNA substrate and elevate the MICs of multiple ribosome-targeting antibiotics in clostridial clinical isolates.\textsuperscript{7} The sequence of this homolog was significantly different from both Cfr(C) and the Cfrs from other clades. The potential clinical relevance of the clostridial Cfr clade warrants further investigation into the rRNA methylation activity and structure of these enzymes.

**Mechanistic characterization of RlmN and Cfr**

The challenging chemistry required to methylate C2 and C8 of A2503, the targets of RlmN\textsuperscript{70} and Cfr\textsuperscript{3}, inspired work to explore the mechanism of C-methylation by these enzymes. An initial study verified the predicted activities with purified proteins and *in vitro* transcribed rRNA fragments and probed protein interaction with substrates of various size.\textsuperscript{80} The fully assembled ribosome (70S), individual subunits (50S, 30S), and protein-free RNA strands (23S, 16S) were all tested for methylation by RlmN and Cfr. Of those options, the 23S rRNA strand was the only substrate modified by each enzyme.
These results suggested that class A enzymes act on the 23S RNA prior to folding and assembly into the protein-bound 50S subunit. Analysis of methylation of 23S fragments showed that helices 90-92 of domain V of the 23S rRNA provide optimal activity, although a 7mer containing A2503 can also be modified to a small extent (Figure 1-3). A different study determined the residues required for class A activity.\(^{60}\) MIC studies with florfenicol and tiamulin and primer extension assays showed that the [4Fe-4S] cluster coordinating residues Cys\(_{112}\), Cys\(_{116}\), and Cys\(_{119}\) of Cfr are all necessary for activity. Two additional conserved cysteine residues, Cys\(_{105}\) and Cys\(_{338}\), are also required. Mutation of these residues led to increased drug susceptibility in \textit{E. coli} AS19 and a lack of a reverse transcriptase stop site at A2503 indicating a lack of C8 methylation.

Insight into the mechanism of C2 methylation by RlmN\(^{71}\) was obtained from activity assays performed using an rRNA substrate in which all adenosines were labeled with deuterium at the C2 position. Product analysis by mass spectrometry indicated that the hydrogen atom initially present at the C2 position transferred to the newly appended methyl group during the reaction. Additional experiments utilizing deuterium-labeled SAM showed transfer of two deuterons to the resulting methyl group and one deuteron to the 5’-dA coproduct. An independent study by another group also showed the presence of two deuterons in the methyl group of the product when labeled SAM was used.\(^{51}\) This result, however, was contingent upon purification of RlmN/Cfr from \textit{E. coli} given L-methionine-(methyl-d\(^3\)), with the assumption that the deuterated methionine was used by the cells to synthesize S-adenosyl-L-methionine-(methyl-d\(^3\)). If unlabeled methionine was used instead in the growth medium, a single turnover of purified RlmN/Cfr with deuterium-labeled SAM would give unlabeled product. These results were most
consistent with initial SAM-dependent methylation of a protein side chain in RlmN or Cfr. This residue serves as a methyl carrier for the C2 or C8 methyl group that winds up on A2503 as the product of the reaction. Trypsin digests of the proteins and subsequent mass spectrometry analysis revealed that RlmN residue Cys\text{355} (Cfr residue Cys\text{338}) was the site of methylation. Also of note, assays of RlmN isolated with a deuterated methyl group at Cys\text{355} resulted in deuterium labeling of 5'-dA, a result consistent with 5'-dA•-mediated hydrogen abstraction from the Cys\text{355} methyl group (Figure 1-5).

X-ray structures of \textit{E. coli} RlmN, reported with and without SAM bound,\textsuperscript{28} support many features of the proposed mechanism (Figure 1-6). The structures show that RlmN has a partial α/β barrel that is characteristic of radical SAM enzymes. The structure is most similar to pyruvate formate-lyase activating enzyme (PFL-AE), another RS enzyme that modifies a macromolecular substrate, a protein in the case of PFL-AE. When SAM is present, a loop containing methylated Cys\text{355} becomes ordered. The
corresponding electron density at that site supports the assignment of a methyl substituent to the residue. The structure fails to reveal a second SAM binding site, supporting the idea that class A RS methylases use the SAM coordinated to the [4Fe-4S] cluster\(^5\) to transfer the methyl group to Cys\(^355\). After methylation, a second SAM equivalent binds to the same site and performs the radical chemistry. In the structure of RlmN with SAM, the Cys\(^355\) and the SAM methyl group are approximately 3.6 Å apart, a juxtaposition that likely represents their orientation during methyl transfer.\(^5\) Use of this binding site, with SAM coordinated to the [4Fe-4S] cluster, is consistent with the finding that RlmN expressed without iron present or with the coordinating cysteines mutated does not contain the methylated cysteine residue when analyzed by mass spectrometry. Iron-sulfur reconstitution of the apo protein yields methylated and active protein.
After activation of a Cys-appended methyl substituent via H-atom abstraction by RlmN and Cfr, the next step in the reaction involves attack of the resulting methylene radical upon the RNA substrate. This step yields a covalent protein-RNA crosslink with an unpaired electron localized to the adenine base. This intermediate state was detected by freeze-quench continuous wave X-band EPR analysis Cfr under native reaction conditions (155mer RNA substrate, SAM, dithionite). The environment of this radical was probed by using protein overexpressed with labeled S-adenosyl-L-methionine-(methyl-$^{13}$C). The resulting Cfr sample contains a $^{13}$C-labeled methyl group at Cys$_{338}$. EPR analysis of $^{13}$C-labeled protein in the freeze-quenched intermediate state results in splitting of the spectrum, indicating that the protein methyl carbon is coupled to the
unpaired electron. Additionally, this study showed that the rate constants of radical formation and decay were on the same timescale as methylated product formation. This result shows that the cross-linked radical intermediate is likely on the reaction pathway.

The same radical species was not detected in wt RlmN under the same conditions. However, because the final conserved Cys residue, Cys$_{118}$, was proposed to be involved in resolution of the covalent linkage, a Cys$_{118}$Ala RlmN variant was produced to detect the cross-linked intermediate in this system.$^{81}$ Analysis of overexpressed Cys$_{118}$Ala RlmN variants revealed two pools of protein, one of which elutes from ion-exchange columns at higher salt concentrations. This second pool also exhibits a higher molecular weight band by SDS-PAGE and an increased 260 nm absorbance, all of which are consistent with persistent RNA crosslink. Additionally, mass spectrometry of in vitro trypsin and RNase digests of RlmN under reaction conditions revealed the presence of a cysteine with a methylene-linked adenosine. All of these findings were consistent with a second independent study validating formation of covalent protein-RNA adduct using the Cys$_{118}$Ala RlmN variant.$^{52}$ As with the Cfr analysis, a radical intermediate was detected by EPR analysis and the use of $^{13}$C-labeled methyl-Cys$_{355}$ revealed an observable difference in the spectrum, consistent with formation of a bond between the mCys side-chain and the nucleobase-centered radical. Additionally, lack of decay in the radical species when generated with a Cys$_{118}$Ala or Cys$_{105}$Ala variant (in RlmN or Cfr, respectively) shows that the last conserved Cys residue plays a critical role in resolving the crosslink.

X-ray crystallographic characterization of the as-isolated covalent protein-RNA crosslinked protein pool for Cys$_{118}$Ala RlmN revealed that this variant becomes linked to
a tRNA (Figure 1-7, panel A).\textsuperscript{82} RlmN is a dual-specificity enzyme, targeting both rRNA and some tRNAs.\textsuperscript{59} \textit{In vitro} transcribed tRNA\textsuperscript{Glu} with purified RlmN Cys\textsubscript{118}Ala was used to capture a structure of the radical SAM enzyme in the crosslinked intermediate state. The density about Cys\textsubscript{355} and the target adenosine clearly indicated a covalent bond between the protein and substrate (Figure 1-7, panel B). The structure supplied evidence for other steps of the mechanism as well. Prior to forming the covalent linkage, a hydrogen atom was proposed to be abstracted from the methylated cysteine. The structure showed that the 5'-dA coproduct is 4.1 Å from Cys\textsubscript{355} which is sufficient for the predicted hydrogen atom abstraction. Additionally, to resolve the crosslink, a proton was theorized to be removed from C2 by Cys\textsubscript{118}. The structure demonstrated that the side chain of Ala\textsubscript{118} is positioned to directly face C2 and that the distance between the two sites, 3.7 Å, is sufficient for deprotonation. Cys\textsubscript{118} therefore acts as a general base by removing a proton from the C2 (or C8 in Cfr) position of the crosslink. This step ultimately induces cleavage of the C-S bond between the protein and product to form a thiyl radical at Cys\textsubscript{355}, which must be reduced prior to subsequent turnover.

Beyond mechanism, the RlmN-tRNA structure also revealed information about the binding mode of RNA to class A RS enzymes. Interestingly, the target adenosine in tRNA adopts an unorthodox base-flipped conformation to allow access to the C2 position in the RlmN active site. Additionally, the structure revealed that the RNA makes multiple contacts with the protein both proximal and distal to the active site. Most of these contacts are sequence-independent, occurring between the protein and the RNA backbone. This binding mode may help rationalize the dual specificity of RlmN.
It was initially assumed that Cys_{105} in Cfr would, like Cys_{118} in RlmN, act as a general base to resolve the RNA crosslink. However, a recent study provides a more nuanced view of the mechanistic differences between these two enzymes, highlighting the need for similarly detailed structural and biochemical studies of different Cfr enzymes. Cfr isolated from cells given L-methionine-(methyl-d³) and reacted with a 7mer produced multiple isotopic products suggesting that C8 methylation utilized a polyprotic general base. To determine if Cys_{105} acts as the general base, reactivity with a 155mer, a
16mer, and a 7mer was probed in the presence of D$_2$O. The 155mer showed only very limited deuterium incorporation into the product, consistent with reduced solvent exchange with this complex. Utilization of the smaller RNA substrates allowed up to two deuterons to be exchanged into the methylated product. These results indicate a small reservoir of solvent molecules in the active site that can rapidly exchange with the bulk with a small RNA substrate but not with a large one. This effect does not occur when assaying RlmN. Consequently, the active site of Cfr was proposed to differ from RlmN by positioning a water molecule proximal to Cys$^{105}$, which would facilitate limited solvent exchange during a proposed imine–enamine tautomerization step prior to product release. This timing would rationalize the appearance of more than one deuteron in the smaller substrate. Although an x-ray structure of *S. aureus* Cfr has not yet been solved, the structure of the RlmN-tRNA complex shows several water molecules near the side chain of residue 118 (Figure 1-8). Cfr enzymes are theorized to exhibit increased flexibility in their active sites, which would explain their ability to target two different positions on A2503. Such a phenomenon could affect water dynamics in the active site, explaining the solvent exchange differences between the two enzymes.
Figure 1-8. Crystal structure of RlmN crosslinked to tRNA\textsuperscript{Glu} showing the distance between Ala\textsubscript{118} and the nearest water molecules which may be involved in proton exchange in related protein Cfr. Nearby cofactors, 5'\textsuperscript{-}dA (green) and the [4Fe-4S] cluster, and important residues, mCys\textsubscript{355} (burgundy) and tRNA A37 (teal), are highlighted.

Integration of these findings has led to a proposed detailed mechanism for RlmN and Cfr-mediated RNA methylation (Figure 1-5).\textsuperscript{55} In Cfr, the reaction begins with SN2 attack of the SAM methyl group by Cys\textsubscript{338} producing methylated protein and SAH. Upon RNA binding, a second molecule of SAM is cleaved reductively to produce methionine and 5'\textsuperscript{-}dA\textsuperscript{-}. The 5'\textsuperscript{-}dA\textsuperscript{-} abstracts a hydrogen from the Cys\textsubscript{338}-methyl group, which then crosslinks to the RNA substrate at C8 forming a covalent protein-substrate intermediate. The crosslink resolves via intervention of Cys\textsubscript{105} as a general base to deprotonate C8. The currently accepted mechanism results in a thyl radical on Cys\textsubscript{338} which requires an exogenous electron donor to resolve. However, the thyl radical has not been observed directly and the source of the extra reducing equivalent is unknown.
Project Objectives

In this dissertation, I aim to expand upon *Staphylococcus aureus* Cfr characterization by analyzing the activity and biophysical properties of phylogenetically distinct Cfrs. In Chapter 2 I show that four distinct Cfr homologs from *Bacillus amyloliquefaciens*, *Enterococcus faecalis*, *Paenibacillus lautus*, and *Clostridoides difficile* are active C8 A2503 methylases. All homologs bind a [4Fe-4S] radical SAM cluster, but an additional unanticipated C-terminal iron-binding site was identified in the clostridial Cfr. Chapter 3 explores the properties and possible functions of auxiliary metal binding site in this enzyme via characterization of truncation variants. The C-terminal domain coordinates a single Fe$^{2+}$ ion in a rubredoxin-type binding site. Truncation of the domain had little impact on activity, showing that it is not essential for methylation activity. However, substitution of the C-terminal domain with zinc diminished activity, suggesting an important role for iron when the domain is present. Bioinformatic analysis of the novel rubredoxin domain showed homology to a large family of stand-alone proteins of unknown function called Cys-rich KTRs. In chapter 4, I analyze several members of the stand-alone Cys-rich KTR family and explore their potential link to Cfr activity. A genome neighborhood analysis of Cys-rich KTRs uncovered a previously unknown set of sequence divergent Cfr homologs in clostridial gut microbiome organisms. The possible connection of these proteins to human health warrants further investigation into their activity. A non-Cfr associated KTR from *Staphylococcus aureus* was found to bind iron, suggesting that all KTRs might share this property. As a result, this group of proteins could represent a new family of metalloproteins. Genome neighbor analysis of Cfrs from the clostridial clade that are not KTR-associated showed that a Lsa
ABC-F type ribosomal protection protein is always present when the KTRs are absent. I speculate that these two types of proteins may act synergistically with clostridial Cfrs to optimize or confer activity in some way inside the cell. I initiated antibiotic resistance studies in E. coli to address this question, but my initial results are inconclusive, suggesting that an organism-specific approach may be required as a next step.
References


[41] Young, A. P., and Bandarian, V. (2011) Pyruvate is the source of the two carbons that are required for formation of the imidazoline ring of 4-demethylwyosine, *Biochemistry* 50, 10573-10575.


CHAPTER 2

Phylogenetic and biochemical analyses of four radical S-adenosylmethionine Cfr homologs of diverse origin

Abstract

Cfr is a radical S-adenosylmethionine (SAM) RNA methylase linked to multi-drug antibiotic resistance in bacterial pathogens. It catalyzes a chemically challenging C-C-bond forming reaction to methylate C8 of A2503 [Escherichia coli (Ec) numbering] of 23S ribosomal RNA during ribosome assembly. The cfr gene has been identified as a mobile genetic element in diverse bacteria and in the genome of select Bacillales and Clostridiales species. Despite the importance of Cfr, few representatives have been purified and characterized in vitro. Here I show that Cfr homologs from Bacillus amyloliquefaciens, Enterococcus faecalis, Paenibacillus lautus, and Clostridioides difficile contain the [4Fe-4S] radical SAM cluster and act as C8 adenine RNA methylases in biochemical assays. Additionally, I demonstrate that these Cfrs are sensitive to the identity of the reductant and promote different product outcomes when exposed to a chemical or protein-based reductant.
Introduction

Radical S-adenosylmethionine (SAM) enzymes compose one of the largest enzyme superfamilies defined to date, with more than 538,000 members.1 Radical SAM enzymes are essential in primary anaerobic metabolism and in the synthesis of diverse secondary metabolites.2,3 Nucleic acid modification is a third central function of radical SAM proteins, and this activity has been demonstrated in all three domains of life.4,5 A pair of adenine RNA methylases, RlmN and Cfr, are among the most widespread and well-characterized radical SAM nucleic acid modification enzymes in bacteria.6-8 Although originally discovered to methylate C2 of A2503 in ribosomal RNA (rRNA), a base located in the peptide exit tunnel of the peptidyltransferase center of the bacterial ribosome, RlmN also modifies six additional tRNA substrates at A37.9,10 In Escherichia coli, RlmN is non-essential, but its activity improves translational fidelity and organism fitness.10 More than 20,000 RlmN homologs have been identified in the sequenced genomes of prokaryotes of diverse origin.1 Related Cfr enzymes are smaller in number, with approximately 200 homologs annotated in Interpro and other protein sequence databases. Cfrs were initially found encoded as mobile genetic elements in staphylococcal pathogens,9 and Staphylococcus aureus (Sa) Cfr has been characterized extensively in vitro. The enzyme methylates C8 of A2503 in 23S rRNA, and this modification confers resistance to multiple classes of antibiotics.11 Although quite relevant to human health, few of the other Cfrs have been characterized in vitro, and the structure of the enzyme is not known.

The common methyl donor SAM is the source of the newly appended methyl carbon in RlmN and Cfr.6 In their function as C-methylases, RlmN and Cfr perform a
chemically challenging C-C bond forming reaction involving functionalization of inert $sp^2$-hybridized C-H bonds.\textsuperscript{12} This activity cannot be accomplished by the standard polar mechanism characteristic of SAM-dependent methyltransferases. Instead, radical chemistry is required and enabled here by a second SAM-dependent activity in RlmN and Cfr – transient formation of a potently oxidizing 5'-deoxyadenosyl radical (5'-dA•).\textsuperscript{6, 7, 13} Nearly all enzymes in the radical SAM superfamily share this intermediate,\textsuperscript{14, 15} generated by reduction of a SAM-coordinated [4Fe-4S]\textsuperscript{2+} cluster that, in turn, triggers reductive cleavage of the coordinated cosubstrate. In RlmN and Cfr, the 5'-dA• abstracts a hydrogen atom from a Cys-appended –CH\textsubscript{3} group (mCys\textsubscript{338}, $Sa$ Cfr numbering) to yield a methylene radical (Figure 2-1).\textsuperscript{7} Radical addition to the adenine nucleobase at either C2 or C8 generates a transient protein-RNA crosslink containing an unpaired electron on the adenine nucleobase.\textsuperscript{16} Proton transfer from C2/C8 to an active site base, a second conserved Cys residue (Cys\textsubscript{105}, $Sa$ Cfr numbering), resolves the crosslink.\textsuperscript{17-19} This step is proposed to yield a thiyl radical at the C-terminal conserved catalytic residue, Cys\textsubscript{338}, which must be reduced prior to subsequent turnover.\textsuperscript{17}
Interestingly, several other redox-neutral radical SAM C-C bond forming enzymes also oxidize an active site Cys side chain\textsuperscript{12} to balance the initial two-electron reductive cleavage of SAM.\textsuperscript{12, 20} In these systems, auxiliary iron-sulfur cluster or cobalamin-binding domains are often present and proposed to provide necessary reducing equivalents.\textsuperscript{12} The RlmN and Cfr homologs characterized to date are distinct in that they do not harbor an obvious universal dedicated thiyl radical reductant.

The proposed mechanism of the RNA methylases has been validated through spectroscopic and biochemical analysis of \textit{E. coli} RlmN and \textit{S. aureus} Cfr.\textsuperscript{7, 16, 17, 21} The two essential Cys residues (Cys\textsubscript{338} and Cys\textsubscript{105}) are strictly conserved in these homologs and all others identified to date. X-ray crystal structures of \textit{E. coli} RlmN have been solved, both alone\textsuperscript{22} and in complex with tRNA,\textsuperscript{18} the latter obtained by trapping the covalent protein-RNA intermediate via substitution of Cys\textsubscript{118} with a non-functional alanine (Cys\textsubscript{105}Ala equivalent).\textsuperscript{18, 22} To date, no structures have been reported of either
enzyme in complex with a ribosomal RNA substrate. Additionally, Cfr homologs have not been structurally characterized in any form, even though such information would be useful in the design of inhibitors that could rescue the activities of existing antibiotics to which Cfr currently confers bacterial resistance.

To identify Cfr candidates suitable for ongoing structural studies, I performed phylogenetic analyses of Sa Cfr homologs identified in database searches. As reported previously, at least four distinct groups of enzymes emerge, including a clostridial clade classified as Cfr-like. Cfr-like homologs from *Clostridium sporogenes*<sup>23</sup> (Cs) and *Clostridium phytofermentas*<sup>24</sup> (Cp) have been proposed to harbor a different, still undefined, activity because they fail to confer antibiotic resistance and lack C8 methylation activity in primer extension assays when overexpressed in *E. coli*.<sup>23, 24</sup>
<table>
<thead>
<tr>
<th>Cfr Proteins</th>
<th>CD65Q42</th>
<th>C. difficile str. 11140508</th>
<th>Cs</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cfr(C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cfr(C25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2-2.** Comparison of previously characterized clostridial Cfrs. The clostridial Cfr characterized in this work, *Clostridioides difficile* (*Cd*) QCD65Q42, was previously shown to confer a Cfr-like antibiotic resistance profile in a *C. difficile* laboratory strain and annotated as Cfr(C25) in that work. The Cfr from *Cd* str. 11140508 was also shown to confer a similar resistance profile. The Cfr enzymes from *Cs* and *Cp* are unable to confer resistance.

Here I show that heterologously expressed and purified Cfr homologs from *Bacillus amyloliquefaciens* (*Ba*), *Enterococcus faecalis* (*Ef*), *Paenibacillus lautus* (*Pl*), and *Clostridioides difficile* are capable of rRNA methylation *in vitro*. All four enzymes contain a full occupancy [4Fe-4S] cluster in their wild-type (wt) forms, as established by Mössbauer spectroscopy and other analytical techniques. All four homologs modify A2503 at C8 but with different rates. The enzymes also vary in their capacity for subsequent C2 methylation of the same base. *C. difficile* Cfr is active as a C8 methylase with limiting RNA substrate but exhibits distinctly slower methylation when challenged.
with excess RNA. None of the homologs conferred antibiotic resistance in *E. coli* BL21 DE3 ΔtolC/ΔaraBAD suggesting further optimization is required to compare *in vitro* RNA methylation to resistance phenotypes.

**Materials and Methods**

*Materials.* Kanamycin was obtained from Teknova (Hollister, CA). Ampicillin was purchased from Dot Scientific (Burton, MI). N-(2-hydroxyethyl)piperazine-N′-2-ethanesulfonic acid (HEPES), DNaseI (RNase-free), Bradford reagent, and His-Pur Co²⁺-TALON resin were obtained from Thermo Fisher Scientific (Waltham, MA). Sodium sulfide, 2-mercaptoethanol, sodium dithionite, glucose, 5′-adenosylmethionine, and p1 nuclease from *Penicillium citrinum* were purchased from Sigma Aldrich (St. Louis, MS). Iron (III) chloride hexahydrate was acquired from BDH Analytical Chemicals, VWR (Radnor, PA). DL-dithiothreitol (DTT) was obtained from Alfa Aesar (Tewksbury, MA). ⁵⁷Fe metal (98%) was purchased from Isoflex USA (San Francisco, CA). Bovine serum albumin (BSA), *XhoI*, *NdeI*, antarctic phosphatase, Q5 DNA polymerase, and T4 DNA ligase were obtained from New England Biolabs (Ipswich, MA). PD-10 and HiPrep Sephacryl 16/60 S200 columns were purchased from GE Healthcare (Little Chalfont, United Kingdom). *E. coli* BL21 DE3 ΔtolC, *E. coli* BL21 DE3 ΔtolC/ΔaraBAD, Falcon 96-well clear culture plates, lincomycin, MOPS mix, ACGU, and supplement EZ were generously provided by the laboratory of Prof. Kenneth Keiler (Pennsylvania State University).

*Phylogenetic analysis of Cfrs.* Homologs were identified by BLASTp analysis using *Sa* Cfr as the query against the non-redundant protein sequence database (NCBI, accessed March 2018) with default search parameters. The top 500 matches were aligned
in MUSCLE after consolidation with ExPASy (<98% minimum identity filter in the
decrease redundancy tool) to remove duplicate sequences. The final pool contained 361
unique sequences.\textsuperscript{27, 28} The program MEGA7\textsuperscript{29} was used to compute a phylogenetic tree
using the minimum evolution method with the JTT model of substitution, allowance of
pairwise deletions, and closest neighbor set to level = 2. Bootstrap values were computed
from 1000 replicates.

\textit{Cloning and overexpression of full-length Cfr proteins.} DNA sequences encoding
genes for the Cfrs selected for study (Table 2-1) (with the exception of \textit{Cd} Cfr) were
codon-optimized for over-expression in \textit{E. coli, synthesized, and inserted into a pMK-T}
vector (GeneArt). PCR primers (Table 2-1) were designed with Primer3 and appended at
the 5’-end with \textit{NdeI} and \textit{XhoI} restriction sites. The coding sequences were amplified by
PCR using Q5 DNA polymerase. The PCR products were digested using \textit{NdeI} and \textit{XhoI}
restriction enzymes, purified by gel electrophoresis, re-isolated using an Omega Biotek
gel extraction kit, and ligated into pET26b. The \textit{C. difficile cfr} overexpression plasmid
was constructed using the Gibson assembly method.\textsuperscript{30} The gene was codon optimized for
overexpression in \textit{E. coli} and synthesized in two pieces, each containing 20 base-pair
regions that overlap with the pET26b vector and with each other. The gene and linearized
vector were combined with a 5’-exonuclease, DNA polymerase, and DNA ligase to allow
for pET26b-\textit{cfr} assembly. All four pET26b-\textit{cfr} sequences were verified by Sanger
sequencing at the Pennsylvania State University Nucleic Acid Facility, and the coding
regions can be found in Table 2-1 of the Supporting Information. All plasmids
additionally encode a C-terminal His\textsubscript{6}-tag with a short two amino acid (-Leu-Glu-) linker.
Overexpression of all constructs was accomplished in *E. coli* BL21(DE3) cells co-transformed with pET26b-*cfr* and pDB1282, the latter a vector containing inducible iron-sulfur cluster assembly machinery. Successful transformants were selected on kanamycin (50 µg/mL) and ampicillin (100 µg/mL) supplemented LB plates. 1.5 L cultures of the overexpression strains were grown shaking (180 rpm) at 37 °C in M9 minimal medium supplemented with kanamycin (50 µg/mL), ampicillin (100 µg/mL), iron (III) chloride (30 µM), and L-cysteine (300 µM) to maximize iron-sulfur cluster incorporation. Growth for samples suitable for Mössbauer analysis utilized $^{57}$Fe dissolved in 1 M H$_2$SO$_4$. At an OD$_{600} = 0.3$, L-arabinose was added [final concentration 0.08 % (w/v)] to induce overexpression of the genes harbored on the pDB1282 plasmid. At an OD$_{600} = 0.6$, IPTG was added to a final concentration of 100 µM to induce overexpression of the pET26b-*cfr*. After incubation overnight at 18 °C at 180 rpm, cells were pelleted by centrifugation at 5000 g for 10 min at 4 °C. Approximately 3.2 g cell paste was obtained per liter of culture. Cell paste was flash-frozen in liquid nitrogen and stored at -80 °C. Flavodoxin and flavodoxin reductase were purified as described previously.

**Purification of full-length Cfrs.** All purification steps were performed anaerobically in an anoxic chamber with a mixed H$_2$/N$_2$ atmosphere (Coy Laboratory Products). Cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 300 mM KCl, 4 mM imidazole, 10 mM MgCl$_2$, 10% (v/v) glycerol, 10 mM β-mercaptoethanol) in a nickel-coated beaker containing 1 mg/mL lysozyme, 100 µg/mL DNase, and 1 mM phenylmethylsulfonyl fluoride (PMSF). The suspension was stirred at room temperature for 30 min and lysed by sonication (QSonica Q500, 10 s pulse, 30 s rest, 60% amplitude).
on ice. The resulting cell debris was pelleted by centrifugation at 11,300 g at 4 °C for 1.25 h in a sealed centrifuge bottle. The supernatant was applied to Co^{2+}-TALON resin. Protein fractions were eluted in 50 mM HEPES, pH 7.5, 300 mM KCl, 300 mM imidazole, 10 mM MgCl$_2$, 30% (v/v) glycerol, 10 mM β-mercaptoethanol. Reconstitution of [4Fe4S]$^{2+}$ clusters was performed on ice for all samples. Protein samples (200 µM) were exchanged into gel-filtration chromatography buffer [10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl$_2$, 25% (v/v) glycerol] and incubated for 30 min at room temperature. Iron (III) chloride (800 µM) was added to the solution, which was subsequently incubated for 1 h at room temperature. Sodium sulfide was added to the solution (to a final concentration of 800 µM) in six equal volumes over the course of 2 h. The solution was allowed to incubate on ice overnight. A PD-10 column was used to remove excess iron and sulfide. Size-exclusion chromatography was performed using a Sephacryl S200 column. Protein concentrations were determined by Bradford analysis with correction factors (Molecular Structural Facility at UC Davis) of 0.891 (Cd), 0.627 (Ba), 0.828 (Ef), and 0.687 (Pl). Iron-sulfur cluster incorporation was assessed by measuring the number of iron equiv per protein with the ferrozine assay.$^{33}$

*Synthesis and purification of RNA.* To generate a 155 nucleotide (nt) fragment of *Ec* 23S rRNA corresponding to A2454-G2608 (Figure 2-1), previously shown to be efficiently methylated by *Sa Cfr*, a DNA template was amplified using primers 2454 and 2608 (Table S1), and transcribed in a reaction with T7 RNA polymerase (500 µg/mL). The transcription reaction was performed in 30 mM Tris, pH 8.4, 20 mM DTT, 36 mM MgCl$_2$, 0.01% (w/v) Triton X-100, 2 mM spermidine, 2 mM NTPs (4 mM GTP). The template DNA PCR product was diluted to a final concentration of 5 ng/µL and incubated
for 3 h at 37 °C with the components above. DNaseI was added to the solution (to a final concentration of 0.005 U/µL) and incubated at 37 °C for an additional hour. EDTA (pH 8.5) was added to a final concentration of 50 mM, and the solution was frozen at -20 °C for at least 2 h. Full-length 155mer RNA transcripts were purified under anaerobic conditions (Coy Laboratory Products) by PD-10 column equilibrated in filtered nanopure water. RNA concentrations were determined by UV-visible spectrophotometry ($\varepsilon_{260} = 1.87 \mu M^{-1} \text{cm}^{-1}$). The A2503G variant RNA was prepared similarly, except the first PCR step was performed with a synthetic template (Integrated DNA Technologies, see Supporting information for sequence) containing the desired base change.

**RNA methylation activity assays with limiting substrate.** In experiments using a protein-based low-potential reducing system [flavodoxin (Flv), flavodoxin reductase (Flx), and NADPH], RNA methylation was assayed in a total volume of 150 µL at room temperature in an anoxic glovebox (Coy Laboratory Products) over a time course of 2.75 h. Samples contained 48 µM Cfr and 50 µM 155mer RNA in 100 mM sodium EPPS, pH 8.0, 10 mM MgCl$_2$, 12.5% (v/v) glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, and 2 mM NADPH. The solution was incubated at room temperature for 5 min before initiating the reaction by addition of SAM to a final concentration of 1 mM. Methylase activity was also assessed using the chemical reductant dithionite in place of Flv, Flx, and NADPH. The conditions were the same as above except SAM was added with the other components and dithionite (2 mM) was used to initiate the reaction. At designated time points, a 10 µL aliquot of the reaction solution was removed and added to 10 µL of quench solution containing 250 µM tryptophan and 95 mM H$_2$SO$_4$. Quenched reactions were diluted 1:1 in P1 nuclease buffer (250 mM sodium acetate, pH 6.0, 45 mM NaCl, 4
mM ZnCl$_2$). Antarctic phosphatase (10 U) and P1 nuclease (1 U) were added to digest the RNA into individual nucleosides. The digestion reactions were incubated at 37 °C overnight and subsequently analyzed by liquid chromatography (Agilent 1960/1990) coupled to mass spectrometry (LC-MS) (Agilent 6460 QQQ) (LC-MS). Assay mixtures were separated on an Agilent Extend-C18 column (4.6 × 50 mm) equilibrated in a 40 mM ammonium acetate, 5% (v/v) methanol (solvent A) / 100% methanol (solvent B) system using the conditions listed in Table 2-2 below.

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 – 5.0</td>
<td>95%</td>
<td>5%</td>
</tr>
<tr>
<td>5.0 – 6.5</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>6.5 – 7.0</td>
<td>73%</td>
<td>27%</td>
</tr>
<tr>
<td>7.0 – 8.0</td>
<td>47%</td>
<td>53%</td>
</tr>
<tr>
<td>8.0 – 12.0</td>
<td>95%</td>
<td>5%</td>
</tr>
</tbody>
</table>

MS detection of products and assignment of peaks were accomplished as described previously.$^{21}$ Results were analyzed with the MassHunter software package, and quantitative analyses were facilitated by comparison to a standard curve calculated from data acquired from the tryptophan internal standard. Unproductive cleavage of SAM was determined by dividing the number of SAM cleavage events, obtained from the concentration of 5′-dA in LC-MS samples, by the estimated number of methylation events ([m8A] + 2 • [m2,8A]). All assays were performed at least three times on different days. Figures show the average product concentration for three selected experiments,
with an error bar showing the standard deviation in these values. The assay metrics in Table 2-5 are the average ± standard deviation over the selected trials.

Time courses for experiments using limiting substrate for *Ba*, *Ef*, and *Pl* Cfrs were initially simulated in KinTek Explorer\textsuperscript{52} with the following model.

**Equation 2-1:** \[ E + RNA \xrightarrow{k_1} EM \xrightarrow{k_2} EP \]

In this model, enzyme (E) and substrate (RNA) bind and react to form the first product, m8A (EM). This monomethylated compound can be further converted to m2,8A (EP), as shown above. Both steps are considered to be irreversible. In initial simulations, I found that the best fit for \( k_1 \) (\( k_{obs} \) m8A) falls below a lower limit set by the steady-state turnover rate (\( v/[E] \)) determined in subsequent experiments with excess RNA. This discrepancy indicated that the simple kinetic model in equation 2-1 is not sufficient. However, in the absence of additional experimental data to constrain the rates of other steps (substrate binding, product release, etc.), I instead fixed \( k_1 \) to the lower limit stated above, obtained from fits of m8A production with excess RNA to a single exponential function (described below). I then used KinTek Explorer to simulate the reaction to obtain the apparent rate constant (\( k_2 \) or \( k_{obs} \) m2,8A) for the second methylation step. Note that in this model, the rate constant for the second dimethylation reaction includes, in addition to the catalysis step, m8A product release and substrate rebinding steps. The fits I obtained are reasonable but, given these complexities, this \( k_2 \) value should be viewed as a lower limit for the rate of the second methylation reaction. For *Ba*/Ef/*Pl* Cfr, the active enzyme fraction was allowed to vary and fell below 48 µM in the simulation. Because the limiting-substrate time course of *Cd* Cfr yields little dimethylated product, product vs.
time traces were fit to a single exponential function to extract the apparent reaction rate ($k_{\text{obs}}$ m8A) with equation 2-2, in which $a$ is the product concentration at completion and $k$ is equivalent to $k_{\text{obs}}$ for m8A formation. I assumed a fully active enzyme pool in this analysis.

**Equation 2-2:** \[ [\text{m8A}] = a \cdot (1 - e^{-kt}) \]

*RNA methylation activity assays with excess substrate.* Activity assays with excess substrate were performed as described above but with 100 µM 155mer RNA substrate and 3 µM Cfr enzyme. The reaction solutions were prepared in 150 µL total volume with 100 mM sodium EPPS, pH 8, 10 mM MgCl$_2$, 12.5 % (v/v) glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, and 2 mM NADPH. Reactions were initiated by addition of 1 mM SAM. Assays using dithionite were also performed, with SAM being added with the other components and dithionite (2mM) initiating the reaction. Time courses were analyzed via linear fit to the initial portion of the progression curve for m8A formation to obtain the initial velocities of the reactions. This approach allowed us to obtain a reaction rate, $v/[E]$, for each Cfr enzyme (Table 2-6). A single exponential fit, using equation 2-2, to the full m8A time course allowed us to obtain the amplitude of each enzymatic reaction. This value reflects the total amount of methylated product generated by each enzyme for the entire experiment. The information was used to determine a total turnover number (TON, Table 2-6), after division by the initial enzyme concentration (3 µM). The goal of the latter analysis was to ascertain whether each enzyme is capable of performing more than one full substrate conversion.

*Spectroscopic analyses of Cfr proteins.* Reconstituted $^{57}$Fe protein samples were transferred to EPR tubes or Mössbauer cups and flash-frozen in liquid nitrogen in an
anoxic chamber (Coy Laboratory Products). Mössbauer spectra were recorded on constant acceleration Mössbauer spectrometers (SEE Co., Edina, MN) equipped either with a Janis SVT-400 variable-temperature cryostat (weak-field) or a Janis 8TMOMS-OM-12SVT variable-temperature cryostat (strong-field) (Janis, Woburn, MA). All isomer shifts are quoted relative to the centroid of the spectrum of α-iron metal at room temperature. Simulations of Mössbauer spectra were carried out using WMOSS (SEE Co., Edina, MN). Some of the simulations are based on the spin Hamiltonian formalism (see equation below), in which the first term describes the electron Zeeman effect, the second and third terms describe the axial and rhombic zero-field splitting of the total electron spin ground state, the fourth term represents the interaction between the electric field gradient and the nuclear quadrupole moment, the fifth term describes the magnetic hyperfine interactions of the $^{57}$Fe nucleus with respect to the total electron spin ground state, and the last term represents the nuclear Zeeman interactions of $^{57}$Fe. All symbols are conventional.$^{34}$ Spectra were calculated in the slow relaxation limit.

$$H = \beta S \cdot g \cdot B + D \left( S_z - \frac{S(S+1)}{3} \right) + E \left( S_x^2 - S_y^2 \right)$$

$$+ \sum_i \frac{eQV_{zz,i}}{4} \left[ I_{z,i}^2 - \frac{1}{3}(I_{i+1}^2) + \frac{\eta}{3}(I_{x,i}^2 - I_{y,i}^2) \right] + \sum_i S \cdot A_i \cdot I_i - \sum_i g_n \beta_n B \cdot I_i$$

EPR samples were analyzed on a Bruker ESP 300 spectrometer (Billerica, MA) equipped with an ER 041 MR microwave bridge and an ER 4116DM resonator.

Cfrs were analyzed spectrophotometrically using an Agilent 8453 G1103A spectrometer. In an anoxic glovebox, a 1 µL sample of each purified Cfr [10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl$_2$, 25% (v/v) glycerol] was diluted to 1
mL in water and sealed in a screw cap quartz cuvette (Starna Cells, Inc). The anoxic protein solution was scanned from 190 nm to 1100 nm with a baseline correction at 800 nm.

Determination of iron content by ferrozine assay and ICP-OES. Iron content was estimated by ferrozine assay with ICP-OES (Pennsylvania State University Laboratory for Isotopes and Metals in the Environment) as a method of confirmation. Cfrs were diluted to 200 µM and a volume corresponding to 12.5 nmol of protein was transferred to a clean eppendorf. Protein was denatured by the addition of 50% TCA equal to 0.2*(volume in eppendorf). Precipitated protein was removed by centrifugation for 2 min at max speed. The supernatant was transferred to a clean eppendorf and ascorbic acid (3 mM), ferrozine (400 µM), and ammonium acetate (~4.5 M) were added. A control without protein was used to blank at 562nm on an Agilent 8453 G1103A spectrometer and the amount of iron was calculated by (25 nmol)*(Abs$_{562}$). This value was divided by the nmol of protein to start to calculate the iron per protein.

Antibiotic resistance profiling of Cfrs by microbroth dilution. To determine a baseline antibiotic resistance profile for each Cfr homolog and to determine if in vitro activity can be correlated to elevated levels of resistance, broth microdilution was performed using Ba, Ef, Pl, Cd, and Cs Cfrs expressed in Escherichia coli. The procedural details of cloning the Cfr genes into pBAD24, E. coli transformation, optimization of broth microdilution conditions, and assessment of antibiotic resistance are found in chapter 4.
Results and Discussion

Identification of Cfr homologs by phylogenetic analysis. To identify Cfrs of diverse origin for biochemical and biophysical studies, I performed BLAST searches of the NCBI non-redundant protein sequence database using Sa Cfr as a query. The resulting set of 361 representative sequences was used to construct a minimum evolution phylogenetic tree (Figure 2-3, panel A) with an unrelated radical SAM class B methylase, TsrM, as the outgroup. The initial search included two groups of annotated RlmN enzymes that share 37% sequence identity with Sa Cfr but belong to distinct sequence clusters. The more distantly branched RlmN group is composed of sequences from Firmicutes and Chloroflexi organisms. The second group includes putative RlmN sequences from Candidatus phyla and obligate intracellular parasites.
Figure 2-3. Selection of phylogenetically diverse Cfr enzymes for biochemical characterization. (A) Minimum evolution phylogenetic tree containing 361 representative sequences obtained by NCBI BLAST. The sequence of Sa Cfr was the initial query. Bootstrap values (1000 replicates) were calculated for branch support and red circles denote values > 85% in selected branches. Dominant genera and the genomic context in each cluster are indicated in colored text/schematics. (B) Cartoon diagram of the domain architecture in Cfrs selected for study. Locations of conserved Cys residues (Sa numbering) in the radical SAM domain are shown.

Annotated Cfrs cluster into a eukaryotic group and two different bacterial clades, as described previously.\textsuperscript{24,35} Plasmid-encoded Cfrs with validated C8 methylation activity from staphylococcal clinical isolates\textsuperscript{9} belong to a group that also includes genomically-encoded Bacillales sequences\textsuperscript{36} and plasmid-encoded Cfr proteins from \textit{Enterococcus faecalis} (\textit{Ef}) and other enterococcal pathogens. From the latter group, I
selected an *Ef* plasmid-borne Cfr and a genomic sequence from a mesophilic strain of *B. amyloliquefaciens (Ba)* for study here. The similarity (Table 2-3) of the *Ef* and *Ba* sequences to *Sa* Cfr suggests that these proteins are likely to share the ability to methylate rRNA at A2503 C8. In support of this hypothesis, several antibiotic-resistant enterococcal clinical isolates have recently been demonstrated to harbor a transposable element with a Cfr that is 99% sequence identical to *Ef* Cfr from strain TX0635 analyzed here.\textsuperscript{37-40} Cfr homologs from an alternative *Ba* strain (92% identity to the Cfr from DSM-7 ATCC 23350 targeted in this work), *B. clausii*, and *B. brevis* all confer multidrug antibiotic resistance when expressed heterologously in *E. coli* AS19, a strain with high antibiotic susceptibility.\textsuperscript{36} In the latter systems, the mechanism of drug resistance has been directly linked to methylation of A2503, confirmed in a primer extension analysis of total rRNA from an *rlmN* knockout of *E. coli* following induction of Bacillales Cfr expression.\textsuperscript{36} In all of these studies, conversion of A2503 to the singly- and di-methylated forms was not confirmed in vitro with purified proteins or analyzed quantitatively.

The second bacterial clade is dominated by genomic sequences from Firmicutes and partitioned into three distinct groups. One of these clusters is composed largely of sequences from the *Paenibacillus* genus.\textsuperscript{23} It includes a representative [*Paenibacillus lautus* sp. Y412MC10 (*Pl*)] originally isolated from the Obsidian hot spring at Yellowstone National Park.\textsuperscript{41} This *Pl* Cfr sequence was selected for characterization, owing to the utility of thermotolerant proteins in structural and biophysical studies. Many of the remaining sequences are found in Clostridial organisms. Some of these proteins have been annotated as Cfr-like and are postulated to be functionally divergent from *Sa*
This analysis shows that this subset breaks down into two smaller groups that consist of genomically-encoded Cfrs from a set of pathogenic *Clostridium difficile* (*Cd*) strains, now reclassified to the *Clostridioides* genus, and a larger group of Cfrs from more diverse clostridial organisms. The second group includes Cfr sequences from *C. sporogenes* and *C. phytofermentans* recently demonstrated to lack rRNA methylation activity.\textsuperscript{23, 24}

Intriguingly, five members of the *Cd* Cfr cluster have an extended C-terminal motif (Figure 2-3, panel B). A BLAST search of the corresponding amino acid sequence shows similarity to an InterPro family (IPR025957)\textsuperscript{1} designated “cysteine-rich KTR” with >2000 members identified as of November 2019. The group was initially named for a common internal Arg-Thr-Lys sequence, although among the current set of IPR025957 sequences this motif is not universally conserved. Members are most often found as stand-alone proteins approximately 60 amino acids in length, with two additional strictly conserved CXXC motifs. To date, none of these proteins has been functionally characterized in purified form. Cys-rich KTRs are found almost exclusively in Firmicutes, most in the order Clostridiales (968 annotated genes). A Cfr-KTR fusion sequence from *Clostridioides difficile* str. QCD-63q42\textsuperscript{42} was selected for *in vitro* characterization in this study. In evaluating the activity and properties of this fusion enzyme, I sought to understand whether it lacks known Cfr activity upon heterologous production in *E. coli*, as shown for the *C. sporogenes* and *C. phytofermentans* homologs. In work by other laboratories, a *C. difficile* Cfr homolog has been shown to confer antibiotic resistance via A2503 modification, but the activity was promoted by a *cfr* on a mobile genetic element that shares only 52% sequence identity with my chromosomally-
encoded Cd homolog.\textsuperscript{26, 43} A Clostridium boltae Cfr-KTR fusion has also been shown to confer moderate drug resistance in cell-based assays.\textsuperscript{25} Neither of these proteins has been purified and characterized \textit{in vitro}, however.

<table>
<thead>
<tr>
<th>Table 2-3. Sequence identity for Cfr homologs tested for RNA methylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homolog</td>
</tr>
<tr>
<td>% ID (Sa Cfr)</td>
</tr>
</tbody>
</table>

\textit{Spectroscopic characterization of metal binding in Cfr homologs.} To ensure that all Cfr homologs assemble full occupancy [4Fe-4S]\textsuperscript{2+} cofactors, UV-visible absorption spectroscopy, Mössbauer spectroscopy, and ferrozine assays were performed. The iron-sulfur cluster in each sample is likely to remain in its as-isolated [4Fe-4S]\textsuperscript{2+} state during all experiments. The only reductant present is DTT. The UV-vis spectrum for each homolog showed the characteristic peaks of an [4Fe-4S]\textsuperscript{2+} cluster at 320 nm and 420 nm, with an additional trailing shoulder around 500 nm (Figure 2-4, panel A). Protein concentrations differ between samples, giving slight differences in peak heights.
Figure 2-4. (A) UV-visible spectra of reconstituted Ba (blue), Ef (red), Pl (green), and Cd (black) Cfrs. Peaks at 320 nm and 420 nm with a trailing shoulder are characteristic of a [4Fe-4S]$^{2+}$ cluster. (B) Comparison of as-isolated Mössbauer spectra of Cfr from Bacillus amyloliquefaciens (Ba), Enterococcus faecalis (Ef), and Paenibacillus lautus (Pl). Each shows the quadrupole doublet characteristic of a [4Fe-4S]$^{2+}$ cluster. Protein was in gel-filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl$_2$, 25% (v/v) glycerol) for both UV-visible and Mössbauer spectroscopies.

Ferrozine assays revealed that reconstituted samples of each homolog had an estimated 2.7-3.9 irons per protein. For some homologs, this value is lower than expected, likely due to error in protein concentration estimation or a fraction of the protein remaining in the apo state. Pl Cfr had the lowest estimated iron content at 2.7 Fe/protein and Cd Cfr had the highest at 3.9 Fe/protein.

In order to verify the [4Fe-4S] cluster content, I examined anoxic samples of the Cfr homologs by Mössbauer spectroscopy (Figure 2-4, panel B; Figure 2-5). Spectra from Ba, Ef, and Pl Cfr are dominated by a single quadrupole doublet with parameters consistent with [4Fe-4S]$^{2+}$ clusters ($\delta = 0.44$-$0.45$ mm/s and $\Delta E_Q = 1.11$ mm/s in each case). From the relative area of the subspectrum associated with [4Fe-4S]$^{2+}$ clusters and
the amount of Fe/protein after reconstitution, the number of clusters per protein can be estimated. After reconstitution, *Ba, Ef*, and *Pl* Cfr contain 4.6, 4.2, and 3.9 Fe per protein, corresponding to 1.1, 1.0, and 0.9 [4Fe4S]²⁺ clusters per protein, respectively (Table 2-4).

### Table 2-4. Iron incorporation for Cfr homologs analyzed by Mössbauer spectroscopy.

<table>
<thead>
<tr>
<th>Homolog</th>
<th>Fe/protein</th>
<th>Cluster</th>
<th>Area of sub-spectrum (%)</th>
<th>Cofactor/protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Clostridioides difficile</em></td>
<td>4.4</td>
<td>[4Fe-4S]²⁺</td>
<td>77</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fe²⁺ Cys₄</td>
<td>22</td>
<td>1.0</td>
</tr>
<tr>
<td><em>Enterococcus faecalis</em></td>
<td>4.2</td>
<td>[4Fe-4S]²⁺</td>
<td>97</td>
<td>1.0</td>
</tr>
<tr>
<td><em>Paenibacillus lautus</em></td>
<td>3.9</td>
<td>[4Fe-4S]²⁺</td>
<td>94</td>
<td>0.9</td>
</tr>
<tr>
<td><em>Bacillus amyloliquefaciens</em></td>
<td>4.6</td>
<td>[4Fe-4S]²⁺</td>
<td>100</td>
<td>1.2</td>
</tr>
</tbody>
</table>

In addition to the features of a [4Fe4S] cluster (δ = 0.45 mm/s and ΔE_Q = 1.14 mm/s), the Mössbauer spectrum of *Cd* Cfr contains a second quadrupole doublet feature with parameters δ = 0.70 mm/s and ΔE_Q = 3.08 mm/s, similar to those of a Fe²⁺-Cys₄ rubredoxin center (Figure 2-5). I propose that this Fe²⁺ center is coordinated by the four conserved cysteine residues in the C-terminal KTR region of the protein (see Chapter 3). Analysis of all four Cfrs by electron paramagnetic resonance spectroscopy did not result in any detectable signals for the as-isolated reconstituted proteins, thus ruling out assembly of any iron-sulfur clusters with an S = 1/2 ground state in these samples.
Figure 2-5. Comparison of as-isolated (top) and reconstituted (bottom) Mössbauer spectra of *Clostridioides difficile* Cfr. In addition to the characteristic [4Fe-4S]$^{2+}$ quadrupole doublet, an additional doublet is visible accounting for about 1/5 of the iron in the sample. Protein samples were prepared in gel-filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl$_2$, 25 % (v/v) glycerol).

*Limiting substrate activity assays using a flavodoxin/flavodoxin reductase/NADPH protein reducing system*. To measure the *in vitro* activity of the purified Cfr homologs and probe possible functions of the fused Cys-rich domain in the *Cd* homolog, I performed two types of experiments. I assessed (i) sequential mono- and di-methylation of the relevant adenine in a previously reported 155mer RNA substrate derived from 23S rRNA under conditions of high enzyme concentration and limiting substrate and (ii) catalytic methylation of the same base/substrate under conditions of low enzyme concentration and excess RNA. I performed each type of experiment with a protein reductant as well as with a chemical reductant. The RNA substrate is a transcribed 155mer of 23S rRNA substrate (nt 2454-2608) from an *Ec* rRNA template. *Ec* rRNA has
been used to characterize the activity of heterologous Cfr enzymes, both in assays with purified protein/RNA components\textsuperscript{7} and in cells.\textsuperscript{11} I elected to use the same substrate to facilitate comparison to \textit{Sa} Cfr enzymes. The sequence identities of these rRNA fragments to the corresponding region of native host organism 23S rRNA range from 87 to 90\% (Figure 2-6).

**Figure 2-6.** Sequence alignment of 23S ribosomal RNA sequences showing conservation of A2503 (red box).

Gel filtration chromatography and agarose electrophoresis analysis of the RNA samples used in dithionite assays indicate that a mixture of two rRNA fragments was present. These RNA samples were all subjected to a gel filtration purification step. I hypothesize that iron contaminants on the gel-filtration column resulted in RNA cleavage. Because RNA fragment size can affect Cfr activity, the level of turnover in the dithionite initiated samples may have been affected. RNA samples in experiments with a
protein-based electron donor were not purified by gel-filtration chromatography, therefore they should be homogeneous in size.

Reactions were initiated by addition of excess SAM to reaction mixtures containing NADPH and a protein-based Ec flavodoxin/flavodoxin reductase (Flv/Flx) system for reduction of the [4Fe-4S]^{2+} cluster of Cfr.\textsuperscript{45, 46} In LC-MS analyses of the digested RNA products in both assays, new peak(s) can be assigned to C8-methyl-adenosine (m8A) or C2,C8-dimethyl-adenosine (m2,8A) based on retention times, \textit{m/z} values, and MS/MS fragmentation (Figure 2-7).\textsuperscript{6, 7}

![Image](image.png)

**Figure 2-7.** Example LC-MS chromatogram showing retention times (in parentheses) of products and internal standard (Trp) used for quantification of Cfr 155mer rRNA methylation reactions.

Assignment of the methylated adenine products to A2503 was confirmed by comparison to an A2503G RNA substrate in \textit{Ba} and \textit{Cd} Cfr assays. The A\textsuperscript{+-}G variant yields negligible turnover (Figure 2-8). Methylation of A2503 by \textit{Pl} Cfr was demonstrated in a previous study.\textsuperscript{23}
Figure 2-8. LC-MS monitoring at m/z = 282.1 (bottom trace, m8A) and m/z = 296.1 (top trace, m2,8A) of Ba and Cd Cfr activity comparing a wt and variant 155mer substrate (A2503G) under limiting substrate conditions. Reactions were run at room temperature for 165 min and used the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5 % glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, 2 mM NADPH, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein.

HPLC-MS analysis of the limiting substrate reactions (i) revealed detectable levels of mono-methylated and dimethylated products for three of the four enzymes (Figure 2-9, panel A). The quantities of the two products were consistent with sequential reactions. Build-up of the m2,8A product initially lagged while the m8A product grew in, and the m2,8A product then accumulated significantly as the m8A product was depleted (Figure 2-9, panels B-D). In general, these enzymes resemble Sa Cfr in their ability to catalyze sequential mono- and dimethylation reactions, albeit with varying efficiencies. The Cd homolog is an outlier in that it produced very little of the dimethylated product.
(Figure 2-9, panel E). As shown previously by others, the Cs Cfr homolog did not demonstrate any substantive methylation activity (Figure 2-10, panel A).

**Figure 2-9.** Methylation of a 155mer rRNA substrate by Cfr enzymes under limiting substrate conditions initiated with a protein-based reductant. Reactions used the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5 % glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, 2 mM NADPH, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein. (A) LC-MS analysis of methylated RNA products via single-ion monitoring at m/z = 282.1 (bottom trace, m8A) and m/z = 296.1 (top trace, m2,8A). Time course experiment monitoring formation and decay of singly-methylated m8A (solid bars) and doubly methylated m2,8A (patterned bars) with (B) Ba Cfr, (C) Ef Cfr, (D) Pl Cfr, and (E) Cd Cfr. Error bars correspond to standard deviation of three averaged trials.

In the limiting substrate time course, Ba Cfr accumulates the greatest amount of singly and dimethylated product and exhibits the fastest initial rate of catalysis for both reactions (Figure 2-9 B). The kinetic profiles of the Ba, Ef, and Pl enzymes were analyzed⁴⁷ to extract \(k_{\text{obs}}\) for the first methylation step (Figure 2-10 and Table 2-5). The rate (0.23 min⁻¹) and extent of Ba Cfr C8 methylation (~80 % substrate consumption within 20 min of initiation) is comparable to previously reported values for Sa Cfr
analyzed under similar (limiting RNA) conditions (0.7 ± 0.2 min⁻¹ with >80% of the substrate consumed within the first 5 min of the reaction). By contrast, the Pl and Ef enzymes exhibit rates of C8 methylation that are 5- to 10-fold lower (0.07-0.14 min⁻¹), while the Cd Cfr reaction proceeds even more slowly (0.02 min⁻¹).

Figure 2-10. Kinetic analysis of methylation of a 155mer rRNA substrate by Cfr enzymes under limiting substrate conditions initiated with a protein-based reductant. Reactions used the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5 % glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, 2 mM NADPH, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein. (A) Time course experiment monitoring formation and decay of singly-methylated m8A. Lines correspond to kinetic simulation/fits, which were used to obtain the rate constants reported in Table 2-5. Error bars correspond to standard deviation of three averaged trials. Cs Cfr was only tested once. (B) Time course experiment monitoring formation of dimethylated product m2,8A. Although the quantities of the two products were generally consistent with a sequential process, build-up of the m2,8A product initially lags while the m8A product grows in, quantitative agreement of simulations (solid lines) generated according to the simplest possible sequential kinetic mechanism [see equation 2.1 in the Materials and Methods section of main manuscript] with the data was imperfect. For the Ba homolog, for example, rate constants for the second step chosen to optimize agreement with the kinetics of the m2,8A product predicted a more rapid depletion of the m8A intermediate than observed in the dataset. No alternative, more complex mechanism that still considered the m8A product as an obligatory precursor to the m2,8A product could achieve better agreement, owing to the requirement for conservation of mass. Better agreement was observed for the first methylation reaction in Ba Cfr and for both transformations by the other three enzymes. On balance, I view the kinetic results as strongly supportive of the conclusion that all four new Cfr homologs are active in vitro, albeit with varying efficiencies.
Table 2-5. Metrics for Cfr limiting substrate assays using protein-based reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Ba</th>
<th>Pl</th>
<th>Ef</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>110</td>
<td>70</td>
<td>80</td>
<td>60</td>
</tr>
<tr>
<td>5’-dA:mA</td>
<td>2.3</td>
<td>2.7</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>SAH:mA</td>
<td>1.00</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>m8A:m2,6A</td>
<td>0.49</td>
<td>1.2</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>$k_{obs}$ m8A (min$^{-1}$)</td>
<td>0.23</td>
<td>0.14</td>
<td>0.07</td>
<td>0.018</td>
</tr>
<tr>
<td>$k_{obs}$ m2,6A (min$^{-1}$)</td>
<td>0.0079</td>
<td>0.0044</td>
<td>0.003</td>
<td>0.002</td>
</tr>
</tbody>
</table>

To understand whether uncoupled SAM consumption could account for these differences, I quantitatively analyzed the 5’-dA and SAH byproducts (Table 2-5). All enzymes exhibited a near 1:1 ratio of SAH:methylated product. Abortive SAM cleavage was assessed by comparing the amount of methylated RNA produced to the 5’-dA coproduct (Figure 2-11).
Figure 2-11. Ratio of 5’-dA to total methylation events ([m8A] + 2 • [m2,8A]) at the initial time points under limiting substrate conditions for Ba, Ef, Pl, and Cd Cfrs using the Flv reducing system (A) or dithionite (B). Final time point (165 min) ratios for (A) are listed in Table 2-5 and for (B) in Table 2-7. Deviation from a 1:1 ratio indicates a greater amount of unproductive SAM cleavage.

All of the homologs tested exhibited slightly elevated levels of 5’-dA (1.3-1.7), suggesting modest uncoupling under these conditions. These values are similar to those of other radical SAM enzymes. For example, MiaB, gives a coupling ratio of ~1.4 with a ferredoxin reducing system. However, the similarities between the enzymes indicate that abortive SAM cleavage cannot account for the differences in initial methylation rate.

Nevertheless, the most active *Ba* enzyme was also the most efficiently coupled with a 1.3:1 ratio of 5’-dA:methylated product. Under limiting substrate conditions, with the exception of *Ba* Cfr, none of the enzymes consumed the rRNA substrate completely (Table 2-5). This outcome could be due to a number of different factors. It may arise from use of a non-native RNA substrate in the assays. It could also reflect a propensity for enzyme inactivation or product inhibition.

Excess substrate activity assays using Flavodoxin/Flavodoxin Reductase/ NADPH protein reducing system. In the second type of activity assay (ii), I incubated low
concentrations of enzyme (3 µM) with excess RNA substrate (100 µM) at room
temperature (Figure 2-12, panels A-D). In these reactions, the m2,8A dimethylated
product was barely detectable (Figure 2-13, panel B), implying that none of the Cfr
homologs methylate the RNA processively and that the first C8 methylation site is vastly
preferred over the second, C2. The excess substrate assays show that all enzymes exhibit
the capacity to undergo more than one m8 methylation event, with 2-13 conversions
observed within the ~3 h assay window (Table 2-6). All are appreciably slower than Sa
Cfr (0.24 ± 0.06 min⁻¹ when assayed under similar multi-turnover conditions)19 with the
exception of Ba Cfr (v/[E] = 0.23 min⁻¹) (Figure 2-13, Table 2-6).
Figure 2-12. Methylation of a 155mer rRNA by Cfr enzymes with excess substrate in reactions initiated with a protein-based reductant. Reactions used the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl$_2$, 12.5 % glycerol, 250mM KCl, 200 µM Flv, 20 µM Flx, 2 mM NADPH, 1 mM SAM, 100 µM 155mer RNA, and 3 µM protein. Time course experiment monitoring formation and decay of singly-methylated m8A with (A) $Ba$ Cfr, (B) $Ef$ Cfr, (C) $Pl$ Cfr, and (D) $Cd$ Cfr. Error bars correspond to standard deviation of three averaged trials. The inset in panel D shows that $Cd$ Cfr can promote multiple turnovers (red line).
Figure 2-13. Kinetic analysis of methylation of a 155mer rRNA by Cfr enzymes with excess substrate in reactions initiated with a protein-based reductant. Reactions used the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5 % glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, 2 mM NADPH, 1 mM SAM, 100 µM 155mer RNA, and 3 µM protein. (A) Time course experiment monitoring formation and decay of singly-methylated m8A. Lines correspond to kinetic simulation/fits, which were used to obtain the rate constants reported in Table 2-5. Error bars correspond to standard deviation of three averaged trials. (B) Time course experiment monitoring formation of dimethylated product m2,8A. In these reactions, linear progress curves (zero-order kinetics) might have been anticipated, but the kinetics of product formation conform more closely to exponential behavior (first-order kinetics). The traces were fit accordingly [see equation 2.2 in the Materials and Methods section of the main manuscript], with the tangents to the exponential fits at time zero taken as the initial rates for calculation of turnover frequencies (Table 2-6).
Table 2-6. Metrics for Cfr excess substrate assays using protein-based reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Ba</th>
<th>Pl</th>
<th>Ef</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>40 ± 5</td>
<td>19 ± 1</td>
<td>11 ± 1</td>
<td>5 ± 2</td>
</tr>
<tr>
<td>5'-dA:mA</td>
<td>2.8 ± 0.8</td>
<td>2.7 ± 0.2</td>
<td>1.9 ± 0.6</td>
<td>1.7 ± 0.2</td>
</tr>
<tr>
<td>SAH:mA</td>
<td>1.1 ± 0.2</td>
<td>1.1 ± 0.2</td>
<td>1.6 ± 0.5</td>
<td>3.1 ± 0.8</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>40 ± 37</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \frac{v}{[E]} ) \text{ (min}^{-1} \text{)}</td>
<td>0.23</td>
<td>0.14</td>
<td>0.05</td>
<td>0.015</td>
</tr>
<tr>
<td>number of turnovers</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

The *Cd* enzyme is the slowest catalyst in this assay, with an m8 methylation rate (\( \frac{v}{[E]} \)) that is 5-30% that of the other homologs and the lowest level (~ 5%) of total RNA substrate consumption. In *Cd* Cfr, I also considered that the low rate of *in vitro* turnover with excess substrate could be linked to the presence of the extra C-terminal domain. This possibility is explored further in Chapter 3.

*Limiting substrate activity assays using chemical reductant dithionite.*

Methylation capacity of each homolog under limiting substrate conditions was also assessed using a chemical reductant, sodium dithionite. The activity of radical SAM enzymes is known to vary as a function of reductant.\(^\textit{19, 50}\) Chemical reducing agents allow for more rapid iron-sulfur cluster reduction, and therefore can provide insight into the role of kinetic control of cofactor activation relative to other steps in a given radical SAM reaction. Small molecule reductants can also increase the rate of abortive SAM cleavage. This phenomenon has been observed in previous *in vitro* studies of *Sa* Cfr and in the work presented here (Table 2-7).\(^7\)
In the experiments with a chemical reductant, the assay conditions largely remained the same as the protein-based reducing system studies, except I initiated the reaction by adding 2 mM dithionite instead of by addition of SAM. As before, all four homologs can produce mono- and di-methylated products sequentially under these conditions (Figure 2-14). In these experiments, I also compared directly my new homologs to the well-characterized Sa Cfr enzyme. As in the assays with protein-based reductant, the enzymes are similar to Sa Cfr in reaction outcome, but the rates of formation and relative amounts of each product differ in each system. Ba and Ef Cfr meet (or exceed) the rate of substrate consumption/product formation of the Sa enzyme. All three enzymes produce similar proportions of mono-/di-methylated products at the assay endpoint, with the dimethylated product dominating at the end of the 2.75 h reaction period. The Cd and Pl systems are less active and produce smaller quantities of m2,8A. Compared to my findings with the protein-based reducing system, the results with dithionite are qualitatively similar, except that all enzymes produce more of the dimethylated product. Surprisingly this trend also applies to Cd Cfr, which previously produced very low levels of m2,8A.
Figure 2-14. Methylation of a 155mer rRNA substrate by Cfr enzymes under limiting substrate conditions initiated with dithionite. All reactions except Cd Cfr contained 100 mM HEPES pH 7.5, 40mM KCl, 10 mM MgCl₂, 12.5% glycerol, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein. Cd Cfr reactions contained 100mM EPPS pH 8, 10mM MgCl₂, 12.5% glycerol, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein. Reactions were initiated by the addition of 2mM dithionite. Plots represent time course experiments monitoring formation and decay of singly-methylated m8A (solid bars) and doubly methylated m2,8A (patterned bars) with (A) Ba Cfr, (B) Ef Cfr, (C) Pl Cfr, (D) Cd Cfr, and (E) Sa Cfr. Datasets are the average of at least two trials.
Ef and Ba Cfr were able to fully convert the RNA substrate to the mono- or di-methylated form within the 2.75 h assay window. Singly-methylated product builds in but is rapidly depleted as the di-methylated product is formed. I note that substrate consumption for these two Cfrs rose above 100%, a phenomenon likely due to variations in the tryptophan standard curve from sample degradation after multiple uses. The issue was largely resolved for the protein-based reducing system assays by using freshly prepared standard solutions for each HPLC / MS analysis. In spite of these complications, the results show that the Ba and Ef Cfrs are robustly active, as observed in the assays with the protein-based reductant, while the Pl and Cd Cfrs are less active overall. Each of the latter two enzymes converted a little over half of the substrate to the mono- or di-methylated forms. In contrast to Ef and Ba Cfrs, the proportion of the singly- and di-methylated forms stayed relatively constant throughout the reaction period. Analysis of SAM cleavage revealed a sharp increase in uncoupled SAM consumption compared to the flavodoxin/flavodoxin reductase/NADPH system during the initial burst phase (Figure 2-11, panel B). Cd Cfr was most well-coupled at 2.2 cleavage events per reaction, while Pl Cfr was the worst at 4.0 cleavage events per reaction. Uncoupling increases throughout the reaction period, with larger ratios of 5’-dA:mA products at the end of the 2.75 h reaction period (Table 2-7). Increased uncoupling with dithionite is also found in related RS RNA methylase, Ec RlmN. Within 1 h, RlmN consumes > 50% of the SAM pool in the absence of substrate in dithionite initiated experiments, compared to almost no unproductive cleavage without substrate with the Flv/Flx reducing system.7
Table 2-7. Metrics for Cfr limiting substrate assays using dithionite reducing system.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Ba</th>
<th>Pl</th>
<th>Ef</th>
<th>Cd</th>
<th>Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>114 ± 21</td>
<td>59 ± 7</td>
<td>134 ± 16</td>
<td>64 ± 18</td>
<td>75 ± 6</td>
</tr>
<tr>
<td>5′-dA:mA</td>
<td>13.9 ± 0.2</td>
<td>22 ± 5</td>
<td>10 ± 0.7</td>
<td>14 ± 9</td>
<td>8.4 ± 0.1</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>0.14 ± 0.01</td>
<td>2 ± 1</td>
<td>0.32 ± 0.01</td>
<td>1.7 ± 0.2</td>
<td>0.29 ± 0.01</td>
</tr>
</tbody>
</table>

I compared the new homologs to previously characterized *Sa* Cfr under these assay conditions. *Sa* Cfr was able to convert ~75% of RNA substrate to singly- or dimethylated product (Figure 2-14, panel E). While the level of substrate consumption is lower in *Sa* Cfr than the most active *Ba* and *Ef* homologs tested here, the product profiles are similar in all three systems, with an initial build-up of singly-methylated product followed by a concomitant decrease as di-methylated product formed. Compared to all of the newly tested homologs, *Sa* Cfr exhibited the lowest level of abortive cleavage of SAM at the conclusion of the assay window with a ratio of 8.4 cleavage events per reaction.

The relatively low activity of the *Pl* homolog, from an organism originally isolated from a hot spring\(^41\), suggested to us that this (and other) homolog(s) might require a different temperature for optimal activity. I tested all four homologs for activity at 37 °C under limiting substrate conditions using dithionite as the reductant. Under these conditions, *Ba* Cfr and *Ef* Cfr are less active than at room temperature (Figure 2-15, Table 2-8). Neither enzyme was able to fully convert the 155mer to singly- or dimethylated product, producing 42 µM and 47 µM total methylated product, respectively. *Pl* Cfr and *Cd* Cfr each produced about the same amount of total methylated
product at high temperature as at room temperature (~29 µM), but the ratio of singly to
dimethylated products shifted considerably to favor the dimethylated products. At room
temperature, the m8A:m2,8A ratio for Pl Cfr was 2.0 and for Cd Cfr was 1.7. At 37 °C,
these homologs had ratios of 0.25 and 0.64, respectively. Although the biological
significance of the dimethylated form of A2503 is not well understood, these results may
also suggest that the enzyme/substrate conformation required to promote the second
methylation event is easier to achieve at high temperature, thus allowing for a greater
population of the RNA to be processively methylated.

Table 2-8. Metrics for high temperature Cfr limiting substrate assays using
dithionite reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Ba</th>
<th>Pl</th>
<th>Ef</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>84</td>
<td>57</td>
<td>94</td>
<td>58</td>
</tr>
<tr>
<td>5'-dA:mA</td>
<td>21.5</td>
<td>31.6</td>
<td>22</td>
<td>40</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>0.1</td>
<td>0.25</td>
<td>0.41</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Figure 2-15. Methylation of a 155mer rRNA substrate by Cfr enzymes at high temperature under limiting substrate conditions initiated with dithionite. All reactions contained 100 mM EPPS pH 8, 10 mM MgCl$_2$, 12.5 % glycerol, 1 mM SAM, 50 µM 155mer RNA, and 48 µM protein. Reaction mixtures were heated to 37 °C for 5 minutes before initiation by the addition of 2 mM dithionite. Reactions were incubated at 37 °C throughout the assay period. Plots represent time course experiments monitoring formation and decay of singly-methylated m8A (solid bars) and doubly methylated m2,8A (patterned bars) with (A) $Ba$ Cfr, (B) $Ef$ Cfr, (C) $Pl$ Cfr, and (D) $Cd$ Cfr. Only one replicate was performed for each homolog.

Excess substrate activity assays using the chemical reductant dithionite. I performed the second type of activity assay (ii), in which I incubated low concentrations of enzyme (3 mM) with excess RNA substrate (100 mM) at room temperature (Figure 2-16, Table 2-9), with dithionite as the reaction-initiating reductant. In contrast to my observations with the Flv/Flx reducing system, in which the m2,8A dimethylated product was barely detectable with all four Cfrs, appreciable dimethylation with $Ba$ and $Ef$ Cfr
was observed. This result indicates that the relative rates of iron-sulfur cluster reduction and product release are a possible mechanism of controlling reaction outcome in some Cfr enzymes. With the exception of *Pl* Cfr, which becomes nearly inactive under these multi-turnover conditions with dithionite, the excess substrate assays show that all of the other enzymes exhibit the capacity to undergo more than one methylation event, with 3-7 conversions observed within the ~3 h assay window (Table S3). When the level of substrate consumption is compared to that of the excess substrate assays with Flv/Flx, each system responded in a different way. *Ba* Cfr is ~50% less active, *Ef* Cfr is similarly active, and *Cd* Cfr is twice as active. These distinctions could reflect effective pairing of the *E. coli* flavin-based reducing system to each Cfr. *Ba* and *Ef* might be well-optimized to use this protein donor while *Cd* could be poorly suited to the Flv/Flx system. This prediction would be consistent with my finding of an additional redox-active iron-binding site in *Cd* Cfr. In Chapters 3 and 4, I explore potential links between this unique cofactor and electron transfer in clostridial Cfrs.

Table 2-9. Metrics for Cfr excess substrate assays using dithionite reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>RNA cons (%)</th>
<th>5’-dA:mA</th>
<th>m8A:m2,8A</th>
<th>number of turnovers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 ± 1</td>
<td>2.0 ± 0.1</td>
<td>11 ± 1</td>
<td>7</td>
</tr>
<tr>
<td><em>Ba</em> Cfr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pl</em> Cfr</td>
<td>1.93 ± 0.03</td>
<td>24.8 ± 0.2</td>
<td>3.16 ± 0.03</td>
<td>1</td>
</tr>
<tr>
<td><em>Ef</em> Cfr</td>
<td>0.75 ± 0.03</td>
<td>33 ± 4</td>
<td>0.34 ± 0.01</td>
<td>4</td>
</tr>
<tr>
<td><em>Cd</em> Cfr</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

75
Figure 2-16. Methylation of a 155mer rRNA substrate by Cfr enzymes under excess substrate conditions initiated with dithionite. All reactions contained 100 mM EPPS pH 8, 10 mM MgCl₂, 12.5 % glycerol, 1 mM SAM, 100 µM 155mer RNA, and 3 µM protein. Reactions were initiated by the addition of 2mM dithionite. Plots represent time course experiments monitoring formation and decay of singly-methylated m8A (solid bars) and doubly methylated m2,8A (patterned bars) with (A) Ba Cfr, (B) Ef Cfr, (C) Pl Cfr, and (D) Cd Cfr. Insets detail 1 turnover (TO) with a red dashed line. Data represents the average of two trials.

Profiling the antibiotic resistance of Cfr homologs in E. coli. To correlate in vitro activity to resistance phenotypes, I performed broth microdilution assays to assess the impact of the Cfr homolog expression on lincomycin resistance in E. coli. Lincomycin was chosen as the initial drug target because the action of a structurally similar antibiotic, clindamycin, was affected by Cfr expression in other systems. I expressed each Cfr from an arabinose-inducible pBAD24 vector in the knockout strain E. coli BL21 DE3.
ΔtolC. These conditions were identified in a series of optimization experiments as the best for visualizing drug resistance phenotypes (see Chapter 4 for additional details). I initially assayed the most active homolog in vitro, Ba Cfr, and the least active, Cs Cfr, for drug resistance against a serial dilution of lincomycin (1 mg/mL – 7.8 µg/mL). For Ba Cfr, a clear difference in growth emerged between the induced and uninduced cells under these conditions (Figure 2-17, panels A). The IC\textsubscript{50} value for the induced cells was >287 µg/mL lincomycin compared to >36.3 µg/mL for the uninduced cells, an ~8-fold increase. A second independent replicate of these conditions produced similar results with an induced IC\textsubscript{50} of 206 µg/mL and an uninduced IC\textsubscript{50} of 41.7 µg/mL, a ~5-fold increase. Cs Cfr expression yielded no detectable difference between the control and induced cells (Figure 2-17, panel B). This result mirrors my in vitro results for the Cs homolog, which showed that it cannot methylate E. coli 23S rRNA and, therefore, cannot confer antibiotic resistance under these conditions.

To determine if increased lincomycin resistance would be observed when expressing the other Cfr homologs, Cd and Pl Cfrs were assayed for their antibiotic resistance profiles. As the amount of expression for optimal growth was unknown, these homologs were first tested against lincomycin with variable concentrations of the inducing agent arabinose. The lowest concentration of arabinose (0.005%) showed the best growth, with higher concentrations of arabinose leading to growth suppression. However, when tested versus an uninduced, drug present (-A, +D) control, the induced cells showed only a small, statistically insignificant increase in growth (Figure 2-17, panels C-D). These results suggest that Pl and Cd Cfr are too toxic to support cell growth
when expressed at high concentrations. At the toxicity threshold that permits growth, I cannot detect any antibiotic resistance-related activity under the current conditions.
Figure 2-17. Antibiotic resistance profiles of (A) Ba Cfr, (B) Cs Cfr, (C) Cd Cfr, and (D) Pl Cfr on pBAD24 in E. coli BL21 DE3 ΔtolC. (E) The resistance profile of Ba Cfr on pBAD24 in E. coli BL21 DE3 ΔtolC/ΔaraBAD. Cell viability normalized to uninduced, drug absent control (-A, -D) is plotted with induced, drug present (+A, +D, closed circles) and uninduced, drug present (-A, +D, open circles).
I also evaluated *Cd* Cfr and *Pl* Cfr with linezolid, in case lincomycin is not ideal for testing Cfr activity under my assay conditions (Figure 2-18).\textsuperscript{11} No change was observed in these trials, suggesting that the identity of the drug is not causing the observed difference in resistance profile. In all homologs, I additionally considered the role of arabinose catabolism in promoting growth, thereby amplifying apparent antibiotic resistance effects. When *Ba* Cfr was assayed using a *tolC* knockout, *araBAD* deletion strain, *E. coli* BL21 DE3 Δ*tolC/ΔaraBAD*, the Cfr-induced growth curve collapsed onto the control. This result suggests that arabinose was contributing significantly to cell growth and that further optimization is required to demonstrate the effect of Cfr activity on antibiotic resistance. Proposals for future work in effective drug resistance profiling for the Cfr homologs can be found in Chapter 4.

**Figure 2-18.** Antibiotic resistance profiles of (A) *Cd* Cfr and (B) *Pl* Cfr on pBAD24 in *E. coli* BL21 DE3 Δ*tolC*. Cell viability normalized to uninduced, drug absent control (-A, -D) is plotted with induced, drug present (+A, +D, closed circles) and uninduced, drug present (-A, +D, open circles). The drug used in these trials was linezolid.
Discussion

In this study, I establish that four different Cfr enzymes from distinct phylogenetic groups assemble a necessary $[4\text{Fe}4\text{S}]^{2+}$ cofactor and exhibit adenine C8 RNA methylation activity \textit{in vitro}, similar to the well-characterized Cfr homolog from \textit{S. aureus}. These results establish that diverse Cfr homologs are viable targets for structural and biophysical study. With a small molecule reductant, all four enzymes exhibit robust mono- and dimethylation activity under limiting substrate conditions (Figure 2-19). With a protein-based reductant, three of the four enzymes analyzed exhibit both mono- and di-methylation activity with limiting RNA substrate. Interestingly, the \textit{Cd} Cfr homolog that lacks dimethylation activity with a protein electron donor also contains an extra C-terminal metal-binding domain. The presence of this domain could explain the differences in reactivity for this homolog. In Chapters 3 and 4, I further analyze the function of the \textit{Cd} C-terminal extension and related stand-alone proteins.

When challenged with excess substrate, additional differences emerge between the homologs. The genomically encoded Bacillales Cfr is the only homolog that retains the ability to robustly consume substrate under multiple turnover conditions. Interestingly, the \textit{Ba} enzyme is more active with a protein-based reducing system in this scenario. It is also the only homolog that operates at a comparable rate to \textit{Sa} Cfr with both limiting and excess RNA.\textsuperscript{48} Initial attempts to correlate \textit{in vitro} RNA methylation activity with the ability to confer drug resistance to 23S-binding compounds, such as lincomycin, appeared to show significant resistance with \textit{Ba} Cfr expression. However, subsequent control experiments to eliminate the effects of inducing agent (arabinose) catabolism showed loss of resistance, suggesting that I need to optimize the assays to
visualize Cfr-mediated growth effects in *E. coli*. Other Cfr drug resistance phenotype studies have been performed either in the native host organism, a close relative, or in the antibiotic-susceptible *E. coli* strain AS19.\textsuperscript{11, 51, 52} Although I attempted to modulate drug susceptibility by using a Δ*tolC* strain of *E. coli*, deficient in an important drug efflux pump, the AS19 strain was recently sequenced, and shown to contain more than 660 DNA-level mutations.\textsuperscript{53} Future work in this more heavily modified strain may allow us to better detect drug resistance phenotypes in all of my systems of interest.
In this chapter, I also compare Cfr reaction outcome and rate/extent of methylation with a flavodoxin-based protein reducing system to experiments with a small molecule reductant. Levels of substrate consumption remained largely the same across trials. The \textit{in vitro} activity of only a few radical SAM enzymes has been analyzed with different reductants. Some systems, such as \textit{B. subtilis} QueE, can be dramatically more active when assayed \textit{in vitro} with their native reducing partners and mediators.\textsuperscript{50} Others, such as \textit{AtsB}, are more active with dithionite.\textsuperscript{54} I also observe a range of outcomes. The
most active homolog with a protein-based reducing system, *Ba* Cfr, is significantly less active with dithionite under multi-turnover conditions. This observation may indicate a particularly good match between the *Ec*-derived protein reductant used in this study and the *Ba* Cfr homolog. By contrast, the least active enzyme with a protein-based donor, *Cd* Cfr, was moderately improved with dithionite under multi-turnover conditions. These differences underscore the importance of the role of the reducing system in control of methylation rate by Cfr.

The identity of the reducing system also has important effects on reaction outcome. In all enzymes except *Pl* Cfr, dithionite reduction yielded significantly more m2,8A production, particularly with limiting substrate. Dithionite is a small, diffusible reductant that can rapidly reduce oxidized metal cofactors. It can likely access the [4Fe4S]$^{2+}$ cluster of Cfr regardless of RNA-binding status. Protein-based donors may instead be optimized to bind selectively to the substrate-bound Cfr. In all radical SAM enzymes, such a phenomenon would ensure efficient use of the SAM cosubstrate by maintaining tight coupling between reductive SAM cleavage and primary substrate transformation. The observation of a marked increase in dimethylated A2503 upon dithionite reduction of the Cfrs shows that the nature of interaction between the reducing system and a radical SAM enzyme can also change the product partition. These results would be consistent with a model in which the relative rates of product release and iron-sulfur cluster reduction dictate the final ratio of mono- and di-methylated products (Figure 2-20). With a rapid and non-specific reducing agent, like dithionite, there is a greater likelihood that the [4Fe4S]$^{2+}$ cluster will be reduced immediately after the first methylation event prior to product release. Dithionite enables the enzyme to act
processively, with rapid appearance of the dimethylated product and lesser accumulation of the monomethylated compound. With a protein-based reductant, these results indicate full release of the monomethylated product prior to cofactor re-reduction.

**Figure 2-20.** Predicted reaction outcomes when using a small chemical reductant like dithionite compared to a bulky protein-based reducing system like flavodoxin. Dithionite can diffuse into the active site and reduce the radical SAM cluster while monomethylated product is present allowing for dimethylation to occur processively. Flavodoxin cannot reduce the cluster rapidly enough to perform processive dimethylation before product leaves and next substrate binds. This effect was most pronounced with *Cd* Cfr, which only showed m2,8A production with the chemical reductant under limiting substrate conditions. With a protein-based donor, the *Cd* enzyme is the only enzyme that remains unable to convert the singly methylated product to the dimethylated product. It also exhibits markedly poor turnover in excess substrate assays under all conditions. I speculate that this difference could be related to the additional C-terminal domain, revealed in this work to contain a
mononuclear Fe$^{2+}$ cofactor. In subsequent chapters, I probe the role of this domain and explore the idea that stand-alone proteins encoded adjacent to clostridial Cfrs could affect their activity. This hypothesis, if true, might explain the failed attempts by other groups to demonstrate RNA methylation activity in other Clostridial homologs.\textsuperscript{23, 24}
Table 2-1. RNA sequence (and associated primer sequences) for the 155mer substrate and protein sequences (with associated source DNA and primer sequences) for the Cfrs used in this study. Methylation site (and its mutants) highlighted in cyan and T7 promoter site shown in red text.

<table>
<thead>
<tr>
<th>rRNA substrate DNA template A2454 – G2608 (E. coli)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGAAATTATTACGACTCTATATACTGGATACCGCCAGATATGCGTCTGCCGGAATATCGTTATAAACAGCTGCTGGATGCAGTTTTCTGCAGG</td>
</tr>
<tr>
<td>GTATTATGCGTTTCGAAGATATGAAACTGCTGCCGAAAACCCTGCGTGAGAAACTGGTTGAACAGTTTGGTGAAACCGTGGTTGAAATTAAAGCCATCCATCACGAAAAAAGCATGCAGACCGATAAAGTTCTGTTTGAACTGAGTGATGGTAATCGTGTTGAAACCGTTGGTCTGTTCTATAAAGAAGGCTGGAATAGCTTTTGTATTAGCAGCCAGAGCGGTTGTGGTTTTGGTTGTAAATTTTGTGCAACCGG</td>
</tr>
<tr>
<td>CACCCTGGGTCTGCGTCGTAATCTGACCGTTGATGAAATTACCGATCAGATCCTGTATTTTATGCAGCAGGGTTGTAGCATTAACAGCATTAGCTTTATGGGTATGGGTGAACCGTTTGCAAATCCGCAGGTTTTTGAAGCACTGCATGATCTGACCGCACCGGAACTGTTTGGTCTGAGCCAGCGTCGTATTACCATTAGCACCATTGGTATTGTTCCGGGTATTCAGAACTGACCCGTAAATATCCTCAGGTT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rRNA substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCUGAUACCGCCCAAGAGAUUCAAUUGCAGCGCCGGUGUUGUGCCACCCUCGUGUCGGGCUCACACACUGG</td>
</tr>
<tr>
<td>GCCUGUAGGUAGUUGCCCAAGGGUAUGGCGUUGCAGCCAUUUAAAAGUGGGUACGCGACUGGUGUUAGAAGCU</td>
</tr>
<tr>
<td>CGUGAGACAGUUCGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PCR primers for 155mer RNA substrate amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer (2545): 5'-CGGAAATTATTACGACTCTATATACTGGATACCGCCAGATATGCGTCTGCCGGAATATCGTTATAAACAGCTGCTGGATGCAGTTTTCTGCAGG</td>
</tr>
<tr>
<td>Reverse Primer (2608): 5'-mCmCGAACTGTCTCACGACGTTCTAAACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A2503G rRNA substrate DNA template</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGAAATTATTACGACTCTATATACTGGATACCGCCAGATATGCGTCTGCCGGAATATCGTTATAAACAGCTGCTGGATGCAGTTTTCTGCAGG</td>
</tr>
<tr>
<td>GCCUGAUACCGCCCAAGAGAUUCAAUUGCAGCGCCGGUGUUGUGCCACCCUCGUGUCGGGCUCACACACUGG</td>
</tr>
<tr>
<td>GCCUGUAGGUAGUUGCCCAAGGGUAUGGCGUUGCAGCCAUUUAAAAGUGGGUACGCGACUGGUGUUAGAAGCU</td>
</tr>
<tr>
<td>CGUGAGACAGUUCGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A2503G rRNA substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCUGAUACCGCCCAAGAGAUUCAAUUGCAGCGCCGGUGUUGUGCCACCCUCGUGUCGGGCUCACACACUGG</td>
</tr>
<tr>
<td>GCCUGUAGGUAGUUGCCCAAGGGUAUGGCGUUGCAGCCAUUUAAAAGUGGGUACGCGACUGGUGUUAGAAGCU</td>
</tr>
<tr>
<td>CGUGAGACAGUUCGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clostridioides difficile str. QCD63q42 Cfr</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSKYKKMKQLIADMRLPEYRYKQLLDAVFLQGIMRFEDMKLLPKTLREKLVQEFGTVEIKAIHEKSMQ</td>
</tr>
<tr>
<td>TDKVLFELSDGNRTVEVLGYFKEGNSFCISSQXSGCFFCKATGTGTLRRLNVTDEIDTQILYFMQXGC</td>
</tr>
<tr>
<td>SINSISFMGMGEFPANQVFQVFHDLTAPELFGLSQRITITSTIGVPQIQKLTREYPVNVLYSLHAPTDLR</td>
</tr>
<tr>
<td>ETLMPITKTYPLQYVLDLQHIRQNTIKVRLMDVNSDHRAEQLTKKHLHKYPHLDDLUPYQX</td>
</tr>
<tr>
<td>TTVTEVMPPSSHTRIKAFRCRIIHNAGISVINRTQFGSDINAAACQGLAGAYRDQKQGERMTSIAIEQFSWEL</td>
</tr>
<tr>
<td>LLCPVCXSKTRKIRDNLPCFKSCMERLICVXXHFIKIIUPERAPSDQXR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clostridioides difficile str. QCD63q42 Cfr Forward Primer: 5'-GACATTCATATGAGCAAATACAAAAAAATGAAACAGCTGATCGCCGATATGCGTCTGCCGGAATATCGTTATAAACAGCTGCTGGATGCAGTTTTCTGCAGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse Primer: 5'-GACATTCTCGAGACGGCTCTG</td>
</tr>
</tbody>
</table>

87
**Bacillus amyloliquefaciens** str. DSM-7 (ATCC 23350) Cfr

Forward Primer: 5’-GGATACCATATGCGACAAAAAAAACAAGTATATAAGAATTCAAGAG-

Reverse Primer: 5’-GCTGACCTCGAGGCTATTATTTTTTTGATAGTTACCATATAATTGCCC-

**Paenibacillus lautus** sp. Y412MC10 Cfr

Forward Primer: 5’-GGATAACATATGAAATACCTGAGCAAATAC-

Reverse Primer: 5’-ATTCTCGAGGATGTTACCATTGCT-

**Enterococcus faecalis** str. TX0635 Cfr

Forward Primer: 5’-GGATAACATATGAAATACCTGAGCAAATAC-

Reverse Primer: 5’-ATTCTCGAGGATGTTACCATTGCT-
### PCR primers for *Enterococcus faecalis* str. TX0635 Cfr

<table>
<thead>
<tr>
<th>PCR primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer</td>
<td>5’-GGATAACATATGAAACAGACCAAAACCAAA-3’</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>5’-ATTCTCGAGCAGTTCATTTTCATAGCA-3’</td>
</tr>
</tbody>
</table>

### Clostridium sporogenes str. ATCC 15579 Cfr

<table>
<thead>
<tr>
<th>PCR primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer</td>
<td>5’-GGATACCATATGCAGCAGAAAAACAAATACATCCGC-3’</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>5’-GCTGACCTCGAGCTGGCTATTGGTTTTCTGATAATTGCC-3’</td>
</tr>
</tbody>
</table>

### Staphylococcus aureus (plasmid) Cfr

<table>
<thead>
<tr>
<th>PCR primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer</td>
<td>5’-GGATAACATATGAAACAGACCAAAACCAAA-3’</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>5’-ATTCTCGAGCAGTCCTACTATG-3’</td>
</tr>
</tbody>
</table>
MNFNNKTYGKIqEQFLRSNNEPDYRIKQITNAIFKQRISRFEDMKVLPKLLREDLINNFGETVLNIKLLAEQN
SEQVTKVLFEEVSKNERVETVNMKYKAGWESFCISSGCNFGCKFCATGDIGLKLKNLTDEITDQVLHYFHL
GHQIDSISFMGMEALANRQVFDAISFTDNLFPSRRLSISTIGIIIPSIIKQTQEPQVNLTFLHSPYSEE
RSKLMPINDRPIDEVMNILDEHILTSRKVIAYIMLPGVNDSEHANEVSSLKRSYKSGKLYHVNLIRYN
PTISAPEMYGEANEQVFVEAFYKVLKSAGIHTIRSQFGIDIDAACGQLYGNQNSQ

ATGAATTTTAATAAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAGTATGGTAAAATACAGGAATTTTTAAGAAGTAAATGAGGCTATTATAGAAATAAACAAAATTGATACGATTTCCCAATGCTGAGATCAATCGCAAGGAACTGCTATAAGCGGCGAGACGATTTTAACTTCCATTTATTAGTACATATTTCTTTTATGGGAATGGGTGAAGCTCTAGCCAACCGTCAAGTTGATGCTCTTGATTCGTTTACGGATCCTAATTTATTTGCATTAAGTCCTCGTAGACTTTCTATATCAACGATTGGTATTATACCTATATCAAAAAATAACCCAGGAATATCCTCAAATATCTTACATTACACTACCTACCTTATAGTGAAGGAAGCCACAGAAAT
TGATGCACAAAATATGATAGATACCCAATAGATGAGGTAATGAATATACTCGATGAACATATAAGATTAACTTCAAGGAAAGTATATATAGCTTATATCATGTTGCCTGGTGTAAATGATTCTCTTGAGCATGCAAACGAAGTTGTTAGCCTTCTTAAAAGTCGCTATAATCAGGGAAGTTATATCATGTAAATTTGATACGATACAATCCTACAATAAGTGCACCTGAGATGTATGGAGAAGCAAACGAGGCGAGTCAGAAGCCTTTTACAAAGTTTTGAAGTCTGCTGGTATCCATGTCACAATTAGAAGTCAATTTGGGATTGATATTGACGCTGCTTTGTGGTCAATTATATGGTAATTATCAAAATAGCCAATAG
References


anaerobic ribonucleotide reduction, cloning of the gene (fpr), and overexpression of the protein, *J Bacteriol* 175, 1590-1595.


CHAPTER 3

Biochemical and spectroscopic characterization of a C-terminal rubredoxin-like domain in Clostridioides difficile Cfr

Abstract

Clostridial Cfr homologs have displayed only sporadic capability to perform C8 methylation of A2503 in 23S rRNA, a modification associated with antibiotic resistance in bacteria. The lack of Cfr-like activity in many homologs suggests a different cellular target or the presence of additional as yet unidentified reaction components. In chapter 2 I identified the presence of an additional C-terminal domain on \textit{Cd} Cfr which contains a metal site. Herein I characterize the biophysical and functional properties of this accessory domain. My results suggest that the C-terminal domain is part of an uncharacterized rubredoxin-like protein family, named cys-rich KTR, and that the domain binds a single iron. Though not necessary to carry out C8 methylation of rRNA, perturbation of the metal site by metal replacement or targeted mutagenesis results in a reduction in Cfr-like activity under multiple turnover conditions. This result and the demonstration of a strong interaction between the domain and RNA suggests that the domain may be involved in electron transfer or substrate positioning during the reaction of this clostridial Cfr \textit{in vivo}. 
**Introduction**

In my initial analysis of phylogenetically diverse Cfrs, I detected a 71-amino-acid C-terminal extension in selected clostridial homologs. I showed by Mössbauer analysis that one of these homologs, *Clostridioides difficile* (*Cd*) Cfr, harbors an additional mononuclear Fe$^{2+}$-binding site. My work is the first demonstration of a Cfr with an additional cofactor. Many radical SAM enzymes contain auxiliary cofactors, most commonly additional [2Fe-2S] or [4Fe-4S] clusters. Functions include provision of reducing equivalents/interaction with reducing systems, substrate binding, and the source of reagents for synthesis of functional groups. The functional role of the auxiliary metal cofactor in *Cd* Cfr is not known, but its nuclearity and location in a Cys-rich domain suggests it might most closely resemble a rubredoxin motif. Rubredoxins are often described as FeS$_4$ proteins, thereby representing the simplest type of iron-sulfur cluster cofactor. In prototypical stand-alone rubredoxins, the iron cofactor is redox active, with a high-potential (-95 - +114 mV) Fe$^{3+}$/2+ couple. These small proteins or domains are often employed in complex multicomponent metalloenzymes as a mediator of electron transfer between catalytic cofactors and protein-based reducing systems. No other example of a rubredoxin fusion exists among characterized radical SAM enzymes, but the gene coding for non-canonical radical SAM enzyme Dph2 neighbors a gene coding for a mononuclear iron-binding protein, Dph3, of unknown function.

Although prior studies indicated that certain clostridial Cfrs cannot methylate A2503 in *E. coli* when overexpressed heterologously, I and others have shown that some clostridial homologs are active rRNA C8 methylases. In work by other laboratories, a *Cd* str. 11140508 Cfr homolog has been shown to confer antibiotic resistance via
A2503 modification, but the activity was promoted by a cfr sequence on a mobile genetic element that shares only 52% identity with my chromosomally-encoded Cd homolog. A Clostridium boltae 90B3 Cfr with a similar C-terminal extension has also been shown to confer moderate drug resistance in cell-based assays. In Chapter 2, I report in vitro characterization of an active clostridial Cfr with a C-terminal iron-binding accessory domain. I also validated reports of inactivity towards rRNA harboring A2503 in other clostridial Cfrs in my analysis of Clostridium sporogenes (Cs) Cfr in crude cell extracts. Here I investigate the properties of N- and C-terminally truncated Cd Cfr to investigate the function of the C-terminal domain and its metal cofactor. The radical SAM domain (residues 1-345) and the C-terminal domain (residues 356-416) were expressed and isolated separately for these analyses (Figure 3-1).

The Cd C-terminal domain contains a pair of conserved CXXC motifs that could coordinate the additional iron ion. In this chapter, I tested this hypothesis by mutating one of these putative ligands. I also characterized the redox and metal-binding properties of the isolated C-terminal domain. I find that the C-terminal Fe$^{2+}$ site can, indeed, be oxidized with a midpoint potential similar to other rubredoxins. These findings would enable the domain to serve as a mediator of electron transfer. The proposed mechanism of Cfr methylation has certain features that could require such a domain. The proposed reaction pathway yields a thiyl radical that must be quenched by an external reducing equivalent to allow for subsequent turnover. In some circumstances, rapid re-reduction might be critical to avoid potentially damaging off-pathway reactions. The purpose of the fused rubredoxin domain may be to act as an electron reservoir for radical reduction. Another possibility is that the domain helps bind or position the rRNA
substrate in the Cfr active site. A structure of related enzyme RlmN bound to tRNA suggests that the RNA backbone undergoes dramatic repositioning to enter the protein active site.\textsuperscript{24} To achieve this configuration, multiple RNA-protein contacts are needed both adjacent and distal to the active site. It is possible that the C-terminal domain of \textit{Cd} Cfr facilitates folding of rRNA in the correct configuration for catalysis in clostridial organisms.

**Figure 3-1.** Truncation variants of the full-length \textit{Cd} Cfr radical SAM / KTR fusion: \textit{Cd}-N (1-345), \textit{Cd}-C (356-416, iron or zinc), and \textit{Cd}-M (C363A).

**Materials and Methods**

\textit{Materials.} Kanamycin was obtained from Teknova (Hollister, CA). Ampicillin was purchased from Dot Scientific (Burton, MI). \textit{N}-(2-hydroxyethyl)piperazine-\textit{N}′-2-
ethanesulfonic acid (HEPES), DNaseI (RNase-free), Bradford reagent, and His-Pur 
Co^{2+}-TALON resin were obtained from Thermo Fisher Scientific (Waltham, MA).
Sodium sulfide, 2-mercaptoethanol, sodium dithionite, glucose, \textit{S}-adenosylmethionine, 
and p1 nuclease from \textit{Penicillium citrinum} were purchased from Sigma Aldrich (St. 
Louis, MS). Iron (III) chloride hexahydrate was acquired from BDH Analytical 
Chemicals, VWR (Radnor, PA). DL-dithiothreitol (DTT) was obtained from Alfa Aesar 
(Tewksbury, MA). \textsuperscript{57}Fe metal (98\%) was purchased from Isoflex USA (San Francisco, 
CA). Bovine serum albumin (BSA), \textit{XhoI}, \textit{NdeI}, antarctic phosphatase, Q5 DNA 
polymerase, and T4 DNA ligase were obtained from New England Biolabs (Ipswich, 
MA). PD-10 and HiPrep Sephacryl 16/60 S200 columns were purchased from GE 
Healthcare (Little Chalfont, United Kingdom).

\textit{Cloning and overexpression of Clostridioides difficile (Cd) Cfr protein variants 
and C363A mutant.} PCR primers (Table 3-1) were designed with Primer3 and appended 
at the 5’-end with \textit{NdeI} and \textit{XhoI} restriction sites. Primers were designed for truncated \textit{Cd} 
Cfr variants consisting of either the radical SAM domain (Cd-N), residues 1-345, or the 
rubredoxin/KTR domain (Cd-C), residues 346-416. Primers for a Cd-C variant without 
the linker region, residues 356-416, were also designed. Amino acids 346-355, while 
initially included as part of the accessory domain, were later removed for all experiments 
as they did not match the sequence of stand-alone KTRs. The coding sequences were 
amplified by PCR using Q5 DNA polymerase. The PCR products were digested using 
\textit{NdeI} and \textit{XhoI} restriction enzymes, purified by gel electrophoresis, re-isolated using an 
Omega Biotek gel extraction kit, and ligated into pET26b.
A C363A Cd Cfr variant was generated by inverse PCR using a new forward primer containing the relevant substitution in the annealing region (Table 3-1). Amplification was performed using the pET26b/Cfr (Cd) plasmid as the DNA template. DpnI digestion was used to remove the original template DNA and the vector was circularized by ligation with T4 DNA ligase. All four Cd variant pET26b-cfr sequences were verified by Sanger sequencing at the Pennsylvania State University Nucleic Acid Facility. All plasmids additionally encode a C-terminal His$_6$-tag with a short two amino acid (-Leu-Glu-) linker.

Overexpression of all constructs was accomplished in *E. coli* BL21(DE3) cells co-transformed with pET26b-cfr and pDB1282, the latter a vector containing inducible iron-sulfur cluster assembly machinery. Successful transformants were selected on kanamycin (50 µg/mL) and ampicillin (100 µg/mL) supplemented LB plates. 1.5 L cultures of the overexpression strains were grown shaking (180 rpm) at 37 °C in M9 minimal medium supplemented with kanamycin (50 µg/mL), ampicillin (100 µg/mL), and iron (III) chloride (30 µM) to maximize iron-sulfur cluster (and/or single iron) incorporation. Growth for samples suitable for Mössbauer analysis utilized $^{57}$Fe dissolved in 1 M H$_2$SO$_4$. Zinc replacement of the C-terminal metal binding site in Cd-C was accomplished by addition of ZnCl$_2$ (7 mM) to minimal medium cultures both upon initial inoculation and at OD$_{600}$ ~ 0.6 in place of iron supplementation. Only for Cd-N and Cd C363A, L-arabinose was added [final concentration 0.08 % (w/v)] to induce overexpression of the genes harbored on the pDB1282 plasmid at an OD$_{600}$ = 0.3. L-cysteine (300µM) and additional iron (III) chloride (30µM) were added at the same OD to support iron-sulfur cluster assembly. At an OD$_{600}$ = 0.6, IPTG was added to a final
concentration of 100 μM to induce overexpression of the pET26b Cd variant gene. After incubation overnight at 18 °C at 180 rpm, cells were pelleted by centrifugation at 5000g for 10 min at 4 °C. Approximately 3.9 g cell paste was obtained per liter of culture. Cell paste was flash-frozen in liquid nitrogen and stored at -80 °C. Flavodoxin and flavodoxin reductase were purified as described previously.27

**Purification of Cd Cfr protein variants and C363A mutant.** All purification steps were performed anaerobically in an anoxic chamber with a mixed H₂/N₂ atmosphere (Coy Laboratory Products). Cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 300 mM KCl, 4 mM imidazole, 10 mM MgCl₂, 10% (v/v) glycerol, 10 mM β-mercaptoethanol) in a nickel-coated beaker containing 1 mg/mL lysozyme, 100 μg/mL DNase, and 1 mM phenylmethylsulfonyl fluoride (PMSF). The suspension was stirred at room temperature for 30 min and lysed by sonication (QSonica Q500, 10 s pulse, 30 s rest, 60% amplitude) on ice. The resulting cell debris was pelleted by centrifugation at 11,300g at 4 °C for 1.25 h in a sealed centrifuge bottle. The supernatant was applied to Co²⁺-TALON resin. Protein fractions were eluted in 50 mM HEPES, pH 7.5, 300 mM KCl, 300 mM imidazole, 10 mM MgCl₂, 30% (v/v) glycerol, 10 mM β-mercaptoethanol. Reconstitution of [4Fe4S]²⁺ clusters was performed on ice for Cd-N and Cd C363A. Protein samples (200 μM) were exchanged into gel-filtration chromatography buffer [10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol] and incubated for 30 min at room temperature. Iron (III) chloride (800 μM) was added to the solution, which was subsequently incubated for 1 h at room temperature. Sodium sulfide was added to the solution (to a final concentration of 800 μM) in six equal volumes over the course of 2 h. The solution was allowed to incubate on ice overnight. A PD-10
column was used to remove excess iron and sulfide. Size-exclusion chromatography was performed using a Sephacryl S200 column. Iron and zinc occupancies were determined by inductively coupled plasma optical emission spectroscopy (ICP-OES) at the Pennsylvania State University Laboratory for Isotopes and Metals in the Environment (LIME). Cd-C contains 0.45 equiv iron, Cd-Z contains 0.36 equiv zinc, and full-length Cd C363A contains 2.53 equiv iron.

Mössbauer analyses of Cd Cfr variants. Cd-N and Cd-C $^{57}$Fe protein samples were transferred to EPR tubes or Mössbauer cups and flash-frozen in liquid nitrogen in an anoxic chamber (Coy Laboratory Products). Mössbauer spectra were recorded on constant acceleration Mössbauer spectrometers (SEE Co., Edina, MN) equipped either with a Janis SVT-400 variable-temperature cryostat (weak-field) or a Janis 8TMOSS-OM-12SVT variable-temperature cryostat (strong-field) (Janis, Woburn, MA). All isomer shifts are quoted relative to the centroid of the spectrum of $\alpha$-iron metal at room temperature. Simulations of Mössbauer spectra were carried out using WMOSS (SEE Co., Edina, MN). Some of the simulations are based on the spin Hamiltonian formalism (see equation below), in which the first term describes the electron Zeeman effect, the second and third terms describe the axial and rhombic zero-field splitting of the total electron spin ground state, the fourth term represents the interaction between the electric field gradient and the nuclear quadrupole moment, the fifth term describes the magnetic hyperfine interactions of the $^{57}$Fe nucleus with respect to the total electron spin ground state, and the last term represents the nuclear Zeeman interactions of $^{57}$Fe. All symbols are conventional. Spectra were calculated in the slow relaxation limit.
\[
H = \beta S \cdot g \cdot B + D \left( S_z^2 - \frac{S(S+1)}{3} \right) + E \left( S_x^2 - S_y^2 \right) \\
+ \sum_i \frac{e O V_{i,z,i}}{4} \left[ I_{z,i}^2 - \frac{1}{3}(I_{i+1} + I_i) \right] + \eta \left( I_{x,i}^2 - I_{y,i}^2 \right)
+ \sum_i S \cdot A_i \cdot I_i - \sum_i g \cdot B \cdot I_i
\]

EPR samples were analyzed on a Bruker ESP 300 spectrometer (Billerica, MA) equipped with an ER 041 MR microwave bridge and an ER 4116DM resonator.

*Synthesis and purification of RNA.* To generate a 155 nucleotide (nt) fragment of *Ec* 23S rRNA corresponding to A2454-G2608, previously shown to be efficiently methylated by *Sa* Cfr,\(^5\) a DNA template was amplified using primers 2454 and 2608 (Table 3-1), and transcribed in a reaction with T7 RNA polymerase (500 \(\mu\)g/mL). The transcription reaction was performed in 30 mM Tris, pH 8.4, 20 mM DTT, 36 mM MgCl\(_2\), 0.01% (w/v) Triton X-100, 2 mM spermidine, 2 mM NTPs (4 mM GTP). The template DNA PCR product was diluted to a final concentration of 5 ng/uL and incubated for 3 h at 37 °C with the components above. DNaseI was added to the solution (to a final concentration of 0.005 U/\(\mu\)L) and incubated at 37 °C for an additional hour. EDTA (pH 8.5) was added to a final concentration of 50 mM, and the solution was frozen at -20 °C for at least 2 h. Full-length 155mer RNA transcripts were purified under anaerobic conditions (Coy Laboratory Products) by PD-10 column equilibrated in filtered nanopure water. RNA concentrations were determined by UV-visible spectrophotometry (\(\varepsilon_{260} = 1.87 \ \mu\text{M}^{-1} \ \text{cm}^{-1}\)).\(^{22}\)

*RNA methylation activity assays.* The four variant Cfr proteins were tested for methylase activity. Assays using *Cd* Cfr in which the C-terminal domain of *Cd* Cfr (*Cd*-C) was removed (giving *Cd*-N) are indicated as such in activity measurement plots. *Cd*-C
was also added back separately at a 1:1 ratio. These samples are renamed Cd Fe res (iron-bound Cd-C) or Cd Zn res (zinc-bound Cd-C) in plots.

In experiments using a protein-based low-potential reducing system [flavodoxin (Flv), flavodoxin reductase (Flx), and NADPH], RNA methylation was assayed in a total volume of 150 μL at room temperature in an anoxic glovebox (Coy Laboratory Products) over a time course of 2.75 h. For substrate limiting conditions, samples contained 48 μM Cfr variant and 50 μM 155mer RNA in 100 mM sodium EPPS, pH 8.0, 10 mM MgCl₂, 12.5% (v/v) glycerol, 250 mM KCl, 200 μM Flv, 20 μM Flx, and 2 mM NADPH. Under excess substrate conditions, assays were performed as described above but with 100 μM 155mer RNA substrate and 3 μM Cfr enzyme. Solutions were incubated at room temperature for 5 min before SAM (1 mM) was used to initiate the reaction. Methylase activity was also assessed under substrate limiting conditions using the chemical reductant dithionite in place of Flv, Flx, and NADPH. The procedure was the same as for the Flv/Flx/NADPH substrate limiting conditions except SAM was added with the other components and dithionite (2 mM) was used to initiate the reaction. At designated time points, a 10 μL aliquot of the reaction solution was removed and added to 10 μL of quench solution containing 250 μM tryptophan and 95 mM H₂SO₄. Quenched reactions were diluted 1:1 in P1 nuclease buffer (250 mM sodium acetate, pH 6.0, 45 mM NaCl, 4 mM ZnCl₂). Antarctic phosphatase (10 U) and P1 nuclease (1 U) were added to digest the RNA into individual nucleosides. The digestion reactions were incubated at 37 °C overnight and subsequently analyzed by liquid chromatography (Agilent 1960/1990) coupled to mass spectrometry (LC-MS) (Agilent 6460 QQQ) (LC-MS). Assay mixtures were separated on an Agilent Extend-C18 column (4.6 × 50 mm) equilibrated in a 40
mM ammonium acetate, 5% (v/v) methanol (solvent A) / 100% methanol (solvent B) system using the conditions listed in Table 3-2. MS detection of products and assignment of peaks were accomplished as described previously. Results were analyzed with the MassHunter software package, and quantitative analyses were facilitated by comparison to a standard curve calculated from data acquired from the tryptophan internal standard. Unproductive cleavage of SAM was determined by dividing the number of SAM cleavage events, obtained from the concentration of 5'-dA in LC-MS samples, by the estimated number of methylation events ([m8A] + 2 • [m2,8A]). All flavodoxin assays were performed three times on different days while all dithionite assays were performed twice in the same manner.

Table 3-2. Liquid chromatography elution gradient

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 – 5.0</td>
<td>95%</td>
<td>5%</td>
</tr>
<tr>
<td>5.0 – 6.5</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>6.5 – 7.0</td>
<td>73%</td>
<td>27%</td>
</tr>
<tr>
<td>7.0 – 8.0</td>
<td>47%</td>
<td>53%</td>
</tr>
<tr>
<td>8.0 – 12.0</td>
<td>95%</td>
<td>5%</td>
</tr>
</tbody>
</table>

UV-visible spectrophotometry. Oxidation of the Cd Cfr C-terminal domain was analyzed spectrophotometrically using an Agilent 8453 G1103A spectrometer. In an anoxic glovebox, a 50 μL sample of 9.4 mM Cd-C [10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol] was diluted to 1 mL in water and sealed in a screw cap quartz cuvette (Starna Cells, Inc). The anoxic protein solution was
scanned from 190 nm to 1100 nm with a baseline correction at 800 nm. The sample was oxidized by opening the cuvette and pipetting the sample intermittently to introduce oxygen into the solution. The cuvette remained open for the duration of the experiment. The solution was scanned at the following times: 1 min, 2.5 min, 5 min, 7.5 min, 10 min, 20 min, and 30 min. A difference spectrum was calculated by subtracting the initial anoxic spectrum from the final spectrum.

**Co-elution of Cd-N/Cd-C and Cd-C/RNA.** A 2 mL solution of 365 µM Cd-N and 365 µM Cd-C in gel-filtration buffer was allowed to sit at room temperature for 15 min. The solution was loaded onto a Sephacryl S-200 column and the elution profile was monitored by UV at λ = 280 nm. For the Cd-C/RNA experiment, a 500 µL solution of 155mer RNA (100 µM) and Cd-C (100 µM) in gel-filtration buffer were allowed to sit at room temperature for 15 min. The solution was loaded onto a Superdex 75 10/300 GL column and 1 mL fractions were collected while UV signal was monitored at λ = 260 nm. Peak fractions were concentrated and tested for the presence of Cd-C by SDS-PAGE gel analysis.

**Protein film voltammetry.** Square-wave (SW) voltammetry experiments were performed on a VersaSTAT 3 (Princeton Applied Research) attached to a three-electrode electrochemical cell housed in an anaerobic chamber (Coy Laboratory Products). Protein films were generated via incubation of 9.4 mM Cd-C [10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol] on a pyrolytic graphite edge (PGE) working electrode for 10 min at room temperature. The working electrode surface was prepared by polishing the electrode surface with 0.3 µm alumina slurry (Allied High Tech) followed by washing in nanopure H₂O in a sonication bath for 6 min. The electrode
was dried and brought into the glovebox. All subsequent steps were performed under anoxic conditions. The background current was measured with the bare PGE working electrode, Ag/AgCl reference electrode (Pine Research), and Pt wire auxiliary electrode in 50 mM HEPES, pH 7.5, 300 mM KCl. SW voltammograms were collected using a step potential of 10 mV, a pulse amplitude of 25 mV, and a SW frequency of 10 Hz. A current range gate of 20 µA was applied and a 1 kHz filter was used as a background filter. A significant decrease in current amplitude was observed between the first and second scans with smaller decreases evident upon successive scans.

*Isoelectric point analysis of cysteine-rich KTR proteins.* To calculate the predicted isoelectric points of cysteine-rich KTRs, the full FASTA file for IPR025957 was pruned with the ExPASy decrease redundancy tool (<98% minimum identity), and the resulting FASTA file was used as input to the protein isoelectric point calculator web server (http://isoelectric.org/). The results for each individual KTR were averaged to obtain the overall predicted isoelectric point for the entire family.

**Results**

*Mössbauer analysis of Cd Cfr truncation variants.* In order to verify the [4Fe-4S] cluster content and interrogate the capacity of the Cd Cfr C-terminal domain to bind iron, I examined Cd Cfr truncation variants by Mössbauer spectroscopy. To test the hypothesis that the [4Fe4S]^{2+} cluster and rubredoxin-like center in Cd Cfr reside in the N-terminal and C-terminal domains, respectively, N-terminal (residues 1-345, Cd-N) and C-terminal (residues 346-416, Cd-C) fragments of Cd Cfr were produced. Mössbauer spectra of these truncated proteins (Figure 3-2) unambiguously demonstrate that the [4Fe4S] cluster
resides in the N-terminal radical SAM domain, whereas the ferrous rubredoxin-like center resides in the C-terminal cysteine-rich KTR domain. In order to gain additional insight into the electronic structure of the rubredoxin-like species in *Cd* Cfr, I examined *Cd*-C using Mössbauer spectroscopy at 4.2 K in variable externally applied magnetic fields (Figure 3-2, panel B). The spectra can be simulated with the following parameters and assuming a \( S = 2 \) electron spin ground state: \( D = 5.7 \text{ cm}^{-1} \), \( E/D = 0.2 \), \( \delta = 0.70 \text{ mm/s} \), \( \Delta E_Q = 3.08 \text{ mm/s} \), \( \eta = 0.9 \), and \( A/g_N\beta_N = [-14.4, -8.3, -23.3] \text{ T} \), similar to those previously reported for ferrous rubredoxin centers coordinated by four cysteines.\(^{31, 32}\) I proposed that the two CXXC motifs in the C-terminal domain bind \( \text{Fe}^{2+} \), identical to the first coordination sphere of rubredoxins. To explore the necessity of this domain for promoting Cfr-like activity, I performed activity assays with *Cd* Cfr with and without the domain.
Figure 3-2. Mössbauer spectra of Cfr enzymes. (A) Comparison of reconstituted samples of Cfr from *Bacillus amyloliquefaciens* (Ba), *Enterococcus faecalis* (Ef), *Paenibacillus lautus* (Pl), and *Clostridioides difficile* (Cd). *Cd* Cfr exhibits two distinct quadrupole doublets, consistent with two different iron-containing cofactors, each of which can be localized to the N-terminal (*Cd-N*) or C-terminal (*Cd-C*) domain of *Cd* Cfr. Features associated with the [4Fe4S]^{2+} cluster and ferrous rubredoxin-like site are denoted by red and blue dashed lines, respectively. (B) Mössbauer spectra of the C-terminal domain of *Cd* Cfr in varying applied magnetic fields. Experimental data are displayed as black vertical bars, with simulations overlaid as blue solid lines.

RNA methylation by *Cd* Cfr variants under limiting and excess substrate conditions. A truncated version of *Cd* Cfr, *Cd-N*, lacking the C-terminal Cys-rich domain, produces comparable amounts of m8A to the wild-type full-length *Cd* Cfr in endpoint assays using the Flv/Flx/NADPH reducing system with limiting RNA (Figure 3-3, panels A-B). Attempts to modulate activity by addition of the isolated Fe^{2+}-loaded C-
terminal domain in trans (Cd Fe res) did not yield any significant difference in product formation (Table 3-3). To investigate the functional role of the Fe$^{2+}$ center in the Cd Cfr C-terminal domain, a variant of one of the putative metal-binding residues in the CXXC motif, a Cys$^{363} \rightarrow$ Ala substitution (Cys$^{363}$Ala, renamed Cd M), was generated in the full-length enzyme. Unexpectedly, this substitution alters [4Fe4S]$^{2+}$ cluster stability, giving a sample with < 3 equiv Fe/protein after reconstitution, as determined by ICP-OES analysis. Similar cluster instability is precedented in [4Fe4S]$^{2+}$ DNA repair enzymes upon mutation of an auxiliary metal binding site Cys ligand.$^{33, 34}$ As a consequence, the Cys$^{363}$Ala Cd Cfr enzyme exhibits negligible activity. A Zn$^{2+}$-loaded version of the C-terminal domain was also tested (Cd Zn res) for rescue of RNA methylation activity. Inclusion of the mismetallated Cd-C domain gives a slight decrease in activity in limiting substrate assays (Figure 3-3, panel B). The Zn-loaded Cd-C slows the formation of m8A during the initial burst phase of activity compared to the Fe-loaded enzyme (Figure 3-3, panel E).

I also evaluated the activity of variant forms of Cd Cfr with excess RNA (3 μM enzyme, 100 μM 155mer rRNA) (Figure 3-3, panel C). The trends observed in the limiting substrate assays are largely maintained, with little effect on endpoint product formation upon truncation of Cd Cfr (Table 3-4). However, at early time points, a detectable lag in product formation is evident in the rescue experiment (Figure 3-3, panel F). Interestingly, the impact of Zn$^{2+}$-substitution in the C-terminal domain is magnified with excess substrate. I observe only a small amount of product after 2.75 h (Figure 3-3, C), underscoring the initial findings from the limiting substrate assays.
Figure 3-3. Methylation of a 155mer rRNA substrate by variant Cd Cfr enzymes under limiting substrate and excess substrate conditions and initiated by a protein-based reductant. Reactions were performed with the Ec Flv/Flx/NADPH reducing system in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5% glycerol, 250 mM KCl, 200 µM Flv, 20 µM Flx, and 2mM NADPH. (A) Limiting substrate LC-MS analysis of methylated RNA products via single-ion monitoring at \( m/z = 282.1 \) (bottom trace, m8A) and \( m/z = 296.1 \) (top trace, m2,8A). Cd-N = radical SAM domain only, Cd-C = C-terminal KTR domain with Fe\(^{2+}\) or Zn\(^{2+}\), Cd-M = C363A variant (full-length), rescue (res) experiments contain Cd-C added to Cd-N in trans. Endpoint analysis (165 min) of the formation of singly-methylated m8A under (B) limiting substrate and
(C) excess substrate conditions. Inset shows the deviation in reaction profile of the assay using the Zn-bound Cd-C. Error bars correspond to standard deviation of three averaged trials. (D) Ratio of 5’-dA to total methylation events ([\(m8A\] + 2 • \([m2,8A]\)) at the initial time points under limiting substrate conditions using the protein-based reducing system. Deviation from a 1:1 ratio indicates a greater amount of unproductive SAM cleavage. (E) Limiting substrate and (F) excess substrate conditions showing expanded view of initial lag in activity for each variant of Cd Cfr.

The C-terminal domain is not essential for activity under the *in vitro* assay conditions, but my work indicates that iron may be required for proper function when the domain is present. Disruption of the C-terminal domain iron-binding site via ligand mutation or metal ion substitution impedes turnover. Metal coordination may also help maintain the structural integrity of the Cd Cfr radical SAM cluster. The Cys363Ala mutation yields low iron content (< 2.5 per protein), likely indicating loss of the radical SAM cluster in this variant.

Table 3-3. Metrics for Cd Cfr variant limiting substrate assays using protein-based reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cd</th>
<th>Cd-N</th>
<th>Cd Fe res</th>
<th>Cd Zn res</th>
<th>Cd-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>60 ± 1</td>
<td>64 ± 6</td>
<td>50 ± 10</td>
<td>44 ± 8</td>
<td>6 ± 3</td>
</tr>
<tr>
<td>5’-dA:mA</td>
<td>1.7 ± 0.2</td>
<td>1.8 ± 0.2</td>
<td>2.0 ± 0.2</td>
<td>2.0 ± 0.2</td>
<td>--</td>
</tr>
<tr>
<td>SAH:mA</td>
<td>1.1 ± 0.2</td>
<td>1.1 ± 0.2</td>
<td>1.2 ± 0.6</td>
<td>1.05 ± 0.04</td>
<td>--</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>(k_{\text{obs}}) m8A (min(^{-1}))</td>
<td>0.018</td>
<td>0.018</td>
<td>0.023</td>
<td>0.029</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 3-3. Metrics for Cd Cfr variant limiting substrate assays using protein-based reducing system.
Table 3-4. Metrics for Cd Cfr variant excess substrate assays using protein-based reducing system

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cd</th>
<th>Cd-N</th>
<th>Cd Fe</th>
<th>Cd Zn</th>
<th>Cd-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>5 ± 2</td>
<td>6 ± 1</td>
<td>7 ± 3</td>
<td>4 ± 1</td>
<td>0.2 ± 0.2</td>
</tr>
<tr>
<td>5'-dA:mA</td>
<td>1.7 ± 0.2</td>
<td>1.8 ± 0.2</td>
<td>1.5 ± 0.3</td>
<td>2.4 ± 0.4</td>
<td>--</td>
</tr>
<tr>
<td>SAH:mA</td>
<td>3.1 ± 0.8</td>
<td>2.4 ± 0.7</td>
<td>3 ± 1</td>
<td>4.9 ± 0.9</td>
<td>--</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>v/[E] (min⁻¹)</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>number of turnovers</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

If the \( Cd \) C-terminal domain mediates electron transfer between a protein-based reductant and active site elements, I hypothesized that the reaction profile and/or outcomes could change significantly when a small chemical reductant, dithionite, is used instead. To test this hypothesis, limiting substrate activity assays were performed with dithionite. As in Chapter 2, it should be noted that the dithionite assays were performed with a mix of full length and truncated rRNA substrate.

Compared to assays with a protein-based reductant, all Cfr variants exhibited a decrease in m8A, an increase in m2,8A, and an increased ratio of unproductive SAM cleavage (Figure 3-4, panel A; Table 3-4). Unproductive SAM cleavage was measured in two different ways. I analyzed the ratio of 5'-dA: mA in a time-dependent fashion, as the slope of a plot of this ratio as a function of reaction time. In the dithionite assays, these values did not deviate significantly from those observed with the protein-reducing system in the initial timepoints (Figure 3-4, panel B). The most significant change is that the
slope of a plot of 5’-dA versus mA products for the full-length Cd Cfr reaction increased from 1.4 to 2.1. I also analyzed the proportion of 5’-dA:mA at the end of the 2.75 h assay. By this measure, the variants showed a 1.5x increase in unproductive SAM cleavage compared to the protein-based reducing system (Figure 3-4, panel C). The full-length protein showed an even larger 8-fold increase in uncoupled SAM cleavage with dithionite (compared to assays with Flv/Flx) when measured at the end of the assay period. The variation in activity is likely due to variable activity in different protein preps. Two different preparations of the full-length Cd protein were used in four separate assays, while only a single preparation of each variant was used to perform two independent trials.
Figure 3-4. (A) Endpoint analysis of methylation of a 155mer rRNA substrate by variant Cd Cfr enzymes under limiting substrate conditions and initiated by dithionite. Reactions were performed in mixtures containing 100 mM EPPS, pH 8, 10 mM MgCl₂, 12.5 % glycerol, and 250 mM KCl. LC-MS analysis of methylated RNA products was accomplished via single-ion monitoring at m/z = 282.1 (solid bars) and at m/z = 296.1 (shaded bars). Cd-N = radical SAM domain only, Cd Fe res = C-terminal KTR domain with Fe^{2+}, Cd Zn res = C-terminal KTR domain with Zn^{2+}, Cd-M = C363A variant (full-length). Rescue (res) experiments contain Cd-C added to Cd-N in trans. Error bars correspond to standard deviation of three averaged trials. (B) Ratio of 5’-dA to total methylation events ([m8A] + 2 • [m2,8A]) at the initial time points under limiting substrate conditions using dithionite. Deviation from a 1:1 ratio indicates a greater amount of unproductive SAM cleavage. (C) Comparison of the protein-based reducing system and dithionite unproductive SAM cleavage ratios at the end of the reaction. (D) Comparison of m8A and m2,8A produced using the protein-based reducing system (Flv) and dithionite (Dith).

The total methylated product produced by full length Cd Cfr and each variant stayed relatively constant irrespective of the reducing agent (Figure 3-4, panel D).
largest difference was exhibited by CD-N which produced 32 µM mA using the protein-based reducing system and 22.5 µM mA using dithionite. CD-C rescue variants CD Fe res and CD Zn res showed comparatively smaller decreases in total methylated product when comparing the protein-based reducing system to dithionite, from 25 µM to 21.5 µM and 22.0 µM to 21.0 µM, respectively. The modest reduction in total methylated product could be a result of the presence of the smaller RNA fragment in the dithionite assays, as a shorter substrate length is known to negatively impact turnover in class A RS methylases.35 However, as with the full-length homologs assayed with dithionite in chapter 2, significant levels of dimethylated m2,8A were formed by each variant with a chemical reductant as the reaction initiator. The wt and variant enzymes were each able to convert ~25% of the m8A product pool to m2,8A. As with the full-length Cfr homologs, I speculate that the chemical reductant can re-reduce the [4Fe-4S] cluster quickly prior to m8A RNA release, allowing the enzyme to act processively. In these assays with dithionite, the CD-M variant, again, showed very little activity.

Table 3-5. Metrics for CD Cfr variant limiting substrate assays using dithionite

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cd</th>
<th>Cd-N</th>
<th>Cd Fe res</th>
<th>Cd Zn res</th>
<th>Cd-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA cons (%)</td>
<td>64 ± 18</td>
<td>45 ± 1</td>
<td>43 ± 2</td>
<td>42.0 ± 0.3</td>
<td>2.30 ± 0.07</td>
</tr>
<tr>
<td>5'-dA:mA</td>
<td>14 ± 9</td>
<td>3.1 ± 0.1</td>
<td>3.33 ± 0.03</td>
<td>3.18 ± 0.1</td>
<td>187 ± 2</td>
</tr>
<tr>
<td>m8A:m2,8A</td>
<td>1.7 ± 0.2</td>
<td>2.8 ± 0.2</td>
<td>2.51 ± 0.07</td>
<td>2.49 ± 0.01</td>
<td>7 ± 1</td>
</tr>
</tbody>
</table>

UV-visible spectroscopy and square wave voltammetry of CD-C. The activity assays do not indicate an essential in vitro catalytic role for the metal-bound C-terminal
domain in \textit{Cd} Cfr, particularly under limiting substrate conditions. However, I speculated that the domain could mediate electron transfer (ET) steps in the rRNA methylation reaction. This function may be dispensable with a non-native reductant \textit{in vitro} but more critical \textit{in vivo} (e.g., with a clostridial protein-based redox donor). The Mössbauer experiments revealed a Rdx-type Cys$_4$-coordinated Fe$^{2+}$ center in the version of this protein isolated from an \textit{E. coli} heterologous host and expressed under iron-replete conditions in minimal medium. Given that Rdx proteins typically serve as high-potential electron-transfer shuttles that operate by cycling between Fe$^{3+}$ and Fe$^{2+}$ oxidation states, I sought to characterize the redox properties of as-isolated \textit{Cd}-C. Anaerobic isolation of \textit{Cd}-C yields a clear/bluish protein solution that is exclusively Fe$^{2+}$ by Mössbauer analysis. Exposure of the domain to O$_2$ results in a visible color change to a red solution that, when monitored spectrophotometrically, exhibits absorption bands at 370, 485, and 582 nm (Figure 3-5, panel A). These spectral features resemble those of oxidized stand-alone Rdx proteins.$^{36}$

Previously characterized Rdx proteins also exhibit reduction potentials in the -95 - +114 mV range (vs. NHE). To measure the Fe$^{3+/2+}$ midpoint potential of the isolated Cd-C domain, protein films were generated on pyrolytic edge graphite (PGE) electrodes under anaerobic conditions by incubation of protein [9.4 mM in 10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl$_2$, 25\% (\textit{v/v}) glycerol] on the electrode surface for 10 min. Modified electrodes were transferred to an electrochemical cell (50 mM HEPES, pH 7.5, and 300 mM KCl buffer), and evaluated by square-wave (SW) voltammetry. Voltammograms obtained at a SW frequency of 10 Hz showed a quasi-reversible waveform with a midpoint potential of +106 mV vs. NHE (Figure 3-5, panel
B). The relatively high midpoint potential compared to typical Rdx proteins can be rationalized by differences in the predicted charge of Cd-C at pH 7.5.

Cd-C has a calculated pI of 9.2, compared to ~4.0 for a canonical Rdx. At physiological pH, Rdx would be largely negative while Cd-C would have a net positive charge. This scenario would destabilize the Fe$^{3+}$ state in Cd-C, rendering the protein more difficult to oxidize and resulting in a higher midpoint potential. Interestingly, subsequent electrochemical scans of the isolated domain show loss of current amplitude over time, suggesting that the domain may not be stable in the oxidized form. In the presence of O$_2$,
Cd-C visibly precipitated over the 30 min time period in which the protein was monitored electrochemically. These observations could preclude a role for the Fe$^{2+}$ center as an explicit electron donor in Cfr catalysis. However, the stability/reduction potential may be further modulated via interaction with the radical SAM domain of Cd Cfr and/or the RNA substrate. The protein could also require binding to the protein reducing system prior to ET to ensure rapid re-reduction to the Fe$^{2+}$ oxidation state.

**Size exclusion determination of Cd-N/Cd-C and Cd-C/RNA co-elution.** I tested the ability of the Cd-C rubredoxin-like domain to form stable complexes with the Cd-N radical SAM domain and RNA by gel filtration chromatography. The size exclusion profile of the Cd-N/Cd-C mixture revealed a new peak with elution volume of 64 mL, consistent with the expected value for the full-length protein (Figure 3-5, panel C). A smaller peak is visible at 72 mL, the predicted elution volume of free Cd-N. Interaction between the two domains would be consistent with related function of the Cd-C and Cd-N domains, including use of the Cd-C domain as an electron transfer partner or substrate-binding domain.
Figure 3-6. Co-elution of RNA and \( Cd-C \) by size exclusion chromatography. (Left) UV visible trace (\( \lambda = 260 \text{nm} \)) of RNA-containing peaks and associated fractions. (Right) SDS-PAGE analysis of \( Cd-C \) presence in RNA fractions. Ladder is in KDa. KTRs appear at slightly higher molecular weights than expected (~8 KDa), likely due to the unusually high pI.

Given the large number of positively charged amino acids in the C-terminal domain, I also speculated that the domain could interact with the RNA substrate for Cfr. The size exclusion chromatography elution profile of a solution of \( Cd-C \) and RNA was examined for co-elution of the two molecules (Figure 3-6). The elution profile for this experiment monitored at 260 nm showed three distinct peaks. The fractions corresponding to those peaks (7, 8-9, 10) were analyzed by SDS-PAGE gel for the presence of \( Cd-C \). Fractions 8-9 and fraction 10 contained a band corresponding to \( Cd-C \). Because \( Cd-C \) is much smaller than the RNA fragment, it is unlikely that the protein would co-elute with the RNA unless it was bound. These results suggest that \( Cd-C \) can independently bind to the 155mer ribosomal RNA fragment containing A2503.

**Sequence similarity and genomic analyses of Cys-rich KTRs.** As detailed in chapter 4, the C-terminal domain of \( Cd \) Cfr is homologous to a large family of stand-
alone uncharacterized open-reading frames termed Cys-rich KTRs. The latter naming scheme comes from an Arg-Thr-Lys sequence that resides in between the two strictly conserved CXXC motifs. Additionally, analysis of the Cys-rich KTR sequence group shows that the stand-alone proteins are also dominated by positively charged residues, resulting in an average isoelectric point of 8.6. Spectroscopic and electrochemical analyses of the isolated Cfr C-terminal domain suggest that Cys-rich KTRs might be similar to rubredoxin (Rdx) electron transfer proteins in function and metal-binding properties. Although traditional Rdx proteins are similar in length (~60 aa) and contain identical CXXC metal binding motifs separated by ~20-30 amino acids, the overall sequence identity between the Cd-C domain and a model Rdx from C. pasteurianum is only 30% (Figure 3-5, panel D). Cp Rdx (and other Rdx proteins with validated function/properties) additionally contain a conserved Gly immediately adjacent to the second Cys residue in each metal binding motif. This sequence pattern enables key H-bonds to form between peptide backbone N-H groups and the metal-coordinated thiolates of Cys residues. These interactions can influence Rdx reduction potential. The rubredoxin Gly motifs also maintain structural integrity. Mutation to alanine perturbs the structure of the cysteine loops and alters the reduction potential by 42 mV. Interestingly, Cd-C deviates from this sequence pattern significantly because it contains a Lys residue immediately adjacent to each CXXC motif. KTRs, in general, conserve either lysine or arginine at these sites. Overall, Cd-C has a high proportion of positively charged residues, accounting for its aforementioned pi > 9, while Cp Rdx has a net negative charge at neutral pH. These observations leave open the possibility that while Cd-C functionally resembles Rdx proteins, it may adopt a distinct structure.
Discussion

In chapter 2, I showed *in vitro* characterization of a clostridial Cfr homolog fused to a Cys-rich C-terminal domain. The full-length protein is an active mono-methylase with limiting substrate. However, with a >30-fold excess of RNA substrate, the clostridial fusion enzyme is markedly less active than other Cfr homologs. Combined with the unexpected detection of a rubredoxin-like Fe$^{2+}$ binding site in the C-terminal domain of *Cd* Cfr, these findings prompted a more detailed investigation of the properties of the C-terminal domain. In this chapter, I show that complete removal of the domain fails to either increase or decrease the rate of RNA methylation in reactions initiated by a protein-based reducing system, but disruption of the Fe$^{2+}$ binding site by ligand mutagenesis or Zn$^{2+}$ substitution decreases turnover. While the effect of Zn$^{2+}$ is modest with limiting substrate, the mismetallated enzyme is nearly inactive when challenged with excess RNA (Figure 3-8, panel A). The spectroscopic and electrochemical analysis of the stand-alone domain suggests a similarity to rubredoxin-type electron transfer.
cofactors, often described as a particularly minimalist member of the iron-sulfur cluster cofactor family.

Many radical SAM enzymes harbor additional iron-sulfur clusters with diverse functions, including electron transfer.\textsuperscript{1,40} The redox-neutral carbon methylation catalyzed by RlmN/Cfr enzymes shares in common with other radical SAM-mediated C-C bond forming systems a requirement for reductive quenching of a substrate radical at the end of the reaction.\textsuperscript{41} In other superfamily members, auxiliary cofactors or redox-active amino acids in the active site can mediate controlled provision of redox equivalents.\textsuperscript{42,43} In Cfr and RlmN, a conserved C-terminal Cys residue is oxidized to a thiyl radical to facilitate product release.\textsuperscript{44,25} The midpoint potential of the Fe\textsuperscript{2+} cofactor in the \textit{Cd} Cfr C-terminal rubredoxin-like domain could enable its use as an electron donor to regenerate the Cys\textsubscript{355} thiolate for subsequent methylation.\textsuperscript{41} I note, however, that the rubredoxin-like domain is not universally conserved in Cfrs. Therefore, a KTR-type rubredoxin cannot be the sole mechanism of Cys\textsubscript{355} reduction and other means to quench the radical must exist in other Cfr/RlmNs. Diversity in reductant usage has precedent in other oxidoreductase enzymes. For example, class III ribonucleotide reductases have adapted to use either protein-based (thioredoxin or ferredoxin) or small molecule (formate) reducing systems to resolve an oxidized active site Cys disulfide or a thiosulfuranyl radical.\textsuperscript{44–46}

Reduction of an intermediate thiyl radical has precedent in other radical SAM enzymes, including the C-C bond forming enzyme PolH and epimerase NeoN. PolH is mechanistically similar to Cfr in its use of a Cys side chain to quench a substrate carbon radical. PolH lacks an additional redox cofactor, and the resulting thiyl radical is proposed to be shuttled to the protein surface through a network of internal Tyr side
chains. NeoN performs a redox-neutral stereoinversion of a chiral carbon in a carbohydrate moiety of a complex aminoglycoside natural product. Mechanistic studies show that NeoN uses a Cys side chain as an H-atom donor, yielding a thiyl radical that must be reductively quenched and protonated for additional turnover. The structure of NeoN has not been reported, but sequence analysis suggests that NeoN harbors an auxiliary iron-sulfur cluster in a C-terminal SPASM domain. This feature of the enzyme suggests that NeoN could operate similarly to my proposal for Cd Cfr, by using an auxiliary redox cofactor to reduce a thiyl radical produced at the end of its redox-neutral reaction. A subsequent study of other radical SAM peptide epimerases suggests a similar requirement for an active site Cys as H-atom donor. In this work, the proposed mechanism of thiyl radical resolution involves a direct coordination interaction between the auxiliary [4Fe-4S] cluster and the Cys harboring the thiyl radical. While this feature of the reaction pathway in the epimerases awaits experimental validation, it is interesting to consider a similar possibility in the Cfr-KTR fusions. In this model, a Cys338-centered radical would temporarily replace one of the KTR Cys ligands. Such a ligand swap in a rubredoxin is unprecedented to my knowledge, but the proximity of the KTR domain and the C-terminal catalytic Cys in Cd Cfr renders this phenomenon at least feasible from a structural perspective.
Figure 3-8. (A) Proposed role of the KTR domain in electron transfer leading to activity differences. The iron-bound form can donate an electron to reduce the residual thiyl radical while the zinc-bound form cannot. (B) Diagram showing possible intermediates in the resolution of the terminal Cys338 thiyl radical wherein the KTR acts as an electron donor. Because of observed instability in the oxidized state, re-reduction of the domain is proposed to be rapid. Methylated product is then lost and the Cfr is subsequently prepared for the next turnover.

I cannot rule out other roles for the KTR-rubredoxin such as RNA-binding or structural stabilization. Co-elution of the Cd-C domain with the 155mer Cfr RNA substrate is consistent with an RNA-binding function. The Cd C-terminal domain and/or stand-alone KTR proteins may interact with ribosomal RNA to improve access to A2503 in Clostridia. I do not know if Cd-C binds specifically to A2503 or a nearby site or if RNA co-elution is due to non-specific electrostatic interaction with nucleic acid. With respect to a structural stabilization role, my work shows that the Cd-N truncated protein
is stable and active in the absence of the C-terminal domain in vitro, suggesting that the domain is not essential for structural stability of Cd Cfr.

The Cd Cfr rubredoxin-like domain may also serve as an intermediary or coordinator of Cfr reaction initiation via interaction with protein reducing partners. The reduction potential of the Cd-C domain is likely too high (+100 mV vs NHE) to donate an electron to the [4Fe4S]^{2+} radical SAM cluster directly. Native radical SAM reducing systems, including ferredoxins and flavodoxins, are low-potential electron donors. The most common protein-based reducing system for in vitro assays is flavodoxin, with a reduction potential of approximately -430 mV. Ferredoxins, which are endogenous reductants of radical SAM proteins such as MiaB, also have reduction potentials close to -400 mV. Interestingly, a flavodoxin/ferredoxin-like mediator role was recently proposed for a stand-alone CSL-family Zn finger protein, Dph3, encoded adjacent to a non-canonical radical SAM enzyme, diphthamide biosynthesis protein Dph2. Dph3 is the only other example of a biochemically characterized mononuclear metal binding protein found as a genome neighbor to a radical SAM enzyme. The Fe^{2+}-bound form of Dph3 was reported to function in iron-sulfur cluster reduction in Dph2, initiating the reaction in the absence of reductant. However, the redox potential of Dph3 has not been experimentally determined, and no known rubredoxin-type electron transfer protein has a sufficiently low reduction potential (< -400 mV) for activation of a radical SAM [4Fe4S] cluster.

My findings also raise the possibility that reported inactivity in RNA methylation in heterologously expressed Cfrs from the clostridial clade could be due to lack of a necessary accessory factor or redox partner. The Cys-rich KTR domain is a candidate for
such a factor, although this protein is not essential for activity in *Cd* Cfr and it is not found universally adjacent to the inactive clostridial enzymes. In chapter 4, I perform more detailed bioinformatics and experimental analysis of the stand-alone KTR proteins and the clostridial Cfrs to test this hypothesis.
Table 3-1. RNA sequence (and associated primer sequences) for the 155mer substrate with the A2503 equivalent residue in teal. *Clostridioides difficile* Cfr protein and nucleotide sequences and PCR primers for truncation variants.

<table>
<thead>
<tr>
<th>rRNA substrate DNA template A2454 – G2608 (E. coli)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGAAAATTAATACGACCTCACTATAGGCCGTGATACCCGACGACCCGACGCTGATGGGTCTTGGCACC</td>
</tr>
<tr>
<td>TCGATTGTCCGTCGCTCAGTAATCTCCGGGCTGATGAATAGCGTCGACGACGACGCTGATGGGTCTTGGCACC</td>
</tr>
<tr>
<td>GTCATTGTCCGTCGCTCAGTAATCTCCGGGCTGATGAATAGCGTCGACGACGACGCTGATGGGTCTTGGCACC</td>
</tr>
<tr>
<td>GTCATTGTCCGTCGCTCAGTAATCTCCGGGCTGATGAATAGCGTCGACGACGACGCTGATGGGTCTTGGCACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rRNA substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCGUGAUAACCGCCCAAGUCAUAUUCGACCCCGGUGUUUUGGACACCUGCCGUGGUACUCAUCAUCUG</td>
</tr>
<tr>
<td>GGGCUGAUAACCGCCCAAGUCAUAUUCGACCCCGGUGUUUUGGACACCUGCCGUGGUACUCAUCAUCUG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PCR primers for 155mer RNA substrate amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer (2454): 5′-CGGAAATTAATACGACCTCACTATAGGCCGTGATACCCGACGACGACGCTGATGGGTCTTGGCACC-3′</td>
</tr>
<tr>
<td>Reverse Primer (2608): 5′-mCmCGAACTGTCTCACGACGTTCTAAACC-3′</td>
</tr>
</tbody>
</table>

**Clostridioides difficile** str. QCD63q42 Cfr

<table>
<thead>
<tr>
<th>PCR primers for <em>Clostridioides difficile</em> str. QCD63q42 Cfr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer: 5′-GACATTCATATGAGCAAATACAAAAAAATG-3′</td>
</tr>
<tr>
<td>Reverse Primer: 5′-GACATTCTCGAGACGCTGATGGGTCTTGGCACC-3′</td>
</tr>
</tbody>
</table>

**PCR primers for *Clostridioides difficile* str. QCD63q42 Cfr N-terminal domain (Cd-N)**

<table>
<thead>
<tr>
<th>PCR primers for <em>Clostridioides difficile</em> str. QCD63q42 Cfr N-terminal domain (Cd-N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer: 5′-GACATTCATATGAGCAAATACAAAAAAATG-3′</td>
</tr>
<tr>
<td>Reverse Primer: 5′-GACATTCTCGAGACGCTGATGGGTCTTGGCACC-3′</td>
</tr>
</tbody>
</table>

**PCR primers for *Clostridioides difficile* str. QCD63q42 Cfr C-terminal domain (Cd-C)**

<table>
<thead>
<tr>
<th>PCR primers for <em>Clostridioides difficile</em> str. QCD63q42 Cfr C-terminal domain (Cd-C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer: 5′-GACATTCATATGAGCAAATACAAAAAAATG-3′</td>
</tr>
<tr>
<td>Reverse Primer: 5′-GACATTCTCGAGACGCTGATGGGTCTTGGCACC-3′</td>
</tr>
</tbody>
</table>
PCR primers for *Clostridioides difficile* str. QCD63q42 Cfr C363A mutant (Cd-M)
Forward Primer: 5’-CTCTTGCCCTGTGTGCAAG-3’
Reverse Primer: 5’-CCATTCAAAAGACTGTTCTTGTAT-3’
References


CHAPTER 4

Defining the properties and functions of Clostridial Cfr-associated proteins
Abstract

In Cfrs from clostridia the ability to catalyze the C8 methylation of A2503 in 23S rRNA is limited or completely lacking indicating that additional reaction components may remain unidentified. In chapter 3, I demonstrated that a C-terminal extension of a *Clostridioides difficile* (Cd) Cfr homolog coordinates iron and binds RNA, suggesting a role in electron transfer and/or substrate positioning. Further bioinformatic analysis revealed that the domain was a member of a large group of uncharacterized standalone proteins named Cys-rich KTR. In this chapter, I show that nearly all clostridial cfrs are proximal to a conserved genetic neighbor, either a Cys-rich *ktr* or a *lsa* family ABC-F type ribosomal protection element. Because in bacterial organisms fused protein domains and conserved genome neighbors are often functionally linked, I explored whether KTR or Lsa proteins might augment the RNA methylation activity of Cfrs. I also characterized *Sa* KTR, a standalone KTR that is not associated with a radical SAM RNA methylase. I showed that this standalone protein binds iron and RNA, as I saw for the C-terminal domain of *Cd* Cfr. The Cys-rich KTR protein family could therefore represent a novel group of metalloproteins that function in electron transfer or bind nucleic acids. The results suggest a previously unappreciated connection between Cfr ribosomal RNA methylases, ribosomal protection proteins, and a new family of small rubredoxin-like iron-binding proteins.
Introduction

In this chapter I seek to understand how Cfrs from clostridial organisms function in their native hosts. I address this question by bioinformatic analysis of clostridial cfr genome neighbors, purification and characterization of conserved putative partner proteins, and investigation of the role of these neighboring proteins/genes in Cfr activity. Work by my group and others has shown that clostridial Cfrs are often problematic to characterize, being difficult to purify and having seemingly sporadic propensities for performing the C8 methylation of rRNA associated with antibiotic resistance.\textsuperscript{1-5} These findings suggest either that these Cfrs catalyze alternative chemistry in their native hosts or that other factors are required for efficient methylation of ribosomal RNA.

In chapters 2 and 3, I report in vitro characterization of Clostridioides difficile (Cd) Cfr and the discovery of a novel rubredoxin-type iron-binding site in its distinctive C-terminal domain.\textsuperscript{6} Although the domain is not essential for enzyme activity, mismetallation with zinc nearly eliminates turnover when an RNA substrate is present in excess. I propose two possible roles for this domain. The rubredoxin site could provide a necessary reducing equivalent at the end of the redox-neutral C-C bond formation catalyzed by Cfr. The positively-charged domain could also bind to RNA. These findings prompted a closer examination of the genomic context of this clade of Cfr enzymes. While selected Cfrs from clostridial organisms have previously been analyzed for RNA methylation activity, the role of genome neighbors in enzyme activity has not been studied systematically.\textsuperscript{2-5}

The initial bioinformatic analysis of the entire cfr family did not reveal any universally conserved genome neighbors. Focusing the query to the clostridial cfrs,
however, uncovered a more defined set of conserved neighboring genes. The clostridial clade splits into two distinct branches, each with its own absolutely conserved genome neighbor. These open reading frames are located directly adjacent to the \textit{cfr} or, in a few cases, elsewhere in the host genome. The Cfr branch containing sequences from pathogenic \textit{Clostridioides} Cfrs, including the \textit{Cd} Cfr-KTR fusion sequence, are often found adjacent to a family of uncharacterized proteins termed Cys-rich KTRs (IPR025957). The remaining clostridial Cfrs are neighbors to a conserved Lsa family ABC-F type ribosomal protection protein.

A metagenomics assessment of KTR host organisms reveals that these proteins are often found in commensal clostridial gut bacteria and human staphylococcal pathogens. Bioinformatic analysis of nearby open reading frames indicates that a significant portion of standalone KTR proteins are found adjacent to putative radical SAM methylases, likely Cfrs. While KTRs have no known function, it is likely that their proximity to Cfr signals a role in antibiotic resistance. The conservation of KTR/Cfr pairs suggests that \textit{Cd}-C does have a functional role in Cfr activity and that the role is likely shared amongst its associated protein family. Additionally, the analyses indicate that a previously unknown reservoir of Cfr-like proteins exist in the human microbiome. \textit{In vitro} characterization of a KTR from \textit{Staphylococcus aureus} affirmed the observations first seen in \textit{Cd}-C that these proteins bind iron and represent a new class of metalloenzyme. The standalone KTR was also shown to interact with the 155mer ribonucleotide strand used in assessment of Cfr activity suggesting that the KTR protein family is involved in RNA binding.
My work uncovers a previously unappreciated link between radical SAM RNA methylation enzymes, ribosomal protection proteins, and a putative new family of iron-binding proteins. I propose that these three systems could work together to help beneficial and pathogenic host-associated microbes survive in the face of antibiotic treatment. Model Cfrs from Staphylococcal pathogens are well characterized in vitro.\textsuperscript{9-14} I show in chapter 2 that Cfr enzymes from Enterococcal sources have similar properties in vitro. Here I provide a roadmap to similar understanding of Cfr activity in clostridial organisms relevant to human health. I have identified genomically conserved neighbors of Cfr in these organisms and made inroads into determining the function of these partner proteins. Such information may ultimately allow us to understand the role of clostridial cfr genes in antibiotic resistance, opening up new avenues for treatment of \textit{C. difficile} and other bacterial gut infections.

\textbf{Materials and Methods}

\textit{Materials.} Kanamycin was obtained from Teknova (Hollister, CA). Ampicillin was purchased from Dot Scientific (Burton, MI). \textit{N}-(2-hydroxyethyl)piperazine-\textit{N}′-2-ethanesulfonic acid (HEPES), DNaseI (RNase-free), Bradford reagent, acetonitrile, ammonium bicarbonate, and His-Pur Co\textsuperscript{2+}-TALON resin were obtained from Thermo Fisher Scientific (Waltham, MA). Nickel (II) sulfate hexahydrate, 2-mercaptoethanol, glucose, iodoacetamide, and formic acid were purchased from Sigma Aldrich (St. Louis, MS). Iron (III) chloride hexahydrate was acquired from BDH Analytical Chemicals, VWR (Radnor, PA). DL-dithiothreitol (DTT) was obtained from Alfa Aesar (Tewksbury, MA). Bovine serum albumin (BSA), \textit{XhoI}, \textit{NdeI}, Q5 DNA polymerase, and T4 DNA
ligase were obtained from New England Biolabs (Ipswich, MA). PD-10 and HiPrep Sephacryl 16/60 S200 columns were purchased from GE Healthcare (Little Chalfont, United Kingdom). Trypsin was obtained from Promega (Madison, WI). Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) was purchased from GoldBio (St. Louis, MO). *E. coli* BL21 DE3 ΔtolC, *E. coli* BL21 DE3 ΔaraBAD/ΔtolC, Falcon 96-well clear culture plates, lincomycin, MOPS mix, ACGU, and supplement EZ were generously provided by the laboratory of Prof. Kenneth Keiler (Pennsylvania State University).

**Cloning and expression of Cys-rich KTRs and Lsa.** PCR primers for *Clostridium botulinum* (Cb) KTR, *Staphylococcus aureus* (Sa) KTR, and *Clostridium sporogenes* (Cs) Lsa (Table 4-1) were designed with Primer3 and appended at the 5’-end with *NdeI* and *XhoI* restriction sites. The coding sequences were amplified by PCR using Q5 DNA polymerase. The PCR products were digested using *NdeI* and *XhoI* restriction enzymes, purified by gel electrophoresis, re-isolated using an Omega Biotek gel extraction kit, and ligated into pET26b. The plasmid additionally encodes a C-terminal His$_6$-tag with a short two amino acid (-Leu-Glu-) linker.

Primers for the Cfr and KTR from *Cd* E1 were designed for insertion into dual expression vector pCOLADuet-1 (Millipore Sigma). *Cd* E1 KTR was designed with *NdeI* and *XhoI* restriction sites while *Cd* E1 Cfr was designed with *EcoRI* and *HindIII* restriction sites. PCR and restriction digestion were performed as above. *Cd* E1 KTR was ligated into *NdeI/XhoI* digested pCOLA and the sequence was verified by Sanger sequencing at the Pennsylvania State University Nucleic Acid Facility. The pCOLA plasmid containing the KTR was re-isolated from cells and digested using *EcoRI* and
HindIII. Cd E1 Cfr was ligated into the digested vector and the sequence was verified as above. The KTR was designed to encode a C-terminal His6-tag with no linker.

Overexpression of the KTR constructs was accomplished in *E. coli* BL21(DE3) cells co-transformed with pET26b/pCOLA and pDB1282, the former carrying the cloned gene and the latter a vector containing inducible iron-sulfur cluster assembly machinery. Successful transformants were selected on kanamycin (50 µg/mL) and ampicillin (100 µg/mL) supplemented LB plates. 1.5 L cultures of the overexpression strains were grown shaking (180 rpm) at 37 °C in M9 minimal medium supplemented with kanamycin (50 µg/mL), ampicillin (100 µg/mL), and iron (III) chloride (30 µM) to maximize iron (or iron-sulfur cluster) incorporation. For *Cd* E1 Cfr/KTR dual expression, L-arabinose was added [final concentration 0.08 % (w/v)] to induce overexpression of the genes harbored on the pDB1282 plasmid at an OD_{600} = 0.3. L-cysteine (300 µM) and additional iron (III) chloride (30 µM) were added at the same OD to support iron-sulfur cluster assembly. The additions at OD_{600} = 0.3 were omitted in overexpression of KTRs using the pET26b vector. *Sal* KTR overexpressed for Mössbauer and EPR analysis used ^{57}Fe dissolved in 1 M H_{2}SO_{4} in place of iron (III) chloride. At an OD_{600} = 0.6, IPTG was added to a final concentration of 100 µM to induce overexpression of the pET26b-KTR gene or both genes in pCOLA. After incubation overnight at 18 °C at 180 rpm, cells were pelleted by centrifugation at 5000g for 10 min at 4 °C. Approximately 3.9 g cell paste was obtained per liter of culture. Cell paste was flash-frozen in liquid nitrogen and stored at -80 °C.

*Cloning of genes for Cfr, KTR, and Lsa into pBAD24 for assessment of E. coli antibiotic resistance.* All four Cfr homologs (*Ba, Ef, Pl, Cd*) as well as *Cs* Cfr and *Cs* Lsa
were re-cloned into the arabinose-inducible overexpression vector, pBAD24\textsuperscript{17} for assessment of antibiotic resistance. Primers were designed using Primer3 (Table 4-1). All primers include an EcoRI site in the forward primer, except Ba Cfr, which contains a XmaI site. Ef Cfr, Cd Cfr, and Pl Cfr added HindIII to the reverse primer while Ba Cfr, Cs Cfr, and Cs Lsa added XbaI. Cfr, KTR, and Lsa sequences were amplified by PCR and digested as described. The sequences were ligated into pBAD24 after restriction digest with the appropriate enzymes and verified by Sanger sequencing (Table 4-1).

**Purification of KTRs.** All purification steps were performed anaerobically in an anoxic chamber with a mixed H\textsubscript{2}/N\textsubscript{2} atmosphere (Coy Laboratory Products). Cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 300 mM KCl, 4 mM imidazole, 10 mM MgCl\textsubscript{2}, 10\% (v/v) glycerol, 10 mM β-mercaptoethanol) in a nickel-coated beaker containing 1 mg/mL lysozyme, 100 µg/mL DNase, and 1 mM phenylmethylsulfonyl fluoride (PMSF). The suspension was stirred at room temperature for 30 min and lysed by sonication (QSonica Q500, 10 s pulse, 30 s rest, 60\% amplitude) on ice. The resulting cell debris was pelleted by centrifugation at 11,300 g at 4 °C for 1.25 h in a sealed centrifuge bottle. The supernatant was applied to Co\textsuperscript{2+}-TALON resin. Protein fractions were eluted in 50 mM HEPES, pH 7.5, 300 mM KCl, 300 mM imidazole, 10 mM MgCl\textsubscript{2}, 30\% (v/v) glycerol, 10 mM β-mercaptoethanol. Size-exclusion chromatography was performed using a Sephacryl S200 column using gel filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl\textsubscript{2}, 25\% (v/v) glycerol). Sa KTR iron occupancy was determined by inductively coupled plasma optical emission spectroscopy (ICP-OES) at the Pennsylvania State University Laboratory for Isotopes and Metals in the Environment (LIME). An additional HiPrep SP FF 16/10 cation exchange column
was used for *Sa* KTR samples for crystallographic characterization. Buffer A (low salt) was composed of 10 mM HEPES pH 7.5, 5 mM DTT, and 25% glycerol. Buffer B (high salt) was composed of 10 mM HEPES pH 7.5, 5 mM DTT, 25% glycerol, and 1.0 M potassium chloride. Protein samples were exchanged into buffer A prior to injection onto the column. After a wash step with three column volumes (CV) of buffer A, a linear 0-100% B gradient was applied over 12 CV (buffers A/B). *Sa* KTR elutes at 61% B as determined by SDS-PAGE gel electrophoresis analysis of gel fractions.

*Genome neighborhood network (GNN) and isoelectric point analysis of cysteine-rich KTR proteins.* The complete FASTA file for the cysteine-rich KTR family was downloaded from Interpro (IPR025957) and used to generate sequence similarity networks (SSNs) with the EFI-EST server\(^8\) via default parameters. An SSN calculated with an alignment score of 31 was used to make a genome neighborhood network with the EFI-GNT server\(^18\) (±10 ORFs, ≥ 5% co-occurrence). Results were visualized and analyzed in Cytoscape.

To calculate the predicted isoelectric points of cysteine-rich KTRs, the full FASTA file for IPR025957 was pruned with the ExPASy decrease redundancy tool (<98% minimum identity), and the resulting FASTA file was used as input to the protein isoelectric point calculator web server (http://isoelectric.org/).\(^19\) The results for each individual KTR were averaged to obtain the overall predicted isoelectric point for the entire family.

*Human microbiome metagenome abundance of Cys-rich KTR.* The EFI-CGFP/shortBRED webtool was used to determine the abundance of KTRs in metagenomic samples obtained as part of the Human Microbiome Project (HMP).\(^8, 20-24\)
The KTR SSN described above (alignment score = 31) was used as the input file for marker identification (UniRef90 database, CD-hit identity = 85, DIAMOND pairwise comparison algorithm). Marker quantification with USEARCH in 380 healthy microbiome datasets yielded a genome-normalized median protein abundance heatmap. The dataset compares the relative quantity of KTR coding sequences (separated by cluster) per microbial genome at six different body sites. To highlight the relative abundance of radical SAM-adjacent KTR proteins in gut microbiome samples, boxplots for clusters 1-3 were generated in Kaleidagraph using the genome-normalized median protein abundance. The shaded area represents the median quartile (Q2), and the error bars extend to the first (Q1) and third (Q3) quartiles. Data points outside of these quartile limits are plotted as open circles.

*Modeling of small Cfr-like protein using crystal structure of RlmN.* Genome neighborhood analysis identified numerous Cfr-like proteins associated with KTRs that differed from typical Cfrs both in sequence identity and in length. To obtain an initial assessment of the impact of the decreased length on structure, a model of representative protein WP_002584956 was constructed using the iTasser server\(^{25, 26}\) using the crystal structure of RlmN with SAM bound (3RFA) as a reference.\(^{27}\) The highest scoring model was chosen for visualization.

*Determinant of antibiotic resistance conferred by Cfr, KTR, and Lsa in E. coli BL21 DE3 ΔtolC.* As with the Cfrs in chapter 2, associated proteins of interest, Cs Lsa, Cb KTR, and Sa KTR, were assessed by broth microdilution to determine if they display antibiotic resistance properties without Cfr present. Broth microdilution was chosen to test antibiotic resistance because it allows for the rapid determination of cell growth.
across a wide range of drug concentrations. Plasmids were inserted into *E. coli* BL21 DE3 Δ*tolC*, a strain selected to avoid immediate drug efflux associated with the *tolC* transporter. *Ba* Cfr and *Sa* KTR were also inserted into *E. coli* BL21 DE3 Δ*aaraBAD*/Δ*tolC* to test the effect of arabinose on cell growth. Initial tests used pET26b-based overexpression plasmids, but later assays were performed with the arabinose-inducible pBAD24 overexpression vector to avoid toxicity identified with some of the proteins. All assays were performed in 96-well (8 x 12) clear culture plates.

 Screens containing fixed concentrations of inducing agent were split by row into eight different drug concentration conditions, each performed in triplicate (Figure 4-1). Inducing agent (A) was present in columns 1-3 and 7-9 (grey boxes) and drug (D) was present in columns 7-12.

![Table](image)

**Figure 4-1.** Fixed concentration broth microdilution screen for antibiotic resistance profiling. Rows represent a serial dilution of lincomycin or linezolid starting at concentration X in row A. Columns denote the presence or absence of inducing agent (A) and drug (D).

Trials were performed with either lysogeny broth (LB) or glycerol-supplemented, glucose-deficient EZ rich medium (Teknova, GS EZ medium – 1 X MOPS mix, 1.32 mM K$_2$HPO$_4$, 1 X ACGU, 1 X supplement EZ, 5 % glycerol). Two preparations of medium were used, Solution A containing either 100 µg/mL kanamycin or 200 µg/mL ampicillin (pET26b, pBAD24-dependent) and Solution B containing 100 µg/mL kanamycin or 200 µg/mL ampicillin plus inducing agent. Unless otherwise noted, the inducing agent
concentrations were 200 µM IPTG (pET26b) and 1.0 % or 2 % arabinose (pBAD24). These concentrations of kanamycin/ampicillin and inducing agent are 2 X the final concentrations. 50 µL of the appropriate solution was added to all wells except row A, Solution A in columns 4-6, 10-12 and Solution B in columns 1-3, 7-9. In row A, wells lacking drug received 100 µL of the appropriate solution while wells containing drug received 96 µL or 95 µL for lincomycin or linezolid screens, respectively. 4 µL of lincomycin (100 mg/mL) or 5 µL linezolid (20 mg/mL) was then added to wells 7-12 in row A. A serial dilution was performed to achieve the appropriate drug concentration in each row resulting in a volume of 50 µL in each well. Overnight cultures were diluted to OD₆₀₀ = 0.004 in the appropriate medium and 50 µL of the cell dilution was added to each well. The plates were placed in an incubator at 37 °C for ~24 hours after which each well OD₆₀₀ was recorded using a SpectraMax i3 (Molecular Devices). In the case of \( \text{Ba Cfr and Sa KTR in E. coli BL21 DE3 } \Delta \text{tolC/ΔaraBAD} \), a more narrow range of lincomycin concentrations were used (350, 250, 213, 175, 138, 100, 62.5, and 31.25 µg/mL). These solutions were individually prepared and added at 2 X concentration before cell addition. Result are conveyed as percent viability (average OD of row / average OD of control).

Initial assays indicated that certain Cfr homologs might be toxic to cell growth. Therefore, in certain cases, I performed additional experiments to determine the optimal concentration of inducing agent. These assays were performed as described for the fixed inducing agent concentration screens, except that conditions were analyzed individually (without triplicates) and the inducing agent concentration varied by column (Figure 4-2). Results are reported as a comparison of growth curves (OD₆₀₀).
**Figure 4-2.** Variable percentage concentration screen for antibiotic resistance profiling. Rows represent lincomycin concentration in mg/mL. Columns denote different concentrations of inducing agent (A). A0 represents the condition where no inducing agent is present. Odd columns do not contain drug (-D) while even columns contain drug (+D).

*Sa KTR and RNA co-elution.* A 500 µL solution of 155mer RNA (100 µM) and *Sa* KTR (100 µM) in gel-filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol) was allowed to sit at room temperature for 15 min. The solution was loaded onto a Superdex 75 10/300 GL column and 1 mL fractions were collected. Fractions containing protein/RNA, as determined by monitoring at λ = 280 nm, were concentrated and analyzed for *Sa* KTR by SDS-PAGE gel analysis.

*Mass-spectrometry analysis of KTR co-purifying proteins.* A 22.5 kDa protein band that is present in all purified *Cd*-C, *Sa* KTR, and *Cb* KTR samples was digested and analyzed by mass spectrometry. The band was excised from each sample lane, sliced into pieces, and destained by the addition of a 50% acetonitrile (ACN) / 8 mg/mL ammonium bicarbonate solution and heating at 37 °C. Disulfides were reduced by incubation of gel pieces in a solution of 5 mM TCEP / 8 mg/mL ammonium bicarbonate at 60 °C. Cysteine residues were alkylated by addition of 100 mM iodoacetamide / 8 mg/mL ammonium bicarbonate and incubation at 37 °C with occasional shaking. Gel slices were washed three times with 50% ACN / 8 mg/mL ammonium bicarbonate solution to remove TCEP and iodoacetamide. The gel slices were shrunk through the addition of ACN and
incubation at room temperature. The ACN was removed during a brief incubation at 37 °C. Proteins were digested overnight at 37 °C using a trypsin solution (0.01 mg/mL trypsin / 8 mg/mL ammonium bicarbonate). Peptides were extracted by incubating the trypsin digests at 37 °C in a 50% ACN / 0.1% formic acid solution. Aliquots were then evaporated using a Savant SPD111V speedvac concentrator (ThermoScientific).

The samples were analyzed by the Pennsylvania State University Proteomics and Mass Spectrometry Core Facility. The peptides were reconstituted in 10 µL of 2% ACN, 0.1% formic acid after which 5 µL was loaded onto an Acclaim PepMap100 trapping column (100 µm × 2 cm, C18, 5 µm, 100 Å, Thermo) at a flow rate of 20 µL/min using 2% aqueous acetonitrile, 0.1% formic acid as a mobile phase. The peptides were separated on an Acclaim PepMap RSLC column (75 µm × 15 cm, C18, 2 µm, 100 Å, Thermo) with a 30-min 2% - 35% linear gradient of acetonitrile in water containing 0.1% formic acid. The gradient was delivered by a Dionex Ultimate 3000 nano-LC system (Thermo) at 300 nL/min.

An LTQ Orbitrap Velos mass spectrometer (Thermo) was operated in ‘2nd order double play’ mode with default AGC target and inject time parameters. The data-dependent settings were as follows: full FT MS scan at R 60,000 followed by 10 ion-trap MS2 scans on most intense precursors with CID activation. Only the precursors with charge states +2 and higher were selected for MS2; monoisotopic precursor selection with non-peptide recognition was enabled, and the isolation window was 2 m/z.

The mass spectra were processed using Proteome Discoverer 1.3 (P.D. Thermo). The proteins were identified by searching the data with SEQUEST against E. Coli K-12 reference proteome database (Uniprot UP000000318) containing 4220 sequences. The
following search parameters were used: precursor tolerance 10 ppm, fragment tolerance 0.8 Da (ion trap), methionine oxidation (+15.995 Da, M) as a dynamic modification, and cysteine carbamidomethylation as a static modification (+57.021 Da, C).

**UV-visible spectrophotometry.** Oxidation of Sa KTR was analyzed spectrophotometrically using an Agilent 8453 G1103A spectrometer. In an anoxic glovebox, a 200 µL solution of Sa KTR (17.1 mM) was diluted 1/10 by gel-filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol). The protein solution was exposed to air, quickly transferred to a cuvette, and scanned from 190 nm to 1100 nm with a baseline correction at 800 nm. The cuvette remained open for the duration of the experiment. The solution was scanned at the following times: 0 min, 1 min, 3 min, 5 min, 10 min, 20 min, 30 min, 45 min, and 1 hour. A difference spectrum was calculated by subtracting the initial spectrum from the final spectrum.

**Mössbauer spectroscopy of Sa KTR.** A Sa KTR ^{57}\text{Fe} protein sample (17.1 mM) in gel-filtration buffer (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl₂, 25% (v/v) glycerol) was transferred to an EPR tube and Mössbauer cup and flash-frozen in liquid nitrogen in an anoxic chamber (Coy Laboratory Products). A Mössbauer spectrum was recorded on constant acceleration Mössbauer spectrometers (SEE Co., Edina, MN) equipped with a Janis SVT-400 variable-temperature cryostat (weak-field) (Janis, Woburn, MA). All isomer shifts are quoted relative to the centroid of the spectrum of α-iron metal at room temperature. Simulations of Mössbauer spectra were carried out using WMOSS (SEE Co., Edina, MN). Some of the simulations are based on the spin Hamiltonian formalism (see equation below), in which the first term describes the
electron Zeeman effect, the second and third terms describe the axial and rhombic zero-field splitting of the total electron spin ground state, the fourth term represents the interaction between the electric field gradient and the nuclear quadrupole moment, the fifth term describes the magnetic hyperfine interactions of the $^{57}\text{Fe}$ nucleus with respect to the total electron spin ground state, and the last term represents the nuclear Zeeman interactions of $^{57}\text{Fe}$. All symbols are conventional.\(^{28}\) Spectra were calculated in the slow relaxation limit.

\[
H = \beta S \cdot g \cdot B + D \left( S_z^2 - \frac{S(S+1)}{3} \right) + E (S_x^2 - S_y^2) \\
+ \sum_i \frac{eQV}{4} \left[ I_{z,i}^2 - \frac{1}{3} (I_{1,i}^2 + I_{2,i}^2) + \frac{\eta}{3} (I_{x,i}^2 - I_{y,i}^2) \right] + \sum_i S \cdot A_i \cdot I_i - \sum_i g_i \beta_n B \cdot I_i
\]

EPR samples were analyzed on a Bruker ESP 300 spectrometer (Billerica, MA) equipped with an ER 041 MR microwave bridge and an ER 4116DM resonator.

**Crystallization of Sa KTR.** All protein preparation and crystallographic screening was performed in a Coy anaerobic chamber (~97% N\(_2\)/3% H\(_2\)). Classics and JCSG+ sparse matrix screens (Qiagen) were used to promote crystal formation by sitting drop vapor diffusion. 1 µL of pure Sa KTR (2.5 mg/mL) was mixed with 1 µL of each well condition. After four months, iridescent plate crystals formed in Classics condition B4 (0.2 M NaCl, 0.1 M sodium acetate pH 4.6, 30% [v/v] MPD).
Results

Sequence similarity and genome neighborhood analyses of Cys-rich KTRs.

The discovery of an Fe$^{2+}$ binding site in the Cd Cfr C-terminal domain raises the possibility that the larger family of stand-alone Cys-rich KTR proteins might have the same property, with widespread functional significance for the metal-bound form. To gain insight into this possibility, I analyzed patterns in genome neighbors (± 10 open reading frames) for the entire InterPro family (IPR025957). The resulting sequence similarity network revealed that 13% of the Cys-rich KTR family can be found in two large clusters (133 sequences, clusters 1 and 3 in Figure 4-3). For the 85 of the 133 sequences from organisms with fully sequenced genomes, 84 are encoded directly adjacent to a putative radical SAM RNA methylase. Consistent with my identification of a Cd Cfr-KTR fusion enzyme, nearly all of these Cys-rich KTRs are from organisms belonging to the clostridiales order (Table 4-3).
I initially presumed that these radical SAM enzymes were Cfr homologs, but I was surprised by the large number of clostridial sequences (>100) identified as KTR neighbors – which was significantly more than the number of clostridial Cfrs identified in the phylogenetic analysis (~35). An alignment of a representative KTR-adjacent radical
SAM enzyme sequence from the GNN to *Sa* or *Cd* Cfr sequences shows the newly identified sequences are uniformly short (327 amino acids vs. 349 amino acids) with only ~30% similarity. A homology model of one of these shortened sequences (WP_002584956) created using the crystal structure of RlmN (3RFA) as a reference predicts that there are differences in the N-terminal domain, an extension loop, and the C-terminal helix (Figure 4-4). With regard to the active site, the only notable difference is the position of the Cfr equivalent of Cys<sub>355</sub> which remains in the α-helical secondary structure of the C-terminal helix and is located farther from the iron-sulfur cluster. However, it is possible that the addition of SAM and other reaction components may allow for similar positioning of the residue in a crystal structure. No other significant differences exist in the active site which suggests that these enzymes catalyze RNA methylase chemistry like other Cfrs. Because a structure of RlmN bound to tRNA demonstrates that multiple contacts exist between the substrate and sites other than the active site, the dissimilarities in the distant parts of the protein may indicate a difference in substrate identity or binding mode. I propose that these KTR-adjacent RNA methylases therefore represent a previously unknown reservoir of putative Cfr enzymes. Interestingly, the organisms with paired *ktr-cfr* gene clusters identified in this analysis are commonly found as commensal human gut microbiome constituents (Table 4-2).
Figure 4-4. (A) Homology model of WP_002584956 created by comparison to the crystal structure of RlmN with SAM bound (3RFA).25-27 (B) Surface representation looking down the partial α/β TIM barrel and (C) a rotated view highlighting the structural differences between the proposed model and the RlmN crystal structure.

The largest group of remaining $ktr$ genes (clusters 2 and 4, Figure 4-4) are encoded adjacent to a ribosomal protection protein tetracycline-resistance element, $tetM$, and drug efflux pumps such as the ABC transporter and multidrug and toxic compound extrusion (MATE) efflux families. TetM binds to intact 70S ribosomes stalled by tetracycline during translation. The protein induces drug release by binding to the same site as elongation factor G (EF-G), a GTPase that enhances the rate of tRNA and mRNA translocation on the ribosome during translation.30 This group of KTR homologs is almost exclusively associated with known bacterial pathogens originating from Clostridiales and Bacillales. The remaining $ktr$ genes are often found near drug efflux systems, with the majority annotated as ABC family transporters or MATE/MFS effluxers. All Cys-rich $ktrs$ are also proximal to other genetic elements including helix-turn-helix proteins (HTH3/HTH16),
RNA polymerase $\sigma^{70}$, and tetR-like repressors. Besides the radical SAM clusters, a functional role for Fe$^{2+}$ and Rdx-like activity is not immediately evident, although it is interesting to note that the protein families listed above all exhibit some overlap with the biological function of Cfr in drug resistance. The prevalence of Cys-rich $ktr$ genes in pathogens and in members of the human microbiome warrants further investigation of their biological function and metal ion requirements.

I further validated the observation that KTRs are abundant in human gut commensal organisms by analyzing the healthy human microbiome metagenome abundance of Cys-rich $ktr$ genes via the Enzyme Function Initiative chemically-guided functional profiling (CGFP) web tool (Figure 4-5).$^{35-37}$ The $ktr$ genes associated with clusters 1 and 3, found adjacent to cfr homologs in anaerobic clostridia, are significantly enriched in human stool microbiome metagenomes. These patterns could indicate a reservoir of Cys-rich $ktr$-cfr genes in the beneficial gut flora to protect against antibiotic exposure with the possibility of transfer to pathogens. Cluster 2 sequences show moderate abundance in the nose and mouth, consistent with their occurrence in aerobic opportunistic pathogens such as *Staphylococcus aureus* and *Streptococcus pneumoniae*.
Figure 4-5. Metagenome abundance of *ktr* genes in healthy human microbiome samples. (A) Heatmap for clusters 1-10 in the KTR SSN (see Fig. 8) showing marker abundance in 380 HMP datasets collected from six different areas of the body. Cluster 4 is omitted due to low abundance in the metagenomic sample set. (B-D) Box plots of the per-site abundance for the top three clusters. This analysis shows that the KTR proteins encoded adjacent to radical SAM enzymes are abundant in the gut microbiome.

Investigating the genome neighbors of other clostridial *Cfrs*. In my prior analysis of *Cfr* phylogenetic relationships, I identified a small number of pathogenic *Clostridioides difficile* sequences containing a C-terminal KTR fusion. In this chapter, I show that >100 Cys-rich KTRs can be found as genome neighbors to radical SAM enzymes, presumably distantly-related *Cfr* homologs, and found predominantly in anaerobic gut flora. Here I investigate whether a KTR neighbor is ubiquitous in
clostridial Cfrs – including sequences annotated as Cfr-like and found to be inactive when expressed heterologously in *E. coli*.

The phylogenetic tree showed that clostridial Cfrs diverge into two clades, with the smaller group containing the KTR-fused *Clostridioides difficile* Cfr homolog that I showed to be active *in vitro* in chapter 2. The larger group of clostridial Cfrs includes inactive homologs from *Clostridium sporogenes* and *Clostridium phytofermentans*. Analysis of neighboring genes in this group revealed a conserved Lsa-family ABC-F type ribosomal protection protein adjacent to the Cfr (Figure 4-6). My findings show that nearly all clostridial Cfrs are found adjacent to either a KTR or an Lsa protein.
Lsa family ABC-F ribosomal protection proteins (RPP) are large antibiotic resistance proteins composed of three domains. Two of these are annotated as nucleotide binding domains, NBD1 and NBD2, and third is termed the antibiotic resistance domain.
These enzymes promote resistance by recognizing fully assembled 70S ribosomes stalled during translation. Structures of related ABC-F proteins VmlR and MrSE show that the RPP anchors to the ribosome through interactions with both subunits. NBD1 binds to the 50S subunit and NBD2 interacts predominantly with the 30S subunit. These interactions position the ARD near the peptidyl transferase center in the 50S subunit, presumably to dislodge bound drugs. The RPP then dissociates from the ribosome and translation is allowed to continue. Interestingly, this mechanism of resistance does not provide a way to prevent subsequent drug binding events.

Figure 4-7. Cryo-EM structure of ABC-F type ribosomal protection protein VmlR complexed to the 70S ribosome of B. subtilis (PDB 6HA8). VmlR is colored in red, the A2503 equivalent residue of the ribosome in purple, the tRNA in teal, the 50S ribosome in green, and the 30S ribosome in orange.

Given this nearly universal co-occurrence with a Lsa or KTR protein, I hypothesized that clostridial Cfrs might require one or more partner proteins for efficient
function. Consequently, lack of these factors during heterologous overexpression in other organisms would yield an apparently inactive or modestly active enzyme, as I found with *Cd* Cfr. KTR proteins may be important for interface with native reducing systems or substrate binding for efficient multi-turnover methylation. The role of Lsa in clostridial Cfr activity has not been analyzed. Proteins related to Lsa are known to bind to the fully assembled 70S ribosome to dislodge antibiotics that function by stalling translation. Subsequent strategies to prevent drug re-binding are not well-understood. All Cfrs characterized to date can modify 23S rRNA but are inactive with fully-assembled 50S or 70S rRNA. I propose that clostridial Cfrs might be capable of methylating intact 50S rRNA after conformational rearrangement of the ribosome by Lsa. Consistent with this hypothesis, the structure of VmlR shows close proximity of the ARD and A2503 (Figure 4-7).

*Optimization and assessment of antibiotic resistance profiles of Cfr, KTR, and Lsa.* The nearly universal conservation of an Lsa or KTR protein adjacent to clostridial Cfrs suggests that that the lack of activity demonstrated in some clostridial Cfr homologs could be due to the absence of partner proteins necessary for RNA modification. I tested this hypothesis by monitoring the drug resistance profiles of KTR or Lsa, alone or when co-expressed with Cfr partners. I determined half maximal inhibitory drug concentrations (IC$_{50}$) by using the broth microdilution method. Cell viability is reported as the average OD$_{600}$ of three test trials divided by the average OD$_{600}$ of the control wells for each drug concentration. Initial tests were performed in *E. coli* BL21 DE3 ΔtolC with candidate genes introduced via IPTG-inducible pET26b vectors. To establish a baseline of antibiotic resistance activity, *Ba* Cfr and *Ef* Cfr were assayed first for activity against
lincomycin (Figure 4-8), a drug for which Cfr is known to confer resistance.\textsuperscript{31} I only observed evidence of drug resistance at the very lowest lincomycin concentrations (< 62.5 µg/mL). In all other conditions, the uninduced \textit{E. coli} grew similarly to or outperformed the cells overexpressing Cfr. These results were unexpected, particularly for \textit{Ba} Cfr which showed \textit{in vitro} activity similar to \textit{Sa} Cfr. I therefore suspected that the assay conditions were suboptimal.

The observed results could be due to toxic overexpression of the Cfr protein. The T7 promoter of pET26b is insensitive to the amount of IPTG used to turn on protein expression. Therefore, to determine if overexpression was causing growth inhibition, a new overexpression vector with a tunable promoter was used. Cfr genes were cloned into the pBAD24\textsuperscript{17} vector with an arabinose-inducible promoter. \textit{Ba} Cfr was again tested for activity but with either 0.5\% or 1\% arabinose as the inducing agent. The results displayed a similar overall trend as with pET26b where the uninduced cells grew to a higher optical density than the induced cells.
As altering expression levels did not yield significant change to the antibiotic resistance profile, the next variable tested was growth medium. The components of LB media cannot be as tightly controlled as forms of minimal medium. To that end, the same expression tests (0.5%, 1.0% arabinose) were performed with GS EZ media (1X MOPS mix, 1.32 mM potassium phosphate dibasic, 1X ACGU, 1X supplement EZ, 5% glycerol). Unlike with LB, a clear difference in growth emerged between the induced and uninduced cells under these conditions (Figure 4-9). As detailed in chapter 2, the IC$_{50}$ values of the induced cells exceeded those of the control by a factor of 5-8.
I used these conditions to test $Cs$ Cfr and $Cs$ Lsa for their antibiotic resistance profiles using GS EZ medium with 0.5% arabinose (Figure 4-10). Unlike $Ba$ Cfr, the two $Cs$ proteins had little effect on the IC$_{50}$, with $Cs$ Cfr starting to impact growth at $\sim$125 µg/mL lincomycin and $Cs$ Lsa showing no deviation from the uninduced cells. In a different study, it was shown that similar proteins, VgaA and LsaA, did not confer antibiotic resistance in $E. coli$. The activity of VgaA in particular was limited to the gram-positive organism $S. aureus$.\textsuperscript{31} This observation suggests that there may be
additional factors required for ABC-F type RPP activity, such as protein partners or rRNA structural components that differ between organisms.

**Figure 4-10.** Antibiotic resistance profiles of (A) *Cs* Cfr and (B) *Cs* Lsa on pBAD24 in *E. coli* BL21 DE3 ΔtolC. Cell viability normalized to uninduced, drug absent control (−A, −D) is plotted with induced, drug present (+A, +D, closed circles) and uninduced, drug present (−A, +D, open circles).

To determine if the lack of activity seen from *Cs* Cfr and *Cs* Lsa was a phenomenon specific to these proteins, the other three original Cfr homologs were tested for their antibiotic resistance profiles, starting with *Ef* Cfr. Using GS EZ medium and 0.5% arabinose, *Ef* Cfr also demonstrated a lack of antibiotic resistance activity. The question of protein toxicity and expression levels was revisited and a variable percentage arabinose screen was performed (Figure 4-11). The results showed that high levels of protein expression were indeed toxic to the cells. The lowest arabinose concentrations also show diminished growth. The optimal inducing agent levels lie between 0.025% and 0.05%.
Figure 4-11. Antibiotic resistance profile of *Ef* Cfr on pBAD24 in *E. coli* BL21 DE3 ΔtolC. Cell viability normalized to uninduced, drug absent control (-A, -D) is plotted with induced, drug present (+A, +D) colored by amount of arabinose inducing agent. An uninduced, drug present (-A, +D) condition was not run.

I also tested *Cd* and *Pl* Cfr for lincomycin resistance at various arabinose concentrations (Figure 4-12, panels A-B). *Cd* Cfr and *Pl* Cfr induce toxicity at high levels of expression, similar to *Ef* Cfr. Any concentration higher than 0.005% arabinose suppressed cell growth. At this level of protein overexpression, I did not observe any beneficial resistance activity compared to an uninduced control (Figure 4-12, panels C-D). I also tested the Cfr proteins with linezolid, a oxazolidinone drug to which Cfr is known to confer protection (Figure 4-13).\textsuperscript{34} The resulting growth curves for Cfr-induced cultures did not deviate from the uninduced controls.
Figure 4-12. Antibiotic resistance profiles of *Cd* Cfr and *Pl* Cfr expressed from a pBAD24 vector in *E. coli* BL21 DE3 ΔtolC. OD of induced, drug present (+A, +D) normalized to uninduced, drug absent control (-A, -D) colored by amount of arabinose inducing agent for (A) *Cd* Cfr and (B) *Pl* Cfr. Uninduced, drug absent (-A, -D) normalized comparison of induced, drug present cells (+A, +D, closed circles) expressing (C) *Cd* Cfr and (D) *Pl* Cfr compared to uninduced, drug present cells (-A, +D, open circles).
I tested the ability of Cys-rich KTRs to induce antibiotic resistance on their own accord. *Sa* KTR and *Cb* KTR were expressed from a pBAD24 vector in *E. coli* BL21 DE3 Δ*tolC* cells. At 0.5% arabinose, both KTRs showed a small positive shift in the IC$_{50}$ curve (Figure 4-14). However, it was unclear whether the increased growth relative to the control was solely the result of protein expression or due to the presence of arabinose as an additional carbon source.
To assess whether arabinose itself was leading to increased growth in induced cells, \textit{Ba} Cfr and \textit{Sa} KTR were inserted into an \textit{ΔaraBAD/ΔtolC} \textit{E. coli} strain and tested for activity against lincomycin with variable amounts of arabinose present (Figure 4-15, panels A-B). Ideal growth was observed at 0.100% arabinose for both proteins. New assays were performed for \textit{Ba} Cfr and \textit{Sa} KTR in the \textit{ΔaraBAD/ΔtolC} strain with 0.100% arabinose. A smaller range of lincomycin was used (31.25 µg/mL – 350 µg/mL) to increase the detection window. At optimal expression levels, the IC\textsubscript{50} curves with drug collapsed into the control curves (Figure 4-15, panels C-D). These results suggest that the additional arabinose was artificially inflating the IC\textsubscript{50} shift seen from \textit{Ba} Cfr and \textit{Sa} KTR in the presence of drug.

The profiles of Cfr- and KTR-mediated drug resistance in \textit{E. coli} reveal a number of complications with the experimental approach. With my current optimized conditions, it is clear that the toxicity of \textit{Ef}, \textit{Pl}, and \textit{Cd} Cfrs dominates growth effects in \textit{E. coli} BL21 DE3 \textit{ΔtolC}. In chapter 2, I showed that while \textit{Ba} Cfr is not toxic, it has a diminished

![Figure 4-14](image-url). Antibiotic resistance profiles of (A) \textit{Sa} KTR and (B) \textit{Cb} KTR on pBAD24 in \textit{E. coli} BL21 DE3 \textit{ΔtolC}. Cell viability normalized to uninduced, drug absent control (-A, -D) is plotted with induced, drug present (+A, +D, closed circles) and uninduced, drug present (-A, +D, open circles).
effect on drug resistance compared to other Cfrs previously tested. Others have shown that *E. coli* expressing *Sa* Cfr grows at >512 µg/mL clindamycin, a lincomycin relative, a significant difference compared to *E. coli* expressing *Ba* Cfr which has almost zero growth above 138 µg/mL lincomycin under these conditions.\textsuperscript{34} It is likely that the difference is due to use of native bacterial strains or antibiotic susceptible strains of *E. coli* AS19 in the previous studies.\textsuperscript{4, 34, 35} If a gram-positive firmicute could be used in these assays, it would likely give a greater understanding of the antibiotic resistance capability (or lack thereof) for each homolog. If a suitable host organism could be found, it would be worthwhile to examine the effect that the KTRs have on antibiotic resistance both alone and in concert with associated Cfrs. Using a native strain would also likely allow for synergistic activity of Cfrs with an Lsa partner to be observed, because previous studies show that the activity of ABC-F RPPs tends to be tailored to the ribosomes of a specific organism.\textsuperscript{31}
Figure 4-15. Antibiotic resistance profiles of *Ba* Cfr and *Sa* KTR on pBAD24 in *E. coli* BL21 DE3 ΔaraBAD/ΔtolC. OD of induced, drug present (+A, +D) normalized to uninduced, drug absent control (-A, -D) colored by amount of arabinose inducing agent for (A) *Ba* Cfr and (B) *Sa* KTR. Uninduced, drug absent (-A, -D) normalized comparison of 0.100% arabinose-induced, drug present cells (+A, +D, closed circles) expressing (C) *Ba* Cfr and (D) *Sa* KTR compared to uninduced, drug present cells (-A, +D, open circles).

Purification of three different stand-alone KTRs. Given that no member of the stand-alone Cys-rich KTR family has been characterized *in vitro*, I selected three different homologs from sequence clusters 1-3 (Figure 4-4) for structural and spectroscopic study. I wanted to determine whether iron binding, redox activity, and RNA interaction were universal properties of this class of small proteins. My initial
targets include Cfr-associated KTRs from *Clostridioides difficile* str. E1 and *Clostridium boltae* and a TetM-adjacent KTR from *Staphylococcus aureus*.

Purification of *Cd* E1 KTR did not initially yield soluble protein. I hypothesized that this protein might exist as part of a complex with the adjacent Cfr, thereby requiring the radical SAM protein for stability. I cloned both genes into a dual expression vector and attempted to isolate the putative protein-protein complex. Soluble expression of both proteins was confirmed by SDS-PAGE gel electrophoresis, but purification of the complex was unsuccessful because the Cfr and KTR components did not bind to the Co$^{2+}$-TALON column. This observation suggests that the interaction between the Cfr and KTR buried the N-terminal 6x-His tag on the KTR component, preventing isolation by affinity chromatography.

*Cb* and *Sa* KTRs are both soluble after overexpression in *E. coli*. Purification by affinity and gel filtration chromatography yielded mostly pure protein for both homologs. However, in all samples (including *Cd*-C), a 22.5 kDa contaminant remained after the second column. The contaminant can be removed by a third cation exchange column but with significant loss of material. Only *Sa* KTR was sufficiently overexpressed to obtain useful amounts of pure protein after this treatment. Yields and other parameters can be found in Table 4-3. ICP-AES analysis of *Sa* KTR reveals that the protein can be isolated with similar amounts of iron as found in *Cd*-C. Given the high yield of pure protein and distinctive genomic context, I focused exclusively on *Sa* KTR for additional biophysical characterization.
Table 4-3. Properties of purified stand-alone KTR proteins

<table>
<thead>
<tr>
<th>KTR</th>
<th>MW</th>
<th>pI</th>
<th>Metal Equivalents</th>
<th>Purification Stage</th>
<th>Soluble Protein</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd-C</td>
<td>8250</td>
<td>9</td>
<td>0.45 Fe</td>
<td>gel filtration (GF)</td>
<td>39 mg</td>
<td>22.5 KDa contaminant</td>
</tr>
<tr>
<td>Cd-C</td>
<td></td>
<td></td>
<td>0.36 Zn</td>
<td>gel filtration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sa</td>
<td>8353</td>
<td>9.2</td>
<td>0.52 Fe</td>
<td>cation exchange (CE)</td>
<td>143 mg (GF)</td>
<td>pure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.45 mg (CE)</td>
<td></td>
</tr>
<tr>
<td>Cd</td>
<td>8072</td>
<td>9.4</td>
<td></td>
<td>affinity</td>
<td>1.23 mg</td>
<td>22.5 KDa contaminant</td>
</tr>
<tr>
<td></td>
<td>8460</td>
<td>9.1</td>
<td></td>
<td>gel filtration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**UV-visible spectroscopy of Sa KTR.** Cd-C undergoes a color change upon exposure to air, consistent with redox cycling between a Fe$^{3+/2+}$ couple. To determine if Sa KTR undergoes the same phenomenon, I analyzed the UV-visible spectrum of the protein after air oxidation (Figure 4-16, panel A). A difference spectrum shows peak maxima at $\lambda = 370$, 490, and 575 nm, similar to Cd-C (370, 485, 582 nm), suggesting that redox active Fe$^{3+}$ incorporation is a shared property among the stand-alone KTRs. The KTR UV-vis features also resemble those of rubredoxins (380, 480, 575 nm), but the major peaks are broader and the shoulder peak at 350 nm is much less pronounced in the two KTRs. KTRs also deviate from rubredoxins in the intensity of the color associated with the oxidized species. Sa KTR is a faint red, even at high protein concentrations. Rubredoxins instead exhibit an intense red color under the same conditions. Additionally, Sa KTR was unstable when exposed to oxygen, precipitating completely after one hour. This timeframe proved sufficient for examination of the oxidized state by UV-visible spectroscopy. The differences in sequence, stability of the Fe$^{3+}$-bound form, and peak
broading could be consistent with a distinct structure for KTRs compared to typical rubredoxins.

*Mössbauer and EPR spectroscopic analysis of Sa KTR.* To further verify that *Sa* KTR can bind a single iron (II) in a rubredoxin-like site, I performed Mössbauer and EPR spectroscopy on the purified protein. The protein was overexpressed in $^{57}$Fe-containing M9 minimal medium and purified anaerobically in the presence of the reductant dithiothreitol to maintain the Fe$^{2+}$ state. In the analogous form of *Cd*-C, I did not observe any EPR signals, as expected for a high-spin Fe$^{2+}$ cofactor. *Cd*-C does exhibit a single quadrupole doublet by Mössbauer analysis ($\Delta Q = 3.08 \text{ mm/s}$ and an isomer shift of $\delta = 0.70$), similar to those of a Fe$^{2+}$-Cys$_4$ rubredoxin center. $^{36}$ *Sa* KTR is also EPR-silent. It displays a single quadrupole doublet with splitting of $\Delta Q = -3.235 \text{ mm/s}$ and an isomer shift of $\delta = 0.70 \text{ mm/s}$ (Figure 4-16, panel B). These results show that *Cd*-C and *Sa* KTR have the same isomer shift and a similar magnitude $\Delta Q$ value.

![Figure 4-16](image_url)

**Figure 4-16.** (A) UV-visible difference spectrum of the oxidation of *Sa* KTR (blue) compared to *Cd*-C (black). *Sa* KTR absorbance maxima are visible at 370 nm, 490 nm, and 575 nm. (B) Mössbauer spectra of *Sa* KTR at (top) $B = 0$ and (bottom) $B = 53 \text{ mT}$. The black trace is the experimental spectrum and the red line is the simulation showing a $\Delta Q = -3.235 \text{ mm/s}$ and a $\delta = 0.70 \text{ mm/s}$.
Size exclusion chromatography co-elution of Sa KTR and RNA. In chapter 3, the 155mer rRNA substrate for Cfr was shown to bind to Cd-C by co-elution during size exclusion chromatography. To determine if Sa KTR can also interact with RNA, co-elution of the 155mer rRNA fragment [A2454-G2608 (E. coli numbering)] with Sa KTR was assessed by gel-filtration chromatography (10 mM HEPES pH 7.5, 500 mM KCl, 5 mM DTT, 10 mM MgCl2, 25% (v/v) glycerol) (Figure 4-17). I observe one major peak and several minor peaks at 280 nm. The major peak fractions elute close to the void volume, consistent with the high molecular weight of the RNA substrate. SDS-PAGE analysis showed that this peak also contains Sa KTR. As with Cd-C, I do not know if the interaction between Sa KTR and the 155mer rRNA fragment is specific. The results could also reflect non-specific electrostatic interaction between the positively-charged KTR and negatively-charged nucleic acid.

Figure 4-17. Co-elution of RNA and Sa KTR by size exclusion chromatography. (Left) UV visible trace (λ= 280 nm) showing region of RNA-containing peaks and associated fractions. (Right) SDS-PAGE analysis of Sa KTR presence in RNA fractions. KTR proteins tend to appear slightly higher than their molecular weight (~8 KDa), likely due to their unusually high pl. Ladder is in KDa.
Mass spectrometry of unknown KTR co-elutants. In my purification of stand-alone KTR homologs, I noted that all samples (including Cd-C) contain a ~22 kDa contaminant after several rounds of purification (Figure 4-18). To identify this contaminant and gain insight into its potential interaction with Cys-rich KTR motifs, I used mass spectrometry to identify the unknown *E. coli* protein(s). *Cb* KTR, *Sa* KTR, and *Cd*-C were heterologously overexpressed, purified by affinity chromatography and size exclusion chromatography, and separated on an SDS-PAGE gel. The 22.5 kDa bands were excised, the protein was removed and digested by trypsin, and the resulting isolations were analyzed by mass spectrometry. Of the resulting seven possibilities, only two proteins scored highly, SlyD and ZinT. Both of these proteins are implicated in metal trafficking in *E. coli*. SlyD is a peptidyl-prolyl cis-trans isomerase\(^{37-39}\) and has been shown to complex with HypB in assembly of the [Ni-Fe] hydrogenase cluster, likely acting as the nickel source.\(^{40,41}\) It has also been demonstrated to assist in protein folding.\(^{42,43}\) ZinT is a zinc receptor associated with the periplasmic Zn\(^{2+}\) ABC transporter, ZnuABC, and proposed to function in metal shuttling to the ion pump.\(^{44-46}\) SlyD and ZinT are known to be common contaminants isolated during metal-affinity chromatography\(^47\) However, I think it is unlikely that these proteins are solely purification artifacts, in part because the characteristic SDS-PAGE bands are not present in any significant amount during Cfr purification. Additionally, during size exclusion chromatography, the KTRs elute at a volume that is more consistent with a higher-order complex (as opposed to a monomer). It remains unknown whether the complexes formed by the co-eluting proteins are functionally relevant. If they are, the interaction could be consistent with a universal requirement for a divalent metal ion in this protein family.
Figure 4-18. (left) SDS-PAGE analysis of Cb KTR, Sa KTR, and Cd-C after gel chromatography. The KTR bands are located slightly above 10 KDa, likely due to their high pI. The co-purified protein band of interest is at ~22.5 KDa. The ladder values are given in kDa. (right) Table of mass spectrometry results for each KTR showing only the highly scoring matches and their properties. The scores represent the sum of all Xcorr values for peptides with a correlation above the threshold.

<table>
<thead>
<tr>
<th>KTR</th>
<th>Proteins Matched</th>
<th>Score</th>
<th>Accession</th>
<th>MW (KDa)</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cb</td>
<td>SlyD</td>
<td>542</td>
<td>P0A9K9</td>
<td>20.8</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>ZinT</td>
<td>384</td>
<td>P76344</td>
<td>24.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Sa</td>
<td>SlyD</td>
<td>356</td>
<td>P0A9K9</td>
<td>20.8</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>ZinT</td>
<td>275</td>
<td>P76344</td>
<td>24.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Cd-C (Fe)</td>
<td>ZinT</td>
<td>226</td>
<td>P76344</td>
<td>24.7</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>SlyD</td>
<td>122</td>
<td>P0A9K9</td>
<td>20.8</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Identification of initial hit in protein crystallography of Sa KTR. Biophysical characterization of Sa KTR and Cd-C indicate that cys-rich KTRs are iron-binding proteins which bind RNA and may have a role in electron transfer. While containing some sequence elements reminiscent of rubredoxins, spectroscopic evidence suggests that the cys-rich KTRs have a distinct absorption profile and may adopt a unique variation of the FeS₄ protein fold. Determination of a cys-rich KTR structure may illuminate the reasons for its unusually high midpoint potential and for the capacity to form a tight interaction with RNA. Sa KTR is the only standalone cys-rich KTR to be purified in sufficient quantities to be screened by crystallographic methods. Accordingly, Sa KTR was assessed for crystal formation using multiple sparse matrix screens. Condition B4 in Qiagen Classics (0.2 M NaCl, 0.1 M sodium acetate pH 4.6, 30 % [v/v] MPD) promoted crystal formation after approximately four months (Figure 4-19). The plate-like crystals produced an iridescent effect visible by the naked eye. Further testing is required to
evaluate if the crystals are protein and if they diffract to a resolution appropriate for structure determination.

Figure 4-19. Sa KTR (2.5 mg/mL) in Classics sparse matrix screen well B4 (0.2 M NaCl, 0.1 M sodium acetate pH 4.6, 30 % [v/v] MPD) under a light polarizer.

Discussion

In this chapter, I identified conserved genome neighbors in clostridial Cfr systems. The corresponding proteins may play an important role in the RNA methylase activity of clostridial Cfrs. GNN analysis of a stand-alone protein family related to Cd-C, Cys-rich KTRs, reveals that members split into three major groups with distinct conserved genome neighbors. The first and largest group (Clusters 1 and 3) contains KTRs found adjacent to putative radical SAM methylases. These putative Cfr proteins are considerably sequence divergent from Sa Cfr. Consequently, they were not identified in previous phylogenetic analyses of the Cfr family. Future work will be necessary to validate the functions of these proteins to determine if they perform the A2503 methylation associated with antibiotic resistance. Analysis of metagenomic data from the human microbiome project shows that the associated KTRs from clusters 1 and 3 are
mostly found in commensal clostridia of the gut microbiome. These results reveal a previously unrecognized reservoir of possible Cfr homologs in the human microbiome.

Cys-rich KTRs are found in genomic contexts that do not involve radical SAM methylases, but are linked in some other way to antibiotic resistance. Clusters 2 and 4 contain KTRs found adjacent to the ribosomal protection protein, TetM, while many of the minor clusters contain KTRs adjacent to drug efflux transporters. In contrast to the radical-SAM-adjacent KTRs which are found predominantly in anaerobic commensal gut microbiome organisms associated with healthy people, the TetM-associated KTRs are found in aerobic pathogens and the transporter-associated KTRs are found in anaerobic pathogens. All members of these gene clusters provide protection against antibiotics that target DNA or RNA. Cfr confers resistance to the PhLOPS\(_A\) antibiotics, agents that bind to rRNA in the peptidyl transferase center of the ribosome.\(^{34}\) TetM provides protection against the action of tetracycline, an antibiotic which binds to the rRNA in the 30S subunit and prevents binding of tRNA to the A-site.\(^{30}\) The MATE efflux proteins confer resistance to fluoroquinolones, antibiotics which are known to form a complex with DNA while bound to type II topoisomerases.\(^{48,49}\) A common function for KTRs in nucleic acid binding would consistent with the finding that the 155mer rRNA substrate of Cfr binds to both the \(Cd-C\) domain and the standalone KTR from \(S.\) \(aureus.\)

My work also suggests a universal role for \(Fe^{2+}\) among the Cys-rich KTRs. If true, these systems could constitute a novel metalloprotein family. The bioinformatics survey shows that KTRs are universally small, averaging \(~60\) amino acids in length. Improvements in methods to detect small open reading frames in metagenomic sequencing datasets has inspired new efforts to understand the biological functions of
small proteins, particularly in host-associated bacterial organisms. Interestingly, one of these studies uncovered a new ribosome-associated protein in *Bacteroides thetaiotaomicron*, a common gut microbiome constituent.\textsuperscript{50} This discovery was surprising given that ribosomes are so well-characterized in model organisms, but it lends credence to my idea that KTR proteins might be associated with ribosomal RNA in anaerobic gut bacteria. As of 2019, more than 4,500 different protein families derived from small open reading frames (50 amino acids or less) have been identified.\textsuperscript{50, 51} Most remain uncharacterized. My work on stand-alone KTRs shows that functional characterization of these small proteins can reveal metal cofactors in novel scaffolds.

I also show that while clostridial Cfrs do not universally conserve a KTR neighbor, other homologs may interact with a different cellular factor in a way that could confer new activities. Clostridial Cfrs that lack a KTR are instead encoded next to an absolutely conserved Lsa ABC-F family ribosomal protection protein (RPP) genome neighbor. Other RPPs bind to the fully-assembled intact 70S ribosome to displace drugs bound to the ribosome. I hypothesize that ribosomal conformational change induced by Lsa may allow clostridial Cfrs to methylate A2503 in fully assembled ribosomes. The action of RPPs alone cannot prevent a drug re-binding event from occurring. Combining the effects of an RPP with a RNA modifying enzyme would offer lasting protection against antibiotic insult.
### Table 4-1. KTR and Lsa protein and primer sequences

<table>
<thead>
<tr>
<th><strong>Staphylococcus aureus</strong> strain 546 KTR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MNYWILCPVCGNKTRLKIREDTLELKKFPLYCPKCRQENLIEIKQFKVTITEPDAKTQSR</td>
<td></td>
</tr>
<tr>
<td>ATGAACACTAAGTGGATCTCTGTCCGGTTTGTGTAATAAAAACCCGTCTGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
<tr>
<td>CCGCCGTATTGTCCCCAATTACGAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
<tr>
<td>CAAAAACCCAGAAGCGTTAA</td>
<td></td>
</tr>
<tr>
<td><strong>PCR primers for Staphylococcus aureus</strong> strain 546 KTR</td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5’- TGATGTCCTATGACTACAAATGGATTCCGGTTTGTGGTAATAAAACCCGTCTGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5’- GGATACCTCGAGAAGCGTTTGGATTCCGGTTTGTGGTAATAAAACCCGTCTGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Clostridium botulinum</strong> KTR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MIEVKWILCPVSNKTRQRIREDTELLNPKLRKVLQNLKVEYNGNSADTFVEYFYPY VQEONSFTIDIIIEISPSMNDSLWEIVKSLMDVYDLRYQFTLSKGEQTKALLVAMLFKELSSFILDEPNTN HLDARQKLSLYKLKKGFLISNDRSDLVHISINKTNIQKGFNSWWRNLKDQFLVEKEKL KKRNIDLSLSAKRTSTWSDVSSESSHHSTGSSGKLDKYVGHAKAKMMFKAKEGQAQQMIESKKLLL</td>
<td></td>
</tr>
<tr>
<td><strong>PCR primers for Clostridium botulinum</strong> KTR</td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5’- CATATGACTCGAGGATGAAAATTCCGGAAGATACCGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5’- CCTCGAGCTGGCTGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Clostridium sporogenes</strong> Lsa family ABC-type ribosomal protection protein</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MSLINVTNLTFAYEVSDFINVFSDVDWKLGFGRNIRGKTTFNLNLKVEYNGNSADTFVEYFYPY VQEONSFTIDIIIEISPSMNDSLWEIVKSLMDVYDLRYQFTLSKGEQTKALLVAMLFKELSSFILDEPNTN HLDARQKLSLYKLKKGFLISNDRSDLVHISINKTNIQKGFNSWWRNLKDQFLVEKEKL KKRNIDLSLSAKRTSTWSDVSSESSHHSTGSSGKLDKYVGHAKAKMMFKAKEGQAQQMIESKKLLL</td>
<td></td>
</tr>
<tr>
<td><strong>PCR primers for Clostridium sporogenes</strong> Lsa family ABC-type ribosomal protection protein</td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5’- CATATGACTCGAGGATGAAAATTCCGGAAGATACCGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5’- CCTCGAGCTGGCTGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATTCCGGAAGATACCGAGCTGGAAAAATT</td>
<td></td>
</tr>
</tbody>
</table>
PCR primers for *Clostridium sporogenes* Lsa family ABC-F type ribosomal protection protein
Forward Primer: 5’-TGATGCCATATGAGTCTGATCAACGTCACCAATCTC-3’
Reverse Primer: 5’-GGATAACCTCAGACGCTTAATAGTCTTCTGTCGCCAC-3’

*Clostridioides difficile* E1 Cfr
Forward Primer: 5’-TGATGCCATATGAGTCTGATCAACGTCACCAATCTC-3’
Reverse Primer: 5’-GGATAACCTCAGACGCTTAATAGTCTTCTGTCGCCAC-3’

*Clostridioides difficile* E1 KTR
Forward Primer: 5’-TGATGCCATATGAGTCTGATCAACGTCACCAATCTC-3’
Reverse Primer: 5’-GGATAACCTCAGACGCTTAATAGTCTTCTGTCGCCAC-3’

*pBAD24* primers
Forward Primer: 5’-CCATAGCATTTTTATCCATAAG-3’
Reverse Primer: 5’-GCGTTCTGATTTAATCTGTATCAG-3’

*Ba* Cfr
Forward Primer: 5’-TGATGCCATATGAGTCTGATCAACGTCACCAATCTC-3’
Reverse Primer: 5’-GGATAACCTCAGACGCTTAATAGTCTTCTGTCGCCAC-3’

*Ef* Cfr
Forward Primer: 5’-TGATGCCATATGAGTCTGATCAACGTCACCAATCTC-3’
Reverse Primer: 5’-GGATAACCTCAGACGCTTAATAGTCTTCTGTCGCCAC-3’
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PI Cfr</strong></td>
<td></td>
<td>5’-GGATAAGAATTCATGAAATACCTGAGCAAATACGAGAAAATC-3’</td>
<td>5’-ATTAAGCTTTTAGATGTTACCATGCTACCATACTGCTG-3’</td>
</tr>
<tr>
<td><strong>Cd Cfr</strong></td>
<td></td>
<td>5’-GACATTGAATTCATGAAACAGACCAAACCAAA-3’</td>
<td>5’-ATTCTAGACACTAGTTTCTTTCATAGCA-3’</td>
</tr>
<tr>
<td><strong>Cs Cfr</strong></td>
<td></td>
<td>5’-GGATAAGAATTCATGAAATACCTGAGCAAATACGAGAAAATC-3’</td>
<td>5’-ATTAAGCTTTTAGATGTTACCATGCTACCATACTGCTG-3’</td>
</tr>
<tr>
<td><strong>Cs Lsa</strong></td>
<td></td>
<td>5’-TGATGCGAATTCATGAGTCTGATCAACGTCACAAATCTC-3’</td>
<td>5’-GGATACTCTAGATCAGCTCAAATAGTCTTCGTGCCAC-3’</td>
</tr>
</tbody>
</table>
### Table 4-2. Comparison of KTR proteins from SSN clusters 1-3 highlighting host organism, gut microbiome habitat, and pathogenicity.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Cluster</th>
<th>Class</th>
<th>Genus</th>
<th>Species</th>
<th>Gut?</th>
<th>Pathogen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0A2W5B8VX8</td>
<td>1</td>
<td>Actinobacteria</td>
<td>Gardnerella</td>
<td>vaginalis</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A0E9F8Y7</td>
<td>1</td>
<td>Chlamydiae</td>
<td>Clostridium</td>
<td>asparagiforme</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>A0A252F7K8</td>
<td>1</td>
<td>Clostridia</td>
<td>Butyricicoccus</td>
<td>porcorum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0CY97</td>
<td>1</td>
<td>Clostridia</td>
<td>Clostridium</td>
<td>siraueum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y4QKQ3</td>
<td>1</td>
<td>Clostridia</td>
<td>Clostridium</td>
<td>symbiosum</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>E7GQ78</td>
<td>1</td>
<td>Clostridia</td>
<td>Clostridium</td>
<td>symbiosum</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A2K2TU88</td>
<td>1</td>
<td>Clostridia</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A174F1H9</td>
<td>1</td>
<td>Clostridia</td>
<td>Anaerostipes</td>
<td>lactatifermentans</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A1Y3U0Z9</td>
<td>1</td>
<td>Clostridia</td>
<td>Anaerotignum</td>
<td>lactatifermentans</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A1Y3Y177</td>
<td>1</td>
<td>Clostridia</td>
<td>Blautia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A174EYP9</td>
<td>1</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1C6IZ17</td>
<td>1</td>
<td>Clostridia</td>
<td>Lachnoclostridium</td>
<td>prausnitzii</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>U2R1J5</td>
<td>1</td>
<td>Clostridia</td>
<td>Flavonifractor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D5Q4B4</td>
<td>1</td>
<td>Clostridia</td>
<td>Lachnospira</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y4MJC7</td>
<td>1</td>
<td>Clostridia</td>
<td>Anaerotruncus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1C6JLL7</td>
<td>1</td>
<td>Clostridia</td>
<td>Oscillibacter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y3WG04</td>
<td>1</td>
<td>Clostridia</td>
<td>Faecalibacterium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y3AL8</td>
<td>1</td>
<td>Clostridia</td>
<td>Faecalibacterium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y3RMK8</td>
<td>1</td>
<td>Clostridia</td>
<td>Flavonifractor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y4LDD3</td>
<td>1</td>
<td>Clostridia</td>
<td>Pseudoflavonifractor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y3U0Z9</td>
<td>1</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N5NS56</td>
<td>1</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1C5USY9</td>
<td>1</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Family</td>
<td>Genus</td>
<td>Species</td>
<td>Pathogenicity</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>A0A0S2W7A</td>
<td>Clostridia</td>
<td>Intestinimonas</td>
<td>butyriciproducens</td>
<td>Yes</td>
<td>A0A0S2W7A</td>
<td></td>
</tr>
<tr>
<td>A0A2K9P0E5</td>
<td>Clostridia</td>
<td>Monoglobus</td>
<td>pectinilyticus</td>
<td>Yes</td>
<td>A0A2K9P0E5</td>
<td></td>
</tr>
<tr>
<td>A0A1H1LC6</td>
<td>Coriobacteria</td>
<td>Olsenella</td>
<td>umbonata</td>
<td>Yes</td>
<td>A0A1H1LC6</td>
<td></td>
</tr>
<tr>
<td>A0A1Y3U6R0</td>
<td>Coriobacteria</td>
<td>Enorma</td>
<td>massiliensis</td>
<td>Yes</td>
<td>A0A1Y3U6R0</td>
<td></td>
</tr>
<tr>
<td>A0A1Y4QQY</td>
<td>Erysipelotrichia</td>
<td>Erysipelatoclitidiu</td>
<td>m</td>
<td>Yes</td>
<td>U5F4I0</td>
<td></td>
</tr>
<tr>
<td>A0A1335J8R</td>
<td>Negativicutes</td>
<td>Megasphaera</td>
<td></td>
<td></td>
<td>U3NBY8</td>
<td></td>
</tr>
<tr>
<td>A0A2I1QEA1</td>
<td>Actinobacteria</td>
<td>Gardnerella</td>
<td>vaginalis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L7VWJ4</td>
<td>Bacilli</td>
<td>Listeria</td>
<td>monocytogenes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2S6DGU</td>
<td>Bacilli</td>
<td>Staphylococcus</td>
<td>aureus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8L45</td>
<td>Bacilli</td>
<td>Staphylococcus</td>
<td>rostri</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2I1K6T2</td>
<td>Bacilli</td>
<td>Aerococcus</td>
<td>christenseni</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q56396</td>
<td>Bacilli</td>
<td>Enterococcus</td>
<td>faecalis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A0D5MB96</td>
<td>Bacilli</td>
<td>Enterococcus</td>
<td>faecium</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E0AD6</td>
<td>Bacilli</td>
<td>Enterococcus</td>
<td>faecium</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N6RVH</td>
<td>Bacilli</td>
<td>Lactobacillus</td>
<td>iners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A089QM07</td>
<td>Bacilli</td>
<td>Lactobacillus</td>
<td>salivarius</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8E016</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>agalactiae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2I1UF9</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>anginosus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2I1G983</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>anginosus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3KU35</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>cristatus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1V0GGQ</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>equinus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E0A3E3</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>equinus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N7TEQ</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>intermedius</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2I1Z7T7</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>mitis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A139QL8</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>oralis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E9NM80</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>oralis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2I1SDS</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>parasanguinis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6F278</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>parauberis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A068TM46</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>pneumoniae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2L1GYD</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>pneumoniae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7D7D6</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>pneumoniae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6Y3W9</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>pneumoniae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4H280</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>pyogenes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1V0H993</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>salivarius</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8CUK3</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td>suis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N6PPU</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N6S4J</td>
<td>Bacilli</td>
<td>Streptococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7BG31</td>
<td>Clostridia</td>
<td>Clostridium</td>
<td>perfringens</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Species 1</td>
<td>Species 2</td>
<td>Presence</td>
<td>Identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2NSUM</td>
<td>Clostridia</td>
<td>Finegoldia</td>
<td>magna</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2S1QHK</td>
<td>Clostridia</td>
<td>Clostridoides</td>
<td>difficile</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4V6N4</td>
<td>Clostridia</td>
<td>Clostridoides</td>
<td>difficile</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q188V8</td>
<td>Clostridia</td>
<td>Clostridoides</td>
<td>difficile</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N6S271</td>
<td>Clostridia</td>
<td>Peptonophilus</td>
<td>lacrimalis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2G9E6C</td>
<td>Clostridia</td>
<td>Fusobacterium</td>
<td>periactonum</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5SU22</td>
<td>Gammaproteobacteri a</td>
<td>Klebsiella</td>
<td>pneumoniae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A265AZE9</td>
<td>Gammaproteobacteri a</td>
<td>Salmonella</td>
<td>enterica</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2S2CJQ3</td>
<td>Gammaproteobacteri a</td>
<td>Photobacterium</td>
<td>damselae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2S0SJX4</td>
<td>Gammaproteobacteri a</td>
<td>Vibrio</td>
<td>harveyi</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1P8DQM</td>
<td>Gammaproteobacteri a</td>
<td>vibrio</td>
<td>parahaemolyticus</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1M4XK68</td>
<td>Clostridium</td>
<td>Clostridium</td>
<td>bolteae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1M7DUL</td>
<td>Clostridium</td>
<td>Clostridium</td>
<td>scindens</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B0G6W8</td>
<td>Clostridium</td>
<td>Eubacterium</td>
<td>callanderi</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A173TX1</td>
<td>Clostridium</td>
<td>Eubacterium</td>
<td>formicigenericans</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1E3ANZ4</td>
<td>Clostridium</td>
<td>Dorea</td>
<td>tayi</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A174RRF6</td>
<td>Clostridium</td>
<td>Blautia</td>
<td>wexlerae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2K4ZQR6</td>
<td>Clostridium</td>
<td>Acetatitcbacter</td>
<td>muris</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1M6MLC</td>
<td>Clostridium</td>
<td>Anaerotignum</td>
<td>lactatfermontans</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A174R8W6</td>
<td>Clostridium</td>
<td>Blautia</td>
<td>wexlerae</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1M6MLC</td>
<td>Clostridium</td>
<td>Anaerotignum</td>
<td>lactatfermontans</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1Y4SG60</td>
<td>Lachnoclostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7G112</td>
<td>Clostridium</td>
<td>Roseburia</td>
<td>intestalis</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0FR00</td>
<td>Clostridium</td>
<td>Roseburia</td>
<td>inulinovich</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X8G0Z0</td>
<td>Clostridium</td>
<td>Shuttleworthia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6FRK5</td>
<td>Clostridium</td>
<td>Tyzzerella</td>
<td>nexilis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3ASB3</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7KIT4</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5XHA3</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R9K671</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1LZ55</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Genus</td>
<td>Species</td>
<td>yes?</td>
<td>y?</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------</td>
<td>------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>J5UNA0</td>
<td>Clostridia</td>
<td>Peptoanaerobacter</td>
<td>stomatis</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>B0PDJ5</td>
<td>Clostridia</td>
<td>Anaerotruncus</td>
<td>collihominis</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A174RU91</td>
<td>Clostridia</td>
<td>Anaerotruncus</td>
<td>collihominis</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>A0A2N0UQ64</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td>bromii</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A2N0UY01</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td>bromii</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A174WS9</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td>torques</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A150Y753</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1C6E036</td>
<td>Clostridia</td>
<td>Ruminococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5EWS3</td>
<td>Clostridia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1ARH1</td>
<td>Erysipelotrichia</td>
<td>Coprobacillus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U5FB22</td>
<td>Erysipelotrichia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0A1E9IN48</td>
<td>Fusobacteria</td>
<td>Fusobacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0WA77</td>
<td>Negativicutes</td>
<td>Acidaminococcus</td>
<td>intestini</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


CHAPTER 5

Summary of Results and Future Directions
Conclusions

The class A radical SAM methylase Cfr has been increasingly implicated in conferring resistance to medically-relevant drug agents in clinical settings.\textsuperscript{1-4} Antibiotic resistance is mediated by methylation of A2503 of the bacterial rRNA at the C8 position.\textsuperscript{5} The reaction is chemically challenging, requiring a protein-based methyl intermediate and formation of a potent 5'-dA radical species.\textsuperscript{6-8} Though the risks of \textit{cfr}-mediated drug resistance continue to grow, \textit{in vitro} characterization of Cfr homologs has remained largely unexplored outside that of model system \textit{Staphylococcus aureus} Cfr.\textsuperscript{5-10} The work contained in this thesis aimed to diversify the profile of characterized Cfrs by isolating homologs from four phylogenetically distinct organisms and assessing their biophysical properties and ability to perform Cfr-like rRNA methylation.

Phylogenetic studies of Cfrs revealed distinct clades separated by organism.\textsuperscript{11} In chapter 2, Cfrs from the bacilli/enterococci, paenibacilli, and clostridia clades were chosen for \textit{in vitro} characterization. Each Cfr was demonstrated to bind a [4Fe-4S] cluster evidencing their radical SAM designation. When assaying for methylation activity, all four homologs were able to produce m8A when given a synthetic 155mer rRNA substrate. Differences in the ratio of singly methylated and di-methylated product as well as the ratio of unproductive SAM cleavage were apparent when comparing the use of a protein-based flavodoxin reductant and the chemical reductant dithionite suggesting an important role for the reductant in determining enzyme kinetics and outcomes. \textit{In vitro} methylation activity could not be correlated to antibiotic resistance phenotypes in \textit{E. coli}, likely due to issues of host identity and protein toxicity. Interestingly, spectroscopic analysis of the clostridial Cfr revealed the presence of an additional metal site.
Chapter 3 aimed to characterize the accessory metal site and its role in Cfr-like activity. Truncation variants revealed that the additional metal was iron and that the metal was located in the C-terminal extension domain. Bioinformatic analysis of the accessory domain led to the discovery of a large group of uncharacterized stand-alone proteins, named cys-rich KTRs, with two rubredoxin-like CXXC pairs and a uniquely high isoelectric point. Removal of the domain did not appear to impact the ability of the radical SAM domain to methylate the 155mer rRNA substrate. However, replacement of the iron by zinc or mutation of one of the coordinating Cys ligands severely impacted methylation activity, especially under multiple turnover conditions. Additionally, electrochemical analysis of the domain revealed a highly positive midpoint potential. These results suggest that the presence of iron is important to the function of the domain and that it may play a role in electron transfer. With regard to mechanism, the most likely target would be the thiyl radical proposed to remain at the termination of the reaction.\textsuperscript{12} The domain may also function to position the substrate, as the abundance of positively charged residues may interact with negatively charged RNA. Whatever the role, it was of note that the clostridial Cfr was able to perform methylation considering previous clostridial Cfrs presenting a lack of Cfr-like activity.\textsuperscript{11, 13}

In chapter 4 I determined that each clostridial \textit{cfr} was adjacent to one of two neighbors, \textit{ktr} or \textit{lsa}, and that these genetic elements could represent the missing link in promoting Cfr-like activity in clostridial Cfrs. For this reason, I characterized both types of proteins. Isolation and spectroscopic analysis of a stand-alone KTR revealed that it too bound a single iron atom suggesting that the cys-rich KTRs represent a novel iron-binding protein family. Due to their presence next to Cfr and other antibiotic resistance
elements, it is likely that they play a role in resistance themselves and are thus medically relevant. This hypothesis is further evidenced by their localization next to transposons and their presence in both human gut commensals and pathogenic bacilli. As with the Cfrs, antibiotic resistance could not be conferred by either KTRs or Lsa in *E. coli* but may with further optimization.

Taken together, the work presented here provides evidence for mechanistic diversity within the Cfr family. Members of all phylogenetic clades were shown to coordinate the radical SAM cluster and perform C8 methylation of a 155mer substrate mimic, but the homologs catalyzed the reaction at varying efficiencies. A C-terminal domain was identified on the clostridial Cfr homolog which was demonstrated to coordinate an additional iron and bind RNA. Further analysis of clostridial Cfrs led to the discovery of Lsa ribosomal protection proteins and cys-rich KTRs, a novel metalloprotein family, as possible partner proteins. Identification of these elements expands the understanding of Cfr and may lead to new methods of treatment.

**Future Directions**

*Structural determination of Cfr and KTR*

Perhaps the most important goal moving forward is structural determination of Cfr and KTR. Visualizing Cfr would allow for a direct comparison to previously solved structures of related Class A enzyme RlmN and may provide the answer as to why Cfr promotes C8 methylation while RlmN catalyzes C2 methylation. Additionally, solving the structure of Cfr would allow for rational chemical inhibitor design using the active site as a template. Determining the structure of a KTR would also be of interest because
of its association with multiple antibiotic resistance determinants. The structure may provide a clue as to the positioning of key residues, such as the numerous positively charged amino acids, which may be involved in substrate binding. Additionally, the dissimilarities in sequence (residues in and around the CXXC pairs) and UV-visible spectroscopy (broad peaks) compared to rubredoxins suggests a unique FeS$_4$ protein structure.

_**Antibiotic resistance profiling of Cfr, KTR, and Lsa**_

Though profiling of antibiotic resistance conferred by Cfr, KTR, and Lsa proved unsuccessful under the current conditions, there are further routes of optimization left to explore. Likely the most impactful change would be the use of a gram-positive host. Lsa, in particular, is related to enzymes which have already shown organism-specific interactions. It is possible that the Cfrs and KTRs are similar in this respect. Using a bacilli or clostridial host may provide a closer in vivo environment to that of the native hosts. Barring the use of a gram-positive organism, the use of an antibiotic susceptible strain of *E. coli*, such as AS19, may be more suitable than the current host. Successful demonstration of Cfr-related antibiotic resistance has been reported previously in AS19, and as an additional benefit, would allow direct comparison to previous works.

_**Determining substrate specificity for Lsa**_

Determining the possible role of Lsa in clostridial Cfr-like activity is also of interest. Proteins related to Lsa are known to bind to fully assembled ribosomes to displace pharmacological translational blockers. However, it is not known how those
proteins prevent drug re-binding. It is possible that Cfr represents a method through which this issue could be addressed. However, unlike other Cfrs, prevention of drug re-binding would require Cfr to methylate the A2503 site in a fully assembled ribosome. The structural re-arrangements catalyzed by Lsa may provide Cfr access to the residue. Thus, isolating Lsa and testing for A2503 methylation of fully assembled ribosomes when both a clostridial Cfr and its associated Lsa are present in vitro is a logical next step.
References


VITA
James David Gumkowski

Education:
The Pennsylvania State University, Department of Chemistry, University Park, PA, Ph.D. 2013-2020
Advisor: Dr. Amie Boal

Muhlenberg College, Allentown, PA, Bachelor of Science in Biochemistry, 2009

Publications:


Awards:

Dalalian Graduate Fellowship in Organic Chemistry, The Pennsylvania State University, 2017

Dan H. Waugh Memorial Teaching Award, The Pennsylvania State University, 2016

Fralin Undergraduate Research Fellowship, Virginia Polytechnic Institute and State University, 2012

Alphonse C. Lova Memorial Award, Muhlenberg College, 2012